Journal of Econometrics 184 (2015) 37–61

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Robust score and portmanteau tests of volatility spillover

Mike Aguilar, Jonathan B. Hill *

Department of Economics, University of North Carolina at Chapel Hill, United States

ARTICLE INFO

Article history:
Received 27 March 2012
Received in revised form
1 September 2014
Accepted 1 September 2014
Available online 16 September 2014

JEL classification:
C13
C20
C22

Keywords:
Volatility spillover
Heavy tails
Tail trimming
Robust inference

ABSTRACT

This paper presents a variety of tests of volatility spillover that are robust to heavy tails generated by large errors or GARCH-type feedback. The tests are couched in a general conditional heteroskedasticity framework 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. Trimming is either simple based on an indicator function, or smoothed. In particular, 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 statistics obtain power of one asymptotically. We discuss the choice of trimming portion, including a smoothed p-value over a window of extreme observations. 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. Finally, based on ideas in Patton (2011) we construct a heavy tail robust forecast improvement statistic, which allows us to demonstrate that our spillover test can be used as a model specification pre-test to improve volatility forecasting.

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 Rogibon, 2002; Caporale et al., 2005, 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. (1999); Longin and Solnik (2001), Finkenstadt and Rootzen (2003), and Poon et al. (2003).

The ability to detect volatility spillovers among asset prices has myriad uses in macroeconomics and finance. For policy makers, knowledge of spillovers may inform policy design (King et al., 1994; Forbes and Rogibon, 2002). For investors, knowledge of spillovers may lead to improved volatility forecasts, which can be embedded inside asset pricing models. Similarly, spillovers might capture information transmission, as per Engle et al. (1990), or the spillover effects can be used to design conditional hedge ratios (Chang et al., 2011).

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

1.1. Proposed methods

In this paper, rather than look for new dependence measures, we exploit robust methods that allow for the use of existing representations of so-called volatility *spillover* or *contagion*[1] where

[1] There is some consensus in the applied literature on the 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 view 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.

* Corresponding author.
E-mail addresses: maguilar@email.unc.edu (M. Aguilar), jbhill@email.unc.edu (J.B. Hill).
URL: http://www.unc.edu/~jbhill (J.B. Hill).

http://dx.doi.org/10.1016/j.jeconom.2014.09.001
0304-4076/

idiosyncratic shocks may be heavy tailed. We use a general model of conditional heteroskedasticity, and deliver test statistics with standard limit distributions under mild regularity conditions.

Let $\{y_{1,t}, y_{2,t}\}$ be a joint process of interest with conditionally heteroskedastic coordinates:

$$y_{i,t} = h_{i,t}(\theta_i)\epsilon_{i,t}(\theta_i) \quad \text{where } \theta_i \in \mathbb{R}^q \; q \geq 1. \tag{1}$$

We assume there exists a unique point θ_i^0 in the interior of a compact subset $\Theta \subset \mathbb{R}^q$ such that $\{y_{i,t}, h_{i,t}(\theta_i^0)\}$ is stationary and ergodic, and $E[\epsilon_{i,t}(\theta_i^0)] = 0$ and $E[\epsilon_{i,t}^2(\theta_i^0)] = 1$. Now drop θ_i^0 and write $h_{i,t} = h_{i,t}(\theta_i^0)$ and $\epsilon_{i,t} = \epsilon_{i,t}(\theta_i^0)$. Volatility $h_{i,t}(\theta_i)$ is measurable with respect to $\{y_{i,t-1}, y_{i,t-2}, \ldots\}$, continuous and differentiable on Θ, and bounded $\inf_{\theta\in\Theta}\{h_{i,t}(\theta_i)\} > 0$ *a.s.* An example of (1) is nonlinear GARCH(1, 1) $h_{i,t}^2 = g(y_{i,t}, h_{i,t-1}^2, \theta_i^0)$ where $g(\cdot, \cdot, \theta_i)$ is continuous (see Francq and Zakoïan, 2010). We restrict attention to models where random volatility $h_{i,t}^2$ satisfies

$$E\left(\sup_{\theta_i\in\mathcal{N}_0}\left\|\frac{\partial}{\partial\theta_i}\ln h_{i,t}^2(\theta_i)\right\|\right)^2 < \infty$$

$$\text{on some compact subset } \mathcal{N}_0 \subseteq \Theta \text{ containing } \theta_i^0. \tag{2}$$

Condition (2) simplifies technical arguments, but it can be relaxed at the expense of lengthier proofs. Since we want to allow for heavy tailed $\epsilon_{i,t}$, notice (2) in general implies $h_{i,t}^2$ is stochastic, since otherwise for many models $(\partial/\partial\theta)\ln h_{i,t}^2(\theta)|_{\theta=\theta_i^0}$ is square integrable only if $E[\epsilon_{i,t}^4] < \infty$ (Francq and Zakoïan, 2004, 2010). This allows us to avoid boundary issues for estimating θ_i^0 (for the GARCH case, see Andrews (2001)). In the standard GARCH model $h_{i,t}^2 = \omega_i^0 + \alpha_i^0 y_{i,t-1}^2 + \beta_i^0 h_{i,t-1}^2$ with $\omega_i^0 > 0$, for example, if $\alpha_i^0 + \beta_i^0 > 0$ then (2) holds (cf. Francq and Zakoïan, 2004), while in general (2) covers linear, Quadratic, GJR, Smooth Transition, Threshold, and Asymmetric GARCH, to name a few. See Engle and Ng (1993); Glosten et al. (1993); Sentana (1995) and Francq and Zakoïan (2010). It is only a matter of notation to allow even greater model generality, including nonlinear ARMA–GARCH and other volatility models (e.g. Meddahi and Renault, 2004).

Cheung and Ng (1996) and Hong (2001) work with a linear GARCH model $h_{i,t}^2 = \omega_i^0 + \alpha_i^0 y_{i,t-1}^2 + \beta_i^0 h_{i,t-1}^2$ and argue volatility spillover reduces to testing whether $y_{1,t}^2/h_{1,t}^2 - 1$ and $y_{2,t-h}^2/h_{2,t-h}^2 - 1$ are correlated for some lag $h \geq 1$, 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 not unique to these works since volatility spillover and contagion methods are typically designed under substantial moment conditions. Forbes and Rogibon (2002) implicitly require VAR errors to have a fourth moment; Caporale 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, in applications there appears to be little in the way of robustness checks, or pre-tests to verify the required moment conditions. See especially de Lima (1997) and Hill and Aguilar (2013). Dungey et al. (2005), for example, study an array of sampling properties of tests of contagion and spillover, but do not treat heavy tails. In Section 6, however, we show a variety of asset return series have conditionally heteroskedastic 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). We construct centered squared errors from the volatility function $h_{i,t}(\theta_i)$,

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

and build test equations over H lags:

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

Under the null of no spillover we have $E[m_{h,t}] = E[(\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-h}^2 - 1)] = 0$. 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 shocks $\epsilon_{i,t}$ by transforming $m_t(\theta)$, $\mathcal{E}_{i,t}(\theta_i)$ or $\epsilon_{i,t}(\theta_i)$. First, since $m_{h,t}(\theta)$ is asymmetrically distributed about zero in general, we need an asymmetric transform to ensure both identification of the hypotheses and a standard distribution limit (cf. Hill, 2012, 2014a; Hill and Aguilar, 2013). We therefore focus on tail-trimming $m_{h,t}(\theta)I(-l \leq m_{h,t}(\theta) \leq u)$ for a robust score test, where $I(\cdot)$ is the indicator function, l and u are positive thresholds, and $l, u \to \infty$ as the sample size $T \to \infty$. In general, this does not allow a portmanteau statistic even if $\epsilon_{i,t}$ are iid, and may still lead to small sample bias that arises from trimming. Further, if l and u are fixed asymptotically then, in general, asymptotic bias in the test statistic prevents a score statistic from detecting spillover. By negligible trimming, however, we can obtain both an asymptotic chi-squared distribution under the null and correctly identify spillover. In principle other transformations can be used, including those discussed below for our portmanteau tests, but the need for asymmetry and negligibility makes tail-trimming an appealing and practical choice.

Our second and third approaches transform $\mathcal{E}_{i,t}(\theta_i)$ and $\epsilon_{i,t}(\theta_i)$, respectively, leading to robust score *and* portmanteau statistics. Small sample bias is eradicated by recentering the transformed variables. We use a class of bounded transformations $\psi : \mathbb{R} \times [0, \infty) \to \mathbb{R}$, $|\psi(u, c)| \leq c$, including the so-called *redescending* functions, which generate decreasing or vanishing values far from a threshold c, e.g. $\psi(u, c) = 0$ if $|u| > c$. We say ψ is *symmetric* if $\psi(-u, c) = -\psi(u, c)$, and we say the transformation ψ or threshold c is *negligible* when $\lim_{c\to\infty}\psi(u, c) = u$ such that there is no transformation asymptotically. We assume symmetry and negligibility throughout.

Redescenders are popularly used in the outlier robust estimation literature where an extreme value is considered aberrant. See Andrews et al. (1972), Hampel et al. (1986), and Jureckova and Sen (1996) for classic treatments, and for use in M-estimation see Kent and Tyler (1991); Shevlyakov et al. (2008) and Hill (2013b, 2014a). Examples of popularly used symmetric transforms ψ are *simple trimming* $uI(|u| \leq c)$, *Tukey's bisquare* $u(1 - (u/c)^2)^2 I(|u| \leq c)$, *exponential* $u\exp\{-|u|/c\}I(|u| \leq c)$, and *truncation sign*$\{u\}\min\{|u|, c\}$. Notice only the first three are redescenders.

By recentering $\psi(\mathcal{E}_{i,t}(\theta_i), c) - E[\psi(\mathcal{E}_{i,t}(\theta_i), c)]$ or $\psi(\epsilon_{i,t}^2(\theta_i), c) - E[\psi(\epsilon_{i,t}^2(\theta_i), c)]$ we can always use a symmetric transformation which is intrinsically easier to implement. Moreover, if $\epsilon_{1,t}$ and $\epsilon_{2,t-h}$ are independent under the null then ψ does not need to be redescending, nor even negligible in the sense that c may be bounded, since the null hypothesis is identified with any bounded ψ or any c. This allows for great generality in terms of possible Q-statistic constructions, and as a bonus ensures infinitesimal robustness when c is fixed. In order to conserve space, we do not formally treat data contamination in this paper. See Section 1.2 for further discussion. In practice, however, unless we know the error distribution for a simulation based bias correction or to model the bias (e.g. Ronchetti and Trojani, 2001; Mancini et al., 2005), only negligibility $c \to \infty$ as $T \to \infty$ ensures we capture spillover $E[m_{h,t}] \neq 0$ when it occurs. For example, we can always use simple trimming $uI(|u| \leq c)$ or the exponential $u\exp\{-|u|/c\}I(|u| \leq c)$ with

$c \to \infty$ as $T \to \infty$ since our test statistics have standard limits under the null and asymptotic power of one when there is spillover. In heavy tailed cases, however, if truncation $sign\{u\} \min\{|u|, c\}$ is used with an increasing threshold $c \to \infty$ as $T \to \infty$ then too many extremes enter the test statistic for a standard limit distribution (cf. Csörgo et al., 1986). Thus, truncation can only be used with a fixed threshold c. This reduces asymptotic power since the value of spillover $E[m_{h,t}] \neq 0$ cannot in general be identified, but it ensures infinitesimal robustness as discussed below (e.g. Künsch, 1984; Ronchetti and Trojani, 2001; Mancini et al., 2005; Ortelli and Trojani, 2005; Muler and Yohai, 2008).

Robust correlation methods based on fixed threshold trimming date at least to Gnanadesikan and Kettenring (1972) and Devlin et al. (1975). We believe this is the first study to explore a general class of heavy tail robust sample correlations for robust inference, allowing for negligible and non-negligible transforms. We derive their Gaussian self-standardized limit under the null of no-spillover, and characterize the probability limit under spillover assuming a fairly general weak dependence property. 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. There are several novelties with this statistic. First, the proper scale is T since $T^{1/2}\hat{\rho}_{T,h}(\hat{\theta}_T) \xrightarrow{d} N(0,1)$ under the null and mild regularity conditions, and $T^{1/2}|\hat{\rho}_{T,h}(\hat{\theta}_T)| \xrightarrow{p} \infty$ if there is spillover at lag h.[2] This is far more convenient to compute than Runde (1997)'s re-scaled Box–Pierce Q-statistic for heavy-tailed data since that requires the tail index of $\epsilon_{i,t}$ 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 parameter-free. Second, due to the self-scaling structure of a sample correlation, the statistic is automatically robust to a plug-in $\hat{\theta}_T$ provided it has a minimal rate of convergence, a rate that may be below $T^{1/2}$ when $E[\epsilon_{i,t}^4] = \infty$. The score statistic, on the other hand, need not be plug-in robust, and is not when $E[\epsilon_{i,t}^4] < \infty$, a well known challenge in the literature on specification testing (e.g. Wooldridge, 1990; Hill, 2012). Third, the tail-trimmed correlation for spillover analysis trivially extends to other model specification tests, including robust tests of lag order selection. See Section 2.4.

Permissible plug-ins include Log-LAD for linear GARCH (Peng and Yao, 2003; Linton et al., 2010), heavy-tail robust method of moments and QML for linear and nonlinear GARCH (Hill, 2014a), and Laplace QML for linear GARCH (Zhu and Ling, 2011). Other non-Gaussian QML estimators are similarly treated in Berkes and Horváth (2004), but all such estimators involve moment conditions not traditionally imposed for model (1). See Hill (2014a). In general 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 (2001)'s test remains invalid, ours is trivially robust, and QML is then valid. Even in this case we find in a Monte Carlo experiment that robust methods still matter for accurate empirical size.

Score tests are feasible under any trimming scheme for $\mathcal{E}_{i,t}(\theta_i)$ and $\epsilon_{i,t}(\theta_i)$, and they do not require error independence (cf. Forbes and Rogibon, 2002). In order to gain access to QML-type plug-ins for θ^0 we assume $\epsilon_{i,t}^2 - 1$ are serially independent for the portmanteau test, and martingale differences for the score test, allowing for semi-strong GARCH models (Drost and Nijman, 1993; Linton et al., 2010). In simulation experiments, however, our score tests are outperformed by portmanteau tests: score tests result in size distortions while power is comparatively low. The reasons for both distortions are (i) $m_{h,t}$ has an asymmetric distribution, while asymmetric trimming of $m_{h,t}$ is ad hoc without detailed information about tail decay, and need not identify $E[m_{h,t}]$ in small samples; and (ii) the added structure due to a required covariance matrix of possibly dependent estimating equations $m_{h,t}$. A wild bootstrap or sub-sampling procedure, or extreme value theory based estimation of the bias incurred by trimming as in Hill (2013a) and Hill and McCloskey (2014), may reasonably correct both distortions, but we do not explore these possibilities here. We therefore focus on portmanteau tests in our simulation experiments, and only perform related experiments for score tests in the supplemental appendix (Aguilar and Hill, 2014).

Although recentering 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. In experiments here and elsewhere, we find a simple rule of thumb leads to sharp results (see Hill, 2012, 2013a,b, 2014a; Hill and Aguilar, 2013). We also follow Hill (2014a) and Hill and Aguilar (2013) 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. The p-value occupation time is easily computed and is interpreted like a p-value complement, hence it is quite simple to use. In simulation experiments we find the p-value occupation time leads to sharper empirical size in many cases, but comes at a cost of incurring lower power since it smooths over very low and relatively greater amounts of trimming, where the latter naturally distorts evidence of spillover. See Section 5 for further discussion.

It is important to stress that our method of heavy tail robustness requires that the analyst pick a transform (e.g. indicator or Tukey's bisquare), a test statistic (score or portmanteau), and the number of trimmed extremes (fixed percent or negligible as the sample size increases; and non-smoothed or smoothed p-value). Test performance appears to be non-trivially dependent on the combination, although simulation experiments reveal a dominant strategy. This is key since "trimming" per se is not sufficient for heavy tail robustness *pre-asymptotically*, although *asymptotically* our tests work irrespective of heavy tails, the choice of transform, statistic, trimming amount and p-value smoothing.

First, Q-tests have sharper size and higher power than score tests, as discussed above. Second, p-value smoothing in many cases leads to a sharper Q-test under the null, although size distortions without p-value smoothing in many cases are not great, e.g. 1%–3% over-rejection. Third, trimming with increasing thresholds (negligible trimming) rather than fixed thresholds (fixed quantile trimming) results in higher power since only negligibility leads asymptotically to identification of spillover $E[m_{h,t}] \neq 0$. Fourth, truncation with a fixed threshold works very well with sharp size and high power in many cases, where in some cases size is quite sharp even without p-value smoothing, and power can be higher than all other transforms. Replacing sample extremes with a large threshold, rather than zero as in trimming, appears to have a non-negligible impact on small sample power. Fifth, indicator or smooth transforms lead to similar results in small samples, but smooth transforms like Tukey's bisquare and the exponential lead to sharper size faster as the sample size increases. Moreover, in many cases the exponential transform leads to higher power than Tukey's bisquare, and has power comparable to the simple trimming case. Sixth, all tests work less well when there is substantial noise in the data due to very heavy tails. In summary, the *Q-test with exponentially smoothed trimming* or *truncation* are optimal, where *p-value smoothing leads to sharp size and lower power*, while *size distortions without p-value smoothing are not great.*

[2] All limits in this paper are as $T \to \infty$.

[3] Runde (1997) does not characterize the re-scaled Box–Pierce Q-statistic under the alternative evidently because the correlation does not exist: the properly standardized sample correlation converges to a random variable (Davis and Resnick, 1986). Thus, whether the re-scaled Q-statistic is consistent is unknown. Hill and Aguilar (2013) show the re-scaled Q-statistic with a required tail index estimator results in large size distortions.

Section 1.2 discusses the robust estimation literature. In Section 2 we construct the various test statistics, and we present the main results in Section 3. We discuss valid plug-ins and the choice of trimming portion in Section 4. A simulation study follows in Section 5 where we inspect the performance of our test statistics, and show that spillovers can be used for improved volatility forecasts in thin or thick tailed processes. An empirical application follows in Section 6, and parting comments are left for Section 7.

1.2. Robust methods literature

Our methods of heavy tail robustness variously relate to, and differ from, extant methods of outlier robust methods for estimation and inference, and hence there are advantages and weaknesses associated with our tests. At the risk of neglecting important papers, leading examples in the broad literature on infinitesimal robustness include Hampel (1974), Devlin et al. (1975), Künsch (1984); Ronchetti and Trojani (2001), Genton and Ronchetti (2003), and Ortelli and Trojani (2005), and in the specialized literature on conditional volatility models see Sakata and White (1998), Muler and Yohai (2008), Mancini et al. (2005); Boudt and Croux (2010) and Boudt et al. (2011).

A common implication from this literature, dating to Hampel (1974) and Künsch (1984), is that bounding or truncating the relevant statistic with a fixed threshold bounds the (standardized) influence function, a measure of a single point infinitesimal contamination. Examples include bounding the QML score equation (e.g. Sakata and White, 1998; Mancini et al., 2005) or sample correlations (Devlin et al., 1975). Identification under such bounding holds in special cases: examples are M-estimator score equations for a symmetrically distributed autoregression (Hill, 2014b), or sample correlations for a symmetrically distributed process. Otherwise, such bounding forces the analyst to use indirect inference to ensure valid inference asymptotically in view of possible bias, which requires knowledge of the error distribution (e.g. Ronchetti and Trojani, 2001; Ortelli and Trojani, 2005), or requires a functional approximation of the bias which again uses the error distribution form (e.g. Mancini et al., 2005). In this literature, Sakata and White (1998) extend Rousseeuw and Yohai (1984)'s robust S-estimator to conditional dispersion models based on maximum likelihood theory. They establish Fisher consistency and asymptotic normality under the assumption of a correctly specified density, but under misspecification they neither measure potential asymptotic bias or attempt to correct for bias. Such corrections in non-likelihood contexts are given in, amongst others, Mancini et al. (2005) and Ortelli and Trojani (2005).

We allow either increasing or fixed bounded transforms. In the increasing case we use a redescending transform for robust spillover inference in both portmanteau and score tests. An increasing bound implies our methods are not robust to single point data contaminations (cf. Hampel, 1974; Künsch, 1984). However, it necessarily implies identification of the true spillover property without a bias correction, as well as heavy tail robustness in the sense of standard asymptotics without knowledge of whether tails are in fact heavy, and an asymptotic power of one for any of our score and portmanteau tests. The same robustness and identification properties for estimation and inference are exploited in Hill (2012, 2013a,b, 2014a), Hill and Aguilar (2013) and Hill and McCloskey (2014).

We allow for fixed bounded transforms only for our portmanteau test since the statistic structure naturally allows for a bias correction under the null, an idea dating to Gnanadesikan and Kettenring (1972). In this case we achieve infinitesimal robustness (Devlin et al., 1975), but at a potential cost of diminished asymptotic power since a fixed bound reduces available information about spillovers unless a bias correction is used. Our score statistic that trims $m_{h,t}$ directly cannot have a fixed bound without a subsequent bias correction in order to achieve a centered chi-squared limit under the null, while a bias correction in general requires knowledge of the error distribution (cf. Ronchetti and Trojani, 2001; Mancini et al., 2005). or Karamata Theory, as in Hill (2013a), which is beyond this paper's scope.

It appears straightforward to allow for data contamination, including additive outliers, such that a small portion of the observed data are not completely characterized by (1). Robustness to contamination in the form of large values, and heavy tails due to a heavy tailed data generating process, is gained by the use of a fixed bounded transform. However, as stated above, asymptotic power will be diminished for our Q-tests. Meanwhile, a bias correction to improve Q-test power, or correct for bias in the score test based on trimming $m_{h,t}$, requires additional assumptions.

Finally, a diffusion-jump process is an attractive alternative to a GARCH model for framing spikes in volatility in view of evidence for common jump arrivals across different assets. See, e.g., Ait-Sahalia and Jacod (2009), Ait-Sahalia et al. (2012), and Bollerslev et al. (2013). An extension to a multivariate continuous time framework, however, is well beyond the scope of this paper.

The following notational conventions are used. The indicator function $I(A) = 1$ if A is true, and 0 otherwise. Write $(\partial/\partial\theta_i)h^2_{i,t} = (\partial/\partial\theta_i)h^2_{i,t}(\theta_i)|_{\theta^0}$. If sequences $\{a_T, b_T\}$ are stochastic and $a_T/b_T \xrightarrow{p} 1$ we write $a_T \overset{p}{\sim} b_T$, and if they are non-stochastic and $a_T/b_T \to 1$ we write $a_T \sim b_T$. 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 $\|A\| = (\lambda_{\max}(A'A))^{1/2}$ where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue. If z is a scalar we write $(z)_+ \equiv \max\{0, z\}$, 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. $L(T)$ is a slowly varying function, $L(T) \to \infty$, the value or rate of which may change from line to line.[4] We say a random variable is *symmetric* if its distribution is symmetric.

2. Robust test statistics

In the following we introduce five tail-trimmed test statistics, where a summary of each is provided in Table 1. Asymptotic theory is developed in Section 3 and all assumptions are presented in Appendix A. 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 $\mathbb{R}^q$ or $\mathbb{R}^{2q}$.

2.1. Tail-trimmed equations: score test

Our first approach is to trim $m_{h,t}(\theta) = (\epsilon^2_{1,t}(\theta_1) - 1)(\epsilon^2_{2,t-h}(\theta_2) - 1)$ by its large values, as in Hill and Aguilar (2013). The null hypothesis of no volatility spillover up to horizon $H \geq 1$ can be written

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

Since $m_{h,t}$ may have an asymmetric distribution, we ensure heavy tail robustness and identification by using a *negligible asymmetric* transformation. We use asymmetric tail-trimming for

[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 $T \to \infty$.

Table 1
Test statistics and transformations.

Test statistic descriptions				
Test statistic	Allow error dependence	Object of interest	Variable transformed	Re-Centering after transform
$\hat{W}_T^{(m)}$	yes	$m_{h,t} = \mathcal{E}_{1,t}\mathcal{E}_{2,t-h}$	$m_{h,t}$	None
$\hat{W}_T^{(\mathcal{E})}$	yes	$\mathcal{E}_{i,t} = \epsilon_{i,t}^2 - 1$	$\mathcal{E}_{i,t}$	Re-center $\mathcal{E}$
$\hat{W}_T^{(\mathfrak{E})}$	yes	$\mathcal{E}_{i,t} = \epsilon_{i,t}^2 - 1$	$\epsilon_{i,t}$	Re-center ϵ or ϵ^2
$\hat{Q}_T^{(\mathcal{E})}$	no	$\mathcal{E}_{i,t} = \epsilon_{i,t}^2 - 1$	$\mathcal{E}_{i,t}$	Re-center $\mathcal{E}$
$\hat{Q}_T^{(\mathfrak{E})}$	no	$\mathcal{E}_{i,t} = \epsilon_{i,t}^2 - 1$	$\epsilon_{i,t}$	Re-center ϵ or ϵ^2
$\hat{Q}_T^{(\epsilon)}$	no	$\epsilon_{i,t}$	$\epsilon_{i,t}$	Re-center $\epsilon_{i,t}$
Allowed transformation ψ and fractile k_T properties				
Test statistic	Transform type	Symmetry	Example	Fractile bound
$\hat{W}_T^{(m)}$	Indicator	Asymmetric	$uI(a \leq u \leq b)$	$k_T/T \to 0$
$\hat{W}_T^{(\mathcal{E})}, \hat{W}_T^{(\mathfrak{E})}$	Redescend	Symmetric	$uI(\lvert u\rvert \leq c)$ $u(1-(u/c)^2)^2 I(\lvert u\rvert \leq c)$ $u\exp\{-\lvert u\rvert/c\}I(\lvert u\rvert \leq c)$	$k_T/T \to 0$
$\hat{Q}_T^{(\mathcal{E})}, \hat{Q}_T^{(\mathfrak{E})}, \hat{Q}_T^{(\epsilon)}$	Redescend	Symmetric	$uI(\lvert u\rvert \leq c)$ $u(1-(u/c)^2)^2 I(\lvert u\rvert \leq c)$ $u\exp\{-\lvert u\rvert/c\}I(\lvert u\rvert \leq c)$	$k_T/T \to [0, 1)$
	Non-redescend	Symmetric	$u\min\{1, c/\lvert u\rvert\}$	$k_T/T \to (0, 1)$

convenience.[5] Define tail specific observations of $m_{h,t}(\theta)$ and their sample order statistics:

$$\begin{aligned} m_{h,t}^{(-)}(\theta) &\equiv m_{h,t}(\theta) \times I\left(m_{h,t}(\theta) < 0\right) \quad \text{and} \\ &\quad m_{h,(1)}^{(-)}(\theta) \leq \cdots \leq m_{h,(T)}^{(-)}(\theta) \leq 0 \\ m_{h,t}^{(+)}(\theta) &\equiv m_{h,t}(\theta) \times I\left(m_{h,t}(\theta) \geq 0\right) \quad \text{and} \\ &\quad m_{h,(1)}^{(+)}(\theta) \geq \cdots \geq m_{h,(T)}^{(+)}(\theta) \geq 0. \end{aligned} \tag{3}$$

Let $\{k_{r,T}^{(m)} : r = 1, 2\}$ be integer sequences representing the number of trimmed left- and right-tailed observations from the sample $\{m_{h,t}(\theta)\}_{t=1}^T$. The tail-trimmed version of $m_{h,t}(\theta)$ is then

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

We ensure trimming is *negligible* by assuming $\{k_{r,T}^{(m)}\}$ are *intermediate order* sequences, hence $k_{r,T}^{(m)} \to \infty$ and $k_{r,T}^{(m)}/T \to 0$, since then the thresholds $|m_{h,(k_{r,T}^{(m)})}^{(\cdot)}(\theta)| \xrightarrow{p} \infty$ given the errors $\epsilon_{i,t}$ have support $\mathbb{R}$. See Leadbetter et al. (1983) for details on order sequences. Trimming asymptotically infinitely many observations $k_{r,T}^{(m)} \to \infty$ that represent a negligible or vanishing tail portion $k_{r,T}^{(m)}/T \to 0$ ensures Gaussian asymptotics and identification of the hypotheses since by dominated convergence $E[\hat{m}_{h,T,t}^*] \to E[m_{h,t}]$. Fixed quantile trimming, by contrast, uses *central order* sequences $k_{r,T}^{(m)}/T \to (0, 1)$ which are non-negligible and therefore cannot ensure identification of the hypotheses since $m_{h,t}$ may be asymmetric (e.g. Hill and Aguilar, 2013). Finally, extreme quantile trimming uses *extreme order* sequences $k_{r,T}^{(m)} \to k \in (0, \infty)$: a fixed number of large values are trimmed asymptotically, which is too few to promote a standard limit theory in the heavy tail case.

5 A negligible *symmetric* transform is also possible, provided some form of bias correction is used. As an example, Hill (2013a) and Hill and McCloskey (2014) assume, as we do here, power law tail decay, and use negligible symmetric tail-trimming. Karamata theory is then used to closely approximate the bias, and tail exponent estimators are used to estimate bias. The same methods can in principle be applied here, although we do not treat the possibility due to space constraints.

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

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

where $\hat{m}_T^*(\theta) \equiv 1/T\sum_{t=1}^T \hat{m}_{T,t}^*(\theta)$, $\mathcal{K}(\cdot)$ is a kernel function, and γ_T is bandwidth where $\gamma_T \to \infty$ and $\gamma_T = o(T)$. The test statistic has a quadratic form as in Hill and Aguilar (2013), where $\hat{\theta}_T$ denotes a consistent estimator of θ^0:

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

2.2. Transformed centered errors: score and portmanteau tests

Our second approach is to transform $\mathcal{E}_{i,t}(\theta) \equiv y_{i,t}^2(\theta)/h_{i,t}^2(\theta) - 1$ symmetrically and then recenter. Define $s_{i,t}(\theta) \equiv (\partial/\partial\theta)\ln h_{i,t}^2(\theta)$ and note that we can write $\mathcal{E}_{i,t}(\hat{\theta}_T) = \mathcal{E}_{i,t} - \epsilon_{i,t}^2 s_{i,t}(\hat{\theta}_T - \theta^0)$ with probability approaching one as $T \to \infty$. We assume $s_{i,t}$ is uniformly square integrable by (2), so trimming $\mathcal{E}_{i,t}(\hat{\theta}_T)$ only requires information from $\epsilon_{i,t}(\theta)$. Recall in general this aligns with non-trivial random volatility.

2.2.1. Redescending transforms

Define a mapping $\check{\psi} : \mathbb{R} \times [0, \infty) \to \mathbb{R}$ that is *bounded* $|\check{\psi}(u, c)| \leq c$, *symmetric* $\check{\psi}(-u, c) = -\check{\psi}(u, c)$, and satisfies $\lim_{c\to\infty} \check{\psi}(u, c) = u$. If $\epsilon_{1,t}$ and $\epsilon_{2,t-h}$ may be dependent under the null, then we enforce *negligibility* $c \to \infty$ as $T \to \infty$ to ensure that we test $E[m_{h,t}] = 0$ and can reveal spillover $E[m_{h,t}] \neq 0$. In this case we must be careful that not too many extreme values enter our statistic, hence we impose a strict version of *redescendence* $\check{\psi}(u, c) = 0$ if $|u| > c$.[6] It is therefore convenient to focus on the

6 In the literature redescendence applies when $\psi(u, c) \to 0$ as $|u| \to \infty$. See Andrews et al. (1972) and Hampel et al. (1986).

following class of functions:

$$\check{\psi}(u,c) = u \times \check{\varpi}(u,c) \times I(|u| \le c) \quad \text{where } \check{\varpi}(-u,c) = \check{\varpi}(u,c). \tag{4}$$

Assume $\check{\varpi}(\cdot, c)$ is for each c a Borel function and

$$\lim_{c\to\infty} \check{\varpi}(u,c) \times I(|u| \le c) = 1 \quad \forall u. \tag{5}$$

As a result, $\check{\psi}(u,c)$ operates like a simple tail-trimming function since

$$\check{\psi}(u,c) = uI(|u| \le c) \times (1 + o(1)) \quad \text{as } c \to \infty. \tag{6}$$

In the simple trimming case $\check{\varpi}(u,c) = 1$, Tukey's bisquare has $\check{\varpi}(u,c) = (1-(u/c)^2)^2$, and the exponential transform has $\check{\varpi}(u,c) = \exp\{-|u|/c\}$.

The theory developed in Section 3 easily extends to related redescending functions $\check{\psi}(u,c)$, like Hampel's three-part trimming function with thresholds $0 < a < b < c$ (see Andrews et al., 1972):

$$\begin{cases} u, & 0 \le |u| \le a \\ a \times sign(u) & a < |u| \le b \\ \dfrac{a \times (c-|u|)}{c-b} \times sign(u) & b < |u| \le c \\ 0 & c < |u|. \end{cases}$$

Definition (4)–(5) is not rich enough to include three thresholds, but, with an abuse of notation, Hampel's function fits (4) with

$$\check{\varpi}(u,c) = I(|u| \le a) + \frac{a}{|u|} \times I(a < |u| \le b) + \frac{a(c-|u|)}{|u|(c-b)} \times I(b < |u| \le c).$$

By construction $\check{\varpi}(u,c) \in [0,1]$ and $\check{\varpi}(u,c) \to 1$ as the smallest threshold $a \to \infty$.

Our test statistic uses in place of $\mathcal{E}_{i,t}(\theta)$ a sample version of

$$\check{\psi}_{i,t}(c) \equiv \check{\psi}(\mathcal{E}_{i,t}, c) - E\left[\check{\psi}(\mathcal{E}_{i,t}, c)\right].$$

If $\epsilon_{1,t}$ and $\epsilon_{2,t-h}$ are mutually independent under the null, then c can be any fixed positive number, or $c \to \infty$ as $T \to \infty$, since in either case the null is identified: for any pair of positive numbers $\{c_1, c_2\}$,

$$E\left[\check{\psi}_{1,t}(c_1)\check{\psi}_{2,t-h}(c_2)\right] = E\left[\check{\psi}_{1,t}(c_1)\right] \times E\left[\check{\psi}_{1,t}(c_2)\right] = 0.$$

2.2.2. General bounded transform

If the threshold c is fixed, then we no longer require redescendence, and we may use the following broader class of bounded symmetric functions:

$$\psi(u,c) = u \times \varpi(u,c) \quad \text{where } \varpi(u,c) = \varpi(-u,c) \quad \text{and} \quad |u \times \varpi(u,c)| \le c \quad \forall u. \tag{7}$$

Redescending $\check{\psi}$ in (4) is nested in (7) with $\varpi(u,c) = \check{\varpi}(u,c) \times I(|u| \le c)$ with fixed c or $c \to \infty$, while (7) also allows truncation $\varpi(0,c) = 0$ and $\varpi(u,c) = \min\{1, c/|u|\}$ for $u \ne 0$. Truncation is non-redescending since $\psi(u,c) = c$ when $|u| \ge c$, hence c *must* be fixed since otherwise too many large values enter the test statistic for a Gaussian limit theory.

Now define two-tailed observations and their sample order statistics

$$\mathcal{E}^{(a)}_{i,t}(\theta) \equiv \left|\mathcal{E}_{i,t}(\theta)\right| \quad \text{and} \quad \mathcal{E}^{(a)}_{i,(1)}(\theta) \ge \mathcal{E}^{(a)}_{i,(2)}(\theta) \ge \cdots \ge \mathcal{E}^{(a)}_{i,(T)}(\theta) \ge 0. \tag{8}$$

We use order statistics $\mathcal{E}^{(a)}_{i,(k^{(\mathcal{E})}_{i,T})}(\theta)$ as thresholds c, where $\{k^{(\mathcal{E})}_{i,T}\}$ are sequences of integers $1 \le k^{(\mathcal{E})}_{i,T} < T$ that must satisfy $k^{(\mathcal{E})}_{i,T} \to \infty$ as $T \to \infty$ to promote standard asymptotics, similar to the above score statistic $\hat{W}^{(m)}_T$. Whether we require negligibility $k^{(\mathcal{E})}_{i,T}/T \to 0$ (which implies an increasing threshold $\mathcal{E}^{(a)}_{i,(k^{(\mathcal{E})}_{i,T})}(\theta) \xrightarrow{p} \infty$) or non-negligibility $k^{(\mathcal{E})}_{i,T}/T \to (0,1)$ (which implies a bounded threshold $\mathcal{E}^{(a)}_{i,(k^{(\mathcal{E})}_{i,T})}(\theta) \xrightarrow{p} (0,\infty)$), is respectively dependent upon whether the errors $\{\epsilon_{1,t}, \epsilon_{2,t}\}$ are possibly dependent or the transform is non-redescending.

If $\epsilon_{1,t}$ and $\epsilon_{2,t-h}$ may be *dependent* under the null, then we impose negligibility $k^{(\mathcal{E})}_{i,T}/T \to 0$ to ensure the hypotheses are identified. Since this implies the threshold is increasing we must use a redescending transform $\psi = \check{\psi}$ to control the number of extremes that enter the test statistic.

However, if $\epsilon_{1,t}$ and $\epsilon_{2,t-h}$ are *independent* under the null then either negligible $k^{(\mathcal{E})}_{i,T}/T \to 0$ or non-negligible $k^{(\mathcal{E})}_{i,T}/T \to (0,1)$ are allowed depending on ψ. In the redescending case $\psi = \check{\psi}$, as in simple trimming $\check{\psi}(u,c) = uI(|u| \le c)$, any negligible or non-negligible $k^{(\mathcal{E})}_{i,T}/T \to [0,1)$ is allowed since the null is always identified under recentering. If the transform is non-redescending $\psi \ne \check{\psi}$, as in truncation $\psi(u,c) = sign\{u\}\min\{|u|, c\}$, then non-negligibility $k^{(\mathcal{E})}_{i,T}/T \to (0,1)$ is required to control the number of extremes that enter the test statistic. See Table 1 for a summary of feasible transforms ψ and fractiles k_T for portmanteau statistics where $\epsilon_{1,t}$ and $\epsilon_{2,t-h}$ are independent under the null, and for score statistics where $\epsilon_{1,t}$ and $\epsilon_{2,t-h}$ may be dependent under the null.

Now define the transformed and centered version of $\mathcal{E}_{i,t}(\theta)$ in the redescending case with $\check{\psi}$ in (4):

$$\hat{\check{\psi}}^{(\mathcal{E})}_{i,T,t}(\theta) \equiv \check{\psi}\left(\mathcal{E}_{i,t}(\theta), \mathcal{E}^{(a)}_{i,(k^{(\mathcal{E})}_{i,T})}(\theta)\right) - \frac{1}{T}\sum_{t=1}^{T} \check{\psi}\left(\mathcal{E}_{i,t}(\theta), \mathcal{E}^{(a)}_{i,(k^{(\mathcal{E})}_{i,T})}(\theta)\right),$$

and in the general case with ψ defined in (7):

$$\hat{\psi}^{(\mathcal{E})}_{i,T,t}(\theta) \equiv \psi\left(\mathcal{E}_{i,t}(\theta), \mathcal{E}^{(a)}_{i,(k^{(\mathcal{E})}_{i,T})}(\theta)\right) - \frac{1}{T}\sum_{t=1}^{T} \psi\left(\mathcal{E}_{i,t}(\theta), \mathcal{E}^{(a)}_{i,(k^{(\mathcal{E})}_{i,T})}(\theta)\right).$$

2.2.3. Test statistics

Now suppose $\epsilon_{i,t}$ are serially independent, and the no spillover hypothesis is

$$H^{(\epsilon)}_0 : \epsilon_{1,t} \quad \text{and} \quad \epsilon_{2,t} \text{ are mutually independent.}$$

Then we can use a general transform ψ with $k^{(\mathcal{E})}_{i,T}/T \to [0,1)$ depending on whether ψ is redescending or not, and a portmanteau statistic. Our heavy-tail robust sample correlation coefficient and portmanteau statistic are

$$\hat{\rho}^{(\mathcal{E})}_{T,h}(\theta) \equiv \frac{\sum_{t=1}^{T} \hat{\psi}^{(\mathcal{E})}_{1,T,t}(\theta_1)\,\hat{\psi}^{(\mathcal{E})}_{2,T,t-h}(\theta_2)}{\left(\sum_{t=1}^{T}\left(\hat{\psi}^{(\mathcal{E})}_{1,T,t}(\theta_1)\right)^2\right)^{1/2}\left(\sum_{t=1}^{T}\left(\hat{\psi}^{(\mathcal{E})}_{2,T,t}(\theta_2)\right)^2\right)^{1/2}}$$

$$\hat{Q}^{(\mathcal{E})}_T(H) = T\sum_{h=1}^{H} \mathcal{W}_T(h)\left(\hat{\rho}^{(\mathcal{E})}_{T,h}(\hat{\theta}_T)\right)^2,$$

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 $T\to\infty$. Examples include deterministic weights used in the Box–Pierce and Ljung–Box tests $\mathcal{W}_T(h)=1,(T+2)/(T-h)$ or $T/(T-h)$. A standardized Q-test as in Hong (2001) can similarly be constructed to allow for an increasing horizon $H\to\infty$ as $T\to\infty$. It is interesting to point out the correct scale is T even if $\hat{\psi}^{(\mathcal{E})}_{i,T,t}(\theta_i)$ are negligible due to $k^{(\mathcal{E})}_{i,T}/T\to 0$. This follows from independence and the self-scaled nature of a sample correlation. See Section 3 for theory details.

If the errors may be serially dependent, as for a semi-strong GARCH model, or mutually dependent under the null of no spillover, then $\hat{\rho}^{(\mathcal{E})}_{T,h}(\hat{\theta}_T)$ are not necessarily asymptotically independent, hence $\hat{Q}^{(\mathcal{E})}_T$ need not have a limiting chi-squared distribution. We can, however, still test $H_0^{(m)}$ by using recentered redescending transforms as test equations $[\hat{\check{\psi}}^{(\mathcal{E})}_{1,T,t}(\theta_1)\hat{\check{\psi}}^{(\mathcal{E})}_{2,T,t-h}(\theta_2)]_{h=1}^{H}$ with negligibility $k^{(\mathcal{E})}_{i,T}/T\to 0$, an associated HAC $\hat{S}^{(\mathcal{E})}_T(\theta)$, and a score statistic:

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

2.2.4. Score and portmanteau statistic differences

It is important to note the crucial differences between the above score and Q-statistics. If $m_{h,t}$ is asymmetrically trimmed by its large values, then recentering is not feasible and small sample bias may arise in $\hat{m}^*_{h,T,t}$. In turn, this can lead to a test based on $\hat{W}^{(m)}_T$ over-rejecting the null. By comparison, with recentering, $\hat{W}^{(\mathcal{E})}_T$ will be close to a sum of squared zero mean and unit variance random variables under the null.

If, however, $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are serially independent, and mutually independent under the null, then both $\hat{W}^{(m)}_T$ and $\hat{W}^{(\mathcal{E})}_T$ have redundant structure through the off-diagonal components of $\hat{S}_T(\hat{\theta}_T)$ and $\hat{S}^{(\mathcal{E})}_T(\hat{\theta}_T)$, which can lead to empirical size distortions. Moreover, Q-statistics appear insensitive to the chosen plug-in under general conditions while the score statistic is not. See Section 3. Hence, the small sample dynamics of $\hat{W}^{(m)}_T$ and $\hat{W}^{(\mathcal{E})}_T$ are a result of approximating both the correlation *and* θ^0 as well as the redundant covariance structure, while $\hat{W}^{(m)}_T$ is further affected by small sample bias induced by asymmetric trimming without the possibility of recentering. We find $\hat{W}^{(m)}_T$ exhibits size distortions due to the delicacy of asymmetric trimming, and has a higher rejection rate than $\hat{W}^{(\mathcal{E})}_T$ under the null. Further, both $\hat{W}^{(m)}_T$ and $\hat{W}^{(\mathcal{E})}_T$ reject more often than $\hat{Q}^{(\mathcal{E})}_T$ under the null, while $\hat{Q}^{(\mathcal{E})}_T$ exhibits sharp size. See the simulation experiments of Section 5.

2.3. Transformed errors: score and portmanteau tests

Lastly, we transform $\epsilon^2_{i,t}(\theta)$ or $\epsilon_{i,t}(\theta)$ and recenter. Write $\epsilon^{(a)}_{i,t}(\theta)\equiv|\epsilon_{i,t}(\theta)|$, define order statistics $\{\epsilon^{(a)}_{i,(j)}(\theta)\}$, and let $\{k^{(\epsilon)}_{i,T}\}$ be order sequences, $1\leq k^{(\epsilon)}_{i,T}<T$ and $k^{(\epsilon)}_{i,T}\to\infty$. The transformed and recentered $\epsilon^2_{i,t}(\theta)$ based on ψ are

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

Alternatively, we may transform $\epsilon_{i,t}(\theta)$, recenter, square and recenter again:

$$\hat{\psi}^{(\mathfrak{E})}_{i,T,t}(\theta)\equiv\left(\psi\left(\epsilon_{i,t}(\theta),\epsilon^{(a)}_{i,(k^{(\epsilon)}_{i,T})}(\theta)\right)-\frac{1}{T}\sum_{t=1}^{T}\psi\left(\epsilon_{i,t}(\theta),\epsilon^{(a)}_{i,(k^{(\epsilon)}_{i,T})}(\theta)\right)\right)^2-\frac{1}{T}\sum_{t=1}^{T}\left(\psi\left(\epsilon_{i,t}(\theta),\epsilon^{(a)}_{i,(k^{(\epsilon)}_{i,T})}(\theta)\right)-\frac{1}{T}\sum_{t=1}^{T}\psi\left(\epsilon_{i,t}(\theta),\epsilon^{(a)}_{i,(k^{(\epsilon)}_{i,T})}(\theta)\right)\right)^2.$$

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

$$\hat{\rho}^{(\mathfrak{E})}_{T,h}(\theta)=\frac{\sum_{t=1}^{T}\hat{\psi}^{(\mathfrak{E})}_{1,T,t}(\theta_1)\hat{\psi}^{(\mathfrak{E})}_{2,T,t-h}(\theta_2)}{\left(\sum_{t=1}^{T}\left(\hat{\psi}^{(\mathfrak{E})}_{1,T,t}(\theta_1)\right)^2\right)^{1/2}\left(\sum_{t=1}^{T}\left(\hat{\psi}^{(\mathfrak{E})}_{2,T,t}(\theta_2)\right)^2\right)^{1/2}}$$

$$\hat{Q}^{(\mathfrak{E})}_T(H)\equiv T\sum_{h=1}^{H}\mathcal{W}_T(h)\,(\hat{\rho}^{(\mathfrak{E})}_{T,h}(\hat{\theta}_T))^2.$$

If the errors are possibly serially dependent, then we use equations $[\hat{\check{\psi}}^{(\mathfrak{E})}_{1,T,t}(\theta_1)\hat{\check{\psi}}^{(\mathfrak{E})}_{2,T,t-h}(\theta_2)]_{h=1}^{H}$ based on a redescending $\check{\psi}$, with negligibility $k^{(\epsilon)}_{i,T}/T\to 0$, HAC $\hat{S}^{(\mathfrak{E})}_T(\theta)$, and a score statistic for a test of $H_0^{(m)}$:

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

There is no theory-based advantage for transforming $\mathcal{E}_{i,t}$ versus $\epsilon_{i,t}$, or in which order we recenter the transformed $\epsilon_{i,t}$. In general, of course, transforming $\mathcal{E}_{i,t}$ or $\epsilon_{i,t}$ with recentering 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). Let $\{k^{(\epsilon)}_T\}$ be a central or intermediate order sequence. A robust test of serial correlation in ϵ_t simply uses

$$\hat{\psi}^{(\epsilon)}_{T,t}(\theta)\equiv\psi\left(\epsilon_t(\theta),\epsilon^{(a)}_{(k^{(\epsilon)}_T)}(\theta)\right)-\frac{1}{T}\sum_{t=1}^{T}\psi\left(\epsilon_t(\theta),\epsilon^{(a)}_{(k^{(\epsilon)}_T)}(\theta)\right).$$

As usual, in the redescending case $\psi=\check{\psi}$ any rate $k^{(\epsilon)}_T/T\to[0,1)$ is allowed, otherwise $k^{(\epsilon)}_T/T\to(0,1)$. The Q-statistic is $\hat{Q}^{(\epsilon)}_T(H)\equiv T\sum_{h=1}^{H}\mathcal{W}_T(h)\,(\hat{\rho}^{(\epsilon)}_{T,h}(\hat{\theta}_T))^2$ with correlations

$$\hat{\rho}^{(\epsilon)}_{T,h}(\theta)\equiv\frac{\sum_{t=1}^{T}\hat{\psi}^{(\epsilon)}_{T,t}(\theta)\,\hat{\psi}^{(\epsilon)}_{T,t-h}(\theta)}{\sum_{t=1}^{T}\left(\hat{\psi}^{(\epsilon)}_{T,t}(\theta)\right)^2}.\tag{9}$$

3. Asymptotic theory

In the following we characterize the limiting properties of the portmanteau statistic $\hat{Q}_T^{(\mathcal{E})}$, and in the supplemental material Aguilar and Hill (2014) we derive similar results for the score statistic $\hat{W}_T^{(m)}$. Each remaining statistic follows similarly. See Appendix A for assumptions related to the Q-test concerning the bounded transforms (B), the DGP (D), error memory and moments (E), rate bounds for the trimming fractiles $k_T^{(\mathcal{E})}$ and $k_T^{(\epsilon)}$ (FE), the plug-in for Q-statistics (PQ) and distribution tails (T). All proofs are presented in Appendix B. Aguilar and Hill (2014) present related assumptions for the score tests.

3.1. Limiting null distribution

The Q-statistic is asymptotically chi-squared if $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are serially and mutually independent, as long as $\hat{\theta}_T \xrightarrow{p} \theta^0$ sufficiently fast. In order to characterize the necessary rate we require the non-random quantiles that the order statistics estimate. Although some aspects of the following are treated in Hill and Aguilar (2013), a portmanteau statistic has unique properties that leads to sharply different conclusions about permissible estimators. Thus, the following should be helpful for anyone interested in heavy tail robust serial correlation computation.

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

$$P\left(\left|\mathcal{E}_{i,t}(\theta)\right| > c_{i,T}^{(\mathcal{E})}(\theta)\right) = \frac{k_{i,T}^{(\mathcal{E})}}{T}, \tag{10}$$

and define transformed equations

$$\psi_{i,T,t}^{(\mathcal{E})}(\theta) \equiv \psi\left(\mathcal{E}_{i,t}(\theta), c_{i,T}^{(\mathcal{E})}(\theta)\right) - E\left[\psi\left(\mathcal{E}_{i,t}(\theta), c_{i,T}^{(\mathcal{E})}(\theta)\right)\right].$$

Notice $c_{i,T}^{(\mathcal{E})}(\theta)$ exists for any $\theta \in \Theta$ and $\{k_{i,T}^{(\mathcal{E})}\}$ since $\epsilon_{i,t}$ have smooth distributions on $\mathbb{R}$ under Assumption D, while by construction $\mathcal{E}_{i,(k_{i,T}^{(\mathcal{E})})}^{(a)}(\theta)$ estimates $c_{i,T}^{(\mathcal{E})}(\theta)$. In all cases $k_{i,T}^{(\mathcal{E})} \to \infty$, hence $c_{i,T}^{(\mathcal{E})}(\theta) \to \infty$ under negligibility $k_{i,T}^{(\mathcal{E})}/T \to 0$, whereas $c_{i,T}^{(\mathcal{E})}(\theta) \to (0,\infty)$ under non-negligibility $k_{i,T}^{(\mathcal{E})}/T \to (0, 1)$.

Now define variance and Jacobian matrices

$$\mathfrak{S}_T \equiv E\left[\left(\psi_{1,T,t}^{(\mathcal{E})}\right)^2\right] \times E\left[\left(\psi_{2,T,t}^{(\mathcal{E})}\right)^2\right] \quad \text{and}$$

$$\mathfrak{J}_{i,T}^{(h)} \equiv \frac{\partial}{\partial\theta_i} E\left[\psi_{1,T,t}^{(\mathcal{E})}(\theta_1)\psi_{2,T,t-h}^{(\mathcal{E})}(\theta_2)\right]|_{\theta^0} \in \mathbb{R}^{q\times 1}$$

$$\begin{aligned}\mathfrak{V}_{i,T} = \max_{1\le h\le H}\Bigg\{\frac{T}{\mathfrak{S}_T}\Big[&\mathbf{1}_q \times I\left(\left\|\mathfrak{J}_{i,T}^{(h)}\right\| \le 1\right) \\ &+ \mathfrak{J}_{i,T}^{(h)\prime} \times I\left(\left\|\mathfrak{J}_{i,T}^{(h)}\right\| > 1\right)\Big]\Bigg\} \in \mathbb{R}^{1\times q},\end{aligned}$$

where $\mathbf{1}_q \equiv [1, \dots, 1]' \in \mathbb{R}^q$, and write

$$\mathfrak{V}_T \equiv \left[\mathfrak{V}_{1,T}, \mathfrak{V}_{2,T}\right] \in \mathbb{R}^{1\times 2q}. \tag{11}$$

Note that $\mathfrak{V}_{i,T}$ and $\mathfrak{V}_T$ implicitly depend on the horizon H. Also note $\mathfrak{S}_T$ is bounded under non-negligibility $k_{i,T}^{(\mathcal{E})}/T \to (0, 1)$, and otherwise is unbounded when $E[\epsilon_{i,t}^4] = \infty$. In general:

$$\text{if } \frac{k_{i,T}^{(\mathcal{E})}}{T} \to (0,1) \quad \text{then } \limsup_{T\to\infty} \mathfrak{S}_T < \infty \tag{12}$$

$$\text{if } \frac{k_{i,T}^{(\mathcal{E})}}{T} \to 0 \quad \text{and}$$

$$\begin{cases} E[\epsilon_{i,t}^4] = \infty \text{ for either } i \in \{1,2\} \quad \text{then } \mathfrak{S}_T \to \infty \\ E[\epsilon_{i,t}^4] < \infty \text{ for both } i \in \{1,2\} \\ \qquad \text{then } \limsup_{T\to\infty} \mathfrak{S}_T < \infty. \end{cases}$$

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

$$\begin{aligned} T^{1/2}&\hat{\rho}_{T,h}^{(\mathcal{E})}(\hat{\theta}_T) \\ &\overset{p}{\sim} \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}\sum_{t=1}^T\left\{\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})} - E\left[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}\right]\right\} \\ &\quad + K\sum_{i=1}^2 \frac{T^{1/2}}{\mathfrak{S}_T^{1/2}}\mathfrak{J}_{i,T}^{(h)\prime}\left(\hat{\theta}_{i,T} - \theta_i^0\right) + \frac{T^{1/2}}{\mathfrak{S}_T^{1/2}}E\left[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}\right]. \end{aligned} \tag{13}$$

By recentering we have under mutual independence $E[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}] = 0$. There are two notable features of (13) under mutual independence. First, $T^{-1/2}\mathfrak{S}_T^{-1/2}\sum_{t=1}^T \psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}$ is a self-standardized partial sum of independent random variables and therefore is asymptotically standard normal. This implies the proper scale for $\hat{\rho}_{T,h}^{(\mathcal{E})}(\hat{\theta}_T)$ is $T^{1/2}$, contrary to the sample correlation for heavy tailed data (cf Davis and Resnick, 1986), and exactly like the classic sample correlation case for thin tailed data (e.g. Box and Pierce, 1970). This is true for any redescending transform $\psi = \check{\psi}$ in view of boundedness (5) under either negligibility $k_{i,T}^{(\mathcal{E})}/T \to 0$ or non-negligibility $k_{i,T}^{(\mathcal{E})}/T \to (0,1)$, and it is true for any non-redescending ψ in (7) as long as non-negligibility $k_{i,T}^{(\mathcal{E})}/T \to (0,1)$ holds.

Second, we show in Appendix B that $\mathfrak{J}_{i,T}^{(h)} = o(1)$ under $H_0^{(\epsilon)}$ since recentering ensures $E[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}] = 0$, hence

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

Thus we only need $(T/\mathfrak{S}_T)^{1/2}(\hat{\theta}_{i,T} - \theta_i^0) = O_p(1)$. If both $E[\epsilon_{i,t}^4] < \infty$, then $T^{1/2}(\hat{\theta}_{i,T} - \theta_i^0) = O_p(1)$, as is conventionally assumed (e.g. Hong, 2001). If $k_{i,T}^{(\mathcal{E})}/T \to (0,1)$ such that an asymptotically bounded threshold is used, then by (12) the plug-in again must be $T^{1/2}$-convergent *irrespective of heavy tails*. Otherwise, suppose an unbounded threshold is used due to negligibility $k_{i,T}^{(\mathcal{E})}/T \to 0$. If tails are heavy $E[\epsilon_{i,t}^4] = \infty$, then by (12) we have $\mathfrak{S}_T = E[\psi_{1,T,t}^{(\mathcal{E})}] \times E[\psi_{2,T,t-h}^{(\mathcal{E})}] \to \infty$, hence the plug-in $\hat{\theta}_{i,T}$ may be sub-$T^{1/2}$-convergent.

Under the alternative we can only say $\mathfrak{J}_{i,T}^{(h)} = O(1)$, hence we require $\mathfrak{V}_{i,T}^{1/2}(\hat{\theta}_{i,T} - \theta_i^0) = O_p(1)$, which again reduces to $T^{1/2}(\hat{\theta}_{i,T} - \theta_i^0) = O_p(1)$ when $E[\epsilon_{i,t}^4] < \infty$ or a bounded threshold is used. As long as $\hat{\theta}_{i,T}$ satisfies Assumption PQ $\mathfrak{V}_T^{1/2}(\hat{\theta}_T - \theta^0) = O_p(1)$ then Gaussian asymptotics follow for any case described above.

The Q-statistic is asymptotically chi-squared when the plug-in is based on Log-LAD in Peng and Yao (2003), Quasi-Maximum Tail-Trimmed Likelihood (QMTTL) in Hill (2014a), or Zhu and Ling (2011)'s Quasi-Maximum Weighted Laplace Likelihood (QMWLL). Other method of moments based estimators in Hill (2014a) have properties similar to QMTTL and are therefore valid, and other non-Gaussian QML estimators in principle apply (e.g. Berkes and Horváth, 2004) but with additional non-standard moment conditions that deviate from the classic GARCH model (see Fan et al., 2014; Hill, 2014a).

Theorem 3.1 (*Portmanteau Test Under H_0*)**.**

a. *Let Assumption* B, D, E, FE, 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: QMTTL if* $E[\epsilon_{i,t}^4] < \infty$, *or trimming is negligible* $k_{i,T}^{(\mathcal{E})}/T \to 0$; *QMWLL if* $\epsilon_{i,t}$ *have a zero median, and* $E|\epsilon_{i,t}| = 1$; *Log-LAD if* $\ln(\epsilon_{i,t}^2)$ *has a zero median; QML if both* $E[\epsilon_{i,t}^4] < \infty$.

Remark 1. 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 for model (1) is valid if $E[\epsilon_{i,t}^4] < \infty$.

Remark 2. A standard GARCH model (1) requires $E[\epsilon_{i,t}^2] = 1$, while QMWLL also requires a zero median and $E|\epsilon_{i,t}| = 1$ (Zhu and Ling, 2011), necessarily restricting the feasible error distribution. Similar identification issues arise with other $T^{1/2}$-convergent non-Gaussian QML estimators. See Berkes and Horváth (2004), Fan et al. (2014) and Hill (2014a).

Remark 3. If $E[\epsilon_{i,t}^4] = \infty$, then QMTTL is sub-$T^{1/2}$-convergent, a rate that is too slow for our portmanteau statistic under non-negligibility $k_{i,T}^{(\mathcal{E})}/T \to (0, 1)$. Since QMTTL dominates other plug-ins in terms of heavy tail robustness (see Hill, 2014a, for simulation evidence), and only negligibility $k_{i,T}^{(\mathcal{E})}/T \to 0$ guarantees identification of spillover, the analyst may want to commit to QMTTL and use a redescending transform $\psi = \check{\psi}$ and negligibility $k_{i,T}^{(\mathcal{E})}/T \to 0$ for robust volatility spillover Q-tests. We show by simulation in Section 5 that this combination works quite well.

3.2. Asymptotic power against spillover

Expansion (13) shows that if there is no spillover, then we must have $(T/\mathfrak{S}_T)^{1/2}E[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}] \to 0$, suggesting a global spillover alternative

$$H_1 : \left|E\left[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}\right]\right| \to (0, \infty].$$

This allows for moment divergence since under spillover H_1 it is possible to have $E[\epsilon_{1,t}^2\epsilon_{2,t-h}^2] = \infty$ if either $E[\epsilon_{i,t}^4] = \infty$. If both $E[\epsilon_{i,t}^4] < \infty$ then by dominated convergence and negligibility $k_T^{(\mathcal{E})}/T \to 0$ the alternative is simply $E[\mathcal{E}_{1,t}\mathcal{E}_{2,t-h}] \neq 0$.

This reveals a shortcoming of non-negligibility $k_{i,T}^{(\mathcal{E})}/T \to (0, 1)$. Under the alternative $E[\mathcal{E}_{1,t}\mathcal{E}_{2,t-h}]$ is non-zero and may be unbounded, but $E[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}] \nrightarrow E[\mathcal{E}_{1,t}\mathcal{E}_{2,t-h}]$ is possible. Hence it is possible that $E[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}] \to 0$ even when there is spillover $E[\mathcal{E}_{1,t}\mathcal{E}_{2,t-h}] \neq 0$. The only way to ensure H_1 identifies spillover is to use *unbounded* thresholds, and hence *negligible* fractiles $k_{i,T}^{(\mathcal{E})}/T \to 0$. This in turn requires a redescending transform $\psi = \check{\psi}$ to ensure robustness to heavy tails.

In general $T/\mathfrak{S}_T \to \infty$ is assured for any case. This follows since with intermediate order thresholds $\mathfrak{S}_T$ is $o(T)$, cf. Hill and Aguilar (2013, Appendix B), and with central order thresholds by construction $\mathfrak{S}_T = O(1)$. This ensures under the global alternative H_1 that $(T/\mathfrak{S}_T)^{1/2}|E[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}]| \to \infty$ for some h irrespective of heavy tails, promoting test consistency if $h \leq H$.

Theorem 3.2. *Under Assumption* B, D, E, FE, PQ, *and* T

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

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

If there is spillover $E[\mathcal{E}_{1,t}\mathcal{E}_{2,t-h}] \neq 0$, and since $\lim_{T\to\infty}\mathfrak{S}_T/T = 0$, then as long as $\liminf_{T\to\infty}|E[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}]| > 0$ for some $h \leq H$ then the Q-statistic is consistent $\hat{Q}_T^{(\mathcal{E})}(H) \xrightarrow{p} \infty$. Further, necessarily $\liminf_{T\to\infty}|E[\psi_{1,T,t}^{(\mathcal{E})}\psi_{2,T,t-h}^{(\mathcal{E})}]| > 0$ under spillover provided trimming is negligible $k_{i,T}^{(\mathcal{E})}/T \to 0$.

Corollary 3.3. *Let Assumption* B, D, E, PQ, *and* T *hold. If* $k_{i,T}^{(\mathcal{E})} \to \infty$ *and* $k_{i,T}^{(\mathcal{E})}/T \to 0$,*and* $E[\mathcal{E}_{1,t}\mathcal{E}_{2,t-h}] \neq 0$ *for some* $h \leq H$,*then* $\hat{Q}_T^{(\mathcal{E})}(H) \xrightarrow{p} \infty$.

4. Plug-in and fractile selection

It remains to decide on a plug-in $\hat{\theta}_T$ and how much trimming k_T to use. We first discuss $\hat{\theta}_T$. We then discuss a method for smoothing over a parametric version of k_T as a way to avoid choosing a particular k_T. In these cases we focus on the portmanteau statistic $\hat{Q}_T^{(\mathcal{E})}$, and relegate to Aguilar and Hill (2014) related details for score statistics. Finally, we discuss how differences in the test statistic affect the fractile choice.

4.1. Plug-in selection

Available plug-ins depend intimately on the type of transform and thresholds used. The score statistic $\hat{W}_T^{(m)}$ uses simple trimming and negligible thresholds. The remaining score statistics $\hat{W}_T^{(\mathcal{E})}$ and $\hat{W}_T^{(\mathfrak{e})}$ have available a broad class of redescending transforms and negligible thresholds, while the portmanteau statistics $\hat{Q}_T^{(\mathcal{E})}$ and $\hat{Q}_T^{(\mathfrak{e})}$ have available an even broader class of transforms and possibly non-negligible thresholds. Further, in order to deduce if a plug-in is valid for a portmanteau statistic, we must characterize the scale $\mathfrak{V}_T$, which depends on the transform, thresholds, and volatility model.

We focus on a linear GARCH model for the sake of brevity, hence

$$h_{i,t}^2 = \omega_i^0 + \alpha_i^0 y_{i,t-1}^2 + \beta_i^0 h_{i,t-1}^2, \quad \omega_i^0 > 0, \quad \alpha_i^0, \beta_i^0 \geq 0,$$
$$\alpha_i^0 + \beta_i^0 > 0, \tag{14}$$

and in terms of robustness we focus on simple trimming $\psi(u, c) = \check{\psi}(u, c) = uI(|u| \leq c)$. Define moment suprema $\kappa_i \equiv \arg\inf\{\alpha > 0 : E|\epsilon_{i,t}|^\alpha < \infty\}$. In order to gauge which plug-ins are valid, expansion (13) shows that we first require the order of $\mathfrak{V}_T$ in (11). The exact rate for $\mathfrak{V}_T$ is complicated by the product structure $E[\mathcal{E}_{1,T,t}^{*2}] \times E[\mathcal{E}_{2,T,t}^{*2}]$. Let $L(T) \to \infty$ be slowly varying, the order of which may change from place to place.

Lemma 4.1 ($\hat{Q}_T^{(\mathcal{E})}$*: Rate for* $\mathfrak{V}_T$). *Let the conditions of Theorem 3.1 hold.*

a. *If* $k_{i,T}^{(\mathcal{E})}/T \to (0, 1)$ *then* $\|\mathfrak{V}_T\| \sim KT$.

b. *Let* $k_{i,T}^{(\mathcal{E})}/T \to 0$. *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* $\|\mathfrak{V}_T\| \sim T^{3-4/\kappa_1-4/\kappa_2}(k_{1,T}^{(\mathcal{E})})^{4/\kappa_1-1}(k_{2,T}^{(\mathcal{E})})^{4/\kappa_2-1}$.

The next result exploits Lemma 4.1 to deduce valid plug-ins for the portmanteau statistic, and forms the basis for Theorem 3.1.b.

Theorem 4.2. *Under the conditions of Theorem* 3.1, $\hat{Q}_T^{(\mathcal{E})}$ *can use* $\hat{\theta}_{i,T}$ *computed by: QMTTL if* $\kappa_i > 4$, *or for any* $\kappa_i > 2$ *if* $k_{i,T}^{(\mathcal{E})}/T \to 0$ *QMWLL if* $\epsilon_{i,t}$ *have a zero median and* $E|\epsilon_{i,t}| = 1$; *Log-LAD if* $\ln(\epsilon_{i,t}^2)$ *has a zero median; and QML only if both* $\kappa_i > 4$.

4.2. Fractile selection: p-value smoothing

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 $k_T = k_{i,T}^{(\mathcal{E})}$ is used for $i = 1, 2$.

Theorem 3.1 shows that if redescending $\check{\psi}$ is used then $\hat{Q}_T^{(\mathcal{E})}$ is robust to any $\mathfrak{V}_T^{1/2}$-convergent $\hat{\theta}_T$ under negligibility $k_T/T \rightarrow 0$, and $\hat{Q}_T^{(\mathcal{E})}$ is robust to any $T^{1/2}$-convergent $\hat{\theta}_T$ under non-negligibility $k_T/T \rightarrow (0, 1)$. Power, however, is optimized only when non-negligibility $k_T/T \rightarrow 0$ is imposed. Conversely, $k_T/T \rightarrow (0, 1)$ implies the thresholds are bounded, which effectively removes important information about spillover. In simulation experiments here and elsewhere[7] we repeatedly find the fractile type $k_T(\lambda) \equiv [\lambda T/\ln(T)]$ with small $\lambda \in (0, 1]$ works exceptionally well 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.[8] See Section 5. If a non-redescending ψ is used, as in truncation, then the threshold must be bounded for heavy tail robustness, hence $k_T/T \rightarrow (0, 1)$. We therefore define $k_T(\lambda) = [(\lambda - \iota)T]$ for $\lambda \in (0, 1]$ and infinitesimal $\iota > 0$ such that $\lambda - \iota \in (0, 1)$.

In theory any trimming parameter value $\lambda \in (0, 1]$ is valid. However, clearly a greater amount of trimming will generate small sample bias and therefore cause a size distortion and weaken power. Similarly, too little trimming can lead to a size distortion due to the presence of influential extreme values. This is clearly shown in our simulation study. Although using any $\lambda \in [.01, .10]$ leads to fairly sharp results in many contexts (see the citations in footnote 7), it may be preferred to use many feasible λ at once. Hill and Aguilar (2013) and Hill (2012) solve the challenge of using many λ at once, while retaining standard asymptotics, by using a p-value occupation time on $\lambda \in (0, 1]$.[9]

Let $p_T(\lambda)$ denote the asymptotic p-value for $\hat{Q}_T^{(\mathcal{E})}$, that is $p_T^{(\mathcal{E})}(\lambda) \equiv P(\hat{Q}_T^{(\mathcal{E})}(H) \leq \chi_H)$ where χ_H is distributed $\chi^2(H)$. The occupation time of $p_T^{(\mathcal{E})}(\lambda) \leq \alpha$ on $[\underline{\lambda}, 1]$ for tiny $\underline{\lambda} > 0$ and significance level $\alpha \in (0, 1)$ is

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

Thus, $\tau_T^{(\mathcal{E})}(\alpha)$ is the uniform measure of the sub-interval of $[\underline{\lambda}, 1]$ on which the null hypothesis of no spillover is rejected.

The following is proved in Hill (2012, Theorem 3.1) for the case of simple trimming with $k_T(\lambda) \equiv [\lambda T/\ln(T)]$. A similar argument extends to all transforms allowed under Theorem 3.1, in particular to the case of bounded thresholds associated with an intermediate order statistic $k_T(\lambda) \equiv [(\lambda - \iota)T]$ on $[\underline{\lambda}, 1]$ where $\underline{\lambda} > \iota > 0$ are tiny values.

Theorem 4.3. *Let Assumption* B, D, E, FE, PQ, *and* T *hold, and let* $\{\mathcal{U}(\lambda) : \lambda \in [0, 1]\}$ *be a stochastic process with almost surely uniformly continuous sample paths, where pointwise* $\mathcal{U}(\lambda)$ *is uniformly distributed on* $[0, 1]$. *If* $H_0^{(\epsilon)}$ *is true then* $\tau_T^{(\mathcal{E})}(\alpha) \xrightarrow{d} (1-\underline{\lambda})^{-1} \int_{\underline{\lambda}}^{1} I(\mathcal{U}(\lambda) \leq \alpha) d\lambda$. *In particular,* $\lim_{T\to\infty} P(\tau_T^{(\mathcal{E})}(\alpha) > \alpha) \leq \alpha$, *while if spillover occurs by* $h \leq H$, *then* $\tau_T^{(\mathcal{E})}(\alpha) \xrightarrow{p} 1$.

Remark 4. Under the null hypothesis, $\tau_T^{(\mathcal{E})}(\alpha)$ tends to be small since we rarely reject the null, where the probability that $\tau_T^{(\mathcal{E})}(\alpha) > \alpha$ is no greater than α as $T \rightarrow \infty$. Further, as long as spillover takes place in some horizon $h \leq H$, then $\tau_T^{(\mathcal{E})}(\alpha) \xrightarrow{p} 1$. Thus, we reject the null if $\tau_T^{(\mathcal{E})}(\alpha) > \alpha$ where the probability of a Type I error is no greater than α. Notice $\tau_T^{(\mathcal{E})}(\alpha)$ operates like a smoothed test statistic since we reject the null when it is big. In addition, $\tau_T^{(\mathcal{E})}(\alpha)$ operates like its own p-value compliment since it is bounded in $[0, 1]$ and the probability it is larger than the nominal size α is no greater than α. In practice a discretized version is used, where a simple form is

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

$$\text{where } T_{\underline{\lambda}} \equiv \sum_{i=1}^{T} I\left(i/T \geq \underline{\lambda}\right). \tag{15}$$

4.3. Fractile selection: test differences

There are two remaining challenges with trimming: trimming $\mathcal{E}_{i,t} = \epsilon_{i,t}^2 - 1$ or $\epsilon_{i,t}$ involves two trimmed objects for each equation $m_{h,t}$, and in general $m_{h,t}$ is asymmetrically distributed.

First, each test statistic is a quadratic form of a standardized sample mean for testing the same moment condition $E[m_{h,t}] = E[\mathcal{E}_{1,t}\mathcal{E}_{2,t-h}] = E[(\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-h}^2 - 1)] = 0$. Trimming $\mathcal{E}_{i,t}$ or $\epsilon_{i,t}$ involves two objects (e.g. $\mathcal{E}_{1,t}$ and $\mathcal{E}_{2,t}$), but trimming $m_{h,t}$ by its large values involves just itself. This implies that if the same $k_T^{(m)} = k_T^{(\mathcal{E})}$ is used throughout, then we trim up to twice as many $m_{h,t}$ when we trim by $\mathcal{E}_{i,t}$ or $\epsilon_{i,t}$, and we trim roughly twice the number if $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are mutually independent. Consequently, setting $k_T^{(m)} = 2k_T^{(\mathcal{E})}$ ensures an equable amount of trimming across tests, at least under the null of mutual independence.

Second, the score statistic $\hat{W}_T^{(m)}$ involves trimming $m_{h,t} = (\epsilon_{1,t}^2 - 1)(\epsilon_{2,t-h}^2 - 1)$, where $m_{h,t}$ may be skewed right. Since there will be in general a propensity for positive values in $m_{h,t}$, we do not want to over-trim the right tail since that leads to bias. See Hill and Aguilar (2013). This implies trimming should favor the left tail: $\liminf_{T\to\infty} k_{1,T}^{(m)}/k_{2,T}^{(m)} > 1$. Thus, in conjunction with the above argument, if equable trimming across cases is desired and bias is to be minimized, then $\{k_{1,T}^{(m)}, k_{2,T}^{(m)}\}$ should be set such that $k_{1,T}^{(m)} > k_{2,T}^{(m)}$ and that the total number of $m_{h,t}'s$ trimmed is roughly equal to $2k_T^{(\mathcal{E})}$.

We show by simulation in Aguilar and Hill (2014) that symmetrically trimming $m_{h,t}$ produces large size distortions, and that values of $k_{r,T}^{(m)} = [\lambda_r T/\ln(T)]$ with $\{\lambda_1, \lambda_2\}$ set at or close to $\{.035, .005\}$ works the best in general for thin and heavy tailed errors. Nevertheless, $\hat{W}_T^{(m)}$ typically results in the largest size distortions of all our statistics. Further, $k_T^{(\mathcal{E})} = [\lambda T/\ln(T)]$ for any $\lambda \in [.01, .10]$ works well for trimming for $\hat{W}_T^{(\mathcal{E})}$ and $\hat{Q}_T^{(\mathcal{E})}$. Finally, in terms of the number of observations trimmed, setting $\{\lambda_1, \lambda_2\} \approx \{.035, .005\}$ for $\hat{W}_T^{(m)}$ aligns with setting $k_T^{(\mathcal{E})} = [.025T/\ln(T)]$ for $\hat{W}_T^{(\mathcal{E})}$ and $\hat{Q}_T^{(\mathcal{E})}$.

5. Simulation study—portmanteau statistic

We now study the small sample performance of the portmanteau statistics summarized in Table 1. We first present a base-case, and then supplement the study with various robustness checks. We then run a separate simulation study showing the usefulness of

7 See Hill (2012, 2013a,b, 2014a) and Hill and Aguilar (2013).

8 *Fast* in the sense $\lambda T/\ln(T) \rightarrow \infty$ is faster than $T^{\delta} \rightarrow \infty$ and $\lambda \ln(T) \rightarrow \infty$ 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.

9 See Hill and Aguilar (2013) for references and details on other methods for choosing λ for robust score statistics, including minimizing the covariance determinant, a technique used in the outlier robust estimation literature. In simulation experiments Hill and Aguilar (2013) show p-value occupation time worked comparatively much better. In our simulations, below, we find p-value occupation time leads to sharp empirical size but lower power, while merely picking $\lambda = .05$ leads to much higher power and a slight size distortion with a magnitude of about 1% in many cases.

Table 2
DGP.

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

using spillovers for volatility forecasts, and present required theory for comparing forecast models.

In general, score tests exhibit size distortions and have low power in various cases treated below, making the tests sub-optimal relative to our portmanteau tests. We therefore study the score tests in the supplemental appendix (Aguilar and Hill, 2014).

5.1. Base-case

The data generating process is a bivariate GARCH(1, 1):

$$y_{i,t} = h_{i,t}\epsilon_{i,t}, \qquad E\left[\epsilon_{i,t}\right] = 0, \qquad E\left[\epsilon^2_{i,t}\right] = 1, \tag{16}$$
$$h^2_{i,t} = .3 + \alpha^0_{i,i}y^2_{i,t-1} + \beta^0_{i,i}h^2_{i,t-1} + \alpha^0_{i,j}y^2_{j,t-1} + \beta^0_{i,j}h^2_{j,t-1},$$
$$i \neq j = 1, 2.$$

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

In the base-case we work with the portmanteau statistic $\hat{Q}^{(\mathcal{E})}_T$, which transforms $\mathcal{E}_{i,t} = \epsilon^2_{i,t} - 1$ and recenters, and $\hat{Q}^{(\mathfrak{E})}_T$, which transforms $\epsilon_{i,t}$, squares and recenters. We compute $\hat{Q}^{(\mathcal{E})}_T$ and $\hat{Q}^{(\mathfrak{E})}_T$ with $H = 5$ lags. We use simple trimming $\psi(u, c) = uI(|u| \leq c)$ with a fractile type $k_T = [\lambda T/\ln(T)]$. We either handpick $\lambda = .05$, which lies in the middle of the range [.01, .10] on which our tests work well, or we smooth the p-value using occupation time on [.01, 1]. In order to avoid sampling error and isolate the efficacy of the proposed tests, the base-case uses the true parameter value θ^0 when constructing the Q-statistics. We explore a null of no spillover from $y_{2,t-h}$ to $y_{1,t}$ at any lag $h \geq 1$, and two alternatives. Alternative one ($Alt1$) is "weak" spillover from $y_{2,t-h}$ to $y_{1,t}$ at lag $h \geq 1$, and alternative two ($Alt2$) is "strong" spillover from $y_{2,t-h}$ to $y_{1,t}$ at lag $h \geq 1$.

Simulation results are presented in Table 3. We find that the tests are reasonably sized regardless of the distribution of the innovations. For instance, at the 10% nominal level the empirical size among the 16 test combinations presented range from 10% to 17.4%. Using the p-value occupation time leads to sharp empirical size, but tends to reduce the empirical power of the tests. Overall $\hat{Q}^{(\mathcal{E})}_T$ has roughly the same size and better power than $\hat{Q}^{(\mathfrak{E})}_T$ under simple trimming.

A few examples can better illustrate specific features of these tests. Consider, for instance, $\hat{Q}^{(\mathcal{E})}_T$ at the 10% level with $\lambda = .05$. When the innovations are Paretian, the empirical size and $Alt2$ power are .123 and .520, respectively. However, size and power fall to .107 and .401 with occupation time. In the Gaussian case for $\hat{Q}^{(\mathcal{E})}_T$ the drop in rejection frequencies is even more stark: at the 10% level size falls from .130 to .114, while $Alt2$ power falls from .800 to .454. A likely reason for this loss of power is the structure of the occupation time, which smooths over a large window of fractiles $k_T = [\lambda T/\ln(T)]$. Using these larger k_T values implies that more large GARCH innovations are removed, which are observations that provide important information about spillovers. Thus, logically, smoothing the p-value over a large window of λ leads to lower power than simply using a small fixed λ.

Table 3
(Base-Case) $\hat{Q}_T$-test rejection frequencies at the (1%, 5%, 10%) levels. We use 10, 000 simulated paths, $T = 1000$, $H = 5$ lags, no plug-in, and simple trimming with fractile $k_T = [\lambda T/\ln(T)]$. We use $\lambda = 0.05$ or p-value occupation time (Oc.Time). The null hypothesis is no spillover from $y_{2,t}$ to $y_{1,t}$. Alternative 1 indicates weak spillover and Alternative 2 indicates strong spillover.

$\epsilon \sim N(0, 1)$		$\hat{Q}^{(\mathcal{E})}_T$	$\hat{Q}^{(\mathfrak{E})}_T$
$\lambda = 0.05$	Null—no spill	(.023,.076,.130)	(.022,.074,.127)
	Alt1—weak	(.410,.593,.684)	(.341,.531,.631)
	Alt2—strong	(.561,.726,.800)	(.488,.665,.749)
Oc.Time	Null—no spill	(.015,.061,.114)	(.013,.056,.107)
	Alt1—weak	(.134,.266,.363)	(.068,.146,.214)
	Alt2—strong	(.195,.351,.454)	(.092,.176,.246))
$\epsilon \sim P_{2.5}$		$\hat{Q}^{(\mathcal{E})}_T$	$\hat{Q}^{(\mathfrak{E})}_T$
$\lambda = 0.05$	Null—no spill	(.044,.085,.123)	(.045,.085,.122)
	Alt1—weak	(.239,.349,.421)	(.215,.321,.386)
	Alt2—strong	(.318,.444,.520)	(.281,.398,.471)
Oc.Time	Null—no spill	(.017,.059,.107)	(.017,.057,.104)
	Alt1—weak	(.108,.223,.309)	(.045,.102,.158)
	Alt2—strong	(.164,.305,.402)	(.057,.120,.180)

We also find that power increases with the strength of the spillover in nearly every case. For example, at the 10% level and with $\lambda = .05$, power against weak spillover ($Alt1$) for Paretian innovations is .421, but it increases to .520 when testing against strong spillover ($Alt2$). In the Gaussian case $Alt1$ power is .684, but $Alt2$ power is .800.

5.2. Robustness checks against the base-case

We now explore several directions of robustness to the base-case.

5.2.1. Alternatives to simple trimming

We first explore Tukey's bisquare $\psi(u, c) = u(1 - (u/c)^2)^2 I(|u| \leq c)$ and the exponential variant $\psi(u, c) = u\exp\{-|u|/c\} I(|u| \leq c)$. The results in Table 7 of Appendix C suggest that the findings in the base-case are quite robust to choice of redescender. Consider $\hat{Q}^{(\mathcal{E})}_T$ under Pareto innovations with occupation time. The size/power combinations for the three hypotheses at the 10% level across simple trimming, Tukey, and exponential cases exhibit only minor differences: (.107, .309, .402), (.104, .201, .244), and

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

$(.106, .298, .382)$. The exponential transform, however, leads to decent size and overall higher power than Tukey's bisquare, and in some cases substantially higher power. In general for $\hat{Q}_T^{(\mathcal{E})}$ the exponential transform is best in terms of size and power across cases and innovation distribution tails. Finally, power is quite poor for the other Q-statistic $\hat{Q}_T^{(\epsilon)}$ when the transform is Tukey's bisquare or exponential, and since power is also lower under simple trimming this statistic is sub-optimal. In summary, so far with negligible trimming *the dominant statistic is* $\hat{Q}_T^{(\mathcal{E})}$ *with an exponential transform.*

5.2.2. Plug-in $\hat{\theta}_{i,T}$ for θ_i^0

We use Hill (2014a)'s QMTTL estimator $\hat{\theta}_{i,T}$ as a plug-in for $\theta_i^0 = [\omega_i^0, \alpha_i^0, \beta_i^0]'$ from univariate GARCH models $y_{i,t} = h_{i,t}\epsilon_{i,t}$ where $h_{i,t}^2 = \omega_i^0 + \alpha_i^0 y_{i,t-1}^2 + \beta_i^0 h_{i,t-1}^2$. Let $\tilde{h}_{i,t}^2(\theta_i)$ be the iterated volatility process with start value $\tilde{h}_{i,1}^2(\theta_i) = \omega_i$, and $\tilde{h}_{i,t}^2(\theta) = \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i \tilde{h}_{i,t-1}^2(\theta_i)$ for $t \geq 2$, and define the error $\tilde{\epsilon}_{i,t}(\theta_i) = y_{i,t}/\tilde{h}_{i,t}(\theta_i)$ and $\tilde{\mathcal{E}}_{i,t}(\theta_i) \equiv \tilde{\epsilon}_{i,t}^2(\theta_i) - 1$. Define left and right tail observations $\tilde{\mathcal{E}}_{i,t}^{(1)}(\theta_i) \equiv \tilde{\mathcal{E}}_{i,t}(\theta_i) I(\tilde{\mathcal{E}}_{i,t}(\theta_i) < 0)$ and $\tilde{\mathcal{E}}_{i,t}^{(2)}(\theta_i) \equiv \tilde{\mathcal{E}}_{i,t}(\theta_i) I(\tilde{\mathcal{E}}_{i,t}(\theta_i) \geq 0)$, their order statistics $\tilde{\mathcal{E}}_{i,(1)}^{(1)}(\theta_i) \leq \tilde{\mathcal{E}}_{i,(2)}^{(1)}(\theta_i) \leq \cdots \tilde{\mathcal{E}}_{i,(n)}^{(1)}(\theta_i) \leq 0$ and $\tilde{\mathcal{E}}_{i,(1)}^{(2)}(\theta_i) \geq \tilde{\mathcal{E}}_{i,(2)}^{(2)}(\theta_i) \geq \cdots \tilde{\mathcal{E}}_{i,(n)}^{(2)}(\theta_i) \geq 0$, and indicators $\widehat{\mathfrak{I}}_{i,T,t}^{(\tilde{\mathcal{E}})}(\theta_i) \equiv I(\tilde{\mathcal{E}}_{i,(\tilde{k}_{1,T}^{(\mathcal{E})})}^{(1)}(\theta_i) \leq \tilde{\mathcal{E}}_{i,t}(\theta_i) \leq \tilde{\mathcal{E}}_{i,(\tilde{k}_{2,T}^{(\mathcal{E})})}^{(2)}(\theta_i))$ and $\widehat{\mathfrak{I}}_{i,T,t}^{(y)} \equiv I(|y_{i,t}| \leq y_{i,(\tilde{k}_T^{(y)})}^{(a)})$ where $\{\tilde{k}_{1,T}^{(\mathcal{E})}, \tilde{k}_{2,T}^{(\mathcal{E})}, \tilde{k}_T^{(y)}\}$ are intermediate order sequences. The criterion is

$$\hat{\mathfrak{Q}}_{i,T}(\theta_i) \equiv \sum_{t=1}^{T} \left\{ \ln \tilde{h}_{i,t}^2(\theta_i) + \tilde{\epsilon}_{i,t}^2(\theta_i) \right\} \widehat{\mathfrak{I}}_{i,T,t}^{(\tilde{\mathcal{E}})}(\theta_i) \widehat{\mathfrak{I}}_{i,T,t-1}^{(y)}. \tag{17}$$

The estimator $\hat{\theta}_{i,T} \equiv \arg\inf_{\theta \in \Theta}\{\hat{\mathfrak{Q}}_{i,T}(\theta)\}$ is computed on $\Theta = [0, 1]^3$, with fractiles $\tilde{k}_{2,T}^{(\mathcal{E})} = \max\{1, [.025T/\ln(T)]\}$ and $\tilde{k}_{1,T}^{(\mathcal{E})} = [1.45(\tilde{k}_{2,T}^{(\mathcal{E})})^{.2}T^{.8}]$, and $\tilde{k}_T^{(y)} = \max\{1, [.1\ln(T)]\}$ since for $\epsilon_{i,t} \sim P_{2.5}$ or $\epsilon_{i,t} \sim N(0, 1)$ this leads to a sharp and approximately normal estimator in small samples, and an asymptotically unbiased and normal estimator.[11]

In Table S.6 and S.7 of the supplemental appendix (Aguilar and Hill, 2014) we find that the Q-tests with the QMTTL plug-in lead to roughly the same results as when the true parameter value is used, irrespective of the redescender. This finding holds for handpicked λ and p-value occupation time. In Table S.11 and S.12 we compare the Q-tests across all transforms (simple trimming, Tukey, exponential, truncation), with and without a plug-in. Again, the use of QMTTL as a plug-in does not qualitatively alter test results.

5.2.3. Tail thickness, spillover direction and a non-robust test

Next, we vary the error tail thickness and compare disparate trimming strategies. The entries in Table S.8 of the supplemental appendix focus on $\hat{Q}_T^{(\mathcal{E})}$ for three types of tail thickness: fat tails $\kappa_1, \kappa_2 = \{2.5, 2.5\}$, moderate tails $\kappa_1, \kappa_2 = \{6, 6\}$, and thin tails $\kappa_1, \kappa_2 = \{\infty, \infty\}$ being the Gaussian case. Moreover, we explore three trimming strategies: $\lambda = 0$, which implies no trimming and amounts to a non-standardized version of Hong (2001)'s test; handpicking $\lambda = .05$; and occupation time. We find that trimming with any of the redescenders, either by handpicking λ 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. Furthermore, these results are robust to the use of a QMTTL plug-in, as depicted in Table S.9.

In Table S.10 of the supplemental appendix we explore the tail differential across the errors and the direction of spillover. We consider four cases. In Case A $y_{2,t}$, which has a fat tailed error, spills over into $y_{1,t}$, which has an equally fat tailed error, with $\kappa_1, \kappa_2 = \{2.5, 2.5\}$. In Case B $y_{2,t}$, with a thin tailed error, spills over into $y_{1,t}$, with a fat tailed error, where $\kappa_1, \kappa_2 = \{2.5, \infty\}$. 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 other aspects of this exercise are identical to the base-case. For brevity, we utilize handpicked $\lambda = .05$ only. Notice Cases A and D reflect the findings previously displayed in Table 3. 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 $\epsilon_{2,t}$. Thus, evidence against spillover may arise solely from a heavy tailed error.

5.2.4. Fixed quantile trimming and truncation

In Table S.11 of the supplemental appendix we explore fixed quantile trimming (simple trimming, Tukey's bisquare, exponential) and truncation, where $k_T = [\lambda T]$ with either $\lambda = .05$ or p-value occupation time.

In the trimming cases we find that using a fractile form $k_T = [\lambda T]$ leads to sharp size, but sacrifices power simply because it entails removing roughly 7 times as many observations relative to $[\lambda T/\ln(T)]$: the removal of more large values than $[.05T/\ln(T)] = 7$ improves the accuracy of $\hat{Q}_T^{(\mathcal{E})}$ and $\hat{Q}_T^{(\epsilon)}$ under the null, but removes important information about spillover under the alternative. Recall that in theory, as the sample size grows $T \to \infty$ only the negligible trimming case $k_T/T \to 0$ can guarantee identification of the alternative, whereas with iid symmetric errors, the null is necessarily identified. It is therefore not surprising that empirical size is sharp and power is weak even when $T = 1000$. Roughly the same outcome occurs with fixed quantile truncation. These findings are robust to a QMTTL plug-in, as depicted in Table S.12.

Truncation works very well. Empirical size is good when $\lambda = .05$, but, contrary to any other non-truncation case, size is slightly skewed upward when p-value occupation time is used in the Gaussian case. Power is quite good in general. Overall, negligibility with exponential smoothing, or truncation, work roughly as well, with slight differences: negligibility with exponential smoothing leads to slightly better size in some cases, while truncation leads to higher power when the innovations are heavy tailed. The power improvement with truncation, though, can be quite large: in the Paretian case with exponential smoothing *Alt*2 power at the 10% level is .524 (size is.086) but with truncation power is .814 (size is .115). The advantage of truncation arguably lies in

[11] Recall $\tilde{\epsilon}_{i,t}^2 - 1$ governs QML asymptotics, cf. Francq and Zakoïan (2004), and has a bounded left tail and unbounded right tail. Thus, a few very large right-tail errors are trimmed to ensure Gaussian asymptotics, and correspondingly far more small left tail errors are trimmed to ensure small and large sample unbiasedness. The asymmetric balance $\tilde{k}_{1,T}^{(\mathcal{E})} = [(2^{1-2/\kappa})(1 - 1/\kappa)^{-1}(\tilde{k}_{2,T}^{(\mathcal{E})})^{1-2/\kappa}T^{2/\kappa}]$ ensures QMTTL identification for the Pareto error class treated in this study, and ensures identification for any $\kappa > 2$ in the Gaussian case (Hill, 2014a, Section 2.3). The fractile balance implies $\tilde{k}_{1,T}^{(\mathcal{E})} > \tilde{k}_{2,T}^{(\mathcal{E})}$ for any $T \geq 1$ and $\kappa > 2$, and if $\kappa = 2.5$ then $\tilde{k}_{1,T}^{(\mathcal{E})} = [1.45(\tilde{k}_{2,T}^{(\mathcal{E})})^{.2}T^{.8}]$. Trimming by $y_{i,t-1}$ is not required in theory, but aids small sample performance since large $y_{i,t-1}$ may adversely affect volatility estimation.

In general using $\tilde{k}_{1,T}^{(\mathcal{E})} = [\lambda T/\ln(T)]$ for small λ and $\tilde{k}_{1,T}^{(\mathcal{E})} > \tilde{k}_{2,T}^{(\mathcal{E})}$ diminishes small sample bias, although using symmetric trimming $\tilde{k}_{1,T}^{(\mathcal{E})} = \tilde{k}_{2,T}^{(\mathcal{E})}$ still leads to an accurate estimator in small sample experiments. In general for either Pareto or Gaussian errors, and any scheme $\tilde{k}_{1,T}^{(\mathcal{E})} \geq \tilde{k}_{2,T}^{(\mathcal{E})}$, QMTTL has better small sample properties than QML, but also inherently heavy tail robust Log-LAD (Peng and Yao, 2003), Laplace QML (Berkes and Horváth, 2004; Zhu and Ling, 2011), and Power Law QML (Berkes and Horváth, 2004, in this case QMTTL is better when $k_{1,T}^{(\mathcal{E})} > k_{2,T}^{(\mathcal{E})}$). This arises since the latter three do not directly counter the negative influence of innovation outliers in small samples. See Hill (2014a).

its use of the threshold as a replacement for sample extremes, while trimming merely removes these large values. The added information leads to a mild size distortion, but much higher power when the innovations are heavy tailed.

5.2.5. Recentering

The statistic $\hat{Q}_T^{(\mathfrak{E})}$ involves one of two methods for recentering the errors $\epsilon_{i,t}$: $\epsilon_{i,t}$ is transformed, squared, and recentered (the base case); or $\epsilon_{i,t}$ is transformed, recentered, squared, and recentered again. Table S.13 of the supplemental appendix explores both cases, and shows in general the latter method leads to over-rejection of the null when spillover is not present, and higher (non-size corrected) power. The reason evidently lies in the amount of sampling error: the first method involves one sample tail-trimmed mean used to recenter, while the second involves two sample means used to recenter.

5.2.6. Further checks: sample size, horizon, and trimming parameter

Finally, in Tables S.14–S.23 of the supplemental appendix we explore sample size T, test horizon H, and λ. We inspect the base-case statistics $\hat{Q}_T^{(\mathcal{E})}$ and $\hat{Q}_T^{(\mathfrak{E})}$. Tables S.14–S.17 suggest that size sharpens and power increases in nearly every case as sample size increases. In Tables S.18–S.21 we find that the probability of rejection seems to rise with the test horizon H. Table S.22 and S.23 suggest the Q-tests appear robust to choice of λ, with size and empirical power changing little as we alter the value of λ from .01 to .10. This finding holds for Pareto tails across redescending transforms.

5.2.7. Summary of test performance

The Q-test performs better than the score test in every case considered, cf. Aguilar and Hill (2014). P-value smoothing in many cases leads to a Q-test with roughly correct size, but in general with lower power. Slight size distortions (around 1%–3%) and potentially much higher power arise with fixed $\lambda = .05$. Further, trimming with increasing thresholds, rather than fixed thresholds, results in higher power since only negligibility leads asymptotically to identification of spillover $E[m_{h,t}] \neq 0$. We find that tail trimming with occupation time can lead to sharp size and reasonable power even under mildly fat tails. Truncation must use a fixed threshold to ensure standard asymptotics, but it does not remove sample information. The replacement of sample extremes with the threshold value leads to higher power than fixed threshold trimming. Non-smooth or smooth transforms lead to similar results in small samples, but smooth transforms like Tukey's bisquare and the exponential lead to sharper size faster as the sample size increases. In many cases the exponential leads to higher power, and has power comparable to the simple trimming case.

Of the three transforms with negligibility (simple trimming, Tukey, exponential), the exponential appears best. Of all transforms (simple trimming, Tukey, exponential, truncation), negligible trimming with exponential smoothing, and truncation, are best. Negligibility with exponential smoothing has a slight empirical size advantage, while truncation can promote much higher power with at most a slight size distortion.

5.3. The importance of spillovers—predictive accuracy

The ability to detect volatility spillover is useful in a range of applications. In this subsection we investigate whether incorporating spillovers between two assets can improve volatility forecasts. Our goal is to abstract from pre-testing for volatility spillover, and show merely that if there is spillover then inclusion of that information does in fact lead to a better forecast model. Our evaluation criterion follows from a class of loss functions presented in Patton (2011) that are robust to the choice of volatility proxy being used as a comparative benchmark. As a separate contribution in this paper, we incorporate robust estimation theory in Hill (2014a) in order to render a particular comparative forecast measure robust to heavy tails, and we present the asymptotic theory.

5.3.1. Heavy tail robust forecast improvement test

We focus on testing whether detecting spillovers from $y_{2,t-h}$ to $y_{1,t}$ can be used to improve on a volatility forecast model of $y_{1,t}$. We first simulate $y_{1,t}$ and $y_{2,t}$ according to the "Alt2—strong" case of strong spillover depicted in the last row of Table 2. The assets have a bivariate GARCH(1, 1) structure (16) with Pareto innovations and tail index $\kappa = 2.5$. We simulate 10,000 sample paths with sample size $T = 1000$. The iterated volatility process for $y_{1,t}$ is $\tilde{h}^2_{1,1}(\theta) = \omega_1$ and $\tilde{h}^2_{1,t}(\theta) = \omega_1 + \alpha_{1,1}y^2_{1,t-1} + \beta_{1,1}\tilde{h}^2_{1,t-1}(\theta) + \alpha_{1,2}y^2_{2,t-1} + \beta_{1,2}\tilde{h}^2_{2,t-1}(\theta_2)$ for $t \geq 2$, where $\tilde{h}^2_{2,1}(\theta_2) = \omega_2$ and $\tilde{h}^2_{2,t}(\theta_2) = \omega_2 + \alpha_{2,2}y^2_{2,t-1} + \beta_{2,2}\tilde{h}^2_{2,t-1}(\theta_2)$. We first estimate $\theta^0_2 = \{\omega^0_2, \alpha^0_{2,2}, \beta^0_{2,2}\}$, denoted $\hat{\theta}_{2,T}$, compute the iterated process $\tilde{h}^2_{2,1}(\hat{\theta}_{2,T})$, and then estimate $\theta^0_1 = \{\omega^0_1, \alpha^0_{1,1}, \beta^0_{1,1}, \alpha^0_{1,2}, \beta^0_{1,2}\}$ using $\tilde{h}^2_{2,1}(\hat{\theta}_{2,T})$ in place of $\tilde{h}^2_{2,t}(\theta_2)$. Denote the resulting estimate of θ^0_1 as $\hat{\theta}^{(S)}_{1,T}$ since this volatility model for $y_{1,t}$ contains *spillovers*, and compute volatilities $\hat{h}^{2(S)}_{1,t} \equiv \tilde{h}^2_{1,t}(\hat{\theta}^{(S)}_{1,T})$. In each case we use QMTTL for parameter estimation: we estimate $\hat{\theta}_{2,T}$ using criterion (17), and $\hat{\theta}_{1,T}$ with the augmented criterion

$$\hat{\mathfrak{Q}}_{1,T}(\theta_1) \equiv \sum_{t=1}^{T}\left\{\ln \tilde{h}^2_{1,t}(\theta_1) + \tilde{\epsilon}^2_{1,t}(\theta_1)\right\} \times \widehat{\mathcal{I}}^{(\tilde{\mathcal{E}})}_{1,T,t}(\theta_1)\widehat{\mathcal{I}}^{(y)}_{1,T,t-1}\widehat{\mathcal{I}}^{(y)}_{2,T,t-1} \tag{18}$$

where $\widehat{\mathcal{I}}^{(y)}_{i,T,t} \equiv I(|y_{i,t}| \leq y^{(a)}_{i,(\tilde{k}^{(y)}_T)})$ and

$$\widehat{\mathcal{I}}^{(\tilde{\mathcal{E}})}_{i,T,t}(\theta_i) \equiv I\left(\tilde{\mathcal{E}}^{(1)}_{i,(\tilde{k}^{(\mathcal{E})}_{1,T})}(\theta_i) \leq \tilde{\mathcal{E}}_{i,t}(\theta_i) \leq \tilde{\mathcal{E}}^{(2)}_{i,(\tilde{k}^{(\mathcal{E})}_{2,T})}(\theta_i)\right).$$

The QMTTL score equations are functions of $y_{1,t-1}$ *and* $y_{2,t-1}$, even if there are no spillovers $\alpha_{1,2} = \beta_{1,2} = 0$. Hence, negligible trimming by $y_{1,t-1}$ *and* $y_{2,t-1}$ helps improve the sample properties of QMTTL. Asymptotically, however, it is irrelevant whether we trim by $y_{1,t-1}$ and $y_{2,t-1}$ since heavy tail robustness does not require it for non-trivial GARCH. See Hill (2014a). The fractiles are $\tilde{k}^{(\mathcal{E})}_{2,T} = \max\{1, [.025T/\ln(T)]\}$, $\tilde{k}^{(\mathcal{E})}_{1,T} = [1.45(\tilde{k}^{(\mathcal{E})}_{2,T})^{.2}T^{.8}]$, and $\tilde{k}^{(y)}_T = \max\{1, [.1\ln(T)]\}$ as in Section 5.2.2.

Next, we use a heavy tail robust version of Patton (2011)'s quasi-likelihood based loss function with spillovers. Define $\hat{\epsilon}^{(S)}_{1,t} = y_{1,t}/\hat{h}^{(S)}_{1,t}$, $\widehat{\mathcal{E}}^{(S)}_{1,t} \equiv \hat{\epsilon}^{(S)2}_{1,t} - 1$, $\widehat{\mathcal{I}}^{(\mathcal{E}^{(S)})}_{1,T,t} \equiv I(\widehat{\mathcal{E}}^{(S)(1)}_{1,(\tilde{k}^{(\mathcal{E})}_{1,T})} \leq \widehat{\mathcal{E}}^{(S)}_{1,t} \leq \widehat{\mathcal{E}}^{(S)(2)}_{1,(\tilde{k}^{(\mathcal{E})}_{2,T})})$ and $\widehat{\mathcal{I}}^{(y)}_{i,T,t} \equiv I(|y_{i,t}| \leq y^{(a)}_{i,(\tilde{k}^{(y)}_T)})$. Patton (2011)'s quasi-likelihood function is $\hat{l}^{(S)}_t \equiv \sigma^2_{1,t}/\hat{h}^{2(S)}_{1,t} + \ln(\hat{h}^{2(S)}_{1,t})$ where $\sigma^2_{1,t}$ is an unbiased proxy for the true volatility process $h^2_{1,t}$. We will use $y^2_{1,t}$ as $\sigma^2_{1,t}$ for simplicity, hence $\hat{l}^{(S)}_t = y^2_{1,t}/\hat{h}^{2(S)}_{1,t} + \ln(\hat{h}^{2(S)}_{1,t})$ is just the QML criterion equation which of course does not lead to standard asymptotics when $E[\epsilon^4_{1,t}] < \infty$. In view of the volatility form, we therefore trim $\hat{l}^{(S)}_t$ by the centered squared error $\widehat{\mathcal{E}}_{1,t} \equiv \hat{\epsilon}^2_{1,t} - 1$ and observed variables $y_{1,t-1}$ and $y_{2,t-1}$, similar to Hill (2014a), for

Table 4
Entries are relative improvement index values $\hat{\mathcal{R}}_T$, with rejection frequencies % in parentheses. All tests are performed with 10,000 simulated paths, each of length 1000.

Hypothesis $\backslash\epsilon_t$	$N(0,1)$	$P_{2.5}$
Null—no spill	.165 (.053)	3.23 (.026)
Alt2—strong	.990 (.991)	8.36 (.694)

a robust quasi-likelihood loss function:

$$\hat{\mathcal{L}}_T^{(S)} = \frac{1}{T}\sum_{t=1}^{T} \hat{l}_t^{(S)} \widehat{\mathcal{I}}_{1,T,t}^{(\mathcal{E}^{(S)})} \widehat{\mathcal{I}}_{1,T,t-1}^{(y)} \widehat{\mathcal{I}}_{2,T,t-1}^{(y)}. \tag{19}$$

The fractiles $\{\tilde{k}_{1,T}^{(\mathcal{E})}, \tilde{k}_{2,T}^{(\mathcal{E})}, \tilde{k}_T^{(y)}\}$ are the same above.

We then estimate GARCH model (16) for $y_{1,t}$ without spillovers, as per the "Null—no spill" case in the fourth row of Table 2. The estimator for the non-spillover [NS] model $\hat{\theta}_{1,T}^{(NS)}$ is computed by the QMTTL criterion (18) using the non-spillover volatility function $\tilde{h}_{1,1}^2(\theta_1) = \omega_1$ and $\tilde{h}_{1,t}^2(\theta_1) = \omega_1 + \alpha_{1,1} y_{1,t-1}^2 + \beta_{1,1}\tilde{h}_{1,t-1}^2(\theta_1)$. Write $\hat{h}_{1,t}^{2(NS)} \equiv \tilde{h}_{1,t}^2(\hat{\theta}_{1,T}^{(NS)})$, and form the quasi-likelihood loss function without spillovers $\hat{l}_t^{(NS)} = y_{1,t}^2/\hat{h}_{1,t}^{2(NS)} + \ln(\hat{h}_{1,t}^{2(NS)})$ and the robust loss function

$$\hat{\mathcal{L}}_T^{(NS)} = \frac{1}{T}\sum_{t=1}^{T} \hat{l}_t^{(NS)} \widehat{\mathcal{I}}_{1,T,t}^{(\mathcal{E}^{(NS)})} \widehat{\mathcal{I}}_{1,T,t-1}^{(y)} \widehat{\mathcal{I}}_{2,T,t-1}^{(y)}, \tag{20}$$

where we now use $\hat{\epsilon}_{1,t}^{(NS)} = y_{1,t}/\hat{h}_{1,t}^{(NS)}$, $\widehat{\mathcal{E}}_{1,t}^{(NS)} \equiv \hat{\epsilon}_{1,t}^{(NS)2} - 1$, and

$$\widehat{\mathcal{I}}_{1,T}^{(\mathcal{E}^{(NS)})} \equiv I\left(\widehat{\mathcal{E}}_{1,(\tilde{k}_{1,T}^{(\mathcal{E})})}^{(NS)(1)} \le \widehat{\mathcal{E}}_{1,t}^{(NS)} \le \widehat{\mathcal{E}}_{1,(\tilde{k}_{2,T}^{(\mathcal{E})})}^{(NS)(2)}\right).$$

Notice for QMTTL and for the loss function $\hat{\mathcal{L}}_T^{(NS)}$ we still trim by both $y_{1,t-1}$ *and* $y_{2,t-1}$, even though $y_{2,t-1}$ does not enter the non-spillover model for $y_{1,t}$. This ensures both estimator and forecast loss criteria are the same across spillover and non-spillover models, hence naturally $\hat{\mathcal{L}}_T^{(NS)} \ge \hat{\mathcal{L}}_T^{(S)}$ for any T, and as discussed in Section 5.2.2 trimming by $y_{i,t-1}$ has no impact on asymptotics.

In order to gauge the importance of including spillovers for forecasting volatility, when they actually do occur, we compute the following statistic which we call the *Relative Improvement* index (see also Brownlees et al., 2011, p. 17):

$$\hat{\mathcal{R}}_T = 100 \times \left(1 - \frac{\hat{\mathcal{L}}_T^{(S)}}{\hat{\mathcal{L}}_T^{(NS)}}\right).$$

Since $\hat{\mathcal{L}}_T^{(NS)} \ge \hat{\mathcal{L}}_T^{(S)}$ by construction of the optimization problems, it follows $\hat{\mathcal{R}}_T \in [0, 100]$. Let $\mathcal{R}$ be the probability limit, $\hat{\mathcal{R}}_T \xrightarrow{p} \mathcal{R} \in [0, 100]$ which exists by Theorem 5.1, below. A small positive value occurs when including spillovers does not significantly improve the forecast model fit, while a large positive value indicates including spillovers matters for forecasting volatility. Indeed, by the asymptotic theory for QMTTL, under the null of no spillovers it follows that the use of spillovers does not lead to an improved volatility forecast hence $\mathcal{R} = 0$, and conversely when spillovers occur there is a volatility forecast improvement by including spillovers $\mathcal{R} \in (0, 100]$.

Notice $\hat{\mathcal{R}}_T$ re-scales a sample average forecast improvement statistic:

$$\frac{\hat{\mathcal{L}}_T^{(NS)}}{100}\hat{\mathcal{R}}_T = \frac{1}{T}\sum_{t=1}^{T}\left\{\hat{l}_t^{(NS)} \widehat{\mathcal{I}}_{1,T,t}^{(\mathcal{E}^{(NS)})} - \hat{l}_t^{(S)} \widehat{\mathcal{I}}_{1,T,t}^{(\mathcal{E}^{(S)})}\right\} \widehat{\mathcal{I}}_{1,T,t-1}^{(y)} \widehat{\mathcal{I}}_{2,T,t-1}^{(y)}.$$

This type of statistic measuring the divergence of forecast models is treated at length in Diebold and Mariano (1995); West (1996) and Patton (2011), amongst many others (e.g. Vuong, 1989; Brownlees et al., 2011). Since the non-spillover volatility model is nested in the spillover volatility model, it follows that the standardized statistic

$$\mathcal{LR}_T = \frac{1}{100}\frac{\hat{\mathcal{L}}_T^{(NS)}}{1/T\sum_{t=1}^{T}\widehat{\mathcal{E}}_{1,t}^{(NS)2}\widehat{\mathcal{I}}_{1,T,t}^{(\mathcal{E}^{(NS)})}}T\hat{\mathcal{R}}_T = \frac{\sum_{t=1}^{T}\left\{\hat{l}_t^{(NS)} \widehat{\mathcal{I}}_{1,T,t}^{(\mathcal{E}^{(NS)})} - \hat{l}_t^{(S)} \widehat{\mathcal{I}}_{1,T,t}^{(\mathcal{E}^{(S)})}\right\} \widehat{\mathcal{I}}_{1,T,t-1}^{(y)} \widehat{\mathcal{I}}_{2,T,t-1}^{(y)}}{1/T\sum_{t=1}^{T}\widehat{\mathcal{E}}_{1,t}^{(NS)2}\widehat{\mathcal{I}}_{1,T,t}^{(\mathcal{E}^{(NS)})}} \tag{21}$$

is a heavy tail robust Likelihood Ratio statistic. Thus, a heavy tail robust test of the predictive ability for nested volatility models reduces to a heavy tail robust quasi-likelihood ratio test. The asymptotic theory hinges on the theory of QMTTL, hence we require stationarity, a mixing condition, and Paretian tails if $E[\epsilon_{i,t}^4] = \infty$. The following is proved in Aguilar and Hill (2014).

Theorem 5.1. *Let Assumptions* D, E.2 *and* T *hold, let* $\epsilon_{i,t} \sim P_{2.5}$ *or* $\epsilon_{i,t} \sim N(0,1)$,*and let* $\tilde{k}_{2,T}^{(\mathcal{E})} = [\lambda T/\ln(T)]$ *for* $\lambda \in (0,1]$ *and* $\tilde{k}_{1,T}^{(\mathcal{E})} = [1.45(\tilde{k}_{2,T}^{(\mathcal{E})})^{.2}T^{.8}]$. *Under the null of no spillover there is, asymptotically, no forecast improvement by including spillovers* $\hat{\mathcal{R}}_T \xrightarrow{p} 0$ *and* $\mathcal{LR}_T \xrightarrow{d} \chi^2(2)$. *If there are spillovers in the sense* $\alpha_{1,2}+\beta_{1,2} > 0$ *then asymptotically there is a forecast improvement by including spillovers* $\hat{\mathcal{R}}_T \xrightarrow{p} (0, 100]$ *and* $\mathcal{LR}_T \xrightarrow{p} \infty$.

Remark 5. Recall from Section 5.2.2 that $\tilde{k}_{2,T}^{(\mathcal{E})} = [\lambda T/\ln(T)]$ for $\lambda \in (0,1]$ and $\tilde{k}_{1,T}^{(\mathcal{E})} = [1.45(\tilde{k}_{2,T}^{(\mathcal{E})})^{.2}T^{.8}]$ leads to an asymptotically unbiased QMTTL estimator in its limit distribution when $\epsilon_{i,t} \sim P_{2.5}$ or $\epsilon_{i,t} \sim N(0,1)$. We can easily broaden Theorem 5.1 to cover a general class of error distributions provided we assume $\{\tilde{k}_{1,T}^{(\mathcal{E})}, \tilde{k}_{2,T}^{(\mathcal{E})}\}$ ensure asymptotic unbiasedness for QMTTL. See Section 5.2.2, and see Hill (2014a).

Remark 6. Let χ_2^2 be distributed $\chi^2(2)$ and let χ_α be the $\alpha \in (0,1)$ tail quantile: $P(\chi_2^2 > \chi_\alpha) = \alpha$. The chi-squared limit for $\mathcal{LR}_T$ implies an asymptotic test of the hypothesis that including spillovers does not improve a volatility forecast: we reject $\mathcal{R} = 0$ at level α when $\mathcal{LR}_T > \chi_\alpha$.

5.3.2. *Forecast improvement test results*

We compute $\hat{\mathcal{R}}_T$ and $\mathcal{LR}_T$ across 10,000 samples under "Null—no spill" and "Alt2—strong" cases using either Gaussian or Paretian errors as usual. The results are summarized in Table 4.

Consider the "Alt2—strong" case. In the Gaussian case the simulation average $\hat{\mathcal{R}}_T$ is .990, and the $\mathcal{LR}_T$ test rejection frequency of the hypothesis of no forecast improvement $\mathcal{R} = 0$ is 99.1% at the 5% level. In the Paretian case the simulation average $\hat{\mathcal{R}}_T$ is 8.36 with an $\mathcal{LR}_T$ test rejection frequency 69.4% at the 5% level. It appears that including spillovers when they truly exist is useful when constructing one-step ahead volatility forecasts.

Finally, under the "Null—no spill" cases for Gaussian errors the simulation average $\hat{\mathcal{R}}_T$ is .165, and the $\mathcal{LR}_T$ test rejection frequency of the hypothesis of no forecast improvement $\mathcal{R} = 0$ is 5.3% at the 5% level. Under Paretian errors the simulation average $\hat{\mathcal{R}}_T$ is 3.23, with a rejection rate of 2.6%. Thus, including spillovers when they do not exist does not significantly improve upon a one-step ahead volatility forecast.

6. Empirical application

We now investigate the presence of volatility spillovers across five Exchange Traded Funds [ETF's] that proxy for specific asset

Table 5
Description 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
GSG	Commodities	S&P GSCI Commodity-Indexed Trust
REM	US Real Estate	FTSE NAREIT Mortgage Plus Fund

classes: US equity, fixed income, real estate, non-US equity, and commodities. We then show how the spillover test results lead to an improved volatility forecast model. Such findings could provide strategic and tactical guidance to fund managers.

Our sample consists of daily log returns from 01/03/2012 through 12/31/2012, resulting in 251 trading days. We restrict attention to 2012 since anomalous data points do not appear to exist, nor do we see clustered extremes in estimated residuals that may require an additional layer of outlier robustness (cf. Muler and Yohai, 2008; Boudt and Croux, 2010). This highlights several limitations of the theory developed in this paper. In order to focus explicitly on heavy tail robustness, we neither treat non-stationary processes, in particular those with breaks as in level or conditional volatility shifts, nor processes that allow for one-off or anomalous events. A more general approach that allows for breaks and anomalous events would allow for a broader window of time and therefore a large sample size. Consequently, we need to restrict our study to roughly a year, and we are interested in data that are current.

6.1. Data

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 Rogibon (2002), 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. 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 Hsieh (1999) for a detailed description of various hedge fund styles.

Second, rather than using asset class indices directly, such as the SP500, we use ETF's as proxies. The ETF's used in this study are offered by the fund family iShares, and are depicted in Table 5.[12]

Our decision 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 of holidays, (iv) acknowledging the fact that different exchanges face different rules such as short sales constraints, and short-circuit mechanisms, 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). This problem is similar to the challenges of testing for the presence of jumps where high frequency data clearly reveals such activity better than low frequency data (see, e.g. Jiang and Oomen, 2008; Ait-Sahalia and Jacod, 2009; Ait-Sahalia et al., 2012).

By using ETF's we are able to address a majority of the issues listed above. For instance, addressing asynchronicity formally is beyond the scope of our paper.[13] However, such concerns should be mitigated by the fact that all of the ETF's we explore are 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, our interest lies with the in-sample behavior of the data in order to test for volatility spillovers. We do not extrapolate that in-sample behavior to out-of-sample periods that may be of varying lengths due to trading asynchronicities.

An additional benefit of using ETF's is that 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 of the chosen ETF's are depicted in Fig. 1 of Appendix C, while Table 6 details the univariate sample statistics as well as the unconditional measures of contemporaneous correlation. The assets exhibit conventional stylized facts, such as non-Gaussian behavior indicated by asymmetries and excess kurtosis that are well established in the literature. Furthermore, there do not appear to be one-off events, which would be depicted as a singular extremal spike. REM has the largest negative skew in the sample chosen, while IVV is nearly symmetric. Each of the assets exhibit some excess kurtosis, with the REM having a kurtosis of 7.034. Raw returns in equities, as proxied by the IVV and EFA, appear closely linked contemporaneously, with a correlation of 0.866 ± 0.06.[16] On the other hand, fixed income and real estate seem less closely linked, with a correlation of -0.148 ± 0.10.

There is an important caveat: kurtosis, skewness, and even variance in y_t may not exist due to heavy tails. In Fig. 1 in Appendix C we plot the Hill (1975) estimator $\hat{\kappa}_T$ of the tail index

12 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.

13 We refer the interested reader to Zhang (2011) for a detailed discussion on the subject.

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 US 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.

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

Table 6
Daily log returns from 1/3/2012 through 12/31/2012 (251 trading days) for five asset classes as represented by their respective iShares ETF's.

Sample statistics					
	IVV	AGG	EFA	GSG	REM
Mean	0.051	0.003	0.055	−0.002	0.031
Med	0.058	0.018	0.083	0.030	0.142
Std	0.805	0.173	1.147	1.092	0.848
Skew	0.063	−0.417	−0.245	0.236	−0.772
Kurt	3.808	3.327	4.385	4.763	7.034
Contemporaneous correlations					
	IVV	AGG	EFA	GSG	REM
IVV	1.000	–	–	–	–
AGG	−0.420	1.000	–	–	–
EFA	0.866	−0.414	1.000	–	–
GSG	0.576	−0.300	0.601	1.000	–
REM	0.422	−0.148	0.295	0.193	1.000

κ_y for each asset returns series $\{y_t\}_{t=1}^T$, with 95% non-parametric bands defined in Hill (2010). The estimator is $\hat{\kappa}_T \equiv (1/\tilde{k}_T \sum_{i=1}^{\tilde{k}_T} \ln(y_{(i)}^{(a)}/y_{(\tilde{k}_T)}^{(a)}))^{-1}$ where $y_t^{(a)} \equiv |y_t|$ and $\tilde{k}_T \in \{5, \ldots, 400\}$. The asymptotic bands are $\hat{\kappa}_T \pm 1.95\hat{v}_T\hat{\kappa}_T^2/\tilde{k}_T^{1/2}$, where $\hat{v}_T^2 = 1/T \sum_{s,t=1}^T \mathcal{K}_{T,s,t}\{\ln(y_s^{(a)}/y_{(\tilde{k}_T+1)}^{(a)})_+ - (\tilde{k}_T/T)\hat{\kappa}_T^{-1}\} \times \{\ln(y_t^{(a)}/y_{(\tilde{k}_T+1)}^{(a)})_+ - (\tilde{k}_T/T)\hat{\kappa}_T^{-1}\}$ 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 $\gamma_T = T^{.225}$.[17] As we can see, a fourth moment is unlikely to exist in any of the assets since $\hat{\kappa}_T < 4$ in general. Moreover, the third and even second moment are in question for IVV and REM, hence unconditional correlations may not exist.[18]

6.2. Model specification

In order to investigate the conditional nature of their relationships and to control for volatility dynamics, we pass each series though simple ARMA(1, 1)–GARCH(1, 1) filters $y_t = ay_{t-1} + bu_{t-1} + u_t$, $|a| < 1$, where $u_t = h_t\epsilon_t$ and $h_t^2 = \omega + \alpha u_{t-1}^2 + \beta h_{t-1}^2$, $\omega > 0$, $\alpha + \beta > 0$, with the goal of using the ARMA residuals $\hat{u}_t$ for spillover analysis. The conditional mean and variance parameters are estimated via QMTTL. 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 $\tilde{\epsilon}_t(\xi) = u_t(\phi)/\tilde{h}_t(\xi)$. The iterated volatility process is $\tilde{h}_1^2(\xi) = \omega$ and $\tilde{h}_t^2(\xi) = \omega + \alpha u_{t-1}^2(\phi) + \beta\tilde{h}_{t-1}^2(\xi)$ for $t \geq 2$.

We gauge the adequacy of a linear filter by performing robust tests of white noise and omitted non-linearity. We first perform tail-trimmed Q-tests by p-value occupation time on the GARCH residuals $\hat{\epsilon}_t = \hat{u}_t/\hat{h}_t$ based on five lags of tail-trimmed serial correlations (9) with simple trimming, fractile $k_T = [\lambda T/\ln(T)]$, and $\lambda \in [.01, 1]$. See Section 2.4. We reject the null hypothesis of zero correlation at level α when the occupation time exceeds α. The occupation times at the 10% level for (IVV, AGG, EFA, GSG, REM) are (.07, .00, .00, .00, .11), suggesting little evidence of remaining serial correlation.

In order to detect omitted non-linearity in the conditional mean and variance models we use a generalization of the conditional moment test of Hill (2012). The test requires a k-vector of conditioning variables or regressors $x_t = [x_{i,t}]_{i=1}^k$ and a regression model residual $\hat{\varepsilon}_t$. The test statistic is $\hat{T}_T(\gamma) \equiv \hat{V}_T^{-1}(\gamma) \sum_{t=1}^T \hat{\varepsilon}_t \hat{I}_{T,t} F(\gamma'\xi(x_t^*))$ with composite trimming indicator $\hat{I}_{T,t} \equiv I(|\hat{\varepsilon}_t| \leq \hat{\varepsilon}_{(k_T)}^{(a)}) \prod_{i=1}^k I(|x_{i,t}| \leq x_{i,(k_T)}^{(a)})$ and the same $k_T = [\lambda T/\ln(T)]$ throughout, where $F(w) = \exp\{w\}$, $\xi(x_t^*) = [1, \arctan(x_t^*)']'$ with $x_{i,t}^* = x_{i,t} - 1/T\sum_{t=1}^T x_{i,t}$, and $\hat{V}_T(\gamma)$ is a Bartlett kernel HAC estimator of $E[\{\sum_{t=1}^T \hat{\varepsilon}_t \hat{I}_{T,t} F(\gamma'\xi(x_t))\}^2]$ with bandwidth $T^{1/3}$. As long as $F(\cdot)$ is real analytic and non-polynomial, $\xi(x_t)$ is a bounded one-to-one transform of x_t, and $k_T = [\lambda T/\ln(T)]$ for any $\lambda \in (0, 1]$, then pointwise $\hat{T}_T(\gamma) \xrightarrow{d} \chi^2(1)$ if $E[\varepsilon_t|x_t] = 0$ *a.s.*, and if $P(E[\varepsilon_t|x_t] = 0) < 1$ then $\hat{T}_T(\gamma) \xrightarrow{p} \infty$ for all γ on any compact set Γ except for countably many γ. We randomly choose γ from [.1, 2], we use an orthogonal transform of $\hat{\varepsilon}_t \hat{I}_{T,t} F\left(\gamma'\xi(x_t^*)\right)$ to ensure plug-in robustness (see Hill, 2012, Section 2.2), and we compute the p-value occupation time over $\lambda \in [.01, 1]$. We reject $E[\varepsilon_t|x_t] = 0$ *a.s.* at level α if the occupation time is above α. We first test for omitted nonlinearity in the conditional mean model by using the ARMA residuals $\hat{u}_t$ and regressor set $x_t \equiv [y_{t-1}, \epsilon_{t-1}]'$ where $\hat{\epsilon}_{t-1}$ is substituted for ϵ_{t-1}. This yields occupation times (.004, .000, .062, .065, .025) for (IVV, AGG, EFA, GSG, REM) which are all under .10. In the test of omitted nonlinearity in the conditional variance model we use the GARCH residuals $\hat{\epsilon}_t = \hat{u}_t/\hat{h}_t$ and regressor set $x_t \equiv [y_{t-1}, h_{t-1}]'$ where $\hat{h}_{t-1}$ is substituted for h_{t-1}. The occupation times for tests at the 10% level for the five ETF's are (.027, .000, .000, .000, .013), which are again all under .10. Together this suggests there is little evidence for omitted nonlinearity in the ARMA–GARCH filter.

The GARCH residuals $\hat{\epsilon}_t$ for each of the ETF's are depicted in Fig. 2 of Appendix C. We also present in Fig. 2 Hill-plots of the two-tailed tail index estimator for each asset's $\hat{\epsilon}_t$, with 95% confidence bands computed as above.[19] 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 the REM case $\hat{\kappa}_T > 2$ over most fractiles, but 2 is near the middle of the confidence band. This suggests this one asset may have an infinite variance GARCH error, which is not allowed in this paper. Our volatility spillover tests for this asset should be interpreted with caution.[20]

6.3. Volatility spillover tests

In Tables 8–11 of Appendix C we present the occupation time values, denoted τ, for tests of spillover at horizon $H = \{1, 2,$

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 sizes T. We simply use the midpoint power $\gamma_T = T^{.225}$.

18 A shortcoming of the Hill (1975) estimator is the arbitrary fractile choice $\tilde{k}_T$, and the rate of convergence $\tilde{k}_T^{1/2} = o(T^{1/2})$ which may be very low depending on the rate of tail decay (see Hill, 2010, for references). An alternative method exists for GARCH processes $y_t = h_t(\theta^0)\epsilon_t$ and $h_t^2 = \omega^0 + \alpha^0 y_{t-1}^2 + \beta^0 h_{t-1}^2$ based on the moment condition $E(\alpha^0\epsilon_t^2 + \beta^0)^{\kappa/2} = 1$ (Mikosch and Starica, 2000; Berkes et al., 2003). Since a plug-in for $[\omega^0, \alpha^0, \beta^0]'$ will be no greater than $n^{1/2}$-convergent, the error ϵ_t, however, must have a finite fourth moment (Berkes et al., 2003, Theorem 2.2).

19 The use of a filtered process $\{\hat{\epsilon}_t\}$ for tail index estimation for $\{\epsilon_t\}$ does not impact asymptotics since the QMTTL plug-in is $T^{1/2}/L(T)$-convergent for some slowly varying $L(T)$ based on our choice of trimming fractiles. See Hill (2014a,c).

20 The Hill (1975) estimator is biased in small samples when the data are not exactly Pareto distributed (see Hill, 2010, for references). Further, $\hat{\kappa}_T \approx \kappa$ for very few fractile values is not uncommon in small samples when the data are weakly dependent and heterogeneous. Here we use the residual series $\{\hat{\epsilon}_t\}_{t=2}^T$ where $\hat{\epsilon}_t$ is weakly dependent by construction, although it is a valid filter for estimating the true index κ of ϵ_t (see Hill, 2014c). Further, Hill (2014c) shows by simulation that the use of an iterated volatility approximation, as is necessary in practice, leads to a slight but systematic under-estimation of the GARCH error tail index when a residual series is used. This follows since the iterated volatility process $\tilde{h}_1^2 = \omega^0$ and $\tilde{h}_t^2 = \omega^0 + \alpha^0 u_{t-1}^2 + \beta^0 \tilde{h}_{t-1}^2$ for $t \geq 2$ under-presents the true volatility process $\{h_t^2\}$ in small samples, even if the true parameter values are used rather than QMTTL $\hat{\theta}_T = [\hat{a}_T, \hat{b}_T; \hat{\omega}_T, \hat{\alpha}_T, \hat{\beta}_T]'$, hence the sample standardized GARCH residual $\hat{\epsilon}_t$ appears slightly heavier tailed than the error.

3, 4, 5} in each of the 20 possible directions across the five ETF's (IVV to AGG, AGG to IVV, etc.). We use the Q-statistic $\hat{Q}_T^{(\mathcal{E})}$ with tail trimmed centered errors and a QMTTL plug-in; we use simple trimming, Tukey's bisquare and exponential with negligibility $k_T = [\lambda T/\ln(T)]$, and truncation with non-negligibility $k_T = [\lambda T]$; and we use occupation time over $\lambda \in [.01, 1]$.[21] Using nominal sizes $\alpha \in \{1\%, 5\%, 10\%\}$ for our tests, we reject the null of no spillover if $\tau > \alpha$.

A few notable findings appear for the simple trimming case in Table 8. First, domestic stocks and bonds tend to have a bi-directional relationship at varying horizons, with the IVV (domestic stocks) spilling into AGG (domestic bonds), while AGG also spills into the IVV. Domestic stocks are a destination for spillovers, where not only bonds, but also international stocks (EFA) tend to spillover into the IVV. Commodities (GSG) seem to be a prolific source of spillovers, particularly at shorter horizons. Meanwhile, commodities (GSG) tend not to be a destination for spillovers. Generally speaking, these findings are supported when the transform is Tukey's bisquare or exponential as shown in Tables 9 and 10, and using truncation shown in Table 11.

6.4. The importance of spillovers: predictive accuracy

We now evaluate the usefulness of detecting spillovers for the purposes of constructing volatility forecasts. Specifically, we are interested in one-step ahead volatility forecasts. As such, we use the spillover tests at horizon $H = 1$ to identify appropriate candidates to illustrate the approach of Section 5.3. According to Tables 8–11, the evidence suggests volatility from international equities (EFA) spills into domestic real estate (REM) at horizon 1, while REM volatility appears not to spill into EFA. The strongest evidence of spillover from EFA to REM exists at the 5% and 10% levels for smooth negligible transforms (Tukey's bisquare and exponential with $k_T = [\lambda T/\ln(T)]$) and for truncation with non-negligibility $k_T = [\lambda T]$, although evidence at the 10% exists using simple trimming. Recall from controlled experiments that our best test statistic is $\hat{Q}_T^{(\mathcal{E})}$ based on negligible trimming with exponential smoothing, or truncation.

Let $y_{1,t}$ and $y_{2,t}$ be REM and EFA, respectively. In all that follows the volatility start condition is $\tilde{h}_{i,1}^2 = \omega_i$. We begin by modeling the conditional mean and variance of REM via an ARMA(1, 1)–GARCH(1, 1), and use the square of the extracted QMTTL conditional mean residuals $\hat{u}_{1,t}^2$ as the volatility proxy used for comparative purposes. Using notation from Sections 5.3 and 6.2, we model the volatility dynamics of REM via a GARCH(1, 1) built only from its own dynamics, resulting in the QMTTL estimate $\hat{\xi}_{1,T}^{(NS)} \equiv [\hat{\phi}_{1,T}', \hat{\theta}_{1,T}^{(NS)\prime}]'$. We extract the conditional volatility $\hat{h}_{1,t}^{2(NS)} \equiv \tilde{h}_{1,t}^2(\hat{\xi}_{1,T}^{(NS)})$ and construct the loss function $\hat{\mathcal{L}}_T^{(NS)}$ in (20). We then expand the GARCH(1, 1) model for REM to include spillover effects from EFA, hence for REM volatility we use the function $\tilde{h}_{1,t}^2(\xi) = \omega_1 + \alpha_{1,1}u_{1,t-1}^2(\phi_1) + \beta_{1,1}\tilde{h}_{1,t-1}^2(\xi) + \alpha_{1,2}u_{2,t-1}^2(\phi_2) + \beta_{1,2}\tilde{h}_{2,t-1}^2(\xi_2)$. In view of the strong evidence that spillover does not occur from REM to EFA, we do not include spillovers in $\tilde{h}_{2,t}^2(\xi_2)$, hence $\tilde{h}_{2,t}^2(\xi_2) = \omega_2 + \alpha_{2,2}u_{2,t-1}^2(\phi_2) + \beta_{2,2}\tilde{h}_{2,t-1}^2(\xi_2)$. We extract the conditional volatility $\tilde{h}_{1,t}^2(\hat{\xi}_{1,T}^{(S)})$ and construct the loss function $\hat{\mathcal{L}}_T^{(S)}$ in (19). The Relative Improvement index is $\hat{\mathcal{R}}_T = 100(1 - \hat{\mathcal{L}}_T^{(S)}/\hat{\mathcal{L}}_T^{(NS)}) = 1.75$, and the accompanying $\mathcal{LR}_T$ test p-value is .089. Finally, we repeat this exercise by modeling EFA with and without spillover from REM. This yields a Relative Improvement index $\hat{\mathcal{R}}_T = 1.055$, and accompanying $\mathcal{LR}_T$ test p-value is .758.

In conclusion, the forecast improvement analysis matches our spillover tests. We did not detect spillover from real estate (REM) to international equities (EFA), and including information about REM does not significantly improve the volatility forecast for EFA (p-value .758). Conversely, we did detect spillover from EFA to REM, and we find that using EFA leads to an improved volatility forecast for REM (p-value .089).

7. Conclusion

We extend and improve available tail-trimming methods in the econometrics literature to tests of volatility spillover for random volatility processes. By transforming test equations in various ways we construct a battery of asymptotically chi-squared portmanteau and score statistics. We only require the innovations to have a finite variance, and for score tests the errors may be martingale differences allowing for volatility spillover in a semi-strong nonlinear GARCH setting. We exploit a broad class of bounded transformations used in the robust estimation literature, including redescending transforms like tail trimming, and non-redescending transforms like truncation.

If the volatility innovations are independent under the null, then essentially any bounded transform can be used in theory for our portmanteau statistics since we recenter the transformed errors and therefore ensure identification of the null. The thresholds used for the transformations may be bounded or increasing as the sample size increases (i.e. trimming may be non-negligible or negligible). In the bounded case we ensure infinitesimal robustness (cf. Hampel, 1974), although we do not formally tackle data contamination in this paper. Spillover, however, is guaranteed to be identified asymptotically only if increasing thresholds are used, in which case a redescending transform *and not truncation* must be used to ensure robustness to heavy tails.

Our simulation experiments suggest across cases that for an asymptotic test portmanteau statistics dominate score statistics; negligible trimming with exponential smoothing, and non-negligible truncation, are optimal; and that smoothing the p-value over a fractile window results in a small empirical size improvement and lower power. Further, trimming with an increasing threshold leads to higher power than if the threshold is fixed, but truncation with a fixed threshold leads to even higher power in many cases due to the substitution of large values with a threshold, rather then their removal. We show that a data transform matters even when tails are only mildly heavy since even then extant methods may not result in standard asymptotic inference.

Finally, we exploit Patton (2011)'s theory for forecast model comparisons that are robust to the choice of volatility proxy, by making his quasi-likelihood criterion robust to heavy tails. We then compute a heavy tail robust forecast improvement index and associated p-value for a test of the hypothesis that including spillovers does not improve a volatility forecast model. We show by simulation that the use of spillovers for forecasting volatility matters if there truly is spillover. This finding implies that our spillover test can be used as a forecast model specification pre-test. We illustrate the use of such an approach in this paper with a pair of Exchange Traded Funds.

Important future directions of research should include spillover tests that allow for both contaminated data *and* heavy tails (in the observed contaminated and latent variables); score tests based on symmetric tail-trimming of $m_{h,t}$ that uses an extreme value theory based bias correction, as in Hill (2013a) and Hill and McCloskey (2014); and a deeper study of volatility forecast comparisons.

Acknowledgments

We thank a referee and co-editor Yacine Aït-Sahalia for feedback that lead to a substantially improved manuscript.

21 Results from the other tests generally support these findings.

Appendix A. Assumptions

We now present all assumptions, focused toward the Q-tests. Parallel conditions for the score tests are presented in Aguilar and Hill (2014). Define the σ-fields $\Im_{i,t} \equiv \sigma(y_{i,\tau} : \tau \leq t)$ and $\Im_t \equiv \sigma(\Im_{1,t} \cup \Im_{2,t})$.

Assumption B (*Bounded Transform*): *Redescending* $\check{\psi}$ *satisfies* (4)–(5), *and generic* ψ *satisfies* (7).

Assumption D (*dgp*):

1. $\{y_{i,t}, \epsilon_{i,t}, h_{i,t}\}$ *defined by* (1) *are stationary and* $\{y_{i,t}, h_{i,t}\}$ *are* L_ι*-bounded for tiny* $\iota > 0$.
2. *The marginal distribution of each* $z_{i,t} \in \{\epsilon_{i,t}, h_{i,t}\}$ *is absolutely continuous with support* $\mathbb{R}$, *and is uniformly bounded:* $\sup_{c\in\mathbb{R}}\{(\partial/\partial c) P(z_{i,t} \leq c)\} < \infty$.
3. $h^2_{i,t}(\theta_i)$ *is* $\Im_{i,t-1}$*-measurable, twice differentiable, and bounded on* Θ: $\inf_{\theta\in\Theta} h_{i,t}(\theta) > 0$ *a.s. Further* $(\partial/\partial\theta_i)h^2_{i,t}$ *is stationary, and satisfies* (2).

Assumption E (*Memory and Moments : Q-test*):

1. $E[\epsilon_{i,t}] = 0$ *and* $E[\epsilon^2_{i,t}] = 1$, *and* $\epsilon_{i,t}$ *are serially, but not necessarily mutually, independent.*
2. $\{\epsilon_{i,t}, h_{i,t}\}$ *are geometrically* β*-mixing.*

Remark 7. 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 $\hat{\theta}_T$ (see Hill, 2012, 2013b, 2014a).

Remark 8. Conditions for stationarity, and geometric ergodicity or the more general geometric β-mixing, exist for a large class of nonlinear random volatility processes. See Carrasco and Chen (2002), and Francq and Zakoïan (2006) for references. For example, for a linear GARCH $h^2_{i,t} = \omega^0_i + \alpha^0_i y^2_{i,t-1} + \beta^0_i h^2_{i,t-1}, \omega^0_i > 0, \alpha^0_i, \beta^0_i \geq 0$, if $\epsilon_{i,t}$ is iid, has a continuous distribution and $E[\ln(\alpha^0_i + \beta^0_i \epsilon^2_{i,t})] < 0$, then $h^2_{i,t}$ has a unique stationary solution and is geometrically β-mixing (Nelson, 1990; Bougerol and Picard, 1992; Carrasco and Chen, 2002).

We now characterize the fractiles used to compute the various thresholds. The Q-tests have independence under the null which allows any $k_T/T \to [0, 1)$ depending on the transform ψ. If truncation is used then too many extreme values enter the test statistic asymptotically, unless non-negligibility $k_T/T \to (0, 1)$ is enforced (cf. Csörgo et al., 1986).

Assumption FE (*Fractile Bound For* $\mathcal{E}_{i,t}$ *or* $\epsilon_{i,t}$ **):** *Let* k_T *denote any particular fractile. Then* $k_T \to \infty$ *as* $T \to \infty$. *If the transform* ψ *is redescending then either* $k_T = o(T/L(T))$ *for some slowly varying* $L(T) \to \infty$ *or* $k_T/T \to (0, 1)$, *and otherwise* $k_T/T \to (0, 1)$.

Recall $\mathfrak{V}_T = [\mathfrak{V}_{1,T}, \mathfrak{V}_{2,T}] \in \mathbb{R}^{1\times 2q}$ defined in (11). The Q-test plug-in for θ^0 must be $\mathfrak{V}^{1/2}_T$-convergent.

Assumption PQ (*Plug-In For Q-Test*): $\mathfrak{V}^{1/2}_T(\hat{\theta}_T - \theta^0) = O_p(1)$.

Finally, we characterize probability tail decay of the errors.

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

$$P\left(\left|\epsilon_{i,t}\right| > c\right) = d_i c^{-\kappa_i}\left(1 + o(1)\right)$$

$$\text{where } d_i > 0 \text{ and } \kappa_i \in (2, 4]. \tag{22}$$

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

$$P\left(\left|\epsilon^2_{1,t}\epsilon^2_{2,t-h}\right| > c\right) = d_m c^{-\kappa/2}\left(1 + o(1)\right)$$

$$\text{where } d_m > 0 \text{ and } \kappa = \min\{\kappa_1, \kappa_2\}. \tag{23}$$

Table 7

(Redescenders) $\hat{Q}_T$-test rejection frequencies at the (1%, 5%, 10%) levels. We use 10, 000 simulated paths, $T = 1,000$, $H = 5$ lags, and no plug-in. We use Tukey's bisquare and exponential transforms with fractile $k_T = [\lambda T/\ln(T)]$ and handpicked $\lambda = 0.05$ or p-value occupation time (Oc.Time). The null hypothesis is no spillover from $y_{2,t}$ to $y_{1,t}$. Alternative 1 indicates weak spillover and Alternative 2 indicates strong spillover.

$\epsilon \sim N(0,1)$		Tukey's bisquare $\hat{Q}^{(\mathcal{E})}_T$	$\hat{Q}^{(\epsilon)}_T$
$\lambda = 0.05$	Null—no spill	(.018,.070,.123)	(.011,.053,.103)
	Alt1—weak	(.265,.455,.565)	(.014,.064,.120)
	Alt2—strong	(.406,.601,.699)	(.017,.074,.135)
Oc.Time	Null—no spill	(.014,.060,.113)	(.012,.054,.106)
	Alt1—weak	(.098,.216,.306)	(.016,.067,.126)
	Alt2—strong	(.146,.285,.383)	(.019,.073,.133)
$\epsilon \sim P_{2.5}$		$\hat{Q}^{(\mathcal{E})}_T$	$\hat{Q}^{(\epsilon)}_T$
$\lambda = 0.05$	Null—no spill	(.019,.058,.104)	(.012,.054,.105)
	Alt1—weak	(.207,.352,.443)	(.067,.174,.251)
	Alt2—strong	(.310,.479,.574)	(.114,.243,.333)
Oc.Time	Null—no spill	(.014,.055,.104)	(.011,.053,.105)
	Alt1—weak	(.060,.134,.201)	(.022,.072,.127)
	Alt2—strong	(.088,.173,.244)	(.029,.082,.139)
$\epsilon \sim N(0,1)$		Exponential $\hat{Q}^{(\mathcal{E})}_T$	$\hat{Q}^{(\epsilon)}_T$
$\lambda = 0.05$	Null—no spill	(.021,.074,.132)	(.018,.067,.118)
	Alt1—weak	(.381,.576,.674)	(.205,.387,.499)
	Alt2—strong	(.547,.721,.798)	(.326,.529,.636)
Oc.Time	Null—no spill	(.015,.062,.115)	(.012,.055,.106)
	Alt1—weak	(.133,.271,.369)	(.042,.107,.171)
	Alt2—strong	(.199,.360,.464)	(.058,.130,.195)
$\epsilon \sim P_{2.5}$		$\hat{Q}^{(\mathcal{E})}_T$	$\hat{Q}^{(\epsilon)}_T$
$\lambda = 0.05$	Null—no spill	(.023,.064,.107)	(.015,.059,.108)
	Alt1—weak	(.244,.393,.484)	(.188,.331,.424)
	Alt2—strong	(.363,.520,.606)	(.296,.461,.557)
Oc.Time	Null—no spill	(.015,.057,.106)	(.011,.053,.103)
	Alt1—weak	(.102,.213,.298)	(.038,.096,.156)
	Alt2—strong	(.155,.289,.382)	(.053,.118,.180)

Remark 9. If $\epsilon_{1,t}$ is independent of $\epsilon_{2,t}$ then (22) implies (23), cf. Breiman (1965) and Cline (1986). In the case of dependent random variables with power law tails (22) there does not exist a formal theory for characterizing product convolution tails like $P(|\epsilon^2_{1,t}\epsilon^2_{2,t-h}| > c)$ (cf. Cline, 1986), unless an explicit model linking $\epsilon_{1,t}$ and $\epsilon_{2,t}$ is entertained (see, e.g., Mikosch and Starica, 2000, for theory and references). We impose (23) with index $\kappa/2$ to simplify tail representations.

Appendix B. Proofs of main results

Recall

$$\mathcal{E}_{i,t}(\theta_i) \equiv \epsilon^2_{i,t}(\theta_i) - 1 \quad \textit{and}$$

$$\psi^{(\mathcal{E})}_{i,T,t}(\theta) \equiv \psi\left(\mathcal{E}_{i,t}(\theta), c^{(\mathcal{E})}_{i,T}(\theta)\right) - E\left[\psi\left(\mathcal{E}_{i,t}(\theta), c^{(\mathcal{E})}_{i,T}(\theta)\right)\right]$$

$$\mathfrak{S}_T \equiv E\left[\psi^{(\mathcal{E})}_{1,T,t}\right] \times E\left[\psi^{(\mathcal{E})}_{2,T,t}\right].$$

In order to keep proofs relatively short, we restrict ourselves to linear GARCH models (14). All test statistics with a redescending transform and intermediate order statistic threshold operate essentially like tail-trimming in view of (6). We therefore restrict

Table 8
(ETF's—simple trimming) Occupation times (τ) for the $\hat{Q}_T^{(\mathcal{E})}$-test of volatility spillover at $\alpha \in \{1\%, 5\%, 10\%\}$ levels using *simple trimming*, at horizon H, and a QMTLL plug-in. We reject the null of no spillover at level α if $\tau > \alpha$.

$H = 1$					
FROM/TO	IVV	AGG	EFA	GSG	REM
IVV	(-,-,-)	(.044,.178,.244)	(.000,.000,.000)	(.000,.000,.000)	(.000,.022,.222)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.111,.200)	(.000,.000,.022)
EFA	(.000,.000,.044)	(.000,.022,.044)	(-,-,-)	(.000,.000,.089)	(.000,.022,.111)
GSG	(.067,.556,.667)	(.022,.289,.489)	(.000,.000,.089)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 2$					
FROM/TO	IVV	AGG	EFA	GSG	REM
IVV	(-,-,-)	(.022,.178,.267)	(.000,.000,.000)	(.000,.000,.000)	(.000,.067,.156)
AGG	(.000,.089,.244)	(-,-,-)	(.000,.000,.000)	(.000,.022,.067)	(.000,.000,.133)
EFA	(.000,.000,.000)	(.000,.000,.022)	(-,-,-)	(.000,.000,.000)	(.000,.000,.200)
GSG	(.000,.267,.578)	(.022,.178,.311)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)
REM	(.000,.022,.022)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 3$					
FROM/TO	IVV	AGG	EFA	GSG	REM
IVV	(-,-,-)	(.000,.244,.333)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.067)
AGG	(.000,.000,.133)	(-,-,-)	(.000,.000,.000)	(.000,.000,.044)	(.000,.000,.000)
EFA	(.000,.044,.089)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)
GSG	(.111,.178,.333)	(.022,.133,.178)	(.089,.111,.111)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.022)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 4$					
FROM/TO	IVV	AGG	EFA	GSG	REM
IVV	(-,-,-)	(.000,.178,.244)	(.000,.000,.000)	(.000,.000,.000)	(.000,.044,.089)
AGG	(.000,.000,.200)	(-,-,-)	(.000,.000,.022)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.022,.044)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.022)
GSG	(.067,.156,.222)	(.022,.089,.156)	(.089,.089,.111)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 5$					
FROM/TO	IVV	AGG	EFA	GSG	REM
IVV	(-,-,-)	(.000,.111,.222)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.067)
AGG	(.000,.000,.156)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.044,.156,.200)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.089)
GSG	(.044,.244,.356)	(.000,.067,.111)	(.067,.089,.089)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.022)	(.000,.000,.000)	(-,-,-)

attention to simple trimming $\psi(u, c) = \check{\psi}(u, c) = uI(|u| \leq c)$ with an intermediate order statistic plug-in for c. Therefore

$$\psi_{i,T,t}^{(\mathcal{E})}(\theta) = \mathcal{E}_{i,t}(\theta) I\left(\left|\mathcal{E}_{i,t}(\theta)\right| \leq c_{i,T}^{(\mathcal{E})}(\theta)\right) - E\left[\mathcal{E}_{i,t}(\theta) I\left(\left|\mathcal{E}_{i,t}(\theta)\right| \leq c_{i,T}^{(\mathcal{E})}(\theta)\right)\right]$$

where $P(|\mathcal{E}_{i,t}(\theta)| \geq c_{i,T}^{(\mathcal{E})}(\theta)) = k_{i,T}^{(\mathcal{E})}/T \to 0$, and

$$\hat{\psi}_{i,T,t}^{(\mathcal{E})}(\theta) = \mathcal{E}_{i,t}(\theta) I\left(\left|\mathcal{E}_{i,t}(\theta)\right| \leq \mathcal{E}_{i,(k_{i,T})}^{(a)}(\theta)\right) - \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}_{i,t}(\theta) I\left(\left|\mathcal{E}_{i,t}(\theta)\right| \leq \mathcal{E}_{i,(k_{i,T})}^{(a)}(\theta)\right).$$

Similar arguments extend to other transforms, and to central order statistic thresholds for the Q-statistics.

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 (22) with index $\kappa > 0$, and trimmed version $w_{T,t}^* \equiv w_t I(|w_t| \leq c_T)$, $P(|w_t| > c_T) = k_T/T = o(T)$, and $k_T \to \infty$. Then

$E\left|w_{T,t}^*\right|^{\kappa} \sim L(T) \to \infty$ is slowly varying,

$$E\left|w_{T,t}^*\right|^{p} \sim c_T^p \left(\frac{k_T}{T}\right) = K\left(\frac{T}{k_T}\right)^{p/\kappa - 1} \quad p > \kappa. \tag{24}$$

The proof of Theorem 3.1 requires two preliminary results. Drop superscripts and write $\psi_{i,T,t} = \psi_{i,T,t}^{(\mathcal{E})}$, $k_{i,T} = k_{i,T}^{(\mathcal{E})}$, $\hat{\rho}_{T,h}(\theta) = \hat{\rho}_{T,h}^{(\mathcal{E})}(\theta)$, etc. First, we require LLNs for $\psi_{i,T,t}$ irrespective of tail thickness. Let $L(T) \to \infty$ be slowly varying.

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

Proof. Write $\kappa = \kappa_i$. Claim (a) follows from independence, Chebyshev's inequality, and (24): $E(1/T \sum_{t=1}^{T} \psi_{i,T,t})^2 = E[\psi_{i,T,t}^2]/T$, where $E[\psi_{i,T,t}^2]/T \sim E[\mathcal{E}_{i,t}^2 I(|\mathcal{E}_{i,t}| \leq c_{i,T})]/T = o(1)$ if $\kappa > 4$, $E[\psi_{i,T,t}^2]/T \sim L(T)/T = o(1)$ if $\kappa = 4$, and if $\kappa \in (2, 4)$ then $E[\psi_{i,T,t}^2]/T = O((T/k_T)^{2/\kappa}/T) = o(1)$.

Consider (b). By Assumption T $\mathcal{E}_{i,t}^2$ has moment supremum $\tilde{\kappa} \equiv \kappa/4$. If $\tilde{\kappa} \geq 1$ then the claim follows as above. Assume $\kappa \in (2, 4)$ hence $\tilde{\kappa} \in (0, 1)$, compactly write $w_{T,t} = \psi_{i,T,t}^2$ and $k_T = k_{i,T}$, and note by independence and (24)

$$E\left(\frac{1}{T}\sum_{t=1}^{T}\left\{\frac{w_{T,t}}{E\left[w_{T,t}\right]} - 1\right\}\right)^2 \leq 2\frac{1}{T}\frac{E\left[w_{T,t}^2\right]}{\left(E\left[w_{T,t}\right]\right)^2} \sim K\frac{1}{T}\frac{(T/k_T)^{2/\tilde{\kappa}-1}}{(T/k_T)^{2/\tilde{\kappa}-2}} = K\frac{1}{k_T} = o(1). \tag{25}$$

Now use Chebyshev's inequality to complete the proof. □

Table 9
(ETF's—Tukey's bisqare) Occupation times (τ) for the $\hat{Q}_T^{(\mathcal{E})}$-test of volatility spillover at $\alpha \in \{1\%, 5\%, 10\%\}$ levels using *Tukey's bisquare* transform, at horizon H, and a QMTLL plug-in. We reject the null of no spillover at level α if $\tau > \alpha$.

$H = 1$					
FROM/TO	IVV	AGG	EFA	GSG	REM
IVV	(-,-,-)	(.000,.000,.089)	(.000,.000,.200)	(.044,.156,.311)	(.022,.422,.533)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.067)
EFA	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.222,.267)
GSG	(.156,.467,.622)	(.000,.133,.244)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 2$					
FROM/TO	IVV	AGG	EFA	GSG	REM
IVV	(-,-,-)	(.000,.000,.022)	(.000,.000,.000)	(.000,.089,.156)	(.000,.289,.422)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.200,.289)
GSG	(.000,.289,.644)	(.000,.089,.178)	(.156,.311,.356)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 3$					
FROM/TO	IVV	AGG	EFA	GSG	REM
IVV	(-,-,-)	(.000,.000,.222)	(.000,.000,.000)	(.000,.022,.089)	(.000,.244,.378)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.022)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.089,.244)
GSG	(.000,.200,.444)	(.000,.000,.067)	(.022,.311,.378)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 4$					
FROM/TO	IVV	AGG	EFA	GSG	REM
IVV	(-,-,-)	(.000,.000,.000)	(.000,.289,.333)	(.000,.000,.089)	(.000,.111,.356)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.044,.156)
GSG	(.000,.289,.756)	(.000,.000,.000)	(.022,.200,.378)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.022)	(-,-,-)
$H = 5$					
FROM/TO	IVV	AGG	EFA	GSG	REM
IVV	(-,-,-)	(.000,.000,.000)	(.000,.333,.333)	(.000,.000,.000)	(.000,.089,.178)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.067)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.111,.222,.311)
GSG	(.000,.156,.578)	(.000,.000,.000)	(.022,.022,.333)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)

Second, stochastically trimmed $\hat{\psi}_{1,T,t}(\hat{\theta}_{1,T})\hat{\psi}_{2,T,t-h}(\hat{\theta}_{2,T})$ is sufficiently close to deterministically trimmed $\psi_{1,T,t}\psi_{2,T,t-h}$, 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) $1/T \sum_{t=1}^T \{\hat{\psi}^2_{i,T,t}(\hat{\theta}_{i,T}) - \psi^2_{i,T,t}\} = o_p(1)$;

(b) $T^{-1/2}\mathfrak{S}_T^{-1/2} \sum_{t=1}^T \{\hat{\psi}_{1,T,t}(\hat{\theta}_{1,T})\hat{\psi}_{2,T,t-h}(\hat{\theta}_{2,T}) - \psi_{1,T,t}\psi_{2,T,t-h}\} = o_p(1) \forall h \geq 1$ *if* $\epsilon_{1,t}$ *and* $\epsilon_{2,t}$ *are mutually independent, and otherwise* $O_p(1)$.

Proof. We only prove claim (b). A nearly identical argument shows

$$\frac{1}{T^{1/2}\left(E\left[\psi^2_{i,T,t}\right]\right)^{1/2}} \sum_{t=1}^T \left\{\hat{\psi}^2_{i,T,t}(\hat{\theta}_{i,T}) - \psi^2_{i,T,t}\right\} = o_p(1),$$

where $E[\psi^2_{i,T,t}]/T = o(1)$ is shown in the proof of Lemma B.1, hence (a) follows. Define

$$I_{i,T,t}(\theta) \equiv I\left(\left|\mathcal{E}_{i,t}(\theta)\right| \leq c_{i,T}^{(\mathcal{E})}(\theta)\right) \quad \textit{and}$$

$$\hat{I}_{i,T,t}(\theta) \equiv I\left(\left|\mathcal{E}_{i,t}(\theta)\right| \leq \mathcal{E}^{(a)}_{i,(k_{i,T})}(\theta)\right).$$

In order to reduce notation we only work with the uncentered variables $\psi^{(\mathcal{E})}_{i,T,t}(\theta) = \mathcal{E}_{i,t}(\theta)I_{i,T,t}(\theta)$ and $\hat{\psi}^{(\mathcal{E})}_{i,T,t}(\theta) = \mathcal{E}_{i,t}(\theta)\hat{I}_{i,T,t}(\theta)$. Identical arguments extend to the centered versions. Rearrange terms to deduce

$$\frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}} \sum_{t=1}^T \left\{\hat{\psi}_{1,T,t}(\hat{\theta}_{1,T})\hat{\psi}_{2,T,t-h}(\hat{\theta}_{2,T}) - \psi_{1,T,t}\psi_{2,T,t-h}\right\}$$
$$= \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}} \sum_{t=1}^T \left\{\hat{\psi}_{1,T,t} - \psi_{1,T,t}\right\}\psi_{2,T,t-h} + \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}$$
$$\times \sum_{t=1}^T \psi_{1,T,t}\left\{\hat{\psi}_{2,T,t-h} - \psi_{2,T,t-h}\right\} + \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}$$
$$\times \sum_{t=1}^T \left\{\hat{\psi}_{1,T,t} - \psi_{1,T,t}\right\}\left\{\hat{\psi}_{2,T,t-h} - \psi_{2,T,t-h}\right\} + \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}$$
$$\times \sum_{t=1}^T \left\{\hat{\psi}_{1,T,t}(\hat{\theta}_{1,T}) - \hat{\psi}_{1,T,t}\right\}\psi_{2,T,t-h} + \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}$$
$$\times \sum_{t=1}^T \psi_{1,T,t}\left\{\hat{\psi}_{2,T,t-h}(\hat{\theta}_{2,T}) - \hat{\psi}_{2,T,t-h}\right\} + \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}$$
$$\times \sum_{t=1}^T \left\{\hat{\psi}_{1,T,t}(\hat{\theta}_{1,T}) - \hat{\psi}_{1,T,t}\right\}\left\{\hat{\psi}_{2,T,t-h} - \psi_{2,T,t-h}\right\}$$
$$+ \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}$$

Table 10

(ETF's - exponential) Occupation times (τ) for the $\hat{Q}_T^{(\mathcal{E})}$-test of volatility spillover at $\alpha \in \{1\%, 5\%, 10\%\}$ levels using the *exponential* transform, at horizon H, and a QMTLL plug-in. We reject the null of no spillover at level α if $\tau > \alpha$.

FROM/TO	IVV	AGG	EFA	GSG	REM
$H = 1$					
IVV	(-,-,-)	(.000,.156,.222)	(.000,.000,.000)	(.000,.156,.222)	(.000,.178,.422)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.067)	(.000,.000,.000)
EFA	(.000,.000,.000)	(.000,.000,.044)	(-,-,-)	(.000,.000,.022)	(.000,.133,.333)
GSG	(.133,.622,.800)	(.000,.178,.400)	(.000,.000,.000)	(-,-,-)	(.000,.000,.022)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 2$					
IVV	(-,-,-)	(.000,.111,.200)	(.000,.000,.000)	(.000,.000,.156)	(.000,.133,.311)
AGG	(.000,.000,.089)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.000)	(.000,.000,.022)	(-,-,-)	(.000,.000,.000)	(.000,.044,.511)
GSG	(.000,.378,.644)	(.000,.156,.222)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 3$					
IVV	(-,-,-)	(.000,.133,.356)	(.000,.000,.000)	(.000,.000,.156)	(.000,.000,.067)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.067)
GSG	(.089,.222,.444)	(.000,.067,.156)	(.067,.089,.089)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 4$					
IVV	(-,-,-)	(.000,.089,.267)	(.000,.000,.000)	(.000,.000,.156)	(.000,.000,.089)
AGG	(.000,.000,.089)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.022,.311)
GSG	(.000,.156,.444)	(.000,.022,.133)	(.044,.067,.089)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.044,.178)	(-,-,-)
$H = 5$					
IVV	(-,-,-)	(.000,.000,.156)	(.000,.000,.067)	(.000,.000,.000)	(.000,.000,.000)
AGG	(.000,.000,.044)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.022,.089)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.311,.422)
GSG	(.000,.267,.556)	(.000,.000,.067)	(.022,.067,.089)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.089)	(-,-,-)

$$\times \sum_{t=1}^{T} \left\{ \hat{\psi}_{1,T,t} - \psi_{1,T,t} \right\} \left\{ \hat{\psi}_{2,T,t-h}(\hat{\theta}_{2,T}) - \hat{\psi}_{2,T,t-h} \right\}$$
$$= \sum_{i=1}^{7} \mathcal{A}_{i,T}.$$

Approximation theory developed in Hill (2013b; 2014a, appendices) 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

$$\mathrm{J}_{i,t}(\theta_i) \equiv \frac{\partial}{\partial \theta_i} \mathcal{E}_{i,t}(\theta_i) = -\epsilon_{i,t}^2(\theta_i) \frac{\partial}{\partial \theta_i} h_{i,t}^2(\theta_i).$$

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

$$\begin{aligned}
\mathcal{A}_{4,T} &= \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}} \sum_{t=1}^{T} \mathcal{E}_{1,t} \left\{ \hat{I}_{1,T,t}(\hat{\theta}_{1,T}) - \hat{I}_{1,T,t} \right\} \psi_{2,T,t-h} \\
&\quad + \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}} \sum_{t=1}^{T} \mathrm{J}_{1,t} \hat{I}_{1,T,t} \psi_{2,T,t-h} \times \left(\hat{\theta}_{1,T} - \theta_1^0 \right) \\
&\quad + \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}} \sum_{t=1}^{T} \mathrm{J}_{1,t} \left\{ \hat{I}_{1,T,t}(\hat{\theta}_{1,T}) - \hat{I}_{1,T,t} \right\} \\
&\quad \times \psi_{2,T,t-h} \times \left(\hat{\theta}_{1,T} - \theta_1^0 \right) \\
&\quad + \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}} \sum_{t=1}^{T} \left\{ \mathrm{J}_{1,t}(\theta_{1.*}) - \mathrm{J}_{1,t} \right\} \\
&\quad \times \hat{I}_{1,T,t}(\hat{\theta}_{1,T}) \psi_{2,T,t-h} \times \left(\hat{\theta}_{1,T} - \theta_1^0 \right) = \sum_{i=1}^{4} \mathcal{B}_{i,T}.
\end{aligned} \tag{26}$$

By assumption $\mathfrak{V}_{i,T}^{1/2}(\theta_{i.*} - \theta_i^0) = O_p(1)$ hence arguments essentially identical to Hill (2014a, Lemmas A.1, A.4, A.6) suffice to prove $\mathcal{B}_{1,T}, \mathcal{B}_{3,T}, \mathcal{B}_{4,T} = o_p(1)$. See also Hill (2013b, Appendix B).

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

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

where $\mathcal{C}_{3,T} = o_p(1)$ by the proof of Lemma A.1 in Hill (2014a).

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

$$\mathrm{J}\psi_{1,T,t} \equiv \mathrm{J}_{1,t} I_{1,T,t} \psi_{2,T,t-h}.$$

Table 11
(ETF's - truncation) Occupation times (τ) for the $\hat{Q}_T^{(\mathcal{E})}$-test of volatility spillover at $\alpha \in \{1\%, 5\%, 10\%\}$ levels using *truncation*, at horizon H, and a QMTLL plug-in. We reject the null of no spillover at level α if $\tau > \alpha$.

FROM/TO	IVV	AGG	EFA	GSG	REM
$H = 1$					
IVV	(-,-,-)	(.000,.000,.089)	(.000,.000,.200)	(.044,.156,.311)	(.022,.422,.533)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.067)
EFA	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.222,.267)
GSG	(.156,.467,.622)	(.000,.133,.244)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 2$					
IVV	(-,-,-)	(.000,.000,.022)	(.000,.000,.000)	(.000,.089,.156)	(.000,.289,.422)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.200,.289)
GSG	(.000,.289,.644)	(.000,.089,.178)	(.156,.311,.356)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 3$					
IVV	(-,-,-)	(.000,.000,.222)	(.000,.000,.000)	(.000,.022,.089)	(.000,.244,.378)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.022)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.089,.244)
GSG	(.000,.200,.444)	(.000,.000,.067)	(.022,.311,.378)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)
$H = 4$					
IVV	(-,-,-)	(.000,.000,.000)	(.000,.289,.333)	(.000,.000,.089)	(.000,.111,.356)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.044,.156)
GSG	(.000,.289,.756)	(.000,.000,.000)	(.022,.200,.378)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.022)	(-,-,-)
$H = 5$					
IVV	(-,-,-)	(.000,.000,.000)	(.000,.333,.333)	(.000,.000,.000)	(.000,.089,.178)
AGG	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)
EFA	(.000,.000,.067)	(.000,.000,.000)	(-,-,-)	(.000,.000,.000)	(.111,.222,.311)
GSG	(.000,.156,.578)	(.000,.000,.000)	(.022,.022,.333)	(-,-,-)	(.000,.000,.000)
REM	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(.000,.000,.000)	(-,-,-)

Recall the definition $\mathfrak{J}_{i,T}^{(h)} \equiv (\partial/\partial\theta_i)E[\psi_{1,T,t}(\theta_1)\psi_{2,T,t-h}(\theta_2)]|_{\theta^0}$. By Lemma A.6.c of Hill (2014a) the Jacobian of the tail-trimmed mean is proportional to the trimmed mean of the Jacobian:

$$\begin{aligned}
\mathfrak{J}_{1,T}^{(h)} &= \frac{\partial}{\partial\theta_1}E\left[\psi_{1,T,t}(\theta_1)\,\psi_{2,T,t-h}(\theta_2)\right]|_{\theta^0} \\
&= E\left[\mathrm{J}_{1,t}I_{1,T,t}\psi_{2,T,t-h}\right] \times \left(1 + o_p(1)\right) \\
&= E\left[\mathrm{J}\psi_{1,T,t}\right] \times \left(1 + o_p(1)\right). \qquad (27)
\end{aligned}$$

Under mutual independence $E[\mathrm{J}\psi_{1,T,t}] = 0$ follows from L_2-boundedness of $\mathrm{J}_{1,t}I_{1,T,t}$ given $(\partial/\partial\theta_i)h_{i,t}^2$ is L_2-bounded under Assumption A, and $E[\psi_{2,T,t-h}] = 0$ in view of recentering. Therefore $\mathfrak{J}_{1,T}^{(h)} = o(1)$, and $\mathcal{C}_{2,T} = o(1)$ given $(T/\mathfrak{S}_T)^{1/2}(\hat{\theta}_{i,T} - \theta_i^0) = O_p(1)$. Otherwise, exploit (27) and the supposition $\mathfrak{V}_{i,T}^{1/2}(\hat{\theta}_{i,T} - \theta_i^0) = O_p(1)$ to deduce $|\mathcal{C}_{2,T}| \leq K|\mathfrak{V}_{1,T}^{1/2}(\hat{\theta}_{1,T} - \theta_1^0)| = O_p(1)$.

Finally, for $\mathcal{C}_{1,T}$ use (27) and $\mathfrak{V}_{i,T}^{1/2}(\hat{\theta}_{i,T} - \theta_i^0) = O_p(1)$ to obtain

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

say. If $\limsup_{T\to\infty}\sup_{1\leq t\leq T}E|\mathrm{J}\psi_{1,T,t}|^{1+\iota} < \infty$ for some $\iota > 0$ then $\mathrm{J}\psi_{1,T,t}$ 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. □

Lemma B.3 (*clt*). *Let $H \in \mathbb{R}$ be arbitrary and define $\mathfrak{S}_T \equiv E[\psi_{1,T,t}^2]E[\psi_{2,T,t-h}^2]$, $\mathcal{Z}_{h,T} \equiv T^{-1/2}\mathfrak{S}_T^{-1/2}\sum_{t=1}^{T}\psi_{1,T,t}\psi_{2,T,t-h}$ and $\mathcal{Z}_T \equiv \left[\mathcal{Z}_{1,T}, \ldots, \mathcal{Z}_{H,T}\right]'$. Under Assumptions D, E, and T if $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are mutually independent then $\mathcal{Z}_{h,T} \xrightarrow{d} N(0, I_H)$.*

Proof. Pick any $r \in \mathbb{R}^H$, $r'r = 1$ and define $\psi_{2,T,t}^{(r)} \equiv \sum_{h=1}^{H} r_h \psi_{2,T,t-h}$ and $\mathcal{Z}_{T,t}^{(r)} \equiv \psi_{1,T,t}\psi_{2,T,t}^{(r)}/\mathfrak{S}_T^{1/2}$. By construction, and serial and mutual independence we have $E[r'\mathcal{Z}_T] = 0$ and $E[(r'\mathcal{Z}_T)^2] = 1$, hence $r'\mathcal{Z}_T = 1/T^{1/2}\sum_{t=1}^{T}\mathcal{Z}_{T,t}^{(r)}$ is a self-standardized sum of independent random variables $\mathcal{Z}_{T,t}^{(r)}$ with zero mean and unit variance. Therefore $r'\mathcal{Z}_T \xrightarrow{d} N(0, 1)$ by the Lindeberg–Feller central limit theorem, provided we demonstrate the Lindeberg condition $\lim_{T\to\infty} 1/T\sum_{t=1}^{T}E[(\mathcal{Z}_{T,t}^{(r)})^2 I(|\mathcal{Z}_{T,t}^{(r)}| > T^{1/2}\varepsilon)] \to 0\ \forall\varepsilon > 0$. Although $\psi_{1,T,t}\psi_{2,T,t}^{(r)}/\mathfrak{S}_T^{1/2}$ is not identically distributed across $1 \leq t \leq T$ and $T \geq 1$, it is iid over $1 \leq t \leq T$. Hence we require $\lim_{T\to\infty}E[(\mathcal{Z}_{T,t}^{(r)})^2 I(|\mathcal{Z}_{T,t}^{(r)}| > T^{1/2}\varepsilon)] \to 0\ \forall\varepsilon > 0$. But this holds by dominated convergence in view of $E[(\mathcal{Z}_{T,t}^{(r)})^2] = 1$. □

Proof of Theorem 3.1. We prove (a) while (b) follows from Theorem 4.3.a. Invoke the null of mutual independence such that $E[\psi_{1,T,t}\psi_{2,T,t-h}] = 0$, and recall $\epsilon_{i,t}$ are serially independent under

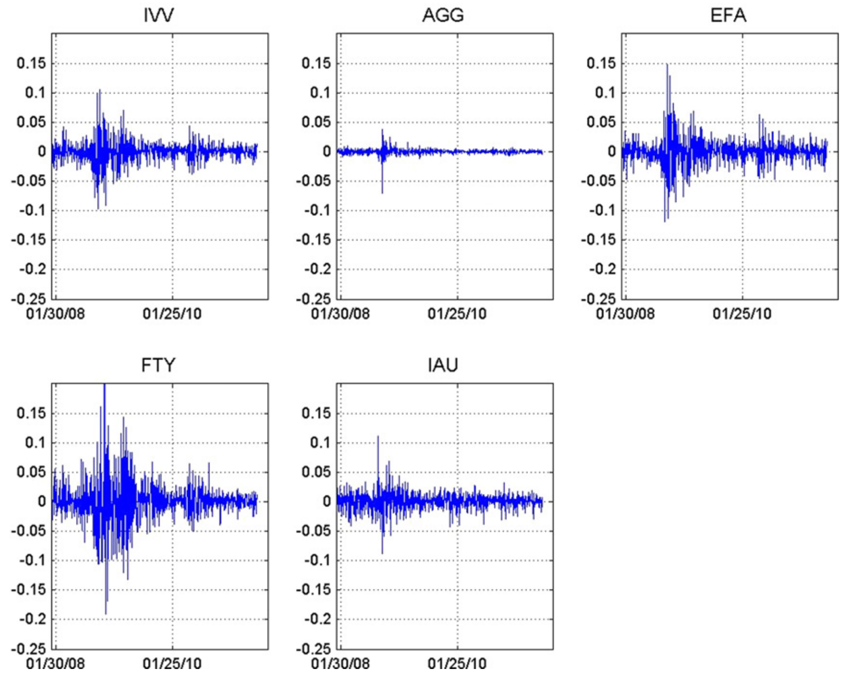

Fig. 1. (ETF's) Top panel: ETF daily log returns: 1/3/12–12/31/12. Bottom panel: Hill plots of daily log returns with robust 95% confidence bands.

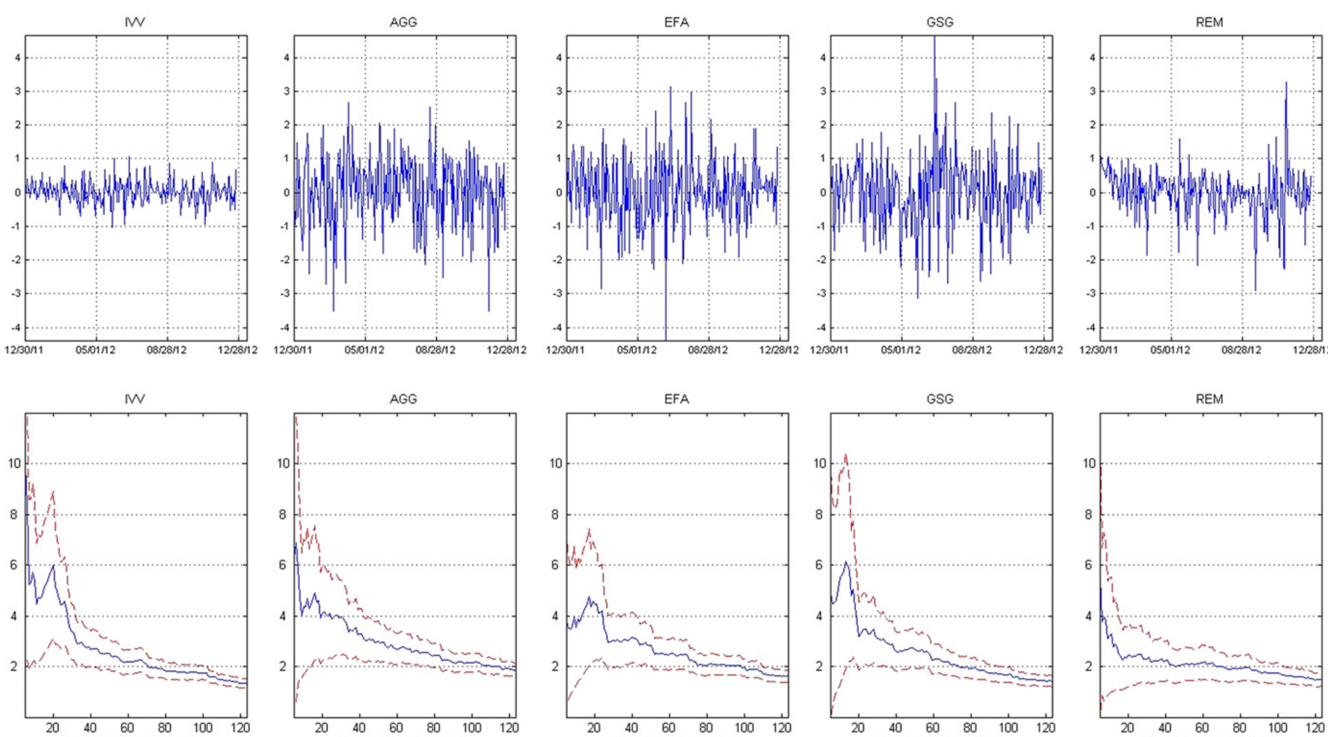

Fig. 2. (ETF) Top panel: ARMA(1, 1)–GARCH(1, 1) residuals $\hat{\epsilon}_t = \hat{u}_t/\hat{h}_t$ for ETF's, where $\hat{u}_t$ are the residuals from an ARMA filter and $\hat{h}_t$ are the GARCH conditional variances. Bottom panel: Hill plots of residuals $\hat{\epsilon}_t$ with robust 95% confidence bands.

Assumption E.1. Note $1/T\sum_{t=1}^{T}\psi_{i,T,t}^2/E[\psi_{i,T,t}^2] \xrightarrow{p} 1$ by Lemma B.1. Therefore, by three applications of Lemma B.2

$$T^{1/2}\hat{\rho}_{T,h}(\hat{\theta}_T)$$

$$= \frac{1/T^{1/2}\sum_{t=1}^{T}\hat{\psi}_{1,T,t}(\hat{\theta}_{1,T})\hat{\psi}_{2,T,t-h}(\hat{\theta}_{2,T})}{\left(1/T\sum_{t=1}^{T}\hat{\psi}_{1,T,t}^2(\hat{\theta}_{1,T})\right)^{1/2}\left(1/T\sum_{t=1}^{T}\hat{\psi}_{2,T,t-h}^2(\hat{\theta}_{2,T})\right)^{1/2}}$$

$$= \frac{1}{T^{1/2}\mathfrak{S}_T^{1/2}}\sum_{t=1}^{T}\psi_{1,T,t}\psi_{2,T,t-h} \times \left(1 + o_p(1)\right) + o_p(1)$$

$$=: \mathcal{Z}_{h,T} \times \left(1 + o_p(1)\right) + o_p(1),$$

say, where $\mathfrak{S}_T \equiv E[\psi_{1,T,t}^2]E[\psi_{2,T,t-h}^2]$. The claim now follows from Lemma B.3 and the mapping theorem. □

Proof of Theorem 3.2. 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

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

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

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

Now take the probability limit to complete the proof. □

Proof of Lemma 4.1. Drop all superscripts. We borrow notation from the proof of Lemma B.2. It suffices to consider $\mathfrak{V}_{i,T}$ in (11). Under mutual independence $\mathfrak{J}_{i,T}^{(h)} = o_p(1)$ follows from the proof of Lemma B.2, hence $\mathfrak{V}_{i,T}$ is identically $T/(E[\psi_{1,T,t}^2]E[\psi_{2,T,t}^2])$ as $T \to \infty$. If $\kappa_l > 4$ for both $l \in \{1, 2\}$ then both $\mathfrak{V}_{i,T} \sim KT$, and if $\kappa_l < 4$ then $E[\psi_{l,T,t}^2] \sim K(T/k_{l,T})^{4/\kappa_l - 1}$. 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 $E[\psi_{l,T,t}^2] \to \infty$ is slowly varying if $\kappa_l = 4$. □

Proof of Theorem 4.2. The proof is similar to the proof of Theorem S1.4 in Aguilar and Hill (2014). □

Appendix C. Simulation and empirical results

See Tables 7–11 and Figs. 1 and 2

Appendix D. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2014.09.001.

References

Aguilar, M., Hill, J.B., 2014. Supplemental Appendix for "Robust Score and Portmanteau Tests of Volatility Spillover". Dept. of Economics, University of North Carolina.

Ait-Sahalia, Y., Jacod, J., 2009. Testing for jumps in a discretely observed process. Ann. Statist. 37, 184–222.

Ait-Sahalia, Y., Jacod, J., Li, J., 2012. Testing for jumps in noisy high frequency data. J. Econometrics 168, 207–222.

Andrews, D.W.K., 1988. Laws of large numbers for dependent non-identically distributed random variables. Econometric Theory 4, 458–467.

Andrews, D.W.K., 2001. Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica 69, 683–734.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P., Rogers, W.H., Tukey, J.W., 1972. Robust Estimates of Location. Princeton Univ. Press.

Baillie, R., Bollerslev, T., 1990. Intra-day and inter-market volatility in foreign exchange markets. Rev. Econom. Stud. 58, 565–585.

Beirne, J., Caporale, G.M., Schulz-Ghattas, M., Spagnolo, N., 2008. Volatility spillovers and contagion from mature to emerging stock markets, iMF Working Paper.

Berkes, I., Horváth, L., 2004. The efficiency of the estimators of the parameters in garch processes. Ann. Statist. 32, 633–655.

Berkes, I., Horváth, L., Kokoszka, P., 2003. Estimation of the maximal moment exponent of a GARCH(1, 1) sequence. Econometric Theory 19, 565–586.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307–327.

Bollerslev, T., Todorov, V., Li, S.Z., 2013. Jump tails, extreme dependencies and the distribution of stock returns. J. Econometrics 172, 307–324.

Boudt, K., Croux, C., 2010. Robust m-estimation of multivariate garch models. Comput. Statist. Data Anal. 54, 2459–2469.

Boudt, K., Danielsson, J., Laurent, S., 2011. Robust Forecasting of Dynamic Conditional Correlation Garch Models. Mimeo.

Bougerol, P., Picard, N., 1992. Stationarity of Garch processes and of some nonnegative time series. J. Econometrics 52, 115–127.

Box, G.E.P., Pierce, D.A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J. Amer. Statist. Assoc. 65, 1509–1526.

Breiman, L., 1965. On some limit theorems similar to the arc-sin law. Theory of Probab. Appl. 10, 323–331.

Brock, W., Dechert, W., Scheinkman, J., LeBaron, B., 1996. A test for independence based on the correlation dimension. Econometric Rev. 15, 197–235.

Brooks, C., 1998. Predicting stock index volatility: Can market volume help? J. Forecast. 17, 59–80.

Brownlees, C., Engle, R., Kelly, B., 2011. A practical guide to volatility forecast through calm and storm. J. Risk 14, 3–22.

Campbell, J., Hentschel, L., 1992. No news is good news: An asymmetric model of changing volatility in stock returns. J. Financ. Econ. 31, 281–318.

Caporale, G.M., Cipollini, A., Spagnolo, N., 2005. Testing for contagion: A conditional correlation analysis. J. Empir. Finance 12, 476–489.

Caporale, G.M., Pittis, N., Spagnolo, N., 2006. Volatility transmission and financial crises. J. Econ. Finance 30, 376–390.

Carrasco, M., Chen, X., 2002. Mixing and moment properties of various garch and stochastic volatility models. Econometric Theory 18, 17–39.

Chang, C.L., McAleer, M., Tansuchat, R., 2011. Crude oil hedging strategies using dynamic multivariate garch. Energy Econ. 33, 912–923.

Cheung, Y.-W., Ng, L.K., 1996. A causality-in variance test and its application to financial market prices. J. Econometrics 72, 33–48.

Cline, D.B.H., 1986. Convolution tails, product tails and domains of attraction. Probab. Theory Related Fields 72, 529–557.

Comte, F., Lieberman, O., 2000. Second-order noncausality in multivariate garch processes. J. Time Ser. Anal. 21, 535–557.

Csörgo, S., Horváth, L., Mason, D., 1986. What portion of the sample makes a partial sum asymptotically stable or normal? Probab. Theory Related Fields 72, 1–16.

Davis, R.A., Mikosch, T., 2009. The extremogram: A correlogram for extreme events. Bernoulli 15, 977–1009.

Davis, R.A., Resnick, S.I., 1986. Limit theory for the sample covariance and correlation function of moving average. Ann. Statist. 14, 533–558.

de Lima, P., 1996. Nuisance parameter free properties of correlation integral based statistics. Econometric Rev. 15, 237–259.

de Lima, P., 1997. On the robustness of nonlinearity tests to moment condition failure. J. Econometrics 76, 251–280.

Devlin, S.J., Gnanadesikan, R., Kettenring, J.R., 1975. Robust estimation and outlier detection with correlation coefficients. Biometrika 62, 531–545.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econom. Statist. 13, 253–263.

Drost, F.C., Nijman, T., 1993. Temporal aggregation of garch processes. Econometrica 61, 909–927.

Dufour, J.M., Farhat, A., Hallin, M., 2006. Distribution-free bounds for serial correlation coefficients in heteroscedastic symmetric time series. J. Econometrics 130, 123–142.

Dungey, M., Fry, R., Gonzlez-Hermosillo, B., Martin, V.L., 2005. Sample Properties of Contagion Tests. Mimeo, Australian National University.

Dungey, M., Fry, R., Gonzlez-Hermosillo, B., Martin, V.L., 2006. Contagion in international bond markets during the Russian and the ltcm crises. J. Financ. Stab. 2, 1–27.

Dungey, M., Fry, R., Gonzlez-Hermosillo, B., Martin, V.L., Tang, C., 2010. Are financial crises alike? IMF Working Paper WP/10/14.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1999. Modelling Extremal Events for Insurance and Finance. Springer, New York.

Engle, R., Ito, I., Lin, W., 1990. Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. Econometrica 58, 525–542.

Engle, R., Ng, V., 1993. Measuring and testing the impact of news on volatility. J. Finance 48, 1749–1778.

Fan, J., Qi, L., Xiu, D., 2014. Quasi maximum likelihood estimation of garch models with heavy-tailed likelihoods. J. Bus. Econom. Statist. 32, 178–191.

B. Finkenstadt, and H. Rootzen (Eds.), 2003. Extreme Values in Finance, Telecommunications, and the Environment. Chapman & Hall/CRC.

Forbes, K., Rogibon, R., 2002. No contagion, only interdependence: Measuring stock market comovements. J. Finance 57, 2223–2261.

Francq, C., Zakoïan, J.-M., 2004. Maximum likelihood estimation of pure garch and ARMA–GARCH processes. Bernoulli 10, 605–637.
Francq, C., Zakoïan, J.-M., 2006. Mixing properties of a general class of GARCH(1, 1) models without moment assumptions on the observed process. Econometric Theory 22, 815–834.
Francq, C., Zakoïan, J.-M., 2010. GARCH Models: Structure, Statistical Inference and Financial Applications. Wiley, New York.
Fung, W., Hsieh, D., 1999. A primer on hedge funds. J. Empir. Finance 6, 209–331.
Genton, M.G., Ronchetti, E., 2003. Robust indirect inference. J. Amer. Statist. Assoc. 98, 67–76.
Glick, R., Rose, A.K., 1999. Contagion and trade: Why are currency crises regional? J. Int. Money Finance 18, 603–617.
Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. The J. Finance 48, 1779–1801.
Gnanadesikan, R., Kettenring, J.R., 1972. Robust estimates, residuals, and outlier detection with multiresponse data. Biometrics 28, 81–124.
Granger, C., Huang, B., Yang, C., 2000. A bivariate causality between stock prices and exchange rates: Evidence from recent asian flu. Quart. Rev. Econ. 40, 337–354.
Hall, P., Yao, Q., 2003. Inference in arch and garch models with heavy-tailed errors. Econometrica 71, 285–317.
Hampel, F.R., 1974. The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. 68, 383–393.
Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics: The Approach Based on Influence Functions. Wiley, New York.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Ann. Statist. 3 (5), 1163–1174.
Hill, J.B., 2009. On functional central limit theorems for dependent, heterogeneous arrays with applications to tail index and tail dependence estimation. J. Statist. Plann. Inference 139, 2091–2110.
Hill, J.B., 2010. On tail index estimation for dependent, heterogeneous data. Econometric Theory 26, 1398–1436.
Hill, J.B., 2012. Heavy-tail and plug-in robust consistent conditional moment tests of functional form. In: Chen, X., Swanson, N. (Eds.), Festschrift in Honor of Hal White. Springer, New York, pp. 241–274.
Hill, J.B., 2013a. Expected shortfall estimation and gaussian inference for infinite variance time series. J. Finan. Econ. in press.
Hill, J.B., 2013b. Least tail-trimmed squares for infinite variance autoregressions. J. Time Ser. Anal. 34, 168–186.
Hill, J.B., 2014a. Robust estimation and inference for heavy tailed garch. Bernoulli in press.
Hill, J.B., 2014b. Robust generalized empirical likelihood for heavy tailed autoregressions with conditionally heteroscedastic errors. Working Paper, Dept. of Economics, University of North Carolina.
Hill, J.B., 2014c. Tail index estimation for a filtered dependent time series. Statist. Sinica in press.
Hill, J.B., Aguilar, M., 2013. Moment condition tests for heavy tailed time series. J. Econometrics 172, 255–274.
Hill, J.B., McCloskey, A., 2014. Heavy tail robust frequency domain estimation. Working Paper, Department of Economics, University of North Carolina, Chapel Hill.
Hong, Y., 2001. A test for volatility spillover with application to exchange rates. J. Econometrics 103, 183–224.
Jiang, G.J., Oomen, R., 2008. Testing for jumps when asset prices are observed with noise–a 'swap variance' approach. J. Econometrics 144, 352–370.
Jureckova, J., Sen, P.K., 1996. Robust Statistical Inference: Asymptotics and Interrelations. Wiley, New York.
Karolyi, G.A., Stulz, R.M., 1996. Why do markets move together? An investigation of U.S.-japan stock return comovements. J. Finance 51, 951–986.
Kent, J.T., Tyler, D.E., 1991. Redescending m-estimates of multivariate location and scatter. Ann. Statist. 19, 2102–2119.
King, M., Sentana, E., Wadhwani, S., 1994. Volatility and links between national stock markets. Econometrica 62, 901–933.
Künsch, H., 1984. Infinitesimal robustness for autoregressive processes. Ann. Statist. 12, 843–863.
Leadbetter, M.R., Lindgren, G., Rootzén, H., 1983. Extremes and Related Properties of Random Sequences and Processes. Springer Series in Statistics, Springer-Verlag, New York.
Linton, O., Pan, J., Wang, H., 2010. Estimation for a non-stationary semi-strong GARCH(1, 1) model with heavy-tailed errors. Econometric Theory 26, 1–28.
Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets. J. Finance 56, 649–676.
Malevergne, Y., Sornette, D., 2004. How to account for extreme co-movements between individual stocks and the market. J. Risk 6, 71–116.
Mancini, L., Ronchetti, E., Trojani, F., 2005. Optimal conditionally unbiased bounded-influence inference in dynamic location and scale. J. Amer. Statist. Assoc. 100, 628–641.
Meddahi, N., Renault, E., 2004. Temporal aggregation of volatility models. J. Econometrics 119, 355–379.
Mikosch, T., Starica, C., 2000. Limit theory for the sample autocorrelations and extremes of a garch (1,1) process. Ann. Statist. 28, 1427–1451.
Muler, N., Yohai, V.J., 2008. Robust estimates for garch models. J. Statist. Plann. Inference 138, 2918–2940.
Nelson, D., 1990. Stationarity and persistence in the GARCH(1, 1) model. Econometric Theory 6, 318–334.
Ng, A., 2000. Volatility spillover effects from japan and the us to the pacific-basin. J. Int. Money Finance 19, 207–233.
Ortelli, C., Trojani, F., 2005. Robust efficient method of moments. J. Econometrics 128, 69–97.
Patton, A., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econometrics 160, 246–256.
Peng, L., Yao, Q., 2003. Least absolute deviations estimation for arch and garch models. Biometrika 90, 967–975.
Poon, S.-H., Rickinger, M., Tawn, J., 2003. Modelling extreme-value dependence in international stock markets. Statist. Sinica 13, 929–953.
Resnick, S., 1987. Extreme Values, Regular Variation and Point Processes. Springer-Verlag, New York.
Ronchetti, E., Trojani, F., 2001. Robust inference with gmm estimators. J. Econometrics 101, 37–69.
Rousseeuw, P.J., Yohai, V.J., 1984. Robust regression by means of s-estimators. In: Franke, W. H., Martin, D. (Eds.), Robust and Nonlinear Time Series Analysis, Springer-Verlag, New York, p. 2562272.
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. J. Econometrics 78, 205–216.
Sakata, S., White, H., 1998. High breakdown point conditional dispersion estimation with application to s&p 500 daily returns volatility. Econometrica 66, 529–567.
Schmidt, R., Stadtmüller, U., 2006. Non-parametric estimation of tail dependence. Scand. J. Stat. 33, 307–335.
Sentana, E., 1995. Quadratic arch models. Rev. Econom. Stud. 62, 639–661.
Shevlyakov, G., Morgenthaler, S., Shurygin, A., 2008. Redescending m-estimators. J. Statist. Plann. Inference 138, 2906–2917.
So, R., 2001. Price and volatility spillovers between interest rate and exchange value of the us dollar. Glob Finance J. 12, 95–107.
Tse, Y., Booth, G.G., 1996. Common volatility and volatility spillovers between u.s. and eurodollar interest rates: Evidence from the futures market. J. Econ. Bus. 48, 299–312.
Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.
West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.
Wooldridge, J.M., 1990. A unified approach to robust, regression-based specification tests. Econometric Theory 6, 17–43.
Yang, S., Doong, S., 2004. Price and volatility spillovers between stock prices and exchange rates: Empirical evidence from the g-7 countries. Int. J. Bus. Econ. 3, 139–153.
Zhang, L., 2011. Estimating covariation: Epps effect, microstructure noise. J. Econometrics 160, 33–47.
Zhu, K., Ling, S., 2011. Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA–GARCH/igarch models. Ann. Statist. 39, 2131–2163.