Robust Score and Portmanteau Tests of Volatility Spillover

Mike Aguilar* and Jonathan B. Hill†

Department of Economics, University of North Carolina at Chapel Hill

February 3, 2012

Abstract

This paper presents a variety of tests of volatility spillover that are robust to heavy tails. The tests are couched in semi-strong and strong GARCH frameworks with idiosyncratic shocks that are only required to have a finite variance if they are independent. We negligibly trim test equations, or components of the equations, and construct heavy tail robust score and portmanteau statistics. We develop the tail-trimmed sample correlation coefficient for robust inference, and prove that its Gaussian limit under the null hypothesis of no spillover has the same standardization irrespective of tail thickness. Further, if spillover occurs within a specified horizon our test obtains a power of one asymptotically. A Monte Carlo study shows our tests provide significant improvements over extant GARCH-based tests of spillover, and we apply the tests to financial returns data.

1 Introduction

A rich literature has emerged on testing for financial market associations, spillover and contagion, and price/volume relationships during volatile periods (King et al 1994, Karolyi and Stulz 1996, Brooks 1998, Comte and Lieberman 2000, Hong 2001, Forbes and Rigobon 2002, Caporale et al 2005, Caporale et al 2006). Similarly, evidence for heavy tails across disciplines is substantial, with a large array of studies showing heavy tails and random volatility effects in financial returns. See Campbell and Hentschel (1992), Engle and Ng (1993), Embrechts et al (1997), Longin and Solnick (2001), Finkenstadt and Rootzén (2003), Poon et al (2003), and Laurini and Tawn (2009).

*Dept. of Economics, University of North Carolina at Chapel Hill, maguilar@email.unc.edu.

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

Keywords: volatility spillover; heavy tails; tail trimming; robust inference.

JEL subject classifications. C13, C20, C22.

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

In lieu of mounting evidence for heavy tails and heteroscedasticity in financial markets, non-correlation based methods have evolved, including distribution free correlation-integral tests (Brock et al 1996, de Lima 1996, Brooks 1998), exact small sample tests based on sharp bounds (Dufour and Roy 1985, Dufour et al 2006), copula-based tests (Schmidt and Stadtmüller 2003), and tail dependence tests (Longin and Solnick 2001, Poon et al 2003, Malevergne and Sornette 2004, Davis and Mikosch 2009, Hill 2009).

In this paper, rather than look for new techniques, we exploit robust methods and asymptotic theory to allow for the use of existing representations of so-called volatility *spillover* or *contagion*[1] where idiosyncratic shocks may be heavy tailed. We use a GARCH(1,1) framework under minimal moment restrictions and deliver test statistics with standard limit distributions. The GARCH(1,1) model allows us to focus ideas, but other frameworks are clearly possible including ARMA-GARCH, nonlinear GARCH and other volatility models (e.g. Meddahi and Renault 2004).

Let $\{y_{1,t}, y_{2,t}\}$ be a joint process of interest with GARCH(1,1) coordinates:

$$y_{i,t} = h_{i,t}\epsilon_{i,t}, \;\; E\left[\epsilon_{i,t}\right] = 0, \;\; E\left[\epsilon_{i,t}^2\right] = 1, \text{ for each } i = 1, 2, \tag{1}$$

$$h_{i,t}^2 = \omega_i^0 + \alpha_i^0 y_{i,t-1}^2 + \beta_i^0 h_{i,t-1}^2, \; \omega_i^0 > 0, \; \alpha_i^0, \beta_i^0 \geq 0.$$

Define the parameter set $\theta = [\theta_1', \theta_2']'$ where $\theta_i = [\omega_i, \alpha_i, \beta_i]' \in \Theta$ a compact subset of $\mathbb{R}^3$. We assume there exist unique interior points $\theta_i^0 \in \Theta$ such that $\{y_{i,t}, h_{i,t}\}$ are stationary and ergodic, and $E[\epsilon_{i,t}] = 0$ and $E[\epsilon_{i,t}^2] = 1$. Cheung and Ng (1996) and Hong (2001) argue volatility spillover reduces to testing whether $y_{1,t}^2/h_{1,t}^2 - 1$ and $y_{2,t-i}^2/h_{2,t-i}^2 - 1$ are correlated, where $\epsilon_{i,t}$ is assumed to be serially independent. Hong (2001) proposes a standardized portmanteau statistic to test for spillover at asymptotically infinitely many lags, and requires $E[\epsilon_{i,t}^8] < \infty$, although $y_{i,t}$ may be IGARCH or mildly explosive GARCH, as long as $y_{i,t}$ is stationary.

The assumption of thin tails is certainly not unique to these works since volatility spillover and contagion methods universally enforce substantial moment conditions. Forbes and Rigobon (2002) implicitly require VAR errors to have a fourth moment; Caparole et al (2005) exploit QML estimates for a GARCH model and therefore need at least $E[\epsilon_{i,t}^4] < \infty$ cf. Francq and Zakoïan (2004). Despite the fixation on thin-tail assumptions, there appears to be little in the way of robustness checks, or pre-tests to verify the required thinness of tails. See especially de Lima (1997) and Hill and Aguilar (2010). Dungey et al (2005), for example, study an array of sampling properties of tests of contagion

[1] There is some consensus in the applied literature on use of the terms "spillover" versus "contagion" in financial markets: spillover concerns "usual" market linkages and contagion suggests "unanticipated transmission of shocks" (e.g. Beirne et al 2008: p. 4). We simply use the term "spillover" for convenience and in lieu of past usage in the volatility literature (e.g. Cheung and Ng 1996, Hong 2001). Since we allow for very heavy tails in the errors, our contributions arguably also apply to the contagion literature since such noise renders anticipating linkages exceptionally difficult.

and spillover, but do not treat heavy tails. In Section 6, however, we show a variety of asset returns series have GARCH components with errors $\epsilon_{i,t}$ that may have an unbounded fourth moment.

Our approach is similar to Cheung and Ng (1996) and Hong (2001). Define a volatility function

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

construct centered squared errors

$$\mathcal{E}_{i,t}(\theta_i) := \frac{y_{i,t}^2}{h_{i,t}^2(\theta_i)} - 1 = \epsilon_{i,t}^2(\theta_i) - 1 \;\; \text{and} \;\; \mathcal{E}_{i,t} = \mathcal{E}_{i,t}(\theta_i^0)$$

and build test equations over H lags

$$m_t(\theta) = \left[\mathcal{E}_{1,t}(\theta_1) \times \mathcal{E}_{2,t-h}(\theta_2)\right]_{h=1}^{H}, \, H \geq 1, \;\; \text{and} \;\; m_t = m_t(\theta^0).$$

Now drop $\theta^0 = [\theta_1^{0\prime}, \theta_2^{0\prime}]'$ throughout. Under the null of no spillover $E[m_{h,t}] = E[(\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-h}^2 - 1)] = 0$. The conventional assumption $E[m_{h,t}^2] < \infty$ requires $E[\epsilon_{i,t}^4] < \infty$ if $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are mutually independent, while $E[\epsilon_{i,t}^8] < \infty$ is imposed to ensure consistency of estimated higher moments $E[m_{h,t}^2]$ given the presence of a plug-in for θ^0.

We conquer the problem of possibly heavy tailed non-iid idiosyncratic shocks $\epsilon_{i,t}$ by exploiting tail-trimming techniques in three ways. First, we tail-trim $m_t(\theta)$ by its large values for a robust score-type test statistic, similar to the statistic developed in Hill and Aguilar (2010). In general this does not allow a portmanteau statistic even if $\epsilon_{i,t}$ are iid, and may incur small sample bias due to asymmetry. Our second method trims $m_t(\theta)$ by trimming $\mathcal{E}_{i,t}(\theta_i)$, leading to robust score *and* portmanteau statistics, while small sample bias is eradicated by re-centering the trimmed $\mathcal{E}_{i,t}(\theta_i)$. Re-centering and trimming negligibly imply we may tail-trim symmetrically, making the decision for trimming relatively easy. Finally, in the third method we symmetrically tail-trim and re-center $\epsilon_{i,t}(\theta_i)$.

We believe this is the first study to explore a tail-trimmed sample correlation for robust inference, and derive its Gaussian self-standardized limit under no-spillover. The Q-statistic form is $T \sum_{h=1}^{H} \mathcal{W}_T(h) \, (\hat{\rho}_{T,h}(\hat{\theta}_T))^2$ where T is the sample size, $\mathcal{W}_T(h)$ are weights, $\hat{\rho}_{T,h}(\hat{\theta}_T)$ is the tail-trimmed sample correlation at lag h, and $\hat{\theta}_T$ estimates θ^0. Notice the proper scale is T, where even in heavy tailed cases $T^{1/2}\hat{\rho}_{T,h}(\hat{\theta}_T)$ $\overset{d}{\rightarrow} N(0,1)$ under the null and mild regularity conditions, and $T^{1/2}|\hat{\rho}_{T,h}(\hat{\theta}_T)| \overset{p}{\rightarrow} \infty$ if there is spillover at lag h[2]. This is more convenient to compute than Runde's (1997) re-scaled Q-statistic for infinite variance data since that requires the true tail index if $E[\epsilon_{i,t}^4] = \infty$, hence a different statistic is required depending on tail thickness[3]. Our tail-trimmed sample correlation, however, is asymptotically nuisance

[2]All limits in this paper are as $T \rightarrow \infty$, while H is always a fixed constant unless otherwise noted.

[3]Runde (1997) does not characterize the re-scaled Q-statistic under the alternative evidently because the correlation

parameter-free, and roughly a non-tail or robust variant of the tail array correlation coefficients in Davis and Mikosch (2009) and Hill (2009), amongst others.

Under any of the three trimming schemes score tests are feasible that do not require error independence (see Forbes and Rigobon 2002). In this case $\epsilon_{i,t}$ only needs to be geometrically mixing and $E[\epsilon_{i,t}^2] = 1$ in the iid case, and up to a finite fourth moment if $\epsilon_{i,t}$ are non-iid. In order to gain access to QML-type plug-ins for θ^0 we assume $\epsilon_{i,t}^2 - 1$ are martingale differences allowing for semi-strong GARCH models (Drost and Nijman 1993, Hill and Renault 2010, Linton et al 2010, Hill 2011).

Although re-centering trimmed $\mathcal{E}_{i,t}(\theta_i)$ or $\epsilon_{i,t}(\theta)$ eradicates bias, we must still decide how many $m_{h,t}$, $\mathcal{E}_{i,t}$ or $\epsilon_{i,t}$ to trim in any one sample. We follow Hill and Aguilar (2010) and Hill (2012) and derive an asymptotic p-value function $p_T(\lambda)$ of a trimming parameter $\lambda \in (0, 1]$ that gauges the number of trimmed observations, and propose a test based on the occupation time of $p_T(\lambda)$ under nominal size α.

Due to the self-scaling structure of a sample correlation, the Q-statistics are robust to any plug-in $\hat{\theta}_T$ for θ^0 with a minimal convergence rate that is under $T^{1/2}$ if at least one $E[\epsilon_{i,t}^4] = \infty$. Allowed plug-ins in general are Log-LAD as in Peng and Yao (2003) and Linton et al (2010), Hill and Renault's (2010) Generalized Method of Tail-Trimmed Moments, Hill's (2011) Quasi-Maximum Tail-Trimmed Likelihood, and Zhu and Ling's (2012) Globally Weighted Quasi-Maximum Exponential Likelihood, while QML converges too slowly when $E[\epsilon_{i,t}^4] = \infty$. If we know both $E[\epsilon_{i,t}^4] < \infty$ and $E[\epsilon_{i,t}^8] = \infty$, then Hong's (2001) test remains invalid, ours is trivially robust, and QML is then valid. Even in this case we find in a Monte Carlo experiment that trimming matters for accurate empirical size.

In Section 2 we construct the various test statistics, and present the main results in Section 3. We discuss valid plug-ins and the choice of trimming portion by occupation time in Section 4. A simulation study and empirical application follow in Sections 5 and 6, and parting comments are left for Section 7.

The following notational conventions are used. If sequences $\{a_T, b_T\}$ are stochastic and $a_T/b_T \overset{p}{\to} 1$ we write $a_T \overset{p}{\sim} b_T$, and if they are non-stochastic and $a_T/b_T \to 1$ we write $a_T \sim b_T$. $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the minimum and maximum eigenvalues of a square matrix A. The L_p-norm of an $M \times N$ matrix A is $||A||_p = (\sum_{i=1,j=1}^{M,N} E|A_{i,j}|^p)^{1/p}$, and the spectral (matrix) norm is $||A|| = (\lambda_{\max}(A'A))^{1/2}$. If z is a scalar we write $(z)_+ := \max\{0, z\}$, and $[z]$ denotes the integer part of z. K denotes a positive finite constant whose value may change from line to line; $\iota > 0$ is a tiny constant; N is a whole number. $\overset{p}{\to}$ and $\overset{d}{\to}$ denote probability and distribution convergence. $L(T)$ is a slowly varying[4] function, $L(T) \to \infty$, the value or rate of which may change from line to line. We say a random variable is *symmetric* if its distribution is symmetric about zero.

does not exist: the properly standardized sample correlation converges to a random variable (Davis and Resnick 1986). Thus, whether the Q-statistic is consistent is unknown.

[4]Recall slowly varying $L(T)$ satisfies $L(\nu T)/L(T) \to 1$ for any $\nu > 0$. Classic examples are constants and powers of the natural logarithm (e.g. $(\ln(n))^a$ for $a > 0$). In this paper always $L(T) \to \infty$ as $T \to \infty$.

2 Robust Test Statistics

In the following we introduce five tail-trimmed test statistics. A summary of each is provided in Table 1 at the end of this section. We drop θ^0 whenever it is understood, and write θ to denote either θ_i or $[\theta_1', \theta_2']'$ when there is no confusion. Thus θ lies in Θ a compact subset of $\mathbb{R}^3$ or $\mathbb{R}^6$.

2.1 Tail-Trimmed Equations: Score Test

Our first approach is to trim $m_{h,t}(\theta) = (\epsilon_{1,t}^2(\theta_1) - 1)(\epsilon_{2,t-h}^2(\theta_2) - 1)$ by its large values, as in Hill and Aguilar [HA] (2010). The null hypothesis of no volatility spillover up to horizon $H \geq 1$ can be written

$$H_0^{(m)} : E\left[m_{h,t}\right] = E\left[\left(\epsilon_{1,t}^2 - 1\right)\left(\epsilon_{2,t-h}^2 - 1\right)\right] = 0 \text{ for all } h = 1, ..., H.$$

Thus if $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are mutually independent under $H_0^{(m)}$ then we only need finite second moments, and otherwise up to a finite fourth moment is required.

Since $m_{h,t}$ may be asymmetric we must trim asymmetrically. Define tail specific observations of $m_{h,t}(\theta)$ and their sample order statistics:

$$m_{h,t}^{(-)}(\theta) := m_{h,t}(\theta) \times I\left(m_{h,t}(\theta) < 0\right) \quad \text{and} \quad m_{h,(1)}^{(-)}(\theta) \leq \cdots \leq m_{h,(T)}^{(-)}(\theta) < 0 \tag{2}$$

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

Let $\{k_{r,T}^{(m)} : r = 1, 2\}$ be integer sequences representing the number of trimmed left- and right-tailed observations from the sample $\{m_{h,t}(\theta)\}_{t=1}^T$. We enforce *negligible* trimming by assuming $\{k_{r,T}^{(m)}\}$ are intermediate order sequences: $k_{r,T}^{(m)} \to \infty$ and $k_{r,T}^{(m)}/T \to 0$ (Leadbetter et al 1983). The tail-trimmed version of $m_{h,t}(\theta)$ is then

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

where the indicator function $I(A) = 1$ if A is true, and 0 otherwise. Trimming asymptotically infinitely many observations $k_{r,T}^{(m)} \to \infty$ that represent a vanishing tail portion $k_{r,T}^{(m)}/T \to 0$ ensures Gaussian asymptotics and identification of the hypotheses. Fixed quantile trimming $k_{i,T}^{(m)}/T \to (0, 1)$, by contrast, cannot ensure identification of the null since $m_{h,t}$ may be asymmetric.

A long-run variance estimator is recommended since even in the strong-GARCH case $\hat{m}_{h,T,t}^*$ may be spuriously correlated due to asymmetric trimming. Let $\hat{S}_T(\theta)$ be a kernel HAC estimator for $\hat{m}^*_{h,T,t}(\theta)$,

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

where $\hat{m}^*_T(\theta) := 1/T\sum_{t=1}^{T}\hat{m}^*_{T,t}(\theta)$, $\mathcal{K}(\cdot)$ is a kernel function and γ_T is bandwidth where $\gamma_T \to \infty$ and $\gamma_T = o(T)$. The tail-trimmed test statistic has a quadratic form as in HA (2010), where $\hat{\theta}_T$ denotes a consistent estimator of θ^0:

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

2.2 Tail-Trimmed Centered Errors: Score and Portmanteau Tests

Our second approach is to tail-trim $\mathcal{E}_{i,t}(\theta) = y^2_{i,t}(\theta)/h^2_{i,t}(\theta) - 1$ symmetrically and then re-center. Define $s_{i,t}(\theta) := (\partial/\partial\theta)\ln(h^2_{i,t}(\theta))$ and note we can write, as $T \to \infty$, $\mathcal{E}_{i,t}(\hat{\theta}_T) = \mathcal{E}_{i,t} - \epsilon^2_{i,t}s_{i,t}(\hat{\theta}_T - \theta_0)$. If there are GARCH effects $\alpha^0_i + \beta^0_i > 0$ then $s_{i,t}$ is uniformly square integrable (Francq and Zakoïan 2004) so trimming $\mathcal{E}_{i,t}(\hat{\theta}_T)$ only requires information from $\epsilon_{i,t}(\theta)$. However, if there are no GARCH effects $\alpha^0_i + \beta^0_i = 0$ and $y_{i,t-1}$ exhibits extremes, then $s_{i,t} = [1, y^2_{i,t-1}/\omega^0_i, 1]$ does as well. Hence, if we allow this case then we must trim $\mathcal{E}_{i,t}(\theta)$ by $\epsilon_{i,t}(\theta)$ *and* $y_{i,t-1}$. By contrast Hill and Renault (2010) show it suffices to trim $m_{h,t}(\theta)$ by its large values[5].

Define two-tailed observations and their sample order statistics

$$\mathcal{E}^{(a)}_{i,t}(\theta) := |\mathcal{E}_{i,t}(\theta)| \quad \text{and} \quad \mathcal{E}^{(a)}_{i,(1)}(\theta) \geq \mathcal{E}^{(a)}_{i,(2)}(\theta) \geq \cdots \geq \mathcal{E}^{(a)}_{i,(T)}(\theta), \tag{3}$$

and intermediate order sequences $\{k^{(\mathcal{E})}_{i,T}\}$. If we assume GARCH effects we use

$$\hat{I}^{(\mathcal{E})}_{i,T,t}(\theta) := I\left(|\mathcal{E}_{i,t}(\theta)| \leq \mathcal{E}^{(a)}_{i,(k^{(\mathcal{E})}_{i,T})}(\theta)\right) \text{ and } \widehat{\mathcal{E}}^*_{i,T,t}(\theta) := \mathcal{E}_{i,t}(\theta) \times \hat{I}^{(\mathcal{E})}_{i,T,t}(\theta) - \frac{1}{T}\sum_{t=1}^{T}\mathcal{E}_{i,t}(\theta) \times \hat{I}^{(\mathcal{E})}_{i,T,t}(\theta),$$

and if we allow for the possibility of no GARCH effects then replace $\hat{I}^{(\mathcal{E})}_{i,T,t}(\theta)$ with $\hat{I}^{(\mathcal{E})}_{i,T,t}(\theta)\hat{I}^{(y)}_{i,T,t-1}$. In practice the analyst can safely use $\hat{I}^{(\mathcal{E})}_{i,T,t}(\theta)\hat{I}^{(y)}_{i,T,t-1}$ because if there are GARCH effects then $\hat{I}^{(y)}_{i,T,t-1}$ does not have any impact on the test statistic asymptotically [6]. Although $\mathcal{E}_{i,t}$ may be asymmetric, trimming negligibility and re-centering ensure both $1/T\sum_{t=1}^{T}\widehat{\mathcal{E}}^*_{i,T,t} \xrightarrow{p} E[\mathcal{E}_{i,t}] = 0$ and $1/T\sum_{t=1}^{T}\widehat{\mathcal{E}}^*_{1,T,t}\widehat{\mathcal{E}}^*_{2,T,t-h} \xrightarrow{p}$

[5] The gradient $(\partial/\partial\theta)m_{h,t}(\theta)$ that appears in a first order expansion of $m_{h,t}(\theta)$ around θ^0 is sufficiently trimmed for Gaussian asymptotics when $m_{h,t}(\theta)$ itself reaches an extreme value. Thus we need only trim $m_{h,t}(\theta)$ by $m_{h,t}(\theta)$ to generate a robust GMM estimator and moment condition test. See Lemma 2.5 in Hill and Renault (2010), and see HA (2010).

[6] This follows from the negligibility of trimming $\hat{I}^{(y)}_{i,T,t-1} \xrightarrow{p} 1$: unnecessary tail-trimming impacts the test statistic neither asymptotically, nor in small samples according to simulation evidence in Section 5.

$E[\mathcal{E}_{1,t}\mathcal{E}_{2,t-h}] = 0$ under the null for *any* intermediate order sequences $\{k_{i,T}^{(\mathcal{E})}\}$.

Define a sample tail trimmed correlation coefficient and portmanteau statistic:

$$\hat{\rho}_{T,h}^{(\mathcal{E})}(\theta) := \frac{\sum_{t=1}^{T} \widehat{\mathcal{E}}_{1,T,t}^{*}(\theta_1)\, \widehat{\mathcal{E}}_{2,T,t-h}^{*}(\theta_2)}{\left(\sum_{t=1}^{T} \widehat{\mathcal{E}}_{1,T,t}^{*2}(\theta_1)\right)^{1/2} \left(\sum_{t=1}^{T} \widehat{\mathcal{E}}_{2,T,t}^{*2}(\theta_2)\right)^{1/2}} \quad \text{and} \quad \hat{Q}_T^{(\mathcal{E})}(H) = T\sum_{h=1}^{H} \mathcal{W}_T(h) \left(\hat{\rho}_{T,h}^{(\mathcal{E})}(\hat{\theta}_T)\right)^2,$$

where $\{\mathcal{W}_T(h)\}$ is a sequences of positive, possibly random weights, satisfying $\max_{1\leq h\leq H} |\mathcal{W}_T(h) - 1| \xrightarrow{p} 0$ as $T \rightarrow \infty$. Examples include deterministic weights used in the Box-Pierce and Ljung-Box tests $\mathcal{W}_T(h) = 1$, $(T+2)/(T-h)$ or $T/(T-h)$.

It is interesting to point out the correct scale is T, although tail-trimmed sums of infinite variance processes lead to non-$T^{1/2}$ asymptotics (e.g. Hahn et al 1990, 1991, Hahn and Weiner 1992, Hill and Renault 2010). Nevertheless, under the null hypothesis of no spillover framed as

$$H_0^{(\epsilon)} : \epsilon_{1,t} \text{ and } \epsilon_{2,t} \text{ are mutually independent,}$$

we have $T^{1/2}\hat{\rho}_{T,h}^{(\mathcal{E})} \xrightarrow{d} N(0,1)$ due to self-normalization through the scales $(T^{-1}\sum_{t=1}^{T} \widehat{\mathcal{E}}_{1,T,t}^{*2})^{1/2} \times (T^{-1}\sum_{t=1}^{T} \widehat{\mathcal{E}}_{2,T,t}^{*2})^{1/2}$, exactly like the classic setting in Box and Pierce (1970). In particular

$$T^{1/2}\hat{\rho}_{T,h}^{(\mathcal{E})}(\hat{\theta}_T) \overset{p}{\sim} \frac{1}{T^{1/2}\left(E\left[\widehat{\mathcal{E}}_{1,T,t}^{*2}\right]\right)^{1/2}\left(E\left[\widehat{\mathcal{E}}_{2,T,t}^{*2}\right]\right)^{1/2}} \sum_{t=1}^{T} \widehat{\mathcal{E}}_{1,T,t}^{*} \widehat{\mathcal{E}}_{2,T,t-h}^{*}$$

is a properly self-standardized tail-trimmed sum, hence a classic Q-statistic with a standard limit is feasible. A standardized Q-test as in Hong (2001) can similarly be constructed to allow for an increasing horizon $H \rightarrow \infty$ as $T \rightarrow \infty$.

If the errors are not mutually independent under the null the no spillover, then $\hat{\rho}_{T,h}^{(\mathcal{E})}(\hat{\theta}_T)$ are not necessarily asymptotically independent hence $\hat{Q}_T^{(\mathcal{E})}(H)$ need not have a limiting chi-squared distribution. In the semi-strong GARCH case we can test $H_0^{(m)}$ by computing test equations $[\widehat{\mathcal{E}}_{1,T,t}^{*}(\theta)\widehat{\mathcal{E}}_{2,T,t-h}^{*}(\theta)]_{h=1}^{H}$, an associated HAC $\hat{S}_T^{(\mathcal{E})}(\theta)$, and a score statistic (cf. HA 2010)

$$\hat{W}_T^{(\mathcal{E})}(H) := \left(\sum_{t=1}^{T}\left[\widehat{\mathcal{E}}_{1,T,t}^{*}(\theta)\widehat{\mathcal{E}}_{2,T,t-h}^{*}(\theta)\right]_{h=1}^{H}\right)' \hat{S}_T^{(\mathcal{E})}(\hat{\theta}_T)^{-1} \left(\sum_{t=1}^{T}\left[\widehat{\mathcal{E}}_{1,T,t}^{*}(\theta)\widehat{\mathcal{E}}_{2,T,t-h}^{*}(\theta)\right]_{h=1}^{H}\right).$$

2.3 Tail-Trimmed Errors: Score and Portmanteau Tests

Lastly, we trim $\epsilon_{i,t}(\theta)$ by its large values. Write $\epsilon_{i,t}^{(a)}(\theta) := |\epsilon_{i,t}(\theta)|$, define order statistics $\{\epsilon_{i,(j)}^{(a)}(\theta)\}$, and let $\{k_{i,T}^{(\epsilon)}\}$ be intermediate order sequences. Define trimmed errors: if GARCH effects are known

$$\hat{\epsilon}_{i,T,t}^{*}(\theta) = \epsilon_{i,t}(\theta) I\left(|\epsilon_{i,t}(\theta)| \leq \epsilon_{i,(k_{i,T}^{(\epsilon)})}^{(a)}(\theta)\right) = \epsilon_{i,t}(\theta) \hat{I}_{i,T,t}^{(\epsilon)}(\theta),$$

else use $\hat{\epsilon}_{i,T,t}^{*}(\theta) = \epsilon_{i,t}(\theta)\hat{I}_{i,T,t}^{(\epsilon)}(\theta)\hat{I}_{i,T,t-1}^{(y)}$. The re-centered squared tail-trimmed errors are

$$\widehat{\mathfrak{E}}_{i,T,t}^{*}(\theta) := \hat{\epsilon}_{i,T,t}^{*2}(\theta) - \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{i,T,t}^{*2}(\theta),$$

or alternatively we may trim $\epsilon_{i,t}(\theta)$, re-center, square and re-center again:

$$\widehat{\mathfrak{E}}_{i,T,t}^{*}(\theta) := \left(\hat{\epsilon}_{i,T,t}(\theta) - \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{i,T,t}^{*}(\theta)\right)^{2} - \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\epsilon}_{i,T,t}(\theta) - \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{i,T,t}^{*}(\theta)\right)^{2}.$$

The tail-trimmed correlation and Q-statistic for a test of $H_0^{(\epsilon)}$ are

$$\hat{\rho}_{T,h}^{(\mathfrak{E})}(\theta) = \frac{\sum_{t=1}^{T}\widehat{\mathfrak{E}}_{1,T,t}^{*}(\theta_1)\widehat{\mathfrak{E}}_{2,T,t-h}^{*}(\theta_2)}{\left(\sum_{t=1}^{T}\widehat{\mathfrak{E}}_{1,T,t}^{*2}(\theta_1)\right)^{1/2}\left(\sum_{t=1}^{T}\widehat{\mathfrak{E}}_{2,T,t}^{*2}(\theta_2)\right)^{1/2}} \quad \text{and} \quad \hat{Q}_T^{(\mathfrak{E})}(H) := T\sum_{h=1}^{H}\mathcal{W}_T(h)(\hat{\rho}_{T,h}^{(\mathfrak{E})}(\hat{\theta}_T))^2.$$

In the semi-strong GARCH case we use equations $[\widehat{\mathfrak{E}}_{1,T,t}^{*}(\theta) \times \widehat{\mathfrak{E}}_{2,T,t-h}^{*}(\theta)]_{h=1}^{H}$, HAC $\hat{S}_T^{(\mathfrak{E})}(\theta)$ and a score statistic for a test of $H_0^{(m)}$:

$$\hat{W}_T^{(\mathfrak{E})}(H) := \left(\sum_{t=1}^{T}\left[\widehat{\mathfrak{E}}_{1,T,t}^{*}(\theta)\widehat{\mathfrak{E}}_{2,T,t-h}^{*}(\theta)\right]_{h=1}^{H}\right)' \hat{S}_T^{(\mathfrak{E})}(\hat{\theta}_T)^{-1}\left(\sum_{t=1}^{T}\left[\widehat{\mathfrak{E}}_{1,T,t}^{*}(\theta)\widehat{\mathfrak{E}}_{2,T,t-h}^{*}(\theta)\right]_{h=1}^{H}\right).$$

There is no theory-based advantage for trimming by $\mathcal{E}_{i,T,t}$ versus $\epsilon_{i,t}$, or in which order we re-center the trimmed $\epsilon_{i,t}$. In general, of course, trimming by $\mathcal{E}_{i,T,t}$ or $\epsilon_{i,t}$ with re-centering eradicates small sample bias, which is supported by our simulation experiments.

2.4 Tail-Trimmed Serial Correlations

Although we focus on testing for volatility spillover, an obvious application of a robust Q-test is for model specification analysis. For example, for a univariate time series $\{y_t\}$ the resulting GARCH errors $\epsilon_t = y_t/h_t$ are orthogonal if the GARCH model is well specified (e.g. Bollerslev 1986). A robust test of serial correlation in ϵ_t simply uses $\hat{\epsilon}_{T,t}^{*}(\theta) := \epsilon_t(\theta)\hat{I}_{T,t}^{(\epsilon)}(\theta)\hat{I}_{T,t-1}^{(y)}$, its sample mean $\hat{\epsilon}_T^{*}(\theta) := 1/T\sum_{t=1}^{T}\hat{\epsilon}_{T,t}^{*}(\theta)$

and the statistic $\hat{Q}_T^{(\epsilon)}(H) := T\sum_{h=1}^{H}\mathcal{W}_T\left(h\right)(\hat{\rho}_{T,h}^{(\epsilon)}(\hat{\theta}_T))^2$ with correlations

$$\hat{\rho}_{T,h}^{(\epsilon)}(\theta) := \frac{\sum_{t=1}^{T}\left(\hat{\epsilon}_{T,t}^{*}\left(\theta\right)-\hat{\epsilon}_{T}^{*}\left(\theta\right)\right)\left(\hat{\epsilon}_{T,t-h}^{*}\left(\theta\right)-\hat{\epsilon}_{T}^{*}\left(\theta\right)\right)}{\sum_{t=1}^{T}\left(\hat{\epsilon}_{T,t}^{*}\left(\theta\right)-\hat{\epsilon}_{T}^{*}\left(\theta\right)\right)^2}. \tag{4}$$

Now use $T\sum_{h=1}^{H}\mathcal{W}_T\left(h\right)(\hat{\rho}_{T,h}^{(\epsilon)}(\theta))^2$ as the Q-statistic for a robust test of serial correlation.

Table 1: Description of Test Statistics

Test Statistic	Object of Interest	Object Trimmed	Trimming Symmetry	Re-Centering After Trimming
$\hat{W}^{(m)}$	$m_{h,t} = \left(\epsilon_{1,t}^2-1\right)\left(\epsilon_{2,t-h}^2-1\right)$	m	Asymmetric	None
$\hat{W}^{(\mathcal{E})}$	$\mathcal{E}_{i,t} = \epsilon_{i,t}^2-1$	$\mathcal{E}$	Symmetric	Re-center $\mathcal{E}$
$\hat{W}^{(\mathfrak{E})}$	$\mathcal{E}_{i,t} = \epsilon_{i,t}^2-1$	ϵ	Symmetric	Re-center ϵ or ϵ^2
$\hat{Q}^{(\mathcal{E})}$	$\mathcal{E}_{i,t} = \epsilon_{i,t}^2-1$	$\mathcal{E}$	Symmetric	Re-center $\mathcal{E}$
$\hat{Q}^{(\mathfrak{E})}$	$\mathcal{E}_{i,t} = \epsilon_{i,t}^2-1$	ϵ	Symmetric	Re-center ϵ or ϵ^2

3 Asymptotic Theory

We only characterize the limiting properties of $\hat{W}_T^{(m)}$ and $\hat{Q}_T^{(\mathcal{E})}$ since each remaining statistic follows similarly. See Appendix A for assumptions concerning the DGP (D), error memory and moments (E), a trivial fractile lower bound (F), the kernel and bandwidth (K), the plug-in (PQ, PW) and distribution tails (T). All proofs are presented in Appendix B.

3.1 Portmanteau Test Asymptotics

The Q-statistic is asymptotically chi-squared if $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are mutually independent, as long as $\hat{\theta}_T \overset{p}{\to} \theta^0$ sufficiently fast. In order to characterize the rate lower bound for $\hat{\theta}_T$, we require non-random thresholds that are associated with the order statistics used in practice. Although some aspects of the following are treated in HA (2010), a portmanteau statistic has unique properties that leads to sharply different conclusions about permissible estimators.

Let $c_{i,T}^{(\mathcal{E})}(\theta)$ be the two-tailed upper $k_{i,T}^{(\mathcal{E})}/T$ quantile of $\mathcal{E}_{i,t}(\theta)$,

$$P\left(|\mathcal{E}_{i,t}(\theta)| > c_{i,T}^{(\mathcal{E})}(\theta)\right) = \frac{k_{i,T}^{(\mathcal{E})}}{T},$$

and defined trimmed equations

$$\mathcal{E}_{T,i,t}^{*}\left(\theta\right) = \mathcal{E}_{i,t}\left(\theta\right)\times I_{i,T,t}^{(\mathcal{E})}\left(\theta\right) - E\left[\mathcal{E}_{i,t}\left(\theta\right)\times I_{i,T,t}^{(\mathcal{E})}\left(\theta\right)\right] \quad \text{where} \quad I_{i,T,t}^{(\mathcal{E})}\left(\theta\right) := I\left(|\mathcal{E}_{i,t}(\theta)| \leq c_{i,T}^{(\mathcal{E})}(\theta)\right).$$

By construction $\mathcal{E}^{(a)}_{i,(k^{(\mathcal{E})}_{i,T})}(\theta)$ estimates $c^{(\mathcal{E})}_{i,T}(\theta)$, while we are guaranteed the existence of $c^{(\mathcal{E})}_{i,T}(\theta)$ for any $\theta \in \Theta$ and any choice of intermediate order sequence $\{k^{(\mathcal{E})}_{i,T}\}$ since we assume $\epsilon_{i,t}$ have smooth distributions under Assumption E. Now define covariance and Jacobian matrices

$$\mathfrak{S}_T := E\left[\mathcal{E}^{*2}_{1,T,t}\right] \times E\left[\mathcal{E}^{*2}_{2,T,t}\right] \quad \text{and} \quad \mathfrak{J}^{(h)}_{i,T} := \frac{\partial}{\partial \theta_i} E\left[\mathcal{E}^{*}_{1,T,t}\left(\theta_1\right) \mathcal{E}^{*}_{2,T,t-h}\left(\theta_2\right)\right]|_{\theta^0} \in \mathbb{R}^{3\times 1}$$

$$\mathfrak{V}_{i,T} = \max_{1\leq h\leq H}\left\{\frac{T}{\mathfrak{S}_T}\left[\mathbf{1}_3 \times I\left(\left\|\mathfrak{J}^{(h)}_{i,T}\right\| \leq 1\right) + \mathfrak{J}^{(h)\prime}_{i,T} \times I\left(\left\|\mathfrak{J}^{(h)}_{i,T}\right\| > 1\right)\right]\right\} \in \mathbb{R}^{1\times 3},$$

and write

$$\mathfrak{V}_T := \left[\mathfrak{V}_{1,T}, \mathfrak{V}_{2,T}\right] \in \mathbb{R}^{1\times 6}. \tag{5}$$

Note $\mathfrak{V}_{i,T}$ and $\mathfrak{V}_T$ implicitly depend on the horizon H which we hide for notational economy.

We show in Appendix B that $T^{1/2}\hat{\rho}^{(\mathcal{E})}_{T,h}(\hat{\theta}_T)$ satisfies the asymptotic expansion (see the proof of Lemma B.2)

$$\begin{aligned} T^{1/2}\hat{\rho}^{(\mathcal{E})}_{T,h}(\hat{\theta}_T) \overset{p}{\sim} & \frac{1}{T^{1/2}\mathfrak{S}^{1/2}_T}\sum_{t=1}^{T}\left\{\mathcal{E}^{*}_{T,1,t}\mathcal{E}^{*}_{T,2,t-h} - E\left[\mathcal{E}^{*}_{T,1,t}\mathcal{E}^{*}_{T,2,t-h}\right]\right\} \\ & + K\sum_{i=1}^{2}\frac{T^{1/2}}{\mathfrak{S}^{1/2}_T}\mathfrak{J}^{(h)\prime}_{i,T}\left(\hat{\theta}_{i,T} - \theta^0_i\right) + \frac{T^{1/2}}{\mathfrak{S}^{1/2}_T}E\left[\mathcal{E}^{*2}_{T,1,t}\mathcal{E}^{*2}_{T,2,t-h}\right]. \end{aligned} \tag{6}$$

Under the null of mutual independence $E[\mathcal{E}^{*2}_{T,1,t}\mathcal{E}^{*2}_{T,2,t-h}] = 0$ but this implies $\mathfrak{J}^{(h)}_{i,T} = o(1)$ as we show in Appendix B, hence

$$T^{1/2}\hat{\rho}^{(\mathcal{E})}_{T,h}(\hat{\theta}_T) \overset{p}{\sim} \frac{1}{T^{1/2}\mathfrak{S}^{1/2}_T}\sum_{t=1}^{T}\mathcal{E}^{*}_{T,1,t}\mathcal{E}^{*}_{T,2,t-h} + o_p\left(\max_{i\in\{1,2\}}\left|\frac{T^{1/2}}{\mathfrak{S}^{1/2}_T}\left(\hat{\theta}_{i,T} - \theta^0_i\right)\right|\right).$$

Since $\mathfrak{J}^{(h)}_{i,T} = o(1)$ we only need $(T/\mathfrak{S}_T)^{1/2}(\hat{\theta}_{i,T} - \theta^0_i) = O_p(1)$. Notice $\mathfrak{S}_T \to \infty$ if either $E[\epsilon^4_{i,t}] = \infty$ hence $\hat{\theta}_{i,T}$ may be sub-$T^{1/2}$-convergent. If both $E[\epsilon^4_{i,t}] < \infty$, then $T^{1/2}(\hat{\theta}_{i,T} - \theta^0_i) = O_p(1)$ as is conventionally assumed. Under the alternative we can only say $\mathfrak{J}^{(h)}_{i,T} = O(1)$, hence we require $\mathfrak{V}^{1/2}_{i,T}(\hat{\theta}_{i,T} - \theta^0_i) = O_p(1)$ which again reduces to $T^{1/2}(\hat{\theta}_{i,T} - \theta^0_i) = O_p(1)$ when $E[\epsilon^4_{i,t}] < \infty$. As long as Assumption PQ $\mathfrak{V}^{1/2}_T(\hat{\theta}_T - \theta^0) = O_p(1)$ holds, then Gaussian asymptotics follow.

The Q-statistic is asymptotically chi-squared when the plug-in is based on Log-LAD in Peng and Yao (2003), Generalized Method of Tail-Trimmed Moments [GMTTM] in Hill and Renault (2010) with QML-type estimating equations $(\epsilon^2_{i,t}(\theta) - 1)z_{i,t}(\theta)$ where $z_{i,t}(\theta) = [h^{-2}_{i,t-j}(\theta)(\partial/\partial\theta)h^2_{i,t-j}(\theta)]^r_{j=0}$ for some $r \geq 0$[7], Quasi-Maximum Tail-Trimmed Likelihood [QMTTL] in Hill (2011), or Zhu and Ling's (2012)

[7] In the GMTTM case other equation forms or weights $z_{i,t}(\theta)$ can be used for estimation, but the rate of convergence

Quasi-Maximum Weighted Exponential Likelihood (QMWEL).

THEOREM 3.1 (Portmanteau Test under H_0)

a. Let Assumptions D, E, F, PQ, and T hold. If $H_0^{(\epsilon)}$ holds then $\hat{Q}_T^{(\mathcal{E})}(H) \xrightarrow{d} \chi^2(H)$.

b. The following plug-ins are valid[8]: GMTTM and QMTTL in general; QMWEL if $\epsilon_{i,t}$ are symmetric and $E|\epsilon_{i,t}| = 1$; Log-LAD if $\ln(\epsilon_{i,t}^2)$ are symmetric; and QML if both $E[\epsilon_{i,t}^4] < \infty$.

Remark: QML converges too slowly when $E[\epsilon_{i,t}^4] = \infty$ due to feedback with the error term (Hall and Yao 2003). Nevertheless, any $T^{1/2}$-convergent plug-in is valid if $E[\epsilon_{i,t}^4] < \infty$. Other estimators like non-Gaussian QML may also have practical value (e.g. Berkes et al 2003, Horváth and Leise 2004).

3.2 Score Test Asymptotics

The W-statistic is more complicated because trimming asymmetric $m_{h,t}(\theta)$ by itself leads to identification of $H_0^{(m)}$ only asymptotically, while the plug-in's asymptotic properties may impact the test statistic even if $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are independent. The following is developed in HA (2010), cf. Hill and Renault (2010).

As above, define non-random threshold sequences $\{l_{h,T}^{(m)}(\theta), u_{h,T}^{(m)}(\theta)\}$ representing lower and upper tail quantiles.

$$P\left(m_{h,t}(\theta) < -l_{h,T}^{(m)}(\theta)\right) = \frac{k_{1,T}^{(m)}}{T} \quad \text{and} \quad P\left(m_{h,t}(\theta) > u_{h,T}^{(m)}(\theta)\right) = \frac{k_{2,T}^{(m)}}{T},$$

and define deterministically trimmed variables

$$m_{T,h,t}^{*}(\theta) = m_{h,t}(\theta) \times I_{h,T,t}^{(m)}(\theta) \quad \text{where} \quad I_{h,T,t}^{(m)}(\theta) := I\left(-l_{h,T}^{(m)}(\theta) \leq m_{h,t}(\theta) \leq u_{h,T}^{(m)}(\theta)\right).$$

The associated covariance, Jacobian and scale matrices are

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

$$J_T(\theta) := \frac{\partial}{\partial\theta} E\left[m_{T,t}^{*}(\theta)\right] \in \mathbb{R}^{H\times 6}$$

$$V_T(\theta) := T J_T(\theta)' S_T^{-1}(\theta) J_T(\theta) \in \mathbb{R}^{6\times 6}. \quad (7)$$

may differ from that encountered here. See Hill and Renault (2010).

[8] We ignore regularity conditions for Log-LAD and QMWEL concerning density smoothness for the sake of brevity. See Assumptions 1-4 in Peng and Yao (2003) and Asssumptions 2.1-2.6 in Zhu and Ling (2012).

Note V_T implicitly depends on the horizon H through $S_T(\theta) \in \mathbb{R}^{H \times H}$ and $J_T(\theta) \in \mathbb{R}^{H \times 6}$.

The score statistic can be asymptotically decomposed by noting (cf. HA 2010, Hill and Renault 2010)

$$T^{-1/2}\hat{S}_T^{-1/2}(\hat{\theta}_T)\sum_{t=1}^{T}\hat{m}^*_{T,t}(\hat{\theta}_T) \overset{p}{\sim} T^{-1/2}S_T^{-1/2}\sum_{t=1}^{T}m^*_{T,t} + V_T^{1/2}\left(\hat{\theta}_T - \theta^0\right). \tag{8}$$

Since in general $m_{h,t} = (\epsilon^2_{1,t} - 1)(\epsilon^2_{2,t-h} - 1)$ may be asymmetric, we must use asymmetric trimming, hence $E[m^*_{T,t}] = 0$ need not hold under the null although $E[m^*_{T,t}] \to 0$ by dominated convergence. A chi-squared limit for $\hat{W}_T^{(m)}$ under $H_0^{(m^*)}$ therefore requires we add and subtract $E[m^*_{T,t}]$ in (8), hence we must strengthen $H_0^{(m^*)}$ to

$$H_0^{(m^*)} : \left\| T^{1/2}S_T^{-1/2}E[m^*_{T,t}]\right\| \to 0.$$

Tail-trimming implies $||T^{1/2}S_T^{-1/2}|| \to \infty$ by Lemma B.1 in HA (2010), hence $H_0^{(m^*)}$ contains $H_0^{(m)}$ by dominated convergence and implies under no spillover the trimmed mean $E[m^*_{T,t}] \to 0$ sufficiently fast.

Expansion (8) shows $V_T^{1/2}(\hat{\theta}_T - \theta^0) = O_p(1)$ must hold. There are two cases. First, fast convergence Assumption PW.1 $V_T^{1/2}(\hat{\theta}_T - \theta^0) \overset{p}{\to} 0$ applies to heavy tailed cases since $V_T^{1/2} = o(T^{1/2})$ while a variety of plug-ins have a faster rate then $V_T^{1/2}$. Second, if $V_T^{1/2}(\hat{\theta}_T - \theta^0)$ is $O_p(1)$ but not $o_p(1)$ then $\hat{\theta}_T$ impacts $\hat{W}_T^{(m)}$ hence we must assume $\hat{\theta}_T$ is asymptotically normal as in HA (2010), or as in Hill (2012) exploit an orthogonal transform of $\hat{m}^*_{T,t}(\theta)$ that is robust to $\hat{\theta}_T$. We follow the former route for brevity and assume under Assumption PW.2 $\hat{\theta}_{i,T}$ are asymptotically linear in stochastic equations $\tilde{m}_{T,t} \in \mathbb{R}^p$ for $p \geq 6$ that satisfy $\tilde{V}_T^{1/2}(\hat{\theta}_T - \theta^0) \overset{p}{\sim} \tilde{A}_T \sum_{t=1}^{T}\{\tilde{m}_{T,t} - E[\tilde{m}_{T,t}]\} \overset{d}{\to} N(0, I_6)$ where $\tilde{V}_T \in \mathbb{R}^{6\times 6}$, $\tilde{V}_{i,i,T} \to \infty$, and $\tilde{A}_T \in \mathbb{R}^{6 \times p}$. The plug-in is again Log-LAD, Quasi-Maximum Tail-Trimmed Likelihood [QMTTL], Quasi-Maximum Weighted Exponential Likelihood [QMWEL] or Generalized Method of Tail-Trimmed Moments [GMTTM] with QML-type equations.

THEOREM 3.2 (Score Test under H_0)

a. Let Assumptions D, E, F, K and T hold, and assume $H_0^{(m^*)}$. Under fast plug-in PW.1 $\hat{W}_T^{(m)}(H) \overset{d}{\to} \chi^2(H)$, and under slow plug-in PW.2 $\hat{W}_T^{(m)}(H) \overset{d}{\to} \chi^2(p + H - 6)$.

b. The following plug-ins are valid: QMTTL and GMTTM in general; QMWEL if $\epsilon_{i,t}$ are iid, symmetric and $E|\epsilon_{i,t}| = 1$; Log-LAD if $\ln(\epsilon^2_{i,t})$ are symmetric and both $E[\epsilon^4_{i,t}] = \infty$; and QML if both $E[\epsilon^4_{i,t}] < \infty$.

Remark 1: Similar to the Q-statistic case, QML is too slow if $E[\epsilon^4_{i,t}] = \infty$. QMWEL in Zhu and Ling (2012) is only treated for iid errors, but likely extends to semi-strong GARCH (e.g. Linton et al 2010).

Remark 2: Asymptotic linearity for a slow plug-in rules out Log-LAD if $E[\epsilon^4_{i,t}] < \infty$.

Remark 3: If the parameters θ are exactly identified ($p = 6$) then $\hat{W}_T^{(m)}(H) \overset{d}{\to} \chi^2(H)$ even for a slow plug-in as in classic contexts. Cf. Newey and McFadden (1994).

3.3 Asymptotic Power Against Spillover

Expansions (6) and (8) show if there is no spillover then we must have $(T/\mathfrak{S}_T)^{1/2}E[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}] \to 0$ and $T^{1/2}S_T^{-1/2}E[m^*_{h,T,t}] \to 0$ respectively, suggesting a global spillover alternative

$$H_1 : \left|E[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}]\right| \to (0,\infty] \text{ or } H_1 : \left|E[m^*_{h,T,t}]\right| \to (0,\infty] \text{ for some } h \geq 1.$$

Either one allows for moment divergence since under spillover H_1 it is possible to have $E[\epsilon^2_{1,t}\epsilon^2_{2,t-h}] = \infty$ if either $E[\epsilon^4_{i,t}] = \infty$. If both $E[\epsilon^4_{i,t}] < \infty$ then by dominated convergence the alternative is simply $E[\mathcal{E}_{1,t}\mathcal{E}_{2,t-h}] = E[m_{h,t}] \neq 0$.

In general $T/\mathfrak{S}_T \to \infty$ and $||TS_T^{-1}|| \to \infty$ is assured since we use intermediate order fractiles (see HA 2010: Appendix B). This ensures under the global alternative $(T/\mathfrak{S}_T)^{1/2}|E[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}]| \to \infty$ and $||T^{1/2}S_T^{-1/2}E[m^*_{h,T,t}]|| \to \infty$ for some h, promoting test consistency if $h \leq H$.

HA (2010: Theorem 2.2) prove $\hat{W}^{(m)}_T(H) \xrightarrow{p} \infty$ under PW.1 or PW.2 if there is spillover $|E[m^*_{h,T,t}]|$ $\to (0,\infty]$ at some $1 \leq h \leq H$, so consider $\hat{Q}^{(\mathcal{E})}_T(H)$.

THEOREM 3.3 *Under Assumptions D, E, F, PQ, and T*

$$\left(\frac{\mathfrak{S}_T}{T}\right) \times \hat{Q}^{(\mathcal{E})}_T(H) \xrightarrow{p} \sum_{h=1}^{H} \lim_{T\to\infty} \left(E[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}]\right)^2$$

where $\lim_{T\to\infty}(E[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}])^2 = \infty$ *is possible.*

Remark: As long as spillover occurs by horizon H then at least one $\liminf \lim_{T\to\infty} |E[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}]|$ > 0, while $\lim_{T\to\infty} \mathfrak{S}^{(h)}_T/T = 0$. The Q-statistic is therefore consistent if such an $h \leq H$ exists: $\hat{Q}^{(\mathcal{E})}_T(H)$ $\xrightarrow{p} \infty$.

4 Plug-In and Fractile Selection

It remains to decide on a plug-in $\hat{\theta}_T$ and how much trimming to use.

4.1 Plug-In Selection

Define moment suprema $\kappa_i := \arg\inf\{\alpha > 0 : E|\epsilon_{i,t}|^\alpha < \infty\}$ and write compactly

$$\kappa := \min\{\kappa_1, k_2\} \text{ and } k_T = \min\{k_{1,T}, k_{2,T}\}.$$

Since none of the following ideas are sensitive to asymmetry, assume symmetric trimming for ease. In order to gauge which plug-ins are valid, expansions (6) and (8) show we first require rates for $\mathfrak{V}_T$ and

V_T in (5) and (7). The exact rate for $\mathfrak{V}_T$ is complicated by the product structure $E[\mathcal{E}_{1,T,t}^{*2}] \times E[\mathcal{E}_{2,T,t}^{*2}]$.

LEMMA 4.1

a. Let the conditions of Theorem 3.1 hold. If both $\kappa_i > 4$ then $||\mathfrak{V}_T|| \sim KT$, and otherwise $||\mathfrak{V}_T|| = o(T)$. In particular, if $\kappa_1 = 4$ and $\kappa_2 \geq 4$ then $||\mathfrak{V}_T|| \sim T/L(T)$. If $\kappa_1 < 4$ and $\kappa_2 > 4$ then $||\mathfrak{V}_T|| \sim T^{2-4/\kappa_1}(k_{1,T}^{(\mathcal{E})})^{4/\kappa_1 - 1}$. If $\kappa_1 < 4$ and $\kappa_2 = 4$ then $||\mathfrak{V}_T|| \sim T^{2-4/\kappa_1}(k_{1,T}^{(\mathcal{E})})^{4/\kappa_1 - 1}/L(T)$. Finally, if $\kappa_1 < 4$ and $\kappa_2 < 4$ then $||\mathfrak{V}_T|| \sim T^{3-4/\kappa_1-4/\kappa_2}(k_{1,T}^{(\mathcal{E})})^{4/\kappa_1-1}(k_{2,T}^{(\mathcal{E})})^{4/\kappa_2-1}$.

b. Let the conditions of Theorem 3.2 hold. If both $\kappa_i > 4$ then $||V_T|| \sim KT$. Otherwise $||V_T|| \sim KT(k_T^{(m)}/T)^{4/\kappa-1} = o(T)$ if $\kappa \in (2,4)$, and $||V_T|| \sim KT/L(T)$ if $\kappa = 4$.

The next result exploits Lemma 4.1 to deduce valid plug-ins for the test statistics, and forms the basis for Theorems 3.1.b and 3.2.b. See the remarks following those results for comments on valid plug-ins.

LEMMA 4.2

a. Under the conditions of Theorem 3.1 $\hat{Q}_T^{(\mathcal{E})}$ can use $\hat{\theta}_{i,T}$ computed by GMTTM, QMTTL, QMWEL if $\epsilon_{i,t}$ is symmetric and $E|\epsilon_{i,t}| = 1$, Log-LAD if $\ln(\epsilon_{i,t}^2)$ is symmetric, and QML if both $\kappa_i > 4$.

b. Under the conditions of Theorem 3.2 $\hat{W}_T^{(m)}$ can use $\hat{\theta}_{i,T}$ computed by GMTTM, QMTTL, QMWEL if $\epsilon_{i,t}$ are iid, symmetric and $E|\epsilon_{i,t}| = 1$, QML if both $\kappa_i > 4$, and Log-LAD if $\ln(\epsilon_{i,t}^2)$ are symmetric and both $\kappa_i \leq 4$.

4.2 Fractile Selection

Trimming $\mathcal{E}_{i,t}$ or $\epsilon_{i,t}$ with re-centering allows intrinsically easier symmetric trimming, so consider trimming $\mathcal{E}_{i,t}$ in $\hat{Q}_T^{(\mathcal{E})}$ for the sake of discussion. In order to simplify notation we assume the same fractile $k_T = k_{i,T}^{(\mathcal{E})}$ for $i = 1, 2$.

Although Theorem 3.1 shows $\hat{Q}_T^{(\mathcal{E})}$ is robust to any $\mathfrak{V}_T^{1/2}$-convergent $\hat{\theta}_T$ for any trimming amount k_T that forms an intermediate order sequence $\{k_T\}$, we repeatedly find $k_T(\lambda) := [\lambda T/\ln(T)]$ with small $\lambda \in (0, 1]$ is optimal across hypotheses, tail thickness, and sample size. Evidently a *fast* but *small* amount of trimming stabilizes $\hat{Q}_T^{(\mathcal{E})}$ in the presence of heavy tails[9]. See Section 5.

Since any $\lambda \in (0, 1]$ is valid, for score-based moment condition tests HA (2010) and Hill (2012) smooth over $\lambda \in (0, 1]$ by computing p-value occupation time. Let $p_T(\lambda)$ denote the asymptotic p-value for $\hat{Q}_T^{(\mathcal{E})}$: $p_T^{(\mathcal{E})}(\lambda) := P(\hat{Q}_T^{(\mathcal{E})}(H) \leq \chi_H)$ where χ_H is distributed $\chi^2(H)$. The occupation time of $p_T^{(\mathcal{E})}(\lambda)$

[9] *Fast* in the sense $\lambda T/\ln(T) \to \infty$ is faster than $T^\delta \to \infty$ and $\lambda \ln(T) \to \infty$ for any $\lambda \in (0, 1]$ and $\delta \in (0, 1)$. *Small* in the sense that $\lambda T/\ln(T)$ is never above 22% of T for $T \geq 100$, and can be made arbitrary small by diminishing λ.

$\leq \alpha$ on $[\underline{\lambda}, 1]$ for tiny $\underline{\lambda} > 0$ and significance level $\alpha \in (0, 1)$ is

$$\tau_T^{(\mathcal{E})}(\alpha) := \int_{\underline{\lambda}}^1 I(p_T^{(\mathcal{E})}(\lambda) \leq \alpha) d\lambda.$$

The following is proved in Hill (2012: Theorem 3.1).

THEOREM 4.3 *Let Assumptions D, E, PQ, and T hold, let* $k_T(\lambda) := [\lambda T / \ln(T)]$ *and let* $\{u(\lambda) : \lambda \in [0,1]\}$ *be a process with uniform laws* $u(\lambda) \sim U[0,1]$ *and a version that has uniformly continuous sample paths. If* $H_0^{(\epsilon)}$ *is true then* $\tau_T^{(\mathcal{E})}(\alpha) \xrightarrow{d} \int_{\underline{\lambda}}^1 I(u(\lambda) \leq \alpha) d\lambda$*, while if spillover occurs by* $h \leq H$ *then* $\tau_T^{(\mathcal{E})}(\alpha) \xrightarrow{p} 1$.

By construction $\int_{\underline{\lambda}}^1 I(u(\lambda) \leq \alpha) d\lambda \leq \alpha$ with probability one because $u(\lambda)$ is a uniform law on $[0,1]$. Thus, a p-value occupation time test is performed by rejecting the null at α-level if $\tau_T^{(\mathcal{E})}(\alpha) > \alpha$. In practice a discretized version is used, where a simple version is

$$\hat{\tau}_T^{(\mathcal{E})}(\alpha) := \frac{1}{T_{\underline{\lambda}}} \sum_{i=1}^T I\left(p_T^{(\mathcal{E})}(i/T) \leq \alpha\right) I\left(i/T \geq \underline{\lambda}\right) \text{ where } T_{\underline{\lambda}} := \sum_{i=1}^T I\left(i/T \geq \underline{\lambda}\right). \quad (9)$$

5 Simulation Study

We now study the small sample performance of the two Q-statistics $\{\hat{Q}_T^{(\mathcal{E})}, \hat{Q}_T^{(\mathfrak{E})}\}$ and three W-statistics $\{\hat{W}_T^{(m)}, \hat{W}_T^{(\mathcal{E})}, \hat{W}_T^{(\mathfrak{E})}\}$ summarized in Table 1 in Section 2. The data generating process for our simulation work is a bivariate GARCH(1,1):

$$y_{i,t} = h_{i,t}(\theta_i^0)\epsilon_{i,t}, \;\; E\left[\epsilon_{i,t}\right] = 0, \;\; E\left[\epsilon_{i,t}^2\right] = 1,$$

$$h_{i,t}^2\left(\theta^0\right) = .3 + \alpha_{i,i}^0 y_{i,t-1}^2 + \beta_{i,i}^0 h_{i,t-1}^2\left(\theta^0\right) + \alpha_{i,j}^0 y_{j,t-1}^2 + \beta_{i,j}^0 h_{j,t-1}^2\left(\theta^0\right), \; i \neq j = 1, 2$$

We simulate 10,000 samples $\{y_{1,t}, y_{2,t}\}_{t=1}^T$ of size $T \in \{100, 500, 1000\}$[10]. The errors $\epsilon_{i,t}$ are iid $N(0,1)$, or symmetric Pareto denoted P_κ with $\kappa = 2.5$. If $\epsilon_t \sim P_\kappa$ then $P(\epsilon_t < -\epsilon) = P(\epsilon_t > \epsilon) = .5(1+\epsilon)^{-\kappa}$ and ϵ_t is standardized to ensure $\epsilon_t \overset{iid}{\sim} (0,1)$. See Table 2 for descriptions of the various DGP's under the null and alternative hypotheses.

We characterize alternative one ($Alt1$) as "weak" spillover of the volatility of $y_{2,t}$ into the volatility of $y_{1,t}$. Similarly, we characterize alternative two ($Alt2$) as "strong" spillover. The main simulation results are presented in Tables 3 and 4. The entries are rejection frequencies across the 10,000 simulated samples evaluated at the $(1\%, 5\%, 10\%)$ levels. We initially use the true parameter values θ^0 as a benchmark to

[10]We use start values $h_{i,1}^2\left(\theta^0\right) = .3$, draw $2T$ observations, and retain the last T.

Table 2: DGP

	$y_{1,t}$				$\epsilon_{1,t}$	$y_{2,t}$				$\epsilon_{2,t}$
Model	$\alpha^0_{1,1}$	$\beta^0_{1,1}$	$\alpha^0_{1,2}$	$\beta^0_{1,2}$		$\alpha^0_{2,2}$	$\beta^0_{2,2}$	$\alpha^0_{2,1}$	$\beta^0_{2,1}$	
Null-no spill	.3	.6	.0	.0	$N_{0,1}$	.3	.6	.0	.0	$N_{0,1}$
Alt1-weak	.3	.6	.1	.3	$N_{0,1}$	.3	.6	.0	.0	$N_{0,1}$
Alt2-strong	.3	.6	.3	.6	$N_{0,1}$	.3	.6	.0	.0	$N_{0,1}$
Null-no spill	.3	.6	.0	.0	$P_{2.5}$	.3	.6	.0	.0	$P_{2.5}$
Alt1-weak	.3	.6	.1	.3	$P_{2.5}$	.3	.6	.0	.0	$P_{2.5}$
Alt2-strong	.3	.6	.3	.6	$P_{2.5}$	.3	.6	.0	.0	$P_{2.5}$

control for plug-in sampling error. See below for robustness checks, including use of a plug-in. The test statistics are constructed with five lags ($H = 5$). The trimming fractile is identical across test equations with $k_T = [\lambda T/\ln(T)]$, where we use a handpicked $\lambda = .05$ in Table 3 and occupation time (9) over $\lambda \in [.01, 1]$ in Table 4[11]. The bandwidth used in each of the W-tests is $T^{.25}$.

Recall for $\hat{Q}_T^{(\mathfrak{E})}$ and $\hat{W}_T^{(\mathfrak{E})}$ the errors ϵ_t are trimmed based on one of two centering methods called here "a" and "b": under "a" we use $\widehat{\mathfrak{E}}^*_{i,T,t}(\theta) := \hat{\epsilon}^{*2}_{i,T,t}(\theta) - 1/T\sum_{t=1}^T \hat{\epsilon}^{*2}_{i,T,t}(\theta)$, and under "$b$" we use $\widehat{\mathfrak{E}}^*_{i,T,t}(\theta) := ((\hat{\epsilon}_{i,T,t}(\theta) - 1/T\sum_{t=1}^T \hat{\epsilon}^*_{i,T,t}(\theta))^2 - (1/T\sum_{t=1}^T \hat{\epsilon}_{i,T,t}(\theta) - 1/T\sum_{t=1}^T \hat{\epsilon}^*_{i,T,t}(\theta))^2$. The associated statistics are written here as $\{\hat{Q}_T^{(\mathfrak{E}:a)}, \hat{W}_T^{(\mathfrak{E}:a)}\}$ and $\{\hat{Q}_T^{(\mathfrak{E}:b)}, \hat{W}_T^{(\mathfrak{E}:b)}\}$.

Consider Table 3 with fixed $\lambda = .05$. We find that the Q-tests generally are superior to the W-tests when the errors are Paretian with $E[\epsilon^4_{i,t}] = \infty$. Both $\hat{Q}_T^{(\mathcal{E})}$and $\hat{Q}_T^{(\mathfrak{E}:a)}$ have sharp empirical size and ample power. As expected, we see that both size and power improve across all the test statistics as the sample size grows. The exception is $\hat{W}_T^{(\mathfrak{E}:b)}$, which is sharply undersized and exhibits almost no power. In fact, when comparing the re-centering methods, method a is generally superior to b.

Although we do not have a theory-based explanation for this finding it is useful to note the order of merit is first $\hat{Q}_T^{(\mathcal{E})}$, then $\hat{Q}_T^{(\mathfrak{E}:a)}$and finally $\hat{Q}_T^{(\mathfrak{E}:b)}$. The statistic $\hat{Q}_T^{(\mathcal{E})}$ involves the most direct trimming of the three since it works with the object on which spillover is measured $\mathcal{E}_{i,t} := \epsilon^2_{i,t} - 1$, while the other two progressively abstract from $\mathcal{E}_{i,t}$.

Thin-tailed errors, of course, do not require trimming, but trimming Gaussian errors does not appear to impede the viability of the tests. The W-tests perform better when the errors are Gaussian rather than Pareto. The exception is $\hat{W}_T^{(m)}$ which exhibits significant size distortions evidently due to test equation asymmetry and the impossibility of re-centering to control for bias. Empirical size for the Q-tests when the errors are Gaussian is comparable to the size under Pareto errors and power is much increased. In fact, $\hat{Q}_T^{(\mathfrak{E}:b)}$ has improved size (e.g. .115 vs .056 at level .10) and higher power (.754 vs .435) when evaluated under strong spillover.

[11]Extensive simulation experiments suggest that $\lambda = 0.05$ works exceptionally well in *any* randomly drawn sample, and similar small values work as well (e.g. $\lambda \in [.03, .08]$). Smaller λ lead to too little trimming for our sample sizes, and larger λ do not always work well depending on the randomly drawn sample. Smoothing over $\lambda \in [0, 1]$ by occupation time, however, allows the data to reveal which λ work well, as we study below.

Table 3: Rejection frequencies reported at the $(1\%, 5\%, 10\%)$ levels. We use $10,000$ samples, $H = 5$ lags, no plug-in, symmetric trimming unless otherwise noted, and handpicked fractile such that $\lambda = 0.05$.

	$\epsilon_{i,t} \sim$ Gaussian			$\epsilon_{i,t} \sim$ Pareto ($\kappa_i = 2.5$)		
	$T = 100$	$T = 500$	$T = 1000$	$T = 100$	$T = 500$	$T = 1000$
$\hat{Q}_T^{(\mathfrak{E}:a)}(H)$						
Null - no spill	(.036,.094,.151)	(.031,.082,.135)	(.027,.084,.136)	(.088,.130,.158)	(.068,.093,.108)	(.066,.092,.110)
Alt1 - weak	(.106,.207,.280)	(.355,.505,.587)	(.573,.725,.791)	(.162,.227,.263)	(.266,.326,.361)	(.327,.404,.448)
Alt2 - strong	(.127,.239,.315)	(.453,.604,.687)	(.713,.832,.882)	(.186,.250,.291)	(.316,.381,.420)	(.396,.480,.526)
$\hat{Q}_T^{(\mathfrak{E}:b)}(H)$						
Null - no spill	(.079,.121,.152)	(.057,.095,.122)	(.048,.085,.115)	(.085,.107,.118)	(.057,.068,.074)	(.041,.050,.056)
Alt1 - weak	(.208,.281,.327)	(.365,.462,.519)	(.517,.624,.677)	(.202,.240,.260)	(.267,.303,.327)	(.305,.348,.372)
Alt2 - strong	(.241,.315,.362)	(.422,.524,.575)	(.598,.697,.754)	(.233,.272,.298)	(.315,.356,.381)	(.365,.409,.435)
$\hat{Q}_T^{(\mathcal{E})}(H)$						
Null - no spill	(.039,.097,.156)	(.032,.084,.137)	(.028,.085,.136)	(.088,.129,.157)	(.068,.093,.108)	(.066,.093,.110)
Alt1 - weak	(.114,.217,.288)	(.366,.515,.597)	(.588,.740,.803)	(.164,.230,.268)	(.268,.329,.364)	(.331,.407,.454)
Alt2 - strong	(.138,.248,.325)	(.465,.616,.697)	(.732,.843,.891)	(.190,.255,.296)	(.319,.384,.423)	(.401,.486,.533)
$\hat{W}_T^{(\mathfrak{E}:a)}(H)^a$						
Null - no spill	(.010,.054,.109)	(.012,.056,.113)	(.014,.058,.108)	(.000,.008,.027)	(.011,.097,.208)	(.058,.200,.306)
Alt1 - weak	(.008,.042,.093)	(.010,.064,.131)	(.038,.159,.283)	(.001,.007,.022)	(.004,.039,.099)	(.012,.059,.120)
Alt2 - strong	(.008,.042,.093)	(.015,.078,.160)	(.069,.252,.395)	(.001,.007,.023)	(.003,.032,.087)	(.009,.050,.107)
$\hat{W}_T^{(\mathfrak{E}:b)}(H)^a$						
Null - no spill	(.001,.009,.027)	(.005,.038,.092)	(.014,.078,.148)	(.003,.004,.006)	(.000,.000,.003)	(.000,.002,.009)
Alt1 - weak	(.001,.008,.023)	(.001,.013,.040)	(.003,.023,.062)	(.005,.007,.010)	(.001,.002,.003)	(.001,.002,.004)
Alt2 - strong	(.001,.008,.021)	(.001,.012,.038)	(.002,.026,.078)	(.006,.009,.012)	(.001,.002,.004)	(.001,.002,.006)
$\hat{W}_T^{(\mathcal{E})}(H)^a$						
Null - no spill	(.010,.049,.107)	(.011,.055,.111)	(.015,.057,.106)	(.001,.009,.027)	(.010,.092,.202)	(.054,.193,.298)
Alt1 - weak	(.010,.045,.093)	(.013,.074,.148)	(.051,.197,.333)	(.001,.009,.026)	(.003,.030,.077)	(.008,.046,.099)
Alt2 - strong	(.010,.045,.096)	(.018,.098,.189)	(.094,.306,.459)	(.001,.009,.025)	(.003,.024,.070)	(.007,.042,.096)
$\hat{W}_T^{(m)}(H)^{a,b}$						
Null - no spill	(.017,.078,.160)	(.028,.101,.172)	(.051,.147,.233)	(.035,.057,.075)	(.038,.063,.091)	(.055,.095,.146)
Alt1 - weak	(.033,.122,.208)	(.012,.076,.161)	(.060,.270,.444)	(.016,.035,.066)	(.027,.087,.165)	(.080,.219,.337)
Alt2 - strong	(.032,.116,.198)	(.011,.080,.177)	(.090,.360,.570)	(.010,.034,.076)	(.037,.135,.234)	(.133,.310,.432)

a: Bandwidth = $T^{.25}$

b: Due to the asymmetric nature of the test equations, trimming is done with left and right tail indices of 0.03 and 0.01, respectively.

Table 4 contains occupation time test results, which support the viability of this smoothing technique for robust inference. Consider, for instance, $\hat{Q}_T^{(\mathcal{E})}$, one of the tests that offered the best balance between size and power when handpicked $\lambda = .05$ was used. The empirical size at the 10% nominal level for $T = 1000$ is .110 using $\lambda = .05$, versus .107 using occupation time. Similarly, power under strong spillover, albeit slightly weaker, remains strong at .402 as compared to .533 when using a handpicked λ.

We provide several robustness checks for these findings in Appendix C. In Tables 8 and 9, we explore the impact of using a plug-in $\hat{\theta}_{i,T}$ for θ_i^0. In each case we use Hill's (2011) QMTTL estimator with criterion

$$\hat{\mathfrak{Q}}_{i,T}(\theta) := \sum_{t=1}^{T} \left\{ \ln h_{i,t}^2(\theta) + \epsilon_{i,t}^2(\theta) \right\} I\left(|\epsilon_{i,t}(\theta)| \leq \epsilon_{i,(\tilde{k}_T^{(\epsilon)})}^{(a)}(\theta \right).$$

The QMTTLE is $\hat{\theta}_{i,T} := \arg\inf_{\theta \in \Theta} \hat{\mathfrak{Q}}_{i,T}(\theta)$ computed on $\Theta = [0,1]^3$. We choose a fractile $\tilde{k}_T^{(\epsilon)} = [.05T/\ln(T)]$ since this leads to a sharp and approximately normal estimator[12].

We find that using a plug-in has little impact on the tests' ability to correctly detect (non)spillover, regardless of the trimming policy. For instance, when $\lambda = .05$ the statistic $\hat{Q}_T^{(\mathcal{E})}$ exhibits a size of .110 at the 10% nominal level with θ^0, and .121 with the QMTTL plug-in. Similarly, with occupation time $\hat{Q}_T^{(\mathcal{E})}$ exhibits an empirical size of .107 with θ^0, and .104 with the QMTTL plug-in. Note that using a plug-in adds sampling error which clouds the statistic's ability to detect spillover, hence power is somewhat lower.

We also explore robustness to the number of lags H. In Table 10 we report rejection frequencies for $H = 1$, 5, and 10, with $H = 5$ being the base case illustrated in Tables 3 and 4. The left hand panel of the table uses occupation time for fractile choice, while the right hand panel uses $\lambda = .05$. As expected, the probability of rejection increases with the number of lags.

In Table 11 we explore spillover across variables that have errors with different tail thickness. We consider four cases. In Case A $y_{1,t}$ with a fat tailed error spills over into $y_{2,t}$ with an equally fat tailed error, hence $\kappa_1, \kappa_2 = \{2.5, 2.5\}$. In Case B $y_{1,t}$ with a thin tailed error spills over into $y_{2,t}$ with a fat tailed error $\kappa_1, \kappa_2 = \{2.5, \infty\}$. In Case C fat spills into thin $\kappa_1, \kappa_2 = \{\infty, 2.5\}$, and in Case D thin spills into thin $\kappa_1, \kappa_2 = \{\infty, \infty\}$. The sample size is $T = 1000$ and we fix $\lambda = .05$ for brevity. Notice Cases A and D reflect the findings previously displayed in Table 2. We find that power generally is higher when volatility spills over into a process with a thin tailed error. Moreover, it is most challenging to detect spillover from thin to fat tails as seen in Case B, due to the low signal to noise ratio: spillover from $y_{1,t}$ cannot be distinguished very well from the noise caused by fat tails in $\epsilon_{2,t}$.

Lastly, we investigate the efficacy of trimming as tail thickness varies. The entries in Table 12 are rejection frequencies at the (1%, 5%, 10%) levels for the $\hat{Q}_T^{(\mathcal{E})}$test with 5 lags, sample size $T = 1000$,

[12] In general QMTTL has better small sample properties than QML and Log-LAD for the present simulation design, and converges faster than QML. See Hill (2011).

Table 4: Rejection frequencies reported at the $(1\%, 5\%, 10\%)$ levels. We use $10,000$ samples, $H = 5$ lags, no plug-in, symmetric trimming unless otherwise noted, and occupation time for fractile selection.

	$\epsilon_{i,t}$ ∼Gaussian			$\epsilon_{i,t}$ ∼Pareto ($\kappa_i = 2.5$)		
	$T = 100$	$T = 500$	$T = 1000$	$T = 100$	$T = 500$	$T = 1000$
$\hat{Q}_T^{(\mathfrak{E}:\mathfrak{a})}(H)$						
Null - no spill	(.019,.072,.133)	(.013,.057,.109)	(.013,.056,.107)	(.034,.086,.136)	(.019,.061,.107)	(.017,.057,.104)
Alt1 - weak	(.026,.089,.154)	(.040,.104,.166)	(.068,.146,.214)	(.044,.100,.154)	(.038,.091,.144)	(.045,.102,.158)
Alt2 - strong	(.029,.093,.159)	(.051,.121,.186)	(.093,.176,.246)	(.047,.105,.159)	(.046,.102,.157)	(.057,.120,.180)
$\hat{Q}_T^{(\mathfrak{E}:\mathfrak{b})}(H)$						
Null - no spill	(.074,.120,.153)	(.054,.095,.128)	(.046,.086,.121)	(.086,.107,.120)	(.057,.069,.077)	(.042,.051,.058)
Alt1 - weak	(.212,.291,.347)	(.406,.515,.573)	(.583,.696,.751)	(.204,.244,.266)	(.275,.313,.337)	(.316,.358,.384)
Alt2 - strong	(.247,.330,.386)	(.477,.581,.636)	(.677,.778,.825)	(.236,.279,.306)	(.324,.367,.395)	(.374,.422,.450)
$\hat{Q}_T^{(\mathcal{E})}(H)$						
Null - no spill	(.021,.079,.142)	(.015,.063,.117)	(.015,.061,.114)	(.029,.082,.139)	(.020,.062,.111)	(.017,.059,.107)
Alt1 - weak	(.036,.108,.176)	(.070,.167,.248)	(.134,.266,.363)	(.048,.116,.178)	(.068,.153,.225)	(.108,.223,.309)
Alt2 - strong	(.042,.120,.190)	(.097,.211,.301)	(.195,.351,.454)	(.055,.127,.192)	(.093,.194,.275)	(.164,.306,.402)
$\hat{W}_T^{(\mathfrak{E}:\mathfrak{a})}(H)^a$						
Null - no spill	(.018,.072,.133)	(.013,.057,.109)	(.012,.054,.106)	(.009,.048,.102)	(.023,.088,.154)	(.029,.094,.156)
Alt1 - weak	(.016,.069,.129)	(.011,.055,.109)	(.015,.071,.136)	(.008,.048,.101)	(.020,.075,.133)	(.019,.069,.123)
Alt2 - strong	(.016,.068,.129)	(.012,.058,.114)	(.020,.087,.157)	(.008,.047,.100)	(.109,.071,.129)	(.018,.065,.119)
$\hat{W}_T^{(\mathfrak{E}:\mathfrak{a})}(H)^a$						
Null - no spill	(.001,.010,.030)	(.008,.053,.113)	(.019,.086,.157)	(.003,.004,.006)	(.000,.001,.004)	(.000,.003,.012)
Alt1 - weak	(.001,.009,.026)	(.001,.017,.049)	(.005,.038,.097)	(.005,.008,.012)	(.001,.002,.004)	(.001,.002,.006)
Alt2 - strong	(.001,.009,.026)	(.001,.017,.051)	(.007,.055,.132)	(.006,.010,.015)	(.001,.002,.005)	(.001,.003,.008)
$\hat{W}_T^{(\mathcal{E})}(H)$						
Null - no spill	(.021,.075,.134)	(.012,.055,.107)	(.011,.051,.101)	(.016,.062,.117)	(.016,.070,.130)	(.021,.078,.137)
Alt1 - weak	(.022,.076,.136)	(.018,.080,.148)	(.038,.137,.232)	(.104,.058,.109)	(.010,.054,.110)	(.019,.084,.158)
Alt2 - strong	(.023,.078,.139)	(.024,.099,.177)	(.061,.194,.307)	(.014,.057,.107)	(.012,.063,.126)	(.030,.120,.210)

a: Bandwidth = $T^{.25}$

and the true θ^0. We vary error tail thickness as follows: $\kappa_1, \kappa_2 = \{2.5, 2.5\}$, $\kappa_1, \kappa_2 = \{6, 6\}$, and $\kappa_1, \kappa_2 = \{\infty, \infty\}$. Moreover, we explore three trimming strategies: $\lambda = 0$, which implies no trimming and amounts to a non-standardized version of Hong's (2001) test; $\lambda = .05$; and occupation time. We find that trimming, either by handpicking the fractile or using occupation time, dominates no trimming when $\kappa_i < 8$. When κ_1, κ_2 is $\{2.5, 2.5\}$ or $\{6, 6\}$ the untrimmed statistic exhibits large size distortions, while our occupation time test has superior empirical size and competitive power. This demonstrates the advantage of trimming even for only mildly heavy tailed data due to the moment constraints of existing methods.

6 Empirical Application

We now investigate the presence of volatility spillover across five asset classes: US equity, fixed income, real estate, non-US equity, and commodities. Our sample consists of daily log returns from 01/02/2008 through 06/30/2011, resulting in 882 trading days encompassing the height of, and recovery from, the Great Recession.

We part from the typical methodology in this line of literature in two ways. First, we do not investigate spillover within a single asset class, as is commonly the case. Seminal examples concerning equities include Forbes and Rigobon (1999), King et al (1994), or Ng (2000). For fixed income, see Tse and Booth (1996) or Dungey et al (2006), and for foreign exchange see Glick and Rose (1999) or Hong (2001). There are fewer studies of cross-class spillover, including Brooks (1998) who studies stock market volume and volatility; Granger et al (2000) and Yang and Doong (2004), who investigate equity/foreign exchange spillover; Dungey et al (2010) who study equity/bond spillover; and So (2001) who studies bond/foreign exchange spillover. In this paper, we investigate spillover *across* asset classes, which is particularly relevant for investors with broad mandates, such as global macro hedge fund managers. See Fung and Hseih (1999) for a detailed description of various hedge fund styles.

Second, rather than using asset class indices directly, such as the S&P 500, we use Exchange Traded Funds (ETF) as proxies. Specifically, we use the ETF's depicted in Table 5, which are offered by the fund family iShares[13]:

Our choice to use ETF's is motivated by logistical ease. For example, consider the myriad empirical challenges facing the typical study that focuses on international equity indices. Some of the issues the researcher must confront include: i) deciding whether to focus on local currencies or translate into a reference currency, typically the US dollar, ii) accommodating for non-synchronicities of trading times and holiday closures across various geographic regions, iii) accommodating for unequal spacing

[13] An ETF is a fund that holds assets such as stocks, bonds, commodities, or currencies. The ETF trades on an exchange just like an individual stock. It differs from a traditional mutual fund in that it typically is managed to track an index. The ETF's generally are managed within a fund family, which offers several funds of varying styles and mandates.

Table 5: Desciption of Asset Classes

Ticker	Asset Class	Description
IVV	US Equity	S&P 500 Index
AGG	US Fixed Income	Lehman Aggregate Bond Fund
EFA	Non-US Equity	MSCI EAFE Fund
FTY	US Real Estate	FTSE NAREIT Real Estate 50 Index Fund
IAU	Commodities	Gold Trust

of holidays, iv) acknowledging the fact that different exchanges face different rules such as short sales constraints, short-circuit mechanisms, and the like, all of which make direct comparisons difficult.

A conventional method to deal with most of these issues is to use relatively low frequency data, such as weekly. Unfortunately, by lowering the frequency, important aspects of the spillover processes might be overlooked. See, for example, the hourly patterns in volatility found by Baillie and Bollerslev (1990).

By using ETF's we are able to address a majority of the issues listed above. The ETF's we use are all traded on a single exchange (NYSE Arca), and thus are all bound to the same market rules and face the same calendar time and holiday schedules. Moreover, they provide an easy way to investigate asset classes that may not have readily available tradeable indices, as is the case with real estate. In addition, we choose iShares as a fund family since they are the dominant issuer of ETF's with relatively liquid trading in each of the securities[14].

Using ETF's, however, is not a panacea. Generally speaking, ETF's have a limited history as compared to the underlying indices, which restricts the horizon of investigation. In addition, the researcher needs to choose not only the assets to track, but also the fund family that represents those assets in the ETF space. Moreover, there may be aspects of the ETFs' data generating processes that differ slightly from the underlying assets. Our choice of iShares mitigates these concerns as they provide liquid trading over a relatively long time frame[15].

The return histories are depicted in Figure 1 of Appendix C, while Table 6 below details the univariate sample statistics for our asset class returns as well as the unconditional measures of contemporaneous correlation (i.e. lag zero). The assets exhibit the conventional stylized facts, such as non-Gaussian behavior indicated by asymmetries and excess kurtosis, that are well established in the literature. The AGG has the largest negative skew in the sample chosen, while IAU is nearly symmetric. Each of the assets exhibit excess kurtosis, with AGG having an outsized 63.94. Raw returns in equities and real

[14]According to company information, iShares commands a 46% market share of assets under management in the US ETF industry. The SPDR fund family offered by State Street Global Advisors is a substantial competitor, with slightly smaller breadth of available U.S. based funds, 113 versus 228 for iShares, according to author calculations.

[15]For instance, the Vanguard SP500 ETF (ticker VOO) has \$1.64bln of assets under management according to Morningstar.com as of 10/24/2011, yet the iShares SP500 ETF (ticker IVV) has \$26.85bln of assets under management. Similarly, Vanguard offers neither commodity nor real estate related ETF's. Other popular fund families, such as Direxion and Powershares offer ETF's that track specialized indices, which often include leverage, and thus are not suitable for our purposes.

estate, as proxied by the IVV, EFA, and FTY ETF's, appear closely linked contemporaneously. For instance, the correlation at lag zero between IVV and EFA is 0.93 ± 0.01[16]. On the other hand, equities and real estate seem less closely linked with fixed income and commodities. For instance, the correlation between the IVV and IAU is 0.03 ± 0.07.

Table 6: Daily log returns from 1/2/2008 through 6/30/2011 (882 trading days) for five asset classes as represented by their respecitve iShares ETF's.

	Sample Statistics				
	IVV	AGG	EFA	FTY	IAU
Mean	-0.0000	0.0002	-0.0002	0.0001	0.0007
Med	0.0008	0.0003	0.0005	0.0006	0.0011
Std	0.0174	0.0048	0.0213	0.0326	0.0146
Skew	-0.1579	-3.0178	0.1151	0.0927	0.0099
Kurt	9.2322	63.944	10.259	10.200	9.4711
	Contemporaneous Correlations				
	IVV	AGG	EFA	FTY	IAU
IVV	1.0000	-	-	-	-
AGG	-0.0936	1.0000	-	-	-
EFA	0.9272	-0.0038	1.0000	-	-
FTY	0.7122	-0.1111	0.6248	1.0000	-
IAU	0.0306	0.1188	0.1374	-0.0075	1.0000

There is an important caveat: kurtosis, skewness, and even variance in y_t may not exist due to heavy tails. In Figure 1 in Appendix C we plot the Hill (1975) estimator $\hat{\kappa}_T$ of the tail index κ_y for each asset returns series $\{y_t\}_{t=1}^{T}$, with 95% non-parametric bands defined in Hill (2010). The estimator is $\hat{\kappa}_T := (1/\tilde{k}_T \sum_{i=1}^{\tilde{k}_T} \ln(y_{(i)}^{(a)}/y_{(\tilde{k}_T)}^{(a)}))^{-1}$ where $y_t^{(a)} := |y_t|$ and $\tilde{k}_T \in \{5, ...400\}$. The asymptotic bands are $\hat{\kappa}_T \pm 1.95\hat{\upsilon}_T\hat{\kappa}_T^2/\tilde{k}_T^{1/2}$ where

$$\hat{\upsilon}_T^2 = \frac{1}{T}\sum_{s,t=1}^{T} \mathcal{K}_{T,s,t}\left\{\ln\left(\frac{y_s^{(a)}}{y_{(\tilde{k}_T+1)}^{(a)}}\right)_+ - \frac{\tilde{k}_T}{T}\hat{\kappa}_T^{-1}\right\} \times \left\{\ln\left(\frac{y_t^{(a)}}{y_{(\tilde{k}_T+1)}^{(a)}}\right)_+ - \frac{\tilde{k}_T}{T}\hat{\kappa}_T^{-1}\right\} \tag{10}$$

is a kernel estimator of $E(\tilde{k}_T^{1/2}(\hat{\kappa}_T^{-1} - \kappa^{-1}))^2$ with Bartlett kernel $\mathcal{K}_{T,s,t} = (1 - |s - t|/\gamma_T)_+$ and bandwidth[17] $\gamma_T = T^{.225}$. As we can see, kurtosis is unlikely to exist in any of the assets. Moreover, skewness and even variance are in question for AGG, hence population correlation may not exist.

In order to investigate the conditional nature of their relationships and to control for volatility dynamics, we first pass each series though ARMA(p,q)-GARCH$(1,1)$ filters $y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{q} b_i u_{t-i} + u_t$, $p, q \in \{0, ..., 5\}$, where $u_t = h_t\epsilon_t$ and $h_t^2 = \omega + \alpha u_{t-1}^2 + \beta h_{t-1}^2$, with the goal of using the ARMA

[16] 95% confidence bands are computed using Fisher's Z-Transformation.

[17] Simulation evidence not reported here suggests $\gamma_T \in \{T^{.20}, T^{.25}\}$ is optimal for a large variety of linear and nonlinear GARCH processes and sample size sizes T. We simply use the midpoint $\gamma_T = T^{.225}$.

residuals $\hat{u}_t$ for spillover analysis. We use QMTTL for estimation[18], and gauge model adequacy by performing tail-trimmed occupation time Q-tests on the GARCH residuals $\hat{\epsilon}_t = \hat{u}_t/\hat{h}_t$ based on five lags of tail-trimmed serial correlations (4) with trimming fractile $k_T = [\lambda T/\ln(T)]$ and $\lambda \in [.01, 1]$. The chosen filters achieve p-values (.22,.28,.65,.56,.19) for (IVV,AGG,EFA,FTY,IAU), respectively, and therefore reasonably fit the data[19]. The GARCH residuals $\hat{\epsilon}_t$ are depicted in Figure 2 in Appendix C. We also present in Figure 2 Hill-plots of the two-tailed tail index estimator for each asset's $\hat{\epsilon}_t$, with 95% confidence bands computed as above. The estimated tail index for each series is predominantly near or below 4, signaling that $E[\epsilon_{i,t}^4] = \infty$ is plausible, and thereby justifying the use of robust methods.

In Table 7 below we present the p-value's for tests of spillover at horizon $H = 5$ in each of the 20 possible directions across the five ETF's (IVV to AGG, AGG to IVV, etc...). Due to its success in the simulation study, we focus on the Q-statistic $\hat{Q}_T^{(\mathcal{E})}$ with tail trimmed centered errors and a QMTTL plug-in, and use occupation time over $\lambda \in [.01, 1]$[20]. Using a 10% nominal size for our tests, two notable findings surface. First, (domestic and international) equities, fixed income, and real estate are unlikely sources of volatility spillover. Only commodities seem to be a prolific source, with evidence of spillover into IVV, EFA, AGG, and FTY. Second, commodities are the most common destination for spillover. Both IVV and FTY appear to spillover into IAU. Such findings should provide strategic and tactical guidance to fund managers.

In Table 13 in Appendix C we examine spillover by the Q-test $\hat{Q}_T^{(\mathcal{E})}$ using shorter horizons $H = \{1, 2, 3, 4\}$. We find that spillover tends to be more apparent over shorter horizons. For instance, at $H = 5$, notice 6 of the 20 possible directions indicate spillover across the assets. Yet for $H = 1$, we find 14 of the 20 indicate spillover. In addition, IAU appears to be an increasingly common source of, and destination for, spillover as the horizon shortens.

7 Conclusion

We extend available tail-trimming methods in the econometrics literature to tests of volatility spillover. By trimming test equations in three ways we construct five asymptotically chi-squared portmanteau and score statistics. We only require GARCH errors to have a finite variance, and in the score test cases the errors may be martingale differences allowing for volatility spillover in a semi-strong GARCH setting. Our simulation experiments which use iid errors suggest across cases and hypotheses that the

[18] Define parameter sets $\phi = [a', b']'$, $\theta = [\omega, \alpha, \beta]'$, and $\xi = [\phi', \theta']'$, define ARMA errors $u_t(\phi) = y_t - \sum_{i=1}^p a_i y_{t-i} - \sum_{i=1}^q b_i u_{t-i}(\phi)$, and GARCH errors $\epsilon_t(\xi) = u_t(\phi)/h_t(\xi)$ where $h_t^2(\xi) = \omega + \alpha u_{t-1}^2(\phi) + \beta h_{t-1}^2(\xi)$. The QMTTL criterion is $\sum_{t=1}^T \{\ln h_t^2(\xi) + u_t^2(\phi)/h_t^2(\xi)\} \hat{I}_{n,t}^{(\epsilon)}(\xi) \prod_{i=1}^q \hat{I}_{n,t-i}^{(u)}(\phi) \prod_{i=1}^{\max\{p,2\}} \hat{I}_{n,t-i}^{(y)}$, with fractiles $\tilde{k}_T^{(\epsilon)}, \tilde{k}_T^{(u)} = [.05T/\ln(T)]$ and $\tilde{k}_T^{(y)} = [\ln(T)]$ the combination of which elevates the convergence rate to $T^{1/2}/L(T)$ for slowly varying $L(T) \to \infty$. See Hill (2011).

[19] The following specifications accomodate the dynamics of the assets sufficiently: ARMA(2,2)-GARCH(1,1) for IVV,AGG,EFA, and ARMA(1,1)-GARCH(1,1) for FTY,IAU.

[20] Results from the other tests generally support these findings.

Table 7: Table entries are p-value occupation times for tests of volatility spillover based on $\hat{Q}_T^{(\mathcal{E})}(5)$ and a QMTLL plug-in. A value under α implies rejection of the null at level α.

FROM/TO	IVV	AGG	EFA	FTY	IAU
IVV	-	0.93	0.96	1.00	0.08
AGG	1.00	-	0.12	1.00	0.19
EFA	0.16	0.18	-	0.56	0.49
FTY	0.99	0.99	0.26	-	0.00
IAU	0.00	0.05	0.05	0.00	-

portmanteau statistics dominate, trimming by $\epsilon_{i,t}^2 - 1$ or $\epsilon_{i,t}$ with re-centering is optimal, and that using p-value occupation time as a device for transcending the need to choose a sample tail portion to trim results in a sharp Q-test. We show trimming matters even when tails are only mildly heavy since even then extant methods may not result in standard asymptotic inference.

8 Appendix A: Assumptions

We now present all assumptions. Define the σ-fields $\Im_{i,t} := \sigma(y_{i,\tau} : \tau \leq t)$ and $\Im_t = \sigma(\Im_{1,t} \cup \Im_{2,t})$.

ASSUMPTION D (*dgp*): *Each $\{y_{i,t}, \epsilon_{i,t}, \sigma_{i,t}\}$ defined by (1) is strictly stationary, geometrically β-mixing, and L_ι-bounded for tiny $\iota > 0$. The marginal distributions of $\epsilon_{i,t}$ are absolutely continuous with respect to Lebesgue measure, and uniformly bounded: $\sup_{c\in\mathbb{R}} |(\partial/\partial c)P(\epsilon_{i,t} \leq c)| < \infty$.*

ASSUMPTION E (*error memory and moments*): *Let $E[\epsilon_{i,t}] = 0$ and $E[\epsilon_{i,t}^2] = 1$. Further, for the Q-test $\epsilon_{i,t}$ are serially, but not necessarily mutually, independent; and for the W-test $\{\epsilon_{i,t}^2 - 1, \Im_{i,t}\}$ are martingale difference sequences.*

Remarks: Stationarity is assured if $E[\ln(\alpha_i + \beta_i\epsilon_{i,t}^2)] < 0$ allowing integrated and explosive cases $\alpha_i + \beta_i \geq 1$ (Nelson 1990, Bougerol and Picard 1992). Distribution smoothness helps with asymptotic expansions under trimming, while β-mixing expedites uniform asymptotic theory for tail-trimmed random variables with a sample plug-in $\hat{\theta}_T$ (Hill and Renault 2010, Hill 2011). Conditions for geometric ergodicity or the more general geometric β-mixing are well known: see Francq and Zakoïan (2006) for references.

ASSUMPTION T (*tail decay*): *Define the moment supremum $\kappa_i := \arg\inf\{\alpha > 0 : E|\epsilon_{i,t}|^\alpha < \infty\}$. If $\kappa_i \leq 4$ then $\epsilon_{i,t}$ has for each t a common power-law tail*

$$P\left(|\epsilon_{i,t}| > \epsilon\right) = d_i\epsilon^{-\kappa_i}\left(1 + o(1)\right) \text{ where } d_i > 0 \text{ and } \kappa_i \in (2, 4]. \tag{11}$$

Further, if either $E[\epsilon_{i,t}^4] = \infty$ *then for any* $h \geq 1$

$$P\left(\left|\epsilon_{1,t}^2\epsilon_{2,t-h}^2\right| > m\right) = d_m\epsilon^{-\kappa_m}\left(1 + o(1)\right) \text{ where } d_m > 0 \text{ and } \kappa_m = \min\left\{\kappa_1, \kappa_2\right\}/2. \tag{12}$$

Remark 1: If $\epsilon_{1,t}$ is independent of $\epsilon_{2,t}$ then (11) implies (12), cf. Cline (1986).

Remark 2: Paretian tails simplify characterizing trimmed moments by Karamata's Theorem (Resnick 1987).

If $\hat{\theta}_T$ converges too slowly then for the score statistic $\hat{W}_T^{(m)}$ based on trimming $m_{i,t}(\theta)$ by $m_{i,t}(\theta)$ we assume $\hat{\theta}_T$ is asymptotically linear. The required $\Im_t$-measurable estimating equations are $\tilde{m}_{T,t} : \Theta \rightarrow \mathbb{R}^p$ for $p \geq 6$[21]. Define the the total set of equations

$$\mathcal{M}_{T,t}^*(\theta) = \left[m_{T,t}^*(\theta)', \tilde{m}_{T,t}(\theta)'\right]' \in \mathbb{R}^{p+H}$$

and long run variances

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

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

We abuse notation since $\tilde{S}_T(\theta)$ and $\mathfrak{C}_T(\theta)$ may not exist for some or any θ due to heavy tails. Recall $\mathfrak{V}_T = [\mathfrak{V}_{1,T}, \mathfrak{V}_{2,T}] \in \mathbb{R}^{1\times 6}$ and $V_T \in \mathbb{R}^{6\times 6}$ defined in (5) and (7).

ASSUMPTION P (*plug-in*):

PQ *(Plug-in for Q-Test):* $\mathfrak{V}_T^{1/2}(\hat{\theta}_T - \theta^0) = O_p(1)$;

PW *(Plug-in for W-Test):* 1. $V_T^{1/2}(\hat{\theta}_T - \theta^0) = o_p(1)$; *or* 2. $V_T^{1/2}(\hat{\theta}_T - \theta^0) = O_p(1)$ *but not* $o_p(1)$, *where* $\tilde{V}_T^{1/2}(\hat{\theta}_T - \theta^0) = \tilde{A}_T\sum_{t=1}^T\{\tilde{m}_{T,t} - E[\tilde{m}_{T,t}]\} \times (1 + o_p(1)) + o_p(1)$ *for* $\Im_t$*-measurable* $\tilde{m}_{T,t}$, *non-stochastic* $\tilde{A}_T \in \mathbb{R}^{6\times p}$ *and* $\tilde{V}_T \sim \mathcal{K}V_T$ *for some positive definite* $\mathcal{K} \in \mathbb{R}^{6\times 6}$. $\mathfrak{C}_T$ *exists and is positive definite for all* $T \geq N$ *and sufficiently large* $N \geq 1$. $\tilde{A}_T$ *has full column rank and* $\tilde{A}_T\tilde{S}_T^{-1}\tilde{A}_T' \rightarrow I_p$. $\mathcal{M}_{T,t}^*$ *belongs to the same domain of attraction as* $m_{T,t}^*$.

Remark: The Q-statistic only requires $\mathfrak{V}_T^{1/2}$-convergence PQ. Since the W-statistic is sensitive to $V_T^{1/2}$-convergent plug-ins we distinguish fast $\hat{\theta}_T$ under PW.1, and slow but asymptotically linear $\hat{\theta}_T$ under PW.2. An orthogonal projection of $\hat{m}_{T,t}^*(\hat{\theta}_T)$ that is robust to the plug-in can similarly be used, allowing for nonlinear estimators, e.g. Log-LAD. See Hill (2012) for details.

[21] Clearly $\tilde{m}_{T,t}(\theta)$ may depend on other parameters. Such generality is easily allowed in practice but is ignored here to simplify notation.

Next, we set a mild upper bound on the trimming fractiles.

ASSUMPTION F (*fractile bound*): *$k_{i,T}^{(\mathcal{E})} = o(T/L(T))$ and $k_{j,T}^{(m)} = o(T/L(T))$ for some slowly varying $L(T) \to \infty$.*

Finally, the W-statistic requires a kernel function $k(\cdot)$ with bandwidth γ_T.

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

Remark: The assumption covers Bartlett, Parzen, Quadratic Spectral, Tukey-Hanning and other kernels (cf. de Jong and Davidson 2000).

9 Appendix B: Proofs of Main Results

We repeatedly use the following implications of Karamata's Theorem (cf. Resnick 1987: Theorem 0.6). Let a scalar random variable w_t have tail (11) with index $\kappa > 0$, and trimmed version $w_{T,t}^* := w_t I(|w_t| \leq c_T)$, $P(|w_t| > c_T) = k_T/T = o(T)$, and $k_T \to \infty$. Then

$$E\left|w_{T,t}^*\right|^{\kappa} \sim L(T) \to \infty \text{ is slowly varying, } E\left|w_{T,t}^*\right|^{p} \sim c_T^p\left(\frac{k_T}{T}\right) = K\left(\frac{T}{k_T}\right)^{p/\kappa - 1} \quad p > \kappa. \tag{13}$$

The proof of Theorem 3.1 requires two preliminary results. Drop superscripts and write $k_{i,T} = k_{i,T}^{(\mathcal{E})}$, $\hat{\rho}_{T,h}(\theta) = \hat{\rho}_{T,h}^{(\mathcal{E})}(\theta)$, etc. First, we require a self-scaled tail-trimmed LLN for $\mathcal{E}_{i,T,t}^{*2}$ irrespective of tail thickness.

LEMMA B.1 *Under Assumptions D, E and T: a. $1/T \sum_{t=1}^{T} \mathcal{E}_{i,T,t}^{*} \xrightarrow{p} 0$; b. $1/T \sum_{t=1}^{T} \mathcal{E}_{i,T,t}^{*2}/E[\mathcal{E}_{i,T,t}^{*2}] \xrightarrow{p} 1$.*

PROOF Write $\kappa = \kappa_i$. Claim (a) follows from independence, Chebyshev's inequality, and (13): $E(1/T \sum_{t=1}^{T} \mathcal{E}_{i,T,t}^{*})^2 = E[\mathcal{E}_{i,T,t}^2]/T$, where $E[\mathcal{E}_{i,T,t}^2]/T \sim E[\mathcal{E}_{i,t}^2]/T = o(1)$ if $\kappa > 4$, $E[\mathcal{E}_{i,T,t}^2]/T \sim L(T)/T = o(1)$ if $\kappa = 4$, and if $\kappa \in (2, 4)$ then $E[\mathcal{E}_{i,T,t}^2]/T = O((T/k_T)^{2/\kappa}/T) = o(1)$.

Consider claim (b). By Assumption T $\mathcal{E}_{i,t}^2$ has moment supremum $\tilde{\kappa} := \kappa/4$. If $\tilde{\kappa} \geq 1$ then the claim follows as above. Assume $\kappa \in (2, 4)$ hence $\tilde{\kappa} \in (0, 1)$, compactly write $w_{T,t} = \mathcal{E}_{i,T,t}^{*2}$ and $k_T = k_{i,T}$, and note by independence and (13)

$$E\left(\frac{1}{T}\sum_{t=1}^{T}\left\{\frac{w_{T,t}}{E\left[w_{T,t}\right]}-1\right\}\right)^{2}\leq 2\frac{1}{T}\frac{E\left[w_{T,t}^{2}\right]}{\left(E\left[w_{T,t}\right]\right)^{2}}\sim K\frac{1}{T}\frac{(T/k_{T})^{2/\tilde{\kappa}-1}}{(T/k_{T})^{2/\tilde{\kappa}-2}}=K\frac{1}{k_{T}}=o(1). \tag{14}$$

Now use Chebyshev's inequality to complete the proof. $\mathcal{QED}$.

Second, stochastically trimmed $\widehat{\mathcal{E}}^{*}_{1,T,t}(\hat{\theta}_{1,T})\widehat{\mathcal{E}}^{*}_{2,T,t-h}(\hat{\theta}_{2,T})$ is sufficiently close to deterministically trimmed $\mathcal{E}^{*}_{1,T,t}\mathcal{E}^{*}_{2,T,t-h}$, and under the null of mutual independence we may simply treat θ^{0} as though it were known.

LEMMA B.2 (*$\mathcal{E}$-approximation*) *Let Assumptions D, E, PQ and T hold. If $\mathfrak{V}_{T}^{1/2}(\hat{\theta}_{T}-\theta^{0})=O_{p}(1)$, then*

a. $1/T\sum_{t=1}^{T}\{\widehat{\mathcal{E}}^{*2}_{i,T,t}(\hat{\theta}_{i,T})-\mathcal{E}^{*2}_{i,T,t}\}=o_{p}(1)$;

b. $T^{-1/2}(\mathfrak{S}_{T})^{-1/2}\sum_{t=1}^{T}\{\widehat{\mathcal{E}}^{*}_{1,T,t}(\hat{\theta}_{1,T})\widehat{\mathcal{E}}^{*}_{2,T,t-h}(\hat{\theta}_{2,T})-\mathcal{E}^{*}_{1,T,t}\mathcal{E}^{*}_{2,T,t-h}\}=o_{p}(1)$ $\forall h\geq 1$ *if* $\epsilon_{1,t}$ *and* $\epsilon_{2,t}$ *are mutually independent, and otherwise* $O_{p}(1)$.

PROOF We only prove claim (b) in the case of known GARCH effects $\alpha_{i}^{0}+\beta_{i}^{0}>0$ since allowing for no GARCH effects is essentially identical. A nearly identical argument shows $T^{-1/2}(E[\mathcal{E}^{*2}_{i,T,t}])^{-1/2}\sum_{t=1}^{T}\{\widehat{\mathcal{E}}^{*2}_{i,T,t}(\hat{\theta}_{i,T})$ $-\ \mathcal{E}^{*2}_{i,T,t}\}=o_{p}(1)$ where $E[\mathcal{E}^{*2}_{i,T,t}]/T=o(1)$ is shown in the proof of Lemma B.1, hence (a) follows.

Drop the lag superscript "(h)" and rearrange terms to deduce

$$\begin{aligned}
&\frac{1}{T^{1/2}\mathfrak{S}_{T}^{1/2}}\sum_{t=1}^{T}\left\{\widehat{\mathcal{E}}^{*}_{1,T,t}(\hat{\theta}_{1,T})\widehat{\mathcal{E}}^{*}_{2,T,t-h}(\hat{\theta}_{2,T})-\mathcal{E}^{*}_{1,T,t}\mathcal{E}^{*}_{2,T,t-h}\right\}\\
&=\frac{1}{T^{1/2}\mathfrak{S}_{T}^{1/2}}\sum_{t=1}^{T}\left\{\widehat{\mathcal{E}}^{*}_{1,T,t}-\mathcal{E}^{*}_{1,T,t}\right\}\mathcal{E}^{*}_{2,T,t-h}+\frac{1}{T^{1/2}\mathfrak{S}_{T}^{1/2}}\sum_{t=1}^{T}\mathcal{E}^{*}_{1,T,t}\left\{\widehat{\mathcal{E}}^{*}_{2,T,t-h}-\mathcal{E}^{*}_{2,T,t-h}\right\}\\
&\quad+\frac{1}{T^{1/2}\mathfrak{S}_{T}^{1/2}}\sum_{t=1}^{T}\left\{\widehat{\mathcal{E}}^{*}_{1,T,t}-\mathcal{E}^{*}_{1,T,t}\right\}\left\{\widehat{\mathcal{E}}^{*}_{2,T,t-h}-\mathcal{E}^{*}_{2,T,t-h}\right\}\\
&\quad+\frac{1}{T^{1/2}\mathfrak{S}_{T}^{1/2}}\sum_{t=1}^{T}\left\{\widehat{\mathcal{E}}^{*}_{1,T,t}(\hat{\theta}_{1,T})-\widehat{\mathcal{E}}^{*}_{1,T,t}\right\}\mathcal{E}^{*}_{2,T,t-h}\\
&\quad+\frac{1}{T^{1/2}\mathfrak{S}_{T}^{1/2}}\sum_{t=1}^{T}\mathcal{E}^{*}_{1,T,t}\left\{\widehat{\mathcal{E}}^{*}_{2,T,t-h}(\hat{\theta}_{2,T})-\widehat{\mathcal{E}}^{*}_{2,T,t-h}\right\}\\
&\quad+\frac{1}{T^{1/2}\mathfrak{S}_{T}^{1/2}}\sum_{t=1}^{T}\left\{\widehat{\mathcal{E}}^{*}_{1,T,t}(\hat{\theta}_{1,T})-\widehat{\mathcal{E}}^{*}_{1,T,t}\right\}\left\{\widehat{\mathcal{E}}^{*}_{2,T,t-h}-\mathcal{E}^{*}_{2,T,t-h}\right\}\\
&\quad+\frac{1}{T^{1/2}\mathfrak{S}_{T}^{1/2}}\sum_{t=1}^{T}\left\{\widehat{\mathcal{E}}^{*}_{1,T,t}-\mathcal{E}^{*}_{1,T,t}\right\}\left\{\widehat{\mathcal{E}}^{*}_{2,T,t-h}(\hat{\theta}_{2,T})-\widehat{\mathcal{E}}^{*}_{2,T,t-h}\right\}=\sum_{i=1}^{7}\mathcal{A}_{i,T}.
\end{aligned}$$

Approximation theory developed in Hill (2011: Appendix B) can be used to prove $\mathcal{A}_{i,T} = o_p(1)$ for $i = 1, 2, 3$.

It suffices to prove $\mathcal{A}_{4,T} = o_p(1)$ since the remaining terms follow similarly. Define $\mathfrak{J}_{i,t}(\theta_i) := (\partial/\partial\theta_i)\mathcal{E}_{i,t}(\theta_i)$. By the mean-value-theorem there exists $\theta_{i,*}$ that satisfies $||\theta_{i,*} - \theta_i^0|| \leq ||\hat{\theta}_{i,T} - \theta_i^0||$ and

$$\begin{aligned}\mathcal{A}_{4,T} &= \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}\sum_{t=1}^{T}\mathcal{E}_{1,t}\left\{\hat{I}_{1,T,t}^{(\mathcal{E})}(\hat{\theta}_{1,T}) - \hat{I}_{1,T,t}^{(\mathcal{E})}\right\}\mathcal{E}_{2,T,t-h}^{*} \\ &+ \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}\sum_{t=1}^{T}\mathfrak{J}_{1,t}\hat{I}_{1,T,t}^{(\mathcal{E})}\mathcal{E}_{2,T,t-h}^{*} \times \left(\hat{\theta}_{1,T} - \theta_1^0\right) \\ &+ \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}\sum_{t=1}^{T}\mathfrak{J}_{1,t}\left\{\hat{I}_{1,T,t}^{(\mathcal{E})}(\hat{\theta}_{1,T}) - \hat{I}_{1,T,t}^{(\mathcal{E})}\right\}\mathcal{E}_{2,T,t-h}^{*} \times \left(\hat{\theta}_{1,T} - \theta_1^0\right) \\ &+ \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}\sum_{t=1}^{T}\left\{\mathfrak{J}_{1,t}(\theta_{1.*}) - \mathfrak{J}_{1,t}\right\}\hat{I}_{1,T,t}^{(\mathcal{E})}(\hat{\theta}_{1,T})\mathcal{E}_{2,T,t-h}^{*} \times \left(\hat{\theta}_{1,T} - \theta_1^0\right) = \sum_{i=1}^{4}\mathcal{B}_{i,T}\end{aligned} \tag{15}$$

By assumption $\mathfrak{V}_{i,T}^{1/2}(\theta_{i.*} - \theta_i^0) = O_p(1)$ hence arguments in Hill (2011: Proof of Lemmas B.3, B.5, B.7) suffice to prove $\mathcal{B}_{1,T}, \mathcal{B}_{3,T}, \mathcal{B}_{4,T} = o_p(1)$.

Finally, consider $\mathcal{B}_{2,T}$ and note

$$\begin{aligned}\mathcal{B}_{2,T} &= \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}\sum_{t=1}^{T}\left\{\mathfrak{J}_{1,t}I_{1,T,t}^{(\mathcal{E})}\mathcal{E}_{2,T,t-h}^{*} - E\left[\mathfrak{J}_{1,t}I_{1,T,t}^{(\mathcal{E})}\mathcal{E}_{2,T,t-h}^{*}\right]\right\} \times \left(\hat{\theta}_{1,T} - \theta_1^0\right) \\ &+ \frac{T^{1/2}}{\mathfrak{S}_T^{1/2}}E\left[\mathfrak{J}_{1,t}I_{1,T,t}^{(\mathcal{E})}\mathcal{E}_{2,T,t-h}^{*}\right] \times \left(\hat{\theta}_{1,T} - \theta_1^0\right) \\ &+ \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}\sum_{t=1}^{T}\mathfrak{J}_{1,t}\left\{\hat{I}_{1,T,t}^{(\mathcal{E})} - I_{1,T,t}^{(\mathcal{E})}\right\}\mathcal{E}_{2,T,t-h}^{*} \times \left(\hat{\theta}_{1,T} - \theta_1^0\right) = \mathcal{C}_{1,T} + \mathcal{C}_{2,T} + \mathcal{C}_{3,T},\end{aligned}$$

where $\mathcal{C}_{3,T} = o_p(1)$ by the proof of Lemma B.3 in Hill (2011).

For $\mathcal{C}_{2,T}$ write compactly

$$\mathfrak{J}_{1,T,t} := \mathfrak{J}_{1,t}I_{1,T,t}^{(\mathcal{E})}\mathcal{E}_{2,T,t-h}^{*}.$$

By Lemma B.7.b of Hill (2011) the Jacobian of the trimmed mean is proportional to the trimmed mean of the Jacobian:

$$\mathfrak{J}_{1,T} := \frac{\partial}{\partial\theta_1}E\left[\mathcal{E}_{1,T,t}^{*}(\theta_1)\,\mathcal{E}_{2,T,t-h}^{*}(\theta_2)\right]|_{\theta^0} \sim E\left[\mathfrak{J}_{1,t}I_{1,T,t}^{(\mathcal{E})}\mathcal{E}_{2,T,t-h}^{*}\right] =: E\left[\mathfrak{J}_{1,T,t}\right]. \tag{16}$$

Under mutual independence $E[\mathfrak{J}_{1,T,t}] = 0$ given re-centering $E[\mathcal{E}_{2,T,t-h}^{*}] = 0$, hence $\mathcal{C}_{2,T} = o(1)$ given $(T/\mathfrak{S}_T)^{1/2}(\hat{\theta}_{i,T} - \theta_i^0) = O_p(1)$. Otherwise, exploit (16) and the supposition $\mathfrak{V}_{i,T}^{1/2}(\hat{\theta}_{i,T} - \theta_i^0) = O_p(1)$

to deduce

$$|\mathcal{C}_{2,T}| \leq K \left|\mathfrak{V}_{1,T}^{1/2} \times \left(\hat{\theta}_{1,T} - \theta_1^0\right)\right| = O_p(1).$$

Finally, for $\mathcal{C}_{1,T}$ use (16) and $\mathfrak{V}_{i,T}^{1/2}(\hat{\theta}_{i,T} - \theta_i^0) = O_p(1)$ to obtain

$$|\mathcal{C}_{1,T}| \leq K \left| \frac{1}{\max\{1, |E[\mathfrak{J}_{1,T,t}]|\}} \frac{1}{T} \sum_{t=1}^{T} \{\mathfrak{J}_{1,T,t} - E[\mathfrak{J}_{1,T,t}]\} \right| =: \mathbf{J}_T.$$

say. If $\limsup_{T\to\infty} \sup_{1\leq t\leq T} E|\mathfrak{J}_{1,T,t}|^{1+\iota} < \infty$ for some $\iota > 0$ then $\mathfrak{J}_{1,T,t}$ is uniformly integrable hence $\mathbf{J}_T \xrightarrow{p} 0$ by Theorem 2 in Andrews (1988). Otherwise $\mathbf{J}_T \xrightarrow{p} 0$ can be shown by the argument used to prove Lemma B.1.b. $\mathcal{QED}$.

PROOF OF THEOREM 3.1 We prove (a) while (b) follows from Lemma 4.1.a. Invoke the null of mutual independence such that $E[\mathcal{E}_{1,T,t}^{*}\mathcal{E}_{2,T,t-h}^{*}] = 0$. Note $1/T\sum_{t=1}^{T} \mathcal{E}_{i,T,t}^{*2}/E[\mathcal{E}_{i,T,t}^{*2}] \xrightarrow{p} 1$ by Lemma B.1. Therefore, by three applications of Lemma B.2

$$\begin{aligned} T^{1/2}\hat{\rho}_{T,h}(\hat{\theta}_T) &= \frac{1/T^{1/2}\sum_{t=1}^{T} \widehat{\mathcal{E}}_{1,T,t}^{*}(\hat{\theta}_{1,T})\widehat{\mathcal{E}}_{2,T,t-h}^{*}(\hat{\theta}_{2,T})}{\left(1/T\sum_{t=1}^{T} \widehat{\mathcal{E}}_{1,T,t}^{*2}(\hat{\theta}_{1,T})\right)^{1/2} \left(1/T\sum_{t=1}^{T} \widehat{\mathcal{E}}_{2,T,t}^{*2}(\hat{\theta}_{2,T})\right)^{1/2}} \\ &= \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}} \sum_{t=1}^{T} \mathcal{E}_{1,T,t}^{*}\mathcal{E}_{2,T,t-h}^{*} \times (1 + o_p(1)) + o_p(1) =: \quad \mathcal{Z}_{h,T} \times (1 + o_p(1)) + o_p(1), \end{aligned}$$

say, where $\mathfrak{S}_T := E[\mathcal{E}_{1,T,t}^{*2}]E[\mathcal{E}_{2,T,t-h}^{*2}]$.

Pick any $r \in \mathbb{R}^H$, $r'r = 1$, and define $\mathcal{Z}_T = [\mathcal{Z}_{1,T}, ..., \mathcal{Z}_{H,T}]'$. By construction, and serial and mutual independence we have $E[r'\mathcal{Z}_T] = 0$ and $E[(r'\mathcal{Z}_T)^2] = 1$, hence $r'\mathcal{Z}_T$ is a self-standardized tail-trimmed sum of geometrically β-mixing $\{\epsilon_{1,t}^2 - 1, \epsilon_{2,t}^2 - 1\}$. Hill's (2011: Lemma B.6) central limit theorem for intermediate order tail-trimmed random variables therefore applies to prove $r'\mathcal{Z}_T \xrightarrow{d} N(0,1)$. Now invoke Cramér-Wold and continuous mapping theorems, and $\max_{1\leq h\leq H} |\mathcal{W}_T(h) - 1| \xrightarrow{p} 1$, to conclude as claimed

$$\hat{Q}_T(H) = T\sum_{h=1}^{H} \mathcal{W}_T(h)\left(\hat{\rho}_{T,h}(\hat{\theta}_T)\right)^2 = \sum_{h=1}^{H} \mathcal{W}_T(h)\,\mathcal{Z}_{h,T}^2 \times (1 + o_p(1)) + o_p(1) \xrightarrow{d} \chi^2(H).\mathcal{QED}.$$

PROOF OF THEOREM 3.2 Claim (a) follows by Theorem 2.1 of HA (2010). Claim (b) is Lemma 4.2.b. $\mathcal{QED}$.

PROOF OF THEOREM 3.3 We maintain serial independence under Assumption E, and Lemma

B.2 holds under either hypothesis. Therefore, by three applications of Lemma B.2 we have

$$T^{1/2}\hat{\rho}_{T,h}(\hat{\theta}_T) \overset{p}{\sim} \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}} \sum_{t=1}^{T} \left\{\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h} - E\left[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}\right]\right\} + \left(\frac{T}{\mathfrak{S}_T}\right)^{1/2} E\left[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}\right] + O_p(1)$$

$$= \mathcal{Z}_{h,T} + \left(\frac{T}{\mathfrak{S}_T}\right)^{1/2} E\left[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}\right] + O_p(1).$$

As in the proof of Theorem 3.1, $\mathcal{Z}_{h,T} = O_p(1)$. Since $\mathfrak{S}_T = o(T)$ and $\max_{1\leq h\leq H} |\mathcal{W}_T(h) - 1| \overset{P}{\rightarrow} 0$ it therefore follows $(\mathfrak{S}_T/T) \times \hat{Q}_T^{(\mathcal{E})}(H)$ is proportional to

$$\sum_{h=1}^{H} \left(\left(\frac{\mathfrak{S}_T}{T}\right)^{1/2} \left\{\mathcal{Z}_{h,T} + \left(\frac{T}{\mathfrak{S}_T^{(h)}}\right)^{1/2} E\left[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}\right] + O_p(1)\right\}\right)^2$$

$$= \sum_{h=1}^{H} \left(\left\{E\left[\mathcal{E}^*_{1,T,t}\mathcal{E}^*_{2,T,t-h}\right] + o_p(1)\right\}\right)^2.$$

Now take the probability limit to complete the proof. $\mathcal{QED}$.

PROOF OF LEMMA 4.1

Claim a ($\mathfrak{V}_T$): Drop all superscripts and note that we borrow notation from the proof of Lemma B.2. It suffices to consider $\mathfrak{V}_{i,T}$ in (5). Under mutual independence both $\mathfrak{V}_{i,T} = \max_{1\leq h\leq H}\{T/(E[\mathcal{E}^{*2}_{1,T,t}]E[\mathcal{E}^{*2}_{2,T,t-h}])\}$. If $\kappa_l > 4$ for both $l \in \{1,2\}$ then both $\mathfrak{V}_{i,T} \sim KT$, and if $\kappa_l < 4$ then $E[\mathcal{E}^{*2}_{l,T,t}] \sim K(T/k_{l,T})^{4/\kappa_l - 1}$. Therefore, if both $\kappa_l \in (2,4)$ then

$$\mathfrak{V}_{i,T} = K\frac{T}{(T/k_{1,T})^{4/\kappa_1 - 1}(T/k_{2,T})^{4/\kappa_2 - 1}} = K \times T^{3-4/\kappa_1 - 4/\kappa_2} k_{1,T}^{4/\kappa_1 - 1} k_{2,T}^{4/\kappa_2 - 1} = o(T),$$

given $k_{i,T} = o(T)$ and $\kappa_l \in (2,4)$. Similarly, if $\kappa_1 < 4$ and $\kappa_2 > 4$ then

$$\mathfrak{V}_{i,T} \sim K\frac{T}{(T/k_{1,T})^{4/\kappa_1 - 1}} = K \times T^{2-4/\kappa_1} k_{1,T}^{4/\kappa_1 - 1} = o(T).$$

Similar results apply if one or both $\kappa_l \leq 4$ since $E[\mathcal{E}^{*2}_{l,T,t}] \rightarrow \infty$ is slowly varying if $\kappa_l = 4$.

Claim b (V_T): Under Assumptions D and T the gradient of the tail trimmed mean is proportional to the mean tail-trimmed gradient (Hill and Renault 2010: Lemma C.4):

$$J_T = E\left[\frac{\partial}{\partial\theta} m_t(\theta)\,|_{\theta^0} I_{T,t}\left(\theta^0\right)\right] \times (1 + o(1)). \tag{17}$$

If either $E[\epsilon^4_{i,t}] = \infty$ then $m_{h,t} = (\epsilon^2_{1,t} - 1)(\epsilon^2_{2,t-h} - 1)$ has tail (12) with index $\kappa/2 := \min\{\kappa_1, k_2\}/2$

≤ 2 by Assumption T. Assume $\kappa < 4$, the case $\kappa = 4$ being similar. Define $c_{h,T} = \max\{l_{h,T}, u_{h,T}\}$ and $k_T = \min\{k_{1,T}, k_{2,T}\}$, and apply (13) to deduce

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

Therefore under $H_0^{(m^*)}$ since $||TS_T^{-1/2}E[m_{T,t}^*]|| \to 0$ it follows $S_{h,h,T} = E[m_{T,h,t}^{*2}] \times (1 + o(1)) \sim K(T/k_T)^{4/\kappa - 1}$.

Now define $x_{i,t} := [1, y_{i,t-1}^2, h_{i,t-1}^2]' + \beta^0(\partial/\partial\theta)h_{i,t-1}^2(\theta)|_{\theta^0}$, and observe under the null

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

If there are GARCH effects then under the null $J_{h,t}$ is integrable since $\epsilon_{i,t}$ is serially and mutually independent and $E[\epsilon_{i,t}^2] = 1$, cf. Francq and Zakoïan (2004). If there are no GARCH effects then $h_{i,t}^2 = K > 0$, and $y_{i,t-1}^2 = K\epsilon_{i,t-1}^2$ is independent of $\epsilon_{i,t}^2$, hence again $J_{h,t}$ is integrable. Therefore $J_T \sim E[J_t] \times (1 + o(1))$ by (17) and dominated convergence. Combine $S_{i,i,T} \sim KT(T/k_T)^{4/\kappa - 1}$ and $J_T \sim E[J_t] \times (1 + o(1))$ to deduce if $\kappa < 4$ then as claimed $V_{i,i,T} \sim KT(k_T/T)^{4/\kappa - 1} = o(T)$. $\mathcal{QED}$.

PROOF OF LEMMA 4.2 We only prove (b) since (a) is similar in view of Lemma 4.1.a and fractile bound Assumption F. Refer to Lemma 4.1.b for the rate $||V_T|| \to \infty$. If both $E[\epsilon_{i,t}^4] < \infty$ then $||V_T|| \sim KT$.

Log-LAD is $T^{1/2}$-convergent for any $\kappa_i > 2$ if $\ln(\epsilon_{i,t}^2)$ is symmetric, but it is not linear. See Peng and Yao (2003) for the strong-GARCH case and Linton et al (2010) for weak GARCH. Therefore under symmetry of $\ln(\epsilon_{i,t}^2)$ it satisfies PW.1 if $\kappa_i \in (2, 4]$ since $||V_T|| = o(T)$, and does not satisfies PW.2 if $\kappa_i > 4$.

QMWEL $T^{1/2}$-convergent when $\kappa_i > 2$, $\epsilon_{i,t}$ is iid, symmetric and $E|\epsilon_{i,t}| = 1$, and obtains an asymptotic linear expansion. Hence PW.1 if $\kappa_i \leq 4$ and PW.2 if $\kappa_i > 4$ under the stated error properties.

If $\kappa_i \in (2, 4]$ then GMTTM and QMTTL have $T^{1/2}/L(T)$ convergence rates for some slowly varying $L(T) \to \infty$ by following simple trimming rules (cf. Hill and Renault 2010, Hill 2011). Each is therefore valid, in particular under Assumption F we can always choose k_T to satisfy $T^{1/2}(k_T/T)^{2/\kappa - 1/2} < T^{1/2}/L(T)$. Thus Assumption PW.1 holds.

If $E[\epsilon_{i,t}^4] = \infty$ then QML has a rate $T^{1-2/\kappa_i}/L(T)$ (Hall and Yao 2003). But $T^{1-2/\kappa_i}/L(T) > T^{1/2}(k_T/T)^{2/\kappa - 1/2}$ as $T \to \infty$ for any $\kappa_i \in (2, 4)$ since $L(T)$ is slowly varying: $T^{1/2}(k_T/T)^{2/\kappa - 1/2} < T^{1/2} < T^{1-2/\kappa_i}/L(T)$. Therefore if at least one $\kappa_i \in (2, 4)$ then QML is too slow. $\mathcal{QED}$.

10 Appendix C: Simulation & Empirical Results

Table 8: Rejection frequencies reported at the $(1\%, 5\%, 10\%)$ levels. We use $10,000$ samples, $H = 5$ lags, QMTTL plug-in, symmetric trimming unless otherwise noted, amd handpicked fractile such that $\lambda = 0.05$.

	$\epsilon_{i,t}$ ∼Gaussian			$\epsilon_{i,t}$ ∼Pareto ($\kappa_i = 2.5$)		
	$T = 100$	$T = 500$	$T = 1000$	$T = 100$	$T = 500$	$T = 1000$
$\hat{Q}_T^{(\mathfrak{E}:\mathfrak{a})}(H)$						
Null - no spill	(.010,.060,.126)	(.006,.056,.104)	(.011,.054,.104)	(.083,.132,.179)	(.052,.089,.121)	(.035,.079,.117)
Alt1 - weak	(.028,.094,.167)	(.090,.226,.327)	(.200,.402,.327)	(.102,.171,.230)	(.119,.211,.276)	(.156,.249,.330)
Alt2 - strong	(.032,.113,.174)	(.130,.292,.413)	(.314,.559,.413)	(.112,.195,.249)	(.147,.248,.331)	(.219,.323,.391)
$\hat{Q}_T^{(\mathfrak{E}:\mathfrak{b})}(H)$						
Null - no spill	(.063,.109,.146)	(.036,.072,.108)	(.026,.061,.108)	(.088,.107,.116)	(.047,.054,.064)	(.03,.038,.043)
Alt1 - weak	(.174,.248,.304)	(.318,.442,.515)	(.476,.631,.515)	(.167,.199,.217)	(.237,.268,.289)	(.253,.294,.322)
Alt2 - strong	(.204,.278,.333)	(.395,.526,.591)	(.589,.741,.591)	(.189,.223,.245)	(.290,.320,.348)	(.321,.361,.387)
$\hat{Q}_T^{(\mathcal{E})}(H)$						
Null - no spill	(.012,.063,.124)	(.007,.056,.107)	(.012,.053,.107)	(.077,.126,.174)	(.047,.090,.121)	(.034,.082,.121)
Alt1 - weak	(.028,.103,.175)	(.114,.252,.361)	(.243,.468,.361)	(.101,.175,.226)	(.136,.232,.298)	(.164,.267,.349)
Alt2 - strong	(.038,.123,.211)	(.165,.349,.461)	(.398,.648,.461)	(.112,.195,.242)	(.171,.279,.362)	(.238,.347,.434)
$\hat{W}_T^{(\mathfrak{E}:\mathfrak{a})}(H)^a$						
Null - no spill	(.017,.066,.140)	(.011,.063,.108)	(.012,.055,.108)	(.001,.017,.048)	(.022,.106,.215)	(.054,.194,.297)
Alt1 - weak	(.016,.058,.127)	(.015,.084,.159)	(.057,.209,.159)	(.002,.013,.041)	(.004,.043,.111)	(.014,.074,.132)
Alt2 - strong	(.017,.060,.127)	(.024,.108,.221)	(.110,.310,.221)	(.002,.014,.039)	(.004,.042,.091)	(.008,.056,.113)
$\hat{W}_T^{(\mathfrak{E}:\mathfrak{b})}(H)^a$						
Null - no spill	(.004,.015,.033)	(.011,.063,.149)	(.024,.111,.149)	(.002,.005,.007)	(.000,.000,.001)	(.000,.001,.017)
Alt1 - weak	(.001,.010,.028)	(.005,.028,.079)	(.011,.059,.079)	(.005,.008,.012)	(.000,.000,.001)	(.001,.001,.008)
Alt2 - strong	(.003,.016,.034)	(.004,.030,.085)	(.018,.094,.085)	(.005,.011,.015)	(.001,.001,.002)	(.001,.002,.009)
$\hat{W}_T^{(\mathcal{E})}(H)^a$						
Null - no spill	(.018,.059,.140)	(.010,.063,.119)	(.012,.062,.119)	(.001,.019,.053)	(.019,.113,.206)	(.051,.185,.291)
Alt1 - weak	(.014,.058,.121)	(.022,.098,.202)	(.088,.272,.202)	(.002,.013,.040)	(.004,.034,.091)	(.012,.054,.101)
Alt2 - strong	(.013,.067,.125)	(.040,.147,.275)	(.177,.406,.275)	(.001,.014,.044)	(.001,.029,.082)	(.007,.046,.098)
$\hat{W}_T^{(m)}(H)^{a,b}$						
Null - no spill	(.042,.162,.253)	(.035,.117,.177)	(.046,.145,.177)	(.005,.018,.061)	(.011,.067,.162)	(.076,.322,.545)
Alt1 - weak	(.030,.098,.187)	(.094,.260,.418)	(.277,.564,.418)	(.004,.014,.040)	(.006,.028,.089)	(.048,.134,.239)
Alt2 - strong	(.026,.096,.182)	(.152,.393,.560)	(.488,.761,.560)	(.003,.012,.041)	(.007,.042,.100)	(.062,.142,.234)

a: Bandwidth = $T^{.25}$

b: Due to the asymmetric nature of the test equations, trimming is done with left and right tail indices of 0.03 and 0.01, respectively.

Table 9: Rejection frequencies reported at the $(1\%, 5\%, 10\%)$ levels. We use $10,000$ samples, $H = 5$ lags, QMTTL plug-in, symmetric trimming, and occupation time for fractile selection.

	$\epsilon_{i,t}$ ∼Gaussian			$\epsilon_{i,t}$ ∼Pareto ($\kappa_i = 2.5$)		
	$T = 100$	$T = 500$	$T = 1000$	$T = 100$	$T = 500$	$T = 1000$
$\hat{Q}_T^{(\mathfrak{E}:\mathfrak{a})}(H)$						
Null - no spill	(.018,.072,.137)	(.009,.050,.103)	(.011,.052,.103)	(.030,.080,.133)	(.018,.060,.108)	(.015,.051,.096)
Alt1 - weak	(.020,.079,.148)	(.028,.092,.156)	(.049,.129,.156)	(.036,.092,.151)	(.032,.083,.135)	(.034,.085,.139)
Alt2 - strong	(.021,.084,.153)	(.036,.108,.179)	(.073,.164,.179)	(.040,.099,.155)	(.038,.092,.147)	(.044,.101,.159)
$\hat{Q}_T^{(\mathfrak{E}:\mathfrak{b})}(H)$						
Null - no spill	(.057,.100,.131)	(.039,.075,.113)	(.025,.067,.113)	(.087,.110,.123)	(.046,.057,.065)	(.033,.040,.046)
Alt1 - weak	(.177,.269,.328)	(.452,.575,.638)	(.694,.807,.638)	(.172,.209,.229)	(.256,.298,.321)	(.275,.326,.353)
Alt2 - strong	(.212,.311,.376)	(.551,.667,.731)	(.804,.889,.731)	(.198,.237,.262)	(.317,.362,.390)	(.348,.400,.430)
$\hat{Q}_T^{(\mathcal{E})}(H)$						
Null - no spill	(.017,.074,.136)	(.009,.054,.108)	(.012,.054,.108)	(.025,.080,.136)	(.017,.061,.112)	(.015,.056,.104)
Alt1 - weak	(.026,.093,.165)	(.061,.168,.259)	(.122,.270,.259)	(.039,.105,.167)	(.059,.142,.214)	(.079,.183,.270)
Alt2 - strong	(.028,.103,.179)	(.091,.222,.324)	(.195,.373,.324)	(.045,.114,.179)	(.082,.180,.261)	(.127,.262,.360)
$\hat{W}_T^{(\mathfrak{E}:\mathfrak{a})}(H)^a$						
Null - no spill	(.022,.085,.146)	(.013,.058,.110)	(.012,.056,.110)	(.012,.059,.117)	(.025,.089,.156)	(.027,.091,.153)
Alt1 - weak	(.018,.075,.138)	(.013,.061,.118)	(.019,.084,.118)	(.011,.058,.116)	(.021,.076,.135)	(.018,.067,.120)
Alt2 - strong	(.020,.073,.137)	(.014,.067,.130)	(.029,.108,.130)	(.011,.056,.116)	(.018,.070,.127)	(.017,.062,.115)
$\hat{W}_T^{(\mathfrak{E}:\mathfrak{b})}(H)^a$						
Null - no spill	(.002,.014,.038)	(.014,.077,.148)	(.026,.107,.148)	(.001,.003,.006)	(.000,.002,.007)	(.001,.005,.024)
Alt1 - weak	(.002,.013,.030)	(.004,.040,.094)	(.025,.150,.094)	(.004,.008,.012)	(.001,.002,.005)	(.000,.001,.010)
Alt2 - strong	(.002,.015,.033)	(.008,.055,.128)	(.055,.248,.128)	(.004,.010,.015)	(.001,.004,.008)	(.000,.002,.013)
$\hat{W}_T^{(\mathcal{E})}(H)^a$						
Null - no spill	(.023,.086,.146)	(.013,.059,.112)	(.012,.057,.112)	(.018,.069,.129)	(.020,.079,.141)	(.021,.080,.141)
Alt1 - weak	(.026,.086,.150)	(.026,.105,.191)	(.059,.185,.191)	(.015,.063,.119)	(.015,.069,.132)	(.021,.089,.162)
Alt2 - strong	(.026,.089,.156)	(.037,.140,.240)	(.102,.270,.240)	(.015,.062,.117)	(.019,.081,.151)	(.035,.126,.213)

a: Bandwidth = $T^{.25}$

Table 10: Rejection frequencies reported at the $(1\%, 5\%, 10\%)$ levels. We use pareto errors $(\kappa_1 = \kappa_2 = 2.5)$, $10,000$ samples, $H = 5$ lags, QMTTL plug-in, symmetric trimming, and fractile selection is either handpicked such that $\lambda = 0.05$ or occupation time (OT)

	OT			Handpicked		
	Lags = 1	Lags = 5	Lags = 10	Lags = 1	Lags = 5	Lags = 10
$\hat{Q}_T^{(\mathfrak{E}:\mathfrak{a})}(H)$						
Null - no spill	(.012,.049,.096)	(.017,.057,.104)	(.019,.063,.104)	(.028,.045,.057)	(.066,.092,.110)	(.094,.123,.144)
Alt1 - weak	(.020,.062,.111)	(.045,.102,.158)	(.046,.106,.158)	(.099,.140,.169)	(.327,.404,.448)	(.355,.429,.475)
Alt2 - strong	(.026,.073,.124)	(.057,.120,.180)	(.055,.119,.180)	(.146,.206,.247)	(.396,.480,.526)	(.404,.486,.531)
$\hat{Q}_T^{(\mathfrak{E}:\mathfrak{b})}(H)$						
Null - no spill	(.014,.020,.026)	(.042,.051,.058)	(.067,.079,.058)	(.014,.019,.025)	(.041,.050,.056)	(.066,.078,.085)
Alt1 - weak	(.104,.130,.149)	(.316,.358,.384)	(.341,.380,.384)	(.101,.125,.142)	(.305,.348,.372)	(.333,.370,.393)
Alt2 - strong	(.152,.187,.212)	(.374,.422,.450)	(.384,.424,.450)	(.147,.180,.203)	(.365,.409,.435)	(.374,.414,.436)
$\hat{Q}_T^{(\mathcal{E})}(H)$						
Null - no spill	(.013,.053,.102)	(.017,.059,.107)	(.020,.068,.107)	(.028,.044,.056)	(.066,.093,.110)	(.094,.123,.144)
Alt1 - weak	(.034,.096,.157)	(.108,.223,.309)	(.101,.214,.309)	(.099,.142,.172)	(.331,.407,.454)	(.359,.435,.477)
Alt2 - strong	(.059,.146,.220)	(.164,.306,.402)	(.141,.275,.402)	(.149,.210,.252)	(.401,.486,.533)	(.408,.491,.537)
$\hat{W}_T^{(\mathfrak{E}:\mathfrak{a})}(H)^a$						
Null - no spill	(.029,.079,.130)	(.029,.094,.156)	(.023,.087,.156)	(.082,.161,.217)	(.058,.200,.306)	(.022,.132,.241)
Alt1 - weak	(.021,.065,.115)	(.019,.069,.123)	(.018,.071,.123)	(.032,.078,.130)	(.012,.059,.120)	(.009,.058,.122)
Alt2 - strong	(.018,.060,.112)	(.018,.071,.129)	(.018,.071,.129)	(.018,.058,.114)	(.009,.050,.107)	(.011,.060,.131)
$\hat{W}_T^{(\mathfrak{E}:\mathfrak{b})}(H)^a$						
Null - no spill	(.015,.112,.228)	(.000,.003,.012)	(.000,.000,.012)	(.014,.111,.233)	(.000,.002,.009)	(.000,.000,.001)
Alt1 - weak	(.002,.024,.062)	(.001,.002,.004)	(.001,.001,.004)	(.002,.026,.068)	(.001,.002,.004)	(.000,.001,.002)
Alt2 - strong	(.001,.011,.041)	(.001,.003,.008)	(.001,.001,.008)	(.001,.013,.043)	(.001,.002,.006)	(.001,.001,.002)
$\hat{W}_T^{(\mathcal{E})}(q)^a$						
Null - no spill	(.024,.074,.126)	(.021,.078,.137)	(.017,.072,.137)	(.079,.156,.210)	(.054,.193,.298)	(.020,.125,.236)
Alt1 - weak	(.015,.067,.129)	(.019,.069,.123)	(.018,.078,.123)	(.021,.062,.115)	(.008,.046,.099)	(.007,.046,.110)
Alt2 - strong	(.021,.092,.171)	(.018,.065,.119)	(.026,.104,.119)	(.010,.047,.114)	(.007,.042,.096)	(.011,.056,.128)
$\hat{W}_T^{(m)}(H)^{a,b}$						
Null - no spill	-	-	-	(.093,.210,.307)	(.055,.095,.146)	(.059,.084,.109)
Alt1 - weak	-	-	-	(.109,.220,.302)	(.080,.219,.337)	(.074,.213,.346)
Alt2 - strong	-	-	-	(.096,.187,.267)	(.133,.310,.432)	(.202,.439,.588)

a: Bandwidth = $T^{.25}$

b: Due to the asymmetric nature of the test equations, trimming is done with left and right tail indices of 0.03 and 0.01, respectively.

Table 11: Rejection frequencies reported at the $(1\%, 5\%, 10\%)$ levels. We use pareto errors, $10,000$ samples, $H = 5$ lags, no plug-in, symmetric trimming, and a handpicked fractile such that $\lambda = 0.05$. Case A: Heavy tail spills into heavy tail $\kappa_1, \kappa_2 = \{2.5, 2.5\}$. Case B: Thin tail spills into heavy tail $\{\kappa_1, \kappa_2\} = \{2.5, \infty\}$. Case C: Heavy tail spills into thin tail $\{\kappa_1, \kappa_2\} = \{\infty, 2.5\}$. Case D: Thin tail spills into thin tail $\kappa_1, \kappa_2 = \{\infty, \infty\}$

	Case A	Case B	Case C	Case D
	heavy ⇒ heavy	light ⇒ heavy	heavy ⇒ light	light ⇒ light
$\hat{Q}_T^{(\mathfrak{E}:\mathfrak{a})}(H)$				
Null - no spill	(.066,.092,.110)	(.025,.070,.120)	(.022,.066,.114)	(.027,.084,.136)
Alt1 - weak	(.327,.404,.448)	(.174,.306,.401)	(.308,.463,.554)	(.573,.725,.791)
Alt2 - strong	(.396,.480,.526)	(.230,.385,.475)	(.449,.605,.685)	(.713,.832,.882)
$\hat{Q}_T^{(\mathfrak{E}:\mathfrak{b})}(H)$				
Null - no spill	(.041,.050,.056)	(.067,.093,.110)	(.063,.089,.105)	(.048,.085,.115)
Alt1 - weak	(.305,.348,.372)	(.165,.213,.243)	(.896,.925,.937)	(.517,.624,.677)
Alt2 - strong	(.365,.409,.435)	(.187,.240,.272)	(.936,.958,.965)	(.598,.697,.754)
$\hat{Q}_T^{(\mathcal{E})}(H)$				
Null - no spill	(.066,.093,.110)	(.025,.071,.120)	(.023,.067,.115)	(.028,.085,.136)
Alt1 - weak	(.331,.407,.454)	(.196,.334,.426)	(.369,.532,.617)	(.588,.740,.803)
Alt2 - strong	(.401,.486,.533)	(.261,.421,.508)	(.528,.679,.749)	(.732,.843,.891)
$\hat{W}_T^{(\mathfrak{E}:\mathfrak{a})}(H)^a$				
Null - no spill	(.058,.200,.306)	(.021,.087,.155)	(.014,.059,.115)	(.014,.058,.108)
Alt1 - weak	(.012,.059,.120)	(.009,.056,.128)	(.005,.037,.093)	(.038,.159,.283)
Alt2 - strong	(.009,.050,.107)	(.013,.072,.159)	(.007,.059,.141)	(.069,.252,.395)
$\hat{W}_T^{(\mathfrak{E}:\mathfrak{b})}(H)^b$				
Null - no spill	(.000,.002,.009)	(.001,.023,.064)	(.000,.008,.026)	(.014,.078,.148)
Alt1 - weak	(.001,.002,.004)	(.001,.008,.033)	(.001,.002,.010)	(.003,.023,.062)
Alt2 - strong	(.001,.002,.006)	(.000,.007,.030)	(.001,.004,.018)	(.002,.026,.078)
$\hat{W}_T^{(\mathcal{E})}(H)^c$				
Null - no spill	(.054,.193,.298)	(.021,.085,.153)	(.006,.050,.119)	(.015,.057,.106)
Alt1 - weak	(.008,.046,.099)	(.013,.059,.114)	(.015,.084,.181)	(.051,.197,.333)
Alt2 - strong	(.007,.042,.096)	(.009,.063,.138)	(.014,.089,.193)	(.094,.306,.459)
$\hat{W}_T^{(m)}(H)^{a,b}$				
Null - no spill	(.055,.095,.146)	(.037,.140,.244)	(.150,.229,.292)	(.051,.147,.233)
Alt1 - weak	(.080,.219,.337)	(.049,.181,.298)	(.050,.231,.397)	(.060,.270,.444)
Alt2 - strong	(.133,.310,.432)	(.043,.184,.322)	(.180,.250,.309)	(.090,.360,.570)

a: Bandwidth = $T^{.25}$

b: Due to the asymmetric nature of the test equations, trimming is done with left and right tail indices of 0.03 and 0.01, respectively.

Table 12: Rejection frequencies reported at the $(1\%, 5\%, 10\%)$ levels for $\hat{Q}_T^{(\mathcal{E})}(H)$. We use pareto errors ($\kappa_1 = \kappa_2 = 2.5$), $10,000$ samples, $H = 5$ lags, $T = 1000$, no plug-in used, and symmetric trimming.

Trimming	Null - no spill	Alt1 - weak	Alt2 - strong
$\{\kappa_1, \kappa_2\} = \{2.5, 2.5\}$			
$\lambda = 0$	(.044,.053,.059)	(.329,.377,.401)	(.392,.444,.475)
$\lambda = 0.5$	(.066,.093,.110)	(.331,.407,.454)	(.401,.486,.533)
OT	(.017,.059,.107)	(.108,.223,.309)	(.164,.306,.402)
$\{\kappa_1, \kappa_2\} = \{6.0, 6.0\}$			
$\lambda = 0$	(.056,.076,.092)	(.488,.570,.614)	(.575,.652,.697)
$\lambda = 0.5$	(.028,.068,.113)	(.330,.485,.575)	(.449,.607,.691)
OT	(.015,.059,.109)	(.140,.279,.376)	(.213,.381.486)
$\{\kappa_1, \kappa_2\} = \{\infty, \infty\}$			
$\lambda = 0$	(.035,.088,.141)	(.693,.808,.857)	(.693,.808,.857)
$\lambda = 0.5$	(.028,.085,.136)	(.588,.740,.803)	(.732,.843,.891)
OT	(.015,.061,.114)	(.134,.266,.363)	(.195,.351,.454)

Table 13: Table entries are p-value occupation times for tests of volatility spillover based on $\hat{Q}_T^{(\mathcal{E})}(5)$ and a QMTLL plug-in. A value under α implies rejection of the null at level α.

$H = 1$					
FROM/TO	IVV	AGG	EFA	FTY	IAU
IVV	-	0.10	0.02	0.31	0.01
AGG	0.31	-	0.03	0.53	0.03
EFA	0.04	0.03	-	0.09	0.02
FTY	0.63	0.43	0.13	-	0.01
IAU	0.00	0.00	0.00	0.00	-
$H = 2$					
FROM/TO	IVV	AGG	EFA	FTY	IAU
IVV	-	0.27	0.90	0.96	0.01
AGG	0.98	-	0.06	0.93	0.02
EFA	0.25	0.09	-	0.07	0.33
FTY	0.83	0.87	0.14	-	0.00
IAU	0.03	0.01	0.13	0.00	-
$H = 3$					
FROM/TO	IVV	AGG	EFA	FTY	IAU
IVV	-	0.67	0.90	0.99	0.09
AGG	1.00	-	0.09	0.96	0.01
EFA	0.19	0.11	-	0.06	0.50
FTY	0.73	0.96	0.13	-	0.00
IAU	0.00	0.24	0.21	0.00	-
$H = 4$					
FROM/TO	IVV	AGG	EFA	FTY	IAU
IVV	-	0.77	0.89	0.99	0.06
AGG	1.00	-	0.14	0.97	0.05
EFA	0.14	0.18	-	0.10	0.38
FTY	0.62	0.98	0.14	-	0.00
IAU	0.00	0.06	0.11	0.00	-

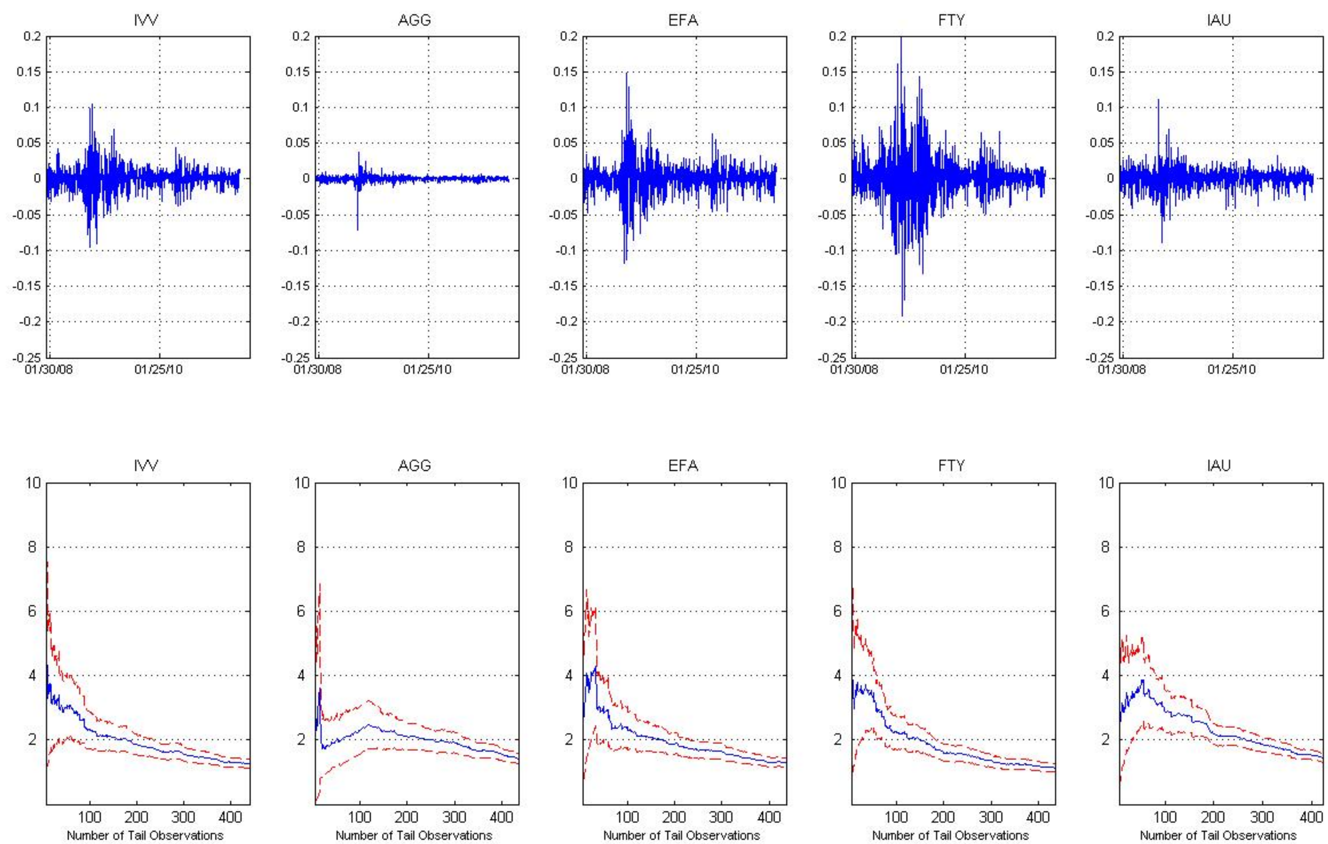

Figure 1: Top Panel: Daily Log Returns: 1/2/08 - 6/30/2011. Bottom Panel: Hill Plots of Daily Log Returns with Robust 95% Confidence Bands.

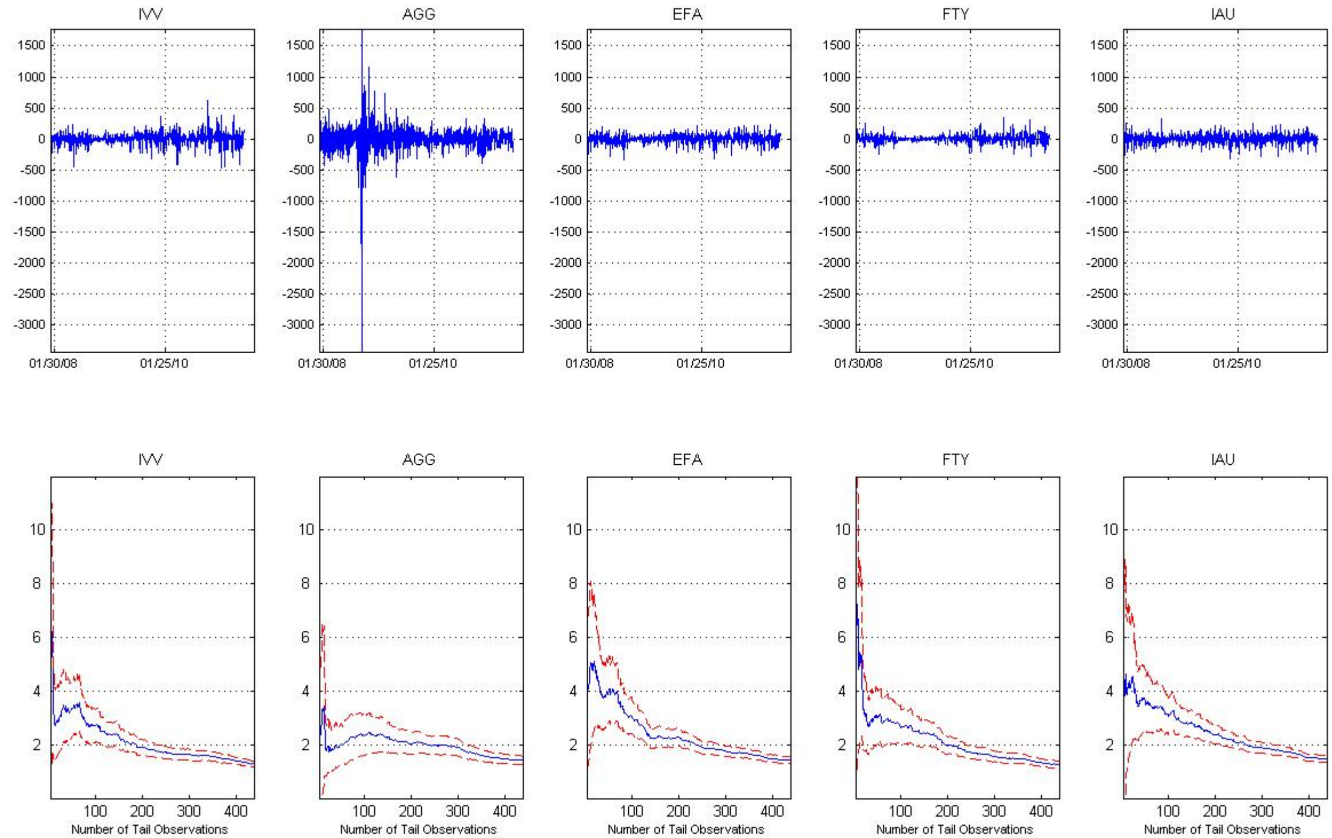

Figure 2: Top Panel: GARCH residuals $\hat{\epsilon}_t = \hat{u}_t/\hat{h}_t$, where $\hat{u}_t$ are the residuals from an ARMA filter and $\hat{h}_t$ are the GARCH conditional variances. Bottom Panel: Hill Plots of residuals $\hat{\epsilon}_t$ with Robust 95% Confidence Bands.

11 References

Andrews, D.W.K. (1988). Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables, Econometric Theory 4, 458-467.

Baillie, R. and T. Bollerslev (1990). Intra-Day and Inter-Market Volatility in Foreign Exchange Markets, Review of Economic Studies 58, 565-585.

Beirne, J., G.M. Caparole, M. Schulz-Ghattas and N. Spagnolo (2008). Volatility Spillovers and Contagion from Mature to Emerging Stock Markets, IMF Working Paper.

Berkes, I., L. Horváth and P. Kokoszka (2003). Garch Processes: Structure and Estimation, Bernoulli 9, 201–227.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity, Journal of Econometrics 31, 307-327.

Bougerol, P. and N. Picard (1992) Stationarity of GARCH Processes and of Some Nonnegative Time Series, Journal of Econometrics 52, 115–127.

Box, G.E.P., and D.A. Pierce (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models, Journal of the American Statistical Association 65, 1509-1526.

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

Brooks, C. (1998). Predicting Stock Index Volatility: Can Market Volume Help? Journal of Forecasting 17, 59-80.

Campbell, J. and L. Hentschel (1992). No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns, Journal of Financial Economics 31, 281-318.

Caporale, G.M, A. Cipollini and N. Spagnolo (2005). Testing for Contagion: A Conditional Correlation Analysis, Journal of Empirical Finance 12, 476-489.

Caporale, G.M., N. Pittis and N. Spagnolo (2006). Volatility Transmission and Financial Crises, Journal of Economics and Finance 30, 376-390.

Cheung, Y-W. and L.K. Ng (1996). A Causality-in Variance Test and its Application to Financial Market Prices, Journal of Econometrics 72, 33-48.

Cline, D.B.H. (1986). Convolution Tails, Product Tails and Domains of Attraction, Probability Theory and Related Fields 72, 529-557.

Comte, F. and O. Lieberman (2000). Second-Order Noncausality in Multivariate GARCH Processes, Journal of Time Series Analysis 21, 535-557.

Davis, R.A. and T. Mikosch (2009). The Extremogram: A Correlogram for Extreme Events, Bernoulli 15, 977–

1009.

Davis, R.A. and S.I. Resnick (1986). Limit Theory for the Sample Covariance and Correlation Function of Moving Averages, Annals of Statistiscs 14, 533-558.

de Lima, P. (1996). Nuisance Parameter Free Properties of Correlation Integral Based Statistics, Econometric Reviews 15, 237-259.

de Lima, P. (1997). On the Robustness of Nonlinearity Tests to Moment Condition Failure, Journal of Econometrics 76, 251-280.

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

Drost, F.C. and T. Nijman (1993). Temporal Aggregation of GARCH Processes, Econometrica 61, 909-927.

Dungey, M., R. Fry, B. González-Hermosillo and V.L. Martin (2005). Sample Properties of Contagion Tests, mimeo, Australian National University.

Dungey, M., R. Fry, B. González-Hermosillo and V.L. Martin (2006). Contagion in International Bond Markets During the Russian and the LTCM Crises, Journal of Financial Stability 2, 1-27.

Dungey, M., R. Fry, B. González-Hermosillo, V.L. Martin and C. Tang (2010). Are Financial Crises Alike?, IMF Working Paper WP/10/14.

Dufour, J.M. and R. Roy (1985). Some Robust Exact Results on Sample Autocorrelations and Tests of Randomness, Journal of Econometrics 29, 257-273.

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

Embrechts, P., C. Klüppleberg and T. Mikosch (1997). Modelling Extremal Events for Insurance and Finance. Springer-Verlag: Berlin.

Engle, R. and V. Ng (1993). Measuring and Testing the Impact of News On Volatility, Journal of Finance 48, 1749-1778.

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

Forbes, K. and R. Rogibon (2002). No Contagion, Only Interdependence: Measuring Stock Market Comovements, Journal of Finance 57, 2223-2261.

Francq, C. and J-M. Zakoian (2004). Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. Bernoulli 10, 605-637.

Francq, C. and J-M. Zakoian (2006). Mixing Properties of a General Class of GARCH(1,1) Models without Moment Assumptions on the Observed Process, Econometric Theory 22, 815-834.

Fung, W. and D. Hsieh (1999). A Primer on Hedge Funds, Journal of Empirical Finance 6, 209-331.

Glick, R. and A.K Rose (1999). Contagion and Trade: Why Are Currency Crises Regional?, Journal of International Money and Finance 18, 603-617.

Granger, C., B. Huang and C. Yang (2000). A Bivariate Causality between Stock Prices and Exchange Rates: Evidence from Recent Asian Flu, Quarterly Review of Economics and Finance 40, 337-354.

Hahn, M.G., J. Kuelbs and D.C. Weiner (1990). The Asymptotic Joint Distribution of Self-Normalized Censored Sums and Sums of Squares, Annals of Probability 18, 1284-1341.

Hahn, M., J. Kuelbs and D. C. Weiner (1991). Sums, Trimmed Sums, and Extremes (Progress in Probability), Birkhäuser.

Hahn, M.G. and D.C. Weiner (1992). Asymptotic Behavior of Self-Normalized Trimmed Sums: Nonnormal Limits, Annals of Probability 20, 455-482.

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

Hill, B.M. (1975). A Simple General Approach to Inference about the Tail of a Distribution, Annals of Mathematical Statistics 3, 1163-1174.

Hill, J.B. (2009). On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays with Applications to Tail Index and Tail Dependence Estimation, Journal of Statistical Planning and Inference 139, 2091-2110.

Hill, J.B. (2010). On Tail Index Estimation for Dependent Heterogeneous Data, Econometric Theory 26, 1398-1436.

Hill, J.B. (2011). Robust M-Estimation for Heavy Tailed Nonlinear AR-GARCH, Working Paper, Dept. of Economics, University of North Carolina.

Hill, J.B. (2012). Heavy-Tail and Plug-In Robust Consistent Conditional Moment Tests of Functional Form, Festschrift in Honor of Hal White: Springer (revised and resubmitted).

Hill, J.B. and M. Aguilar (2010). Moment Condition Tests for Heavy Tailed Time Series, Journal of Econometrics: Annals Issue on Extreme Value Theory: forthcoming.

Hill, J.B. and E. Renault (2010). Generalized Method of Moments with Tail Trimming, Dept. of Economics, University of North Carolina.

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

Hong, Y. (2001). A Test for Volatility Spillover with Application to Exchange Rates, Journal of Econometrics 103, 183-224.

King, M., E. Sentana and S. Wadhwani (1994). Volatility and Links between National Stock Markets, Econometrica 62, 901-933.

Karolyi, G.A. and R.M. Stulz (1996). Why Do Markets Move Together? An Investigation of U.S.-Japan Stock Return Comovements, Journal of Finance 51, 951-986.

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

Linton, O., J. Pan and H. Wang (2010). Estimation for a Non-Stationary Semi-Strong GARCH(1,1) Model with Heavy-Tailed Errors, Econometric Theory: forthcoming.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets, Journal of Finance 56, 649–676.

Malevergne, Y. and D. Sornette (2004). How to Account for Extreme Co-Movements Between Individual Stocks and the Market, Journal of Risk 6, 71–116.

Meddahi, N. and E. Renault (2004). Temporal Aggregation of Volatility Models, Journal of Econometrics 119, 355-379.

Nelson, D.B. (1990). Stationarity and Persistence in the GARCH(1,1) Model, Econometric Theory 6, 318–334.

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

Ng, A. (2000). Volatility Spillover Effects from Japan and the US to the Pacific-Basin, Journal of International Money and Finance 19, 207-233.

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

Poon, S-H., M. Rickinger and J. Tawn (2003). Modelling Extreme-Value Dependence in International Stock Markets, Statistica Sinica 13, 929-953.

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

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

Schmidt R. and U. Stadtmüller (2006). Non-Parametric Estimation of Tail Dependence, Scandinavian Journal of Statistics 33, 307-335.

So, R. (2001). Price and Volatility Spillovers Between Interest Rate and Exchange Value of the US Dollar, Global Finance Journal 12, 95-107.

Tse, Y. and G.G. Booth (1996). Common Volatility and Volatility Spillovers between U.S. and Eurodollar Interest Rates: Evidence from the Futures Market, Journal of Economics and Business 48, 299-312.

Yang, S. and S. Doong (2004). Price and Volatility Spillovers between Stock Prices and Exchange Rates: Empirical Evidence from the G-7 Countries, International Journal of Business and Economics 3, 139-153.

Zhu, K. and S. Ling (2012). Global Self-Weighted and Local Quasi-Maximum Exponential Likelihood Estimators for ARMA-GARCH/IGARCH Models, Annals of Statistics: forthcoming.