ELSEVIER

Journal of Econometrics 103 (2001) 183–224

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

A test for volatility spillover with application to exchange rates

Yongmiao Hong[a,b,*]

[a] *Department of Economics and Department of Statistical Science, Cornell University, Ithaca, NY 14853, USA*

[b] *Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

Accepted 24 October 2000

Abstract

This paper proposes a class of asymptotic N(0, 1) tests for volatility spillover between two time series that exhibit conditional heteroskedasticity and may have infinite unconditional variances. The tests are based on a weighted sum of squared sample cross-correlations between two squared standardized residuals. We allow to use all the sample cross-correlations, and introduce a flexible weighting scheme for the sample cross-correlation at each lag. Cheung and Ng (1996) test and Granger (1969)-type regression-based test can be viewed as uniform weighting because they give equal weighting to each lag. Non-uniform weighting often gives better power than uniform weighting, as is illustrated in a simulation study. We apply the new tests to study Granger-causalities between two weekly nominal U.S. dollar exchange rates—Deutschemark and Japanese yen. It is found that for causality in mean, there exists only simultaneous interaction between the two exchange rates. For causality in variance, there also exists strong simultaneous interaction between them. Moreover, a change in past Deutschemark volatility Granger-causes a change in current Japanese yen volatility, but a change in past Japanese yen volatility does not Granger-cause a change in current Deutschemark volatility. © 2001 Elsevier Science S.A. All rights reserved.

JEL classification: C12, C32, F31

Keywords: Causality in variance; Cross-correlation; Exchange rate; GARCH; Granger causality; Standardized residual; Volatility spillover

* Tel.: +1-607-255-5130; fax: +1-607-255-2818.
E-mail address: yh20@cornell.edu (Y. Hong).

0304-4076/01/$ - see front matter © 2001 Elsevier Science S.A. All rights reserved.
PII: S0304-4076(01)00043-4

1. Introduction

Detection and elucidation of volatility spillover across different assets or markets is important in finance and macroeconomics (e.g., Baillie and Bollerslev, 1990; Cheung and Ng, 1990, 1996; Engle et al., 1990; Engle and Susmel, 1993; Granger et al., 1986; Hamao et al., 1990; King and Wadhwani, 1990; King et al., 1994; Lin et al., 1994; Schwert, 1989). Absence of volatility spillover implies that the major sources of disturbances are changes in asset- or market-specific fundamentals, and one large shock increases the volatility only in that specific asset or market. In contrast, existence of volatility spillover implies that one large shock increases the volatilities not only in its own asset or market but also in other assets or markets as well.

Volatility is often related to the rate of information flow (e.g., Ross, 1989). If information comes in clusters, asset returns or prices may exhibit volatility even if the market perfectly and instantaneously adjusts to the news. Therefore, study on volatility spillover can help understand how information is transmitted across assets and markets. Alternatively, the existence of volatility spillover may be consistent with the market dynamics which exhibits volatility persistence due to private information or heterogeneous beliefs (e.g., Admati and Pfleiderer, 1988; Kyle, 1985; Shalen, 1993). Here, whether volatilities are correlated across markets is important in examining the speed of market adjustment to new information. It is also hypothesized that the changes in market volatility are related to the volatilities of macroeconomic variables. In present value models such as those of Shiller (1981a, b), for example, changes in the volatility of either future cash flows or discount rates cause changes in the volatility of stock returns. Such a macroeconomic hypothesis can be checked by testing volatility spillover.

Nearly all the existing empirical studies on volatility spillover use techniques much like a Granger (1969)-type test, namely by regressing the squared residual of one variable on the squared residuals from its own lagged and other lagged variables in the framework of multivariate GARCH models. Cheung and Ng (1996) (see also Cheung and Ng, 1990) recently proposed a new test for volatility spillover using the sample cross-correlation function between two squared residuals standardized by their conditional variance estimators respectively. Specifically, Cheung and Ng (1996) test is based on the sum of finitely many (M say) squared sample cross-correlations, which has a null asymptotically χ^2_M distribution. This test is relatively simple and convenient to implement, and can provide valuable information in building multivariate GARCH models (cf. Cheung and Ng, 1996).

In this paper, we propose a class of new tests for volatility spillover. Essentially we test for causality in variance in the sense of Granger (1969, 1980), who introduces the concept of causality in terms of incremental predictive ability of one time series for another, as opposed to the more conventional definition of cause and effect. Our tests are a properly standardized version of a weighted sum of squared sample cross-correlations between two squared standardized residuals, and have a null asymptotic $N(0, 1)$ distribution. We do not assume any specific innovation distribution (e.g., normality), and the tests apply to time series that exhibit conditional heteroskedasticity and may have infinite unconditional variances. We permit M, the number of the used sample cross-correlations, to increase with the sample size T (say). In fact, all $T - 1$ sample cross-correlations can be used. This enhances good power against the alternatives with slowly decaying cross-correlations. For such alternatives, the cross-correlation at each lag may be small but their joint effect is substantial. We also introduce a flexible weighting scheme for the sample cross-correlation at each lag. Typically, larger weights are given to lower order lags. In contrast, Cheung and Ng's (1996) test gives uniform weighting to each lag. Non-uniform weighting is expected to give better power against the alternatives whose cross-correlations decay to zero as the lag order increases. Such alternatives often arise in practice, because economic agents normally discount past information. We note that the idea of using non-uniform weighting can also be found in Engle (1982), who uses linearly declining weighting to improve the power of his popular Lagrange Multiplier (LM) test for ARCH effects.

We apply our tests to study volatility spillover between two weekly nominal U.S. dollar exchange rates—Deutschemark and Japanese yen. It is found that there exists strong simultaneous volatility interaction between them. Also a change in past Deutschemark volatility Granger-causes a change in current Japanese yen volatility, but a change in past Japanese yen volatility does not Granger-cause a change in current Deutschemark volatility. These findings differ somewhat from such studies as Baillie and Bollerslev (1990), who find no volatility spillover between these two exchange rates recorded on an hourly basis.

We state hypotheses of interest in Section 2, and introduce the test statistics in Section 3. In Section 4, we derive the null asymptotic distribution of the test statistics and discuss their asymptotic power property under a general class of alternatives. Section 5 reports a simulation study comparing the new tests with Cheung and Ng (1996) test. In Section 6, we apply the tests to study spillover between Deushemark and Japanese yen. The last section provides conclusions and directions for further research. All the proofs are collected in the appendix. Throughout the paper, all convergencies, unless indicated, are taken as the sample size $T \to \infty$.

2. Hypotheses of interest

In modeling two strictly stationary time series $\{Y_{1t}, Y_{2t}\}_{t=-\infty}^{\infty}$, one may be interested in their cross-dependence patterns, especially various Granger causalities (cf. Granger, 1969, 1980). In this paper, we focus on Granger causalities between time-varying conditional variances of Y_{1t} and Y_{2t}, whose unconditional variances may not exist. To state the hypothesis, it may be helpful to briefly review the concept of causality introduced by Granger (1969, 1980). Let I_{it}, $i = 1,2$, be the information set of time series $\{Y_{it}\}$ available at period t, and let $I_t = (I_{1t}, I_{2t})$. As defined in Granger (1980), Y_{2t} is said to Granger-cause Y_{1t} with respect to I_{t-1} if

$$Pr(Y_{1t}|I_{1t-1}) \neq Pr(Y_{1t}|I_{t-1}). \tag{1}$$

Granger (1980) points out that (1) is too general to be operational. [1] In practice, a less general but more easily testable definition is that Y_{2t} Granger-causes Y_{1t} in mean with respect to I_{t-1} if

$$\mathrm{E}(Y_{1t}|I_{1t-1}) \neq \mathrm{E}(Y_{1t}|I_{t-1}) \equiv \mu_{1t}^0. \tag{2}$$

Granger (1969) proposed a convenient regression-based test for (2), assuming conditional homoskedasticity for both Y_{1t} and Y_{2t}. [2]

As introduced in Granger et al. (1986, p. 2), it is also natural to define the "causality in variance" hypotheses as well, which can be stated as

$$\mathrm{H}_0\text{: } \mathrm{E}\{(Y_{1t} - \mu_{1t}^0)^2|I_{1t-1}\} = \mathrm{E}\{(Y_{1t} - \mu_{1t}^0)^2|I_{t-1}\} \equiv Var(Y_{1t}|I_{t-1}) \tag{3}$$

versus

$$\mathrm{H}_A\text{: } \mathrm{E}\{(Y_{1t} - \mu_{1t}^0)^2|I_{1t-1}\} \neq Var(Y_{1t}|I_{t-1}). \tag{4}$$

Note that $\mathrm{E}\{(Y_{1t} - \mu_{1t}^0)^2|I_{1t-1}\} \neq Var(Y_{1t}|I_{1t-1})$ because $\mu_{1t}^0 \neq \mathrm{E}(Y_{1t}|I_{1t-1})$ in general, but we can write H_0 vs. H_A equivalently as

$$\mathrm{H}_0\text{: } \mathrm{E}\{Var(Y_{1t}|I_{t-1})|I_{1t-1}\} = Var(Y_{1t}|I_{t-1})$$

vs.

$$\mathrm{H}_A\text{: } \mathrm{E}\{Var(Y_{1t}|I_{t-1})|I_{1t-1}\} \neq Var(Y_{1t}|I_{t-1}).$$

We say that Y_{2t} does not Granger-cause Y_{1t} in variance with respect to I_{t-1} if H_0 holds, and Y_{2t} Granger-causes Y_{1t} in variance with respect to I_{t-1} if H_A holds. Feedback in variance occurs if Y_{1t} Granger-causes Y_{2t} in variance with

[1] In Section 7 below, we discuss a plausible approach to testing the general causality (1).

[2] Although we focus on testing causality in variance, our approach also immediately delivers tests for causality in mean in the presence of GARCH effects, where the unconditional variances of the time series may not exist. For further details, see the discussion at the end of Section 4.

respect to I_{t-1} and Y_{2t} Granger-causes Y_{1t} in variance with respect to I_{t-1}. There exists simultaneous causality in variance if

$$\mathrm{E}\{(Y_{1t}-\mu_{1t}^0)^2|I_{t-1}\} \neq \mathrm{E}\{(Y_{1t}-\mu_{1t}^0)^2|I_{1t-1},I_{2t}\}. \tag{5}$$

Note that causality in mean, if any, has been filtered out in defining H_0. This ensures that existence of causality in mean will not affect causality in variance.

Clearly, no causality in mean and variance does not necessarily imply no general causality, but if causation is found in mean or variance, then the general causation (1) has been found. From the econometric perspective, detection of causality in variance is particularly important when the test for causality in mean fails to reject the null hypothesis, because it is possible that the general causality (1) exists but there is no causality in mean. In finance and macroeconomics, causality in variance has its own interest, as it is directly related to volatility spillover across different assets or markets.

3. Test statistics and procedures

We now propose a test for H_0. Consider the disturbance processes

$$\varepsilon_{it} = Y_{it}-\mu_{it}^0,\ i=1,2, \tag{6}$$

where $\mu_{it}^0=\mathrm{E}(Y_{it}|I_{t-1})$. To test H_0, we specify the following processes

$$\varepsilon_{it}=\xi_{it}(h_{it}^0)^{1/2}, \tag{7}$$

where h_{it}^0 is a positive time-varying measurable function with respect to I_{it-1}, and $\{\xi_{it}\}$ is an innovation process with

$$\mathrm{E}(\xi_{it}|I_{it-1})=0 \text{ a.s.}, \qquad \mathrm{E}(\xi_{it}^2|I_{it-1})=1 \text{ a.s.} \tag{8}$$

By construction, $\mathrm{E}(\varepsilon_{it}|I_{it-1})=0$ a.s. and $\mathrm{E}(\varepsilon_{it}^2|I_{it-1})=h_{it}^0$ is the univariate conditional variance of ε_{it}^2. Moreover, because $\mathrm{E}(\varepsilon_{it}|I_{t-1})=0$ a.s., it follows that

$$\mathrm{E}(\xi_{it}|I_{t-1})=0 \text{ a.s.} \tag{9}$$

This implies that neither ξ_{2t} Granger-causes ξ_{1t} in mean with respect to I_{t-1} nor ξ_{1t} Granger-causes ξ_{2t} in mean with respect to I_{t-1}.

Now, hypotheses H_0 vs. H_A can be equivalently written as

$$\mathrm{H}_0\text{: } Var(\xi_{1t}|I_{1t-1})=Var(\xi_{1t}|I_{t-1}) \tag{10}$$

vs.

$$\mathrm{H}_A\text{: } Var(\xi_{1t}|I_{1t-1})\neq Var(\xi_{1t}|I_{t-1}). \tag{11}$$

Thus, we can test H_0 by checking if ξ_{2t} Granger-causes ξ_{1t} in variance with respect to I_{t-1}.

The squared innovations $\{\xi_{it}^2\}$ are unobservable, but they can be estimated consistently using squared residuals standardized by their conditional variance estimators, respectively. Throughout, we assume that the conditional mean is parameterized as

$$\mu_{it}^0 = \mu_{it}(b_i^0), \ i = 1, 2, \tag{12}$$

for some finite dimensional parameter vector b_i^0, and the conditional variance h_{it}^0 follows a GARCH(p,q) process (cf. Bollerslev, 1986)

$$h_{it}^0 = \omega_i^0 + \sum_{j=1}^{q} \alpha_{ij}^0 \varepsilon_{it-j}^2 + \sum_{j=1}^{p} \beta_{ij}^0 h_{it-j}^0, \tag{13}$$

where $\omega_i^0 > 0$, and α_{ij}^0 and β_{ij}^0 satisfy appropriate conditions to ensure the strict positivity of h_{it}^0 (cf. Drost and Nijman, 1993; Nelson and Cao, 1992).[3] Given the data $\{Y_t\}_{t=1}^T$, where $Y_t = (Y_{1t}, Y_{2t})'$, let $\hat{\theta}_i = (\hat{b}_i', \hat{\omega}_i, \hat{\alpha}_i', \hat{\beta}_i')'$ be any $\sqrt{T}$-consistent estimator for $\theta_i^0 = (b_i^{0'}, \omega_i^0, \alpha_i^{0'} \beta_i^{0'})'$, where $\alpha_i^0 = (\alpha_{1i}^0, \ldots, \alpha_{qi}^0)'$ and $\beta_i^0 = (\beta_{1i}^0, \ldots, \beta_{pi}^0)'$. For example, we permit (but do not require) $\hat{\theta}_i$ to be a quasi-maximum likelihood estimator (QMLE) of θ_i^0 (e.g., Bollerslev and Wooldridge, 1992; Lee and Hansen, 1994; Lumsdaine, 1996). Then the centered squared standardized residuals can be obtained as

$$\hat{u}_t \equiv u_t(\hat{\theta}_1) = \hat{\varepsilon}_{1t}^2/\hat{h}_{1t} - 1, \quad \hat{v}_t \equiv v_t(\hat{\theta}_2) = \hat{\varepsilon}_{2t}^2/\hat{h}_{2t} - 1, \tag{14}$$

where $\hat{\varepsilon}_{it} \equiv \varepsilon_{it}(\hat{\theta}_i)$, $\hat{h}_{it} \equiv h_{it}(\hat{\theta}_i)$, with

$$\varepsilon_{it}(\theta_i) = Y_{it} - \mu_{it}(b_i), \tag{15}$$

$$h_{it}(\theta_i) = \omega_i + \sum_{j=1}^{q} \alpha_{ij} \varepsilon_{it-j}^2(\theta_i) + \sum_{j=1}^{p} \beta_{ij} h_{it-j}(\theta_i). \tag{16}$$

Here, $\theta_i = (b_i', \omega_i, \alpha_i', \beta_i')'$, the start-up values $h_{it}(\theta_i) \equiv h_{it}^* \leqslant \Delta < \infty$ for $-p + 1 \leqslant t \leqslant 0$ and some constant Δ, and $\varepsilon_{it}(\theta_i) = 0$ for $-q + 1 \leqslant t \leqslant 0$. Lee and Hansen (1994) and Lumsdaine (1996) show that the initial condition is asymptotically negligible for the consistency and asymptotic normality of the QMLE of θ_i^0 for GARCH(1,1) processes. In fact, it also has asymptotically negligible impact on the limiting distribution of our test statistic, as is shown in the appendix.

[3] We use GARCH processes for simplicity and concreteness. Alternatively, other functional forms such as Bera and Higgins (1993) NGARCH, Nelson's (1991) EGARCH, and Sentana's (1995) QARCH Models could be used as well. However, it remains open whether the asymptotic properties of the proposed test statistics would still be valid under these alternative functional forms.

Cheung and Ng (1996) recently proposed a test for H_0 by using the sample cross-correlation function between $\hat{u}_t$ and $\hat{v}_t$, which is defined as

$$\hat{\rho}_{uv}(j) = \{\hat{C}_{uu}(0)\hat{C}_{vv}(0)\}^{-1/2}\hat{C}_{uv}(j), \tag{17}$$

where the sample cross-covariance function

$$\hat{C}_{uv}(j) = \begin{cases} T^{-1}\sum\limits_{t=j+1}^{T} \hat{u}_t\hat{v}_{t-j}, & j \geqslant 0, \\ T^{-1}\sum\limits_{t=-j+1}^{T} \hat{u}_{t+j}\hat{v}_t, & j < 0, \end{cases} \tag{18}$$

and $\hat{C}_{uu}(0) = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2$ and $\hat{C}_{vv}(0) = T^{-1}\sum_{t=1}^{T}\hat{v}_t^2$. Cheung and Ng's statistic is based on the sum of the first M squared cross-correlations

$$S = T\sum_{j=1}^{M}\hat{\rho}_{uv}^2(j), \tag{19}$$

which is asymptotically χ_M^2 under H_0.[4] They also proposed a modified test statistic

$$S^* = T\sum_{j=1}^{M}\omega_j\hat{\rho}_{uv}^2(j), \tag{20}$$

where $\omega_j = T/(T-j)$ or $\omega_j = (T+2)/(T-j)$. The introduction of ω_j gives better match between the moments of S^* and χ_M^2, and therefore is expected to have better sizes in small samples (cf. Ljung and Box, 1978). It does not affect the asymptotic power, however.

The key feature of volatility clustering is that a high volatility "today" tends to be followed by another high volatility "tomorrow", and a low volatility "today" tends to be followed by another low volatility "tomorrow". Recent past volatility often has greater impact on current volatility than distant past volatility. In general, this property carries over to volatility spillover between two assets or markets: the current volatility of an asset or market is more affected by the recent volatility of the other asset or market than by the remote past volatility of that asset or market. Indeed, empirical studies often find that cross-correlations between financial assets or markets generally decay to zero as the lag order j increases. Therefore, when a large M is used, the S test may not be fully efficient, because it gives equal weighting to each of the M sample cross-correlations. The same is true of the S^* test because it has the same asymptotic power as the S test. For large M, a more efficient test may be obtained by giving a larger weight to a lower lag order j. On the other hand, some financial time series may exhibit strong cross-correlation. Such

[4] Cheung and Ng (1996) use the sample mean of $\hat{\varepsilon}_{it}^2/\hat{h}_{it}$ rather than 1 to construct the sample cross-correlations. In our simulation study, we find that there is virtually no difference in both size and power when either definition is used.

processes can have a long distributed lag such that the cross-correlation at each lag is small but their joint effect is substantial. Tests based on a small number of sample cross-correlations (i.e., a small M) may fail to detect such alternatives. In such situations, it is desirable to let M grow with T or to include all $T-1$ sample cross-correlations $\hat{\rho}_{uv}(j)$.

Motivated by these considerations, we suggest a class of new tests based on a generalized version of Cheung and Ng (1996) statistic, namely,

$$T\sum_{j=1}^{T-1} k^2(j/M)\hat{\rho}_{uv}^2(j), \tag{21}$$

where $k(\cdot)$ is a weighting function and M is a positive integer. Examples of $k(\cdot)$ include the truncated, Bartlett, Daniell, Parzen, quadratic-spectral (QS) and Tukey–Hanning kernels:

Truncated:

$$k(z)=\begin{cases}1, & |z|\leqslant 1,\\ 0, & \text{otherwise,}\end{cases}$$

Bartlett:

$$k(z)=\begin{cases}1-|z|, & |z|\leqslant 1,\\ 0, & \text{otherwise,}\end{cases}$$

Daniell:

$$k(z)=\sin(\pi z)/\pi z,\quad -\infty<z<\infty,$$

Parzen:

$$k(z)=\begin{cases}1-6z^2+6|z|^3, & |z|\leqslant 0.5,\\ 2(1-|z|)^3, & 0.5<|z|\leqslant 1,\\ 0, & \text{otherwise,}\end{cases}$$

QS:

$$k(z)=\frac{3}{\sqrt{5}(\pi z)^2}\{\sin(\pi z)/\pi z-\cos(\pi z)\},\quad -\infty<z<\infty,$$

Tukey–Hanning:

$$k(z)=\begin{cases}\frac{1}{2}(1+\cos(\pi z)), & |z|\leqslant 1,\\ 0, & \text{otherwise.}\end{cases}$$

See, e.g., Priestley (1981) for details. Here, the truncated, Bartlett, Parzen and Tukey–Hanning kernels have compact support, i.e., $k(z)=0$ for $|z|>1$. For these functions, M is the "lag truncation number", because the lags of order $j>M$ receive zero weight. In contrast, the Daniell and QS kernels have unbounded support. For these kernels, all $T-1$ sample cross-correlations are used, and M is no longer a lag truncation number. Except for the truncated kernel, all of these kernels have the typical shape of giving a larger weight to a lower lag order j. In contrast, the truncated kernel gives equal weighting to each of the M sample cross-correlations. Cheung and Ng's (1996) S test can thus be viewed as a test based on the truncated kernel. We note that the weight $k^2(j/M)$ is always positive, and differs fundamentally from the weight ω_j for the S^* test. The asymptotic power of our test depends on $k(\cdot)$, while ω_j does not affect the asymptotic power of the S^* test. [5]

Our test statistic is an appropriately standardized version of (21), namely,

$$Q_1 = \left\{ T \sum_{j=1}^{T-1} k^2(j/M)\hat{\rho}_{uv}^2(j) - C_{1T}(k) \right\} \Big/ \{2D_{1T}(k)\}^{1/2}, \tag{22}$$

where

$$C_{1T}(k) = \sum_{j=1}^{T-1} (1-j/T)k^2(j/M),$$

$$D_{1T}(k) = \sum_{j=1}^{T-1} (1-j/T)\{1-(j+1)/T\}k^4(j/M).$$

Both $C_{1T}(k)$ and $D_{1T}(k)$ are approximately the mean and variance of (21). The factors $(1-j/T)$ and $(1-j/T)\{1-(j+1)/T\}$ are finite sample corrections. They are asymptotically negligible, but they give better matches to the mean and variance of (21), respectively.

Alternatively, we can also consider a modified version

$$Q_1^* = \left\{ T \sum_{j=1}^{T-1} (1-j/T)^{-1} k^2(j/M)\hat{\rho}_{uv}^2(j) - C_{1T}^*(k) \right\} \Big/ \{2D_{1T}^*(k)\}^{1/2}, \tag{23}$$

where

$$C_{1T}^*(k) = \sum_{j=1}^{T-1} k^2(j/M),$$

$$D_{1T}^*(k) = \sum_{j=1}^{T-1} \{1-(T-j)^{-1}\}k^4(j/M).$$

It can be shown that Q_1 and Q_1^* are asymptotically equivalent. The weight $(1-j/T)^{-1}$ is the same as that for S^*. This gives better match between the

[5] In addition to the introduction of weighting, our framework also differs from that of Cheung and Ng (1996) in some other aspects. For example, Cheung and Ng consider univariate conditional mean specifications for Y_{1t} and Y_{2t}, respectively. In contrast, we consider a jointly bivariate conditional mean specification for (Y_{1t}, Y_{2t}). This ensures that causality in mean, if any, is filtered out so that it will have no impact on causality in variance. Moreover, we explicitly provide regularity conditions under which our results hold (see Section 4 for details).

moments of $\{T/(1-j/T)\}\hat{\rho}_{uv}^2(j)$ and χ_1^2 in finite samples. While S^* is always larger than S, Q_1^* need not be larger than Q_1, due to the standardization by $C_{1T}(k)$ and $D_{1T}(k)$. Our simulation shows that Q_1 and Q_1^* perform almost the same, while S^* has better sizes than S in small samples. This suggests that the correction by factor $(1-j/T)^{-1}$ is not necessary for Q_1 when it has been properly standardized by $C_{1T}(k)$ and $D_{1T}(k)$.

The constants $C_{1T}(k)$, $D_{1T}(k)$, $C_{1T}^*(k)$ and $D_{1T}^*(k)$ are readily computable given M and $k(\cdot)$. As $M \to \infty$, we have $M^{-1}C_{1T}(k) \to \int_0^\infty k^2(z)\,\mathrm{d}z$ and $M^{-1}D_{1T}(k) \to \int_0^\infty k^4(z)\,\mathrm{d}z$. Consequently, $C_{1T}(k)$ $(C_{1T}^*(k))$ and $D_{1T}(k)$ $(D_{1T}^*(k))$ can be replaced by $M\int_0^\infty k^2(z)\,\mathrm{d}z$ and $M\int_0^\infty k^4(z)\,\mathrm{d}z$, respectively when M is large.

Under appropriate regularity conditions (see Section 4), it can be shown that under H_0,

$$Q_1 \to \mathrm{N}(0,1) \text{ in distribution.}$$

On the other hand, Q_1 diverges to positive infinity in probability as $T \to \infty$ under a general class of alternatives. This implies that asymptotically, negative values of Q_1 occur only under H_0. Therefore, Q_1 is a one-sided test; upper-tailed $\mathrm{N}(0,1)$ critical values should be used. For example, the asymptotic critical value at the 5% level is 1.645.

When the truncated kernel is used, our approach delivers a test statistic

$$Q_{1\mathrm{TRUN}} = (S-M)/(2M)^{1/2}. \tag{24}$$

This is a standardized version of Cheung and Ng (1996) S test. Intuitively, S is a χ_M^2 test. When the degree of freedom M is large, we can transform the S test into a $\mathrm{N}(0,1)$ test by subtracting the mean M and dividing by standard deviation $(2M)^{1/2}$. Similarly, although Cheung and Ng's S^* test uses a weighting function ω_j to improve the size of the S test in finite samples, the weight ω_j does not affect its asymptotic power. Consequently, $Q_{1\mathrm{TRUN}}$ is also asymptotically equivalent to a standardized version of S^*, namely,

$$Q_{1\mathrm{TRUN}} - (S^*-M)/(2M)^{1/2} \to^p 0. \tag{25}$$

Indeed, our simulation below shows that $Q_{1\mathrm{TRUN}}$, S and S^* have similar power.

In fact, the truncated kernel-based test is also asymptotically equivalent to a Granger (1969)-type regression-based test for H_0. To see this, consider the regression model

$$\hat{u}_t = \sum_{j=1}^{M}\gamma_j\hat{v}_{t-j} + w_t. \tag{26}$$

Intuitively, under H_0, the $\hat{v}_{t-j}$ should have no significant explanatory power for $\hat{u}_t$, so the coefficients γ_j for $1 \leqslant j \leqslant M$ should be jointly equal to 0.

If at least one coefficient is significantly different from zero, then there is evidence that $\hat{v}_t$ Granger-causes $\hat{u}_t$, with respect to I_{t-1}. Thus, one can test H_0 by testing whether the coefficients γ_j are jointly equal to zero. Granger (1980) suggests a test for causality in mean based on a regression similar to (26), with a fixed but arbitrarily large M (see also Pierce and Haugh, 1977). To ensure that the test has power against a large class of alternatives, we can let M grow with the sample size T properly. This delivers a R^2-based test statistic

$$Q_{1\mathrm{REG}} = (TR^2 - M)/(2M)^{1/2}, \tag{27}$$

where R^2 is the centered squared multi-correlation coefficient from the regression (26). This is essentially a generalized version of the Granger (1969)-type test for H_0. It can be shown that $Q_{1\mathrm{REG}}$ is asymptotically equivalent to $Q_{1\mathrm{TRUN}}$ under proper conditions. In other words, the Granger-type causality test is asymptotically equivalent to a uniform weighting-based test. Therefore, when a large M is used, we expect that $Q_{1\mathrm{REG}}$ will be less powerful than non-uniform weighting-based tests against the alternatives with decaying cross-correlations.

We now summarize our test procedures:

(1) Estimate univariate GARCH(p,q) models for $\{\hat{\varepsilon}_{1t}\}$ and $\{\hat{\varepsilon}_{2t}\}$ respectively, by the QMLE method, and save the conditional variance estimators $\{\hat{h}_{1t},\ \hat{h}_{2t}\}$.
(2) Compute the sample cross-correlation function $\hat{\rho}_{uv}(j)$ between the centered squared standardized residuals $\{\hat{u}_t = \hat{\varepsilon}^2_{1t}/\hat{h}_{1t} - 1\}$ and $\{\hat{v}_t = \hat{\varepsilon}^2_{2t}/\hat{h}_{2t} - 1\}$.
(3) Choose a weighting function $k(\cdot)$ and an integer M, and compute $C_{1T}(k)$ and $D_{1T}(k)$. For the choice of $k(\cdot)$ and M, see discussion below.
(4) Compute the test statistic Q_1 and compare it to the upper-tailed critical value of $N(0,1)$ at an appropriate level. If Q_1 is larger than the critical value, then the null hypothesis H_0 is rejected. Otherwise, H_0 is not rejected.

In our simulation below, we study the sensitivity of the size and power to the choice of $k(\cdot)$ and M. It is found that some commonly used non-uniform kernels behave rather similarly in terms of size and power, and they have better power than uniform weighting in most cases. Therefore, the choice of $k(\cdot)$, as long as it is non-uniform (e.g., the Daniell kernel), has little impact on the size and power. The choice of M also has little impact on the size, but it has some impact on power (although not substantially). In practice, one may try several different M or use some simple "rule-of-thumb". Because non-uniform weighting discounts higher order lags, we expect that the use of non-uniform weighting will alleviate the loss of power due to choosing too large a M. This is confirmed in the simulation study below. Ideally, M should be chosen to maximize the power, which clearly depends on the data generating process. One may use some data-driven methods to choose M. This is beyond the scope of this paper and has to be left to other work.[6]

It should be emphasized that our formal results in Section 4 are proved only under a simple regression model with GARCH(1,1) errors. These results might still be valid under a more general conditional mean model with higher order GARCH errors, but this conjecture remains to be verified. As asymptotic analysis is extremely complicated, we use the simulation method to study this conjecture in Section 5. We also emphasize that throughout the paper, we assume correct specification of the underlying conditional volatility processes, which is crucial for Q_1 and all the other tests based on $\hat{\rho}_{uv}(j)$.

4. Asymptotic theory

We now provide regularity conditions to support the heuristics given in Section 3. Because GARCH processes are nonlinear functions of the underlying innovations and they may have infinite unconditional variances, asymptotic analysis involved is rather demanding. Following Lee and Hansen (1994) and Lumsdaine (1996), we consider only a simple data generating process with GARCH(1,1) errors:

$$Y_{it} = b_i^0 + \varepsilon_{it}, \quad i = 1,2, \ \ t = 1,\dots,T, \tag{28}$$

$$\varepsilon_{it} = \xi_{it}(h_{it}^0)^{1/2}, \tag{29}$$

$$h_{it}^0 = \omega_i^0 + \alpha_i^0 \varepsilon_{it-1}^2 + \beta_i^0 h_{it-1}^0. \tag{30}$$

The asymptotic properties of the QMLE for this model have been studied by Lee and Hansen (1994) and Lumsdaine (1996).

We first provide some regularity conditions under the model described by (28)–(30).

Assumption A.1. For $i = 1,2$, $\{\xi_{it}\}$ is i.i.d. with $E(\xi_{it}) = 0$, $E(\xi_{it}^2) = 1$ and $E(\xi_{it}^8) < \infty$.

Assumption A.2. $E\{\ln(\alpha_i^0 + \beta_i^0 \xi_{it}^2)\} < 0$, $i = 1,2$.

[6] From the viewpoint of frequency domain, the statistic (21) can be viewed as the quadratic form $2\pi \int_{-\pi}^{\pi} |\hat{H}_{uv}(\omega)|^2 \, d\omega$, where $\hat{H}_{uv}(\omega) = (2\pi)^{-1} \sum_{j=1}^{T-1} k(j/M)\hat{\rho}_{uv}(j)e^{-ij\omega}$ is a kernel estimator for the "one-sided" coherency $H_{uv}(\omega) \equiv (2\pi)^{-1} \sum_{j=1}^{\infty} \rho_{uv}(j)e^{-ij\omega}$, with $\rho_{uv}(j)$ being the cross-correlation function between $u_t^0 = \xi_{1t}^2 - 1$ and $v_t^0 = \xi_{2t}^2 - 1$. One may choose M via some data-driven methods that deliver a reasonable coherency estimator. Such a choice, of course, does not necessarily maximize the power of Q_1.

Assumption A.3. $\sqrt{T}(\hat{\theta}_i - \theta_i^0) = O_P(1)$, where $\hat{\theta}_i = (\hat{b}_i, \hat{\omega}_i, \hat{\alpha}_i, \hat{\beta}_i)'$, $\theta_i^0 = (b_i^0, \omega_i^0, \alpha_i^0, \beta_i^0)'$, and $0 < \omega_i^0 < \infty$, $0 \leqslant \alpha_i^0 < \infty$, $0 \leqslant \beta_i^0 < 1$.

Assumption A.4. $k: \mathbb{R} \to [-1, 1]$ is symmetric about 0, and is continuous at 0 and at all points except for a finite number of points, with $k(0) = 1$ and $\int_0^\infty k^2(z)\,\mathrm{d}z < \infty$.

Assumption A.5. $M \equiv M(T)$ is such that $M^{-1} + M/T \to 0$ as $T \to \infty$.

Like Cheung and Ng (1996) test, we do not assume any specific distribution for ξ_{1t} and ξ_{2t}. Assumption A.1 includes N(0, 1), the generalized error distribution (e.g., Nelson, 1991) and the t-distribution with degrees of freedom larger than 8. The i.i.d. assumption on $\{\xi_{it}\}$ corresponds to a "Strong GARCH" process (cf. Drost and Nijman, 1993). Estimation and inference of GARCH models is frequently done in practice under this assumption. In the present context, the i.i.d. assumption on $\{\xi_{it}\}$ ensures the condition (8) that $E(\xi_{it}|I_{it-1})=0$ a.s. and $E(\xi_{it}^2|I_{it-1})=1$ a.s., and simplifies much the asymptotic analysis. It seems possible to relax this condition so that $\{\xi_{it}\}$ is a martingale difference sequence, but we do not pursue this possibility here.

The GARCH(1,1) model has been the workhorse in the literature, with the largest number of applications. It is found that the GARCH(1,1) model is quite robust and does most of the work for financial time series. Assumption A.2, introduced first by Nelson (1990), ensures that the GARCH(1,1) process is strictly stationary and ergodic. As pointed out by Nelson (1990), $\alpha_i^0 + \beta_i^0 \leqslant 1$ implies Assumption A.2 but not vice versa. Thus, Assumption A.2 permits IGARCH (1,1) processes (i.e., $\alpha_i^0 + \beta_i^0 = 1$; cf. Engle and Bollerslev, 1986) and mildly explosive GARCH (1,1) processes (i.e., $\alpha_i^0 + \beta_i^0 > 1$). These processes are not covariance-stationary, since they have infinite unconditional variances. In Assumption A.3, we allow for any $\sqrt{T}$-consistent estimator $\hat{\theta}_i$ for θ_i^0. In particular, we permit (but do not require) $\hat{\theta}_i$ to be a QMLE. Lee and Hansen (1994) and Lumsdaine (1996) show that under appropriate conditions, a locally QMLE for θ_i^0 exists, and is consistent and $\sqrt{T}$-asymptotically normal, thus satisfying Assumption A.3.[7] Under Assumption A.3, the sampling effects of $\hat{\theta}_i$ have negligible impact on the limiting distribution of Q_1, as is shown in the appendix. Consequently, we can form our test statistic as if the true parameter θ_i^0 were known and were equal to its estimate $\hat{\theta}_i^0$.

Assumption A.4 is a standard condition on the kernel $k(\cdot)$. The examples given in Section 3 all satisfy this assumption. Assumption A.5 is rather weak,

[7] In fact, if $\hat{\theta}_i$ is a QMLE, Assumption A.3 becomes redundant because it will be implied by Assumptions A.1 and A.2 under the model described by (28)–(30). Cf. Lee and Hansen (1994).

requiring only that M grow to infinity as T increases, but at a slower rate than T.

We first derive the null asymptotic distribution of Q_1.

Theorem 1. Suppose Assumptions A.1–A.5 *hold under the model described by* (28)–(30)*. If* $\{\xi_{1t}\}$ *and* $\{\xi_{2t}\}$ *are mutually independent, then* $Q_1 \to \mathrm{N}(0,1)$ *in distribution.*

There exists a gap between $\mathrm{H}_0: Var(\xi_{1t}|I_{1t-1}) = Var(\xi_{1t}|I_{t-1})$ and independence between $\{\xi_{1t}\}$ and $\{\xi_{2t}\}$. The latter implies the former but the converse is not true. There exist processes for which $\{\xi_{1t}\}$ and $\{\xi_{2t}\}$ are not mutually independent but H_0 holds. The assumed independence between $\{\xi_{1t}\}$ and $\{\xi_{2t}\}$ in Theorem 1 renders the asymptotic analysis much simpler. It is possible to relax this assumption and impose conditions closer to H_0, but this would inevitably complicate the asymptotic analysis. [8]

Next, we consider the asymptotic behavior of Q_1 under a general alternative. Put

$$u_t^0 = \xi_{1t}^2 - 1 \quad \text{and} \quad v_t^0 = \xi_{2t}^2 - 1.$$

We let $\rho_{uv}(j)$ denote the cross-correlation function between $\{u_t^0\}$ and $\{v_t^0\}$; namely,

$$\rho_{uv}(j) = Cov(u_t^0, v_{t-j}^0), \qquad j = 0, \pm 1, \pm 2, \dots .$$

Also, let $\kappa_{uvuv}(i,j,l)$ be the fourth order cumulant of the time series $\{u_t^0, v_{t-i}^0, u_{t-j}^0, v_{t-l}^0\}$; namely,

$$\kappa_{uvuv}(i,j,l) = \mathrm{E}(u_t^0 v_{t-i}^0 u_{t-j}^0 v_{t-l}^0) - \mathrm{E}(\bar{u}_t \bar{v}_{t-i} \bar{u}_{t-j} \bar{v}_{t-l}),$$

where $\{\bar{u}_t, \bar{v}_t\}$ is a bivariate zero-mean Gaussian process with the same variance and covariance structure as $\{u_t^0, v_t^0\}$.

Theorem 2. Suppose Assumptions A.1–A.5 *hold under the model described by* (28)–(30)*, and* $\{\xi_{1t}, \xi_{2t}\}$ *is an eighth order stationary bivariate process with* $\sum_{j=1}^{\infty} \rho_{uv}^2(j) < \infty$ *and* $\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\kappa_{uvuv}(i,j,l)| < \infty$*. Then*

$$\frac{M^{1/2}}{T} Q_1 \to \sum_{j=1}^{\infty} \rho_{uv}^2(j) \Bigg/ \left[2 \int_0^{\infty} k^4(z)\,\mathrm{d}z \right]^{1/2} \quad \text{in probability.}$$

[8] We conjecture that $Q_1 \to \mathrm{N}(0,1)$ in distribution under Assumptions A.1–A.5, and (a) $\mathrm{E}(\xi_{1t}^2|I_{t-1}) = 1$ a.s. and (b) $\mathrm{E}(\xi_{1t}^4|I_{t-1}) = \mathrm{E}(\xi_{1t}^4)$ a.s., where (a) is much closer to H_0 than independence between $\{\xi_{1t}\}$ and $\{\xi_{2t}\}$, and (b) is a conditional homokurtosis condition. If (b) does not hold, then Q_1 has to be modified to be heterokurtosis consistent. The proofs under these general conditions would be much more involved.

The fourth order cumulant condition holds when $\{u_t^0, v_t^0\}$ is a bivariate linear process with absolutely summable coefficients and i.i.d. innovations whose fourth order moments are finite (cf. Hannan, 1970, p. 211). It also holds under a proper mixing condition (cf. Andrews, 1991, Lemma 1). We do not impose more primitive conditions here because the cumulant condition and $\sum_{j=1}^{\infty}\rho_{uv}^2(j)<\infty$ allow for some strongly cross-dependent alternatives whose cross-correlation $\rho_{uv}(j)$ decays to zero so slowly that $\rho_{uv}(j)$ is not be absolutely summable.

Theorem 2 implies that Q_1 has asymptotic unit power whenever $\rho_{uv}(j)\neq 0$ for some $j>0$. In other words, Q_1 is able to detect any linear volatility spillover from Y_{2t} to Y_{1t} with respect to I_{t-1} if the sample size T is sufficiently large. It should be noted, though, that Q_1 has no power against the alternatives with zero cross-correlations between u_t^0 and v_{t-j}^0 for all $j>0$ (i.e., $\rho_{uv}(j)=0$ for all $j>0$). Therefore, Q_1 may have no power against some types of nonlinear volatility spillover from Y_{2t} to Y_{1t} with respect to I_{t-1}.

In addition to Q_1, test statistics for other causality hypotheses can be obtained immediately. For example, when no prior information about the direction of causalities is available, it is more appropriate to test the bidirectional hypothesis that neither Y_{2t} Granger-causes Y_{1t} in variance with respect to (I_{1t}, I_{2t-1}) nor Y_{1t} Granger-causes Y_{2t} in variance with respect to (I_{1t-1}, I_{2t}). For this, an appropriate test statistic is

$$Q_2 = \left\{ T\sum_{j=1-T}^{T-1} k^2(j/M)\hat{\rho}_{uv}^2(j) - C_{2T}(k) \right\} \Big/ \{2D_{2T}(k)\}^{1/2}, \tag{31}$$

where

$$C_{2T}(k) = \sum_{j=1-T}^{T-1}(1-|j|/T)k^2(j/M),$$

$$D_{2T}(k) = \sum_{j=1-T}^{T-1}(1-|j|/T)\{1-(|j|+1)/T\}k^4(j/M).$$

The Q_2 test has a null asymptotic N(0,1) distribution, and has asymptotic unit power whenever $\sum_{j=-\infty}^{\infty}\rho_{uv}^2(j)>0$. Like Q_1, upper-tailed N(0,1) critical values should be used for Q_2.

Our approach can also be extended immediately to test causality in mean, by using the sample cross-correlation function between standardized residuals $\hat{\varepsilon}_{1t}/\hat{h}_{1t}^{1/2}$ and $\hat{\varepsilon}_{2t}/\hat{h}_{2t}^{1/2}$.[9] This substantially extends Hong's (1996) test for independence between two covariance-stationary time series with homoskedastic errors whose fourth moments are finite. Such an extension is useful for

[9] To test causality in mean, the eighth moment condition $E(\xi_{it}^8)<\infty$ in Assumption A.1 can be relaxed to the fourth moment condition $E(\xi_{it}^4)<\infty$.

financial and macroeconomic time series, which often exhibit conditional heteroskedasticity and may have infinite unconditional variances.

5. Monte Carlo evidence

To investigate the finite sample performance of the proposed tests, we consider the following data generating process:

$$Y_{it} = X_{it}'b_i^0 + \varepsilon_{it}, \quad i = 1,2, \ t = 1,\ldots,T,$$
$$\varepsilon_{it} = \xi_{it}(h_{it}^0)^{1/2},$$

where $X_{it} = (1, m_{it})'$, $m_{it} = 0.8m_{it-1} + w_{it}$, $w_{it} \sim$ NID(0,4), $b_i^0 = (1,1)'$, and $\xi_{it} \sim$ NID(0,1). Although our theory in Section 4 only allows for a constant regressor, we include an exogenous time series variable m_{it} to examine its effect in finite samples. We first consider the following conditional variance processes

$$h_{it}^0 = \omega_i^0 + \alpha_i^0\varepsilon_{it-1}^2 + \beta_i^0 h_{it-1}^0 + \delta_{ij}\varepsilon_{jt-d}^2 + \gamma_{ij}h_{jt-d}^0, \ d > 0, i \neq j, \ i,j = 1,2,$$

with six parameter combinations, respectively:

$$\text{NULL1(A)}: \ \begin{cases} (\alpha_1^0, \beta_1^0, \delta_{12}, \gamma_{12}) = (0.2, 0.5, 0, 0), \\ (\alpha_2^0, \beta_2^0, \delta_{21}, \gamma_{21}) = (0.2, 0.5, 0, 0), \end{cases}$$

$$\text{NULL1(B)}: \ \begin{cases} (\alpha_1^0, \beta_1^0, \delta_{12}, \gamma_{12}) = (0.2, 0.8, 0, 0), \\ (\alpha_2^0, \beta_2^0, \delta_{21}, \gamma_{21}) = (0.2, 0.8, 0, 0), \end{cases}$$

$$\text{ALTER1(A)}: \ \begin{cases} (\alpha_1^0, \beta_1^0, \delta_{12}, \gamma_{12}) = (0.2, 0.5, 0.2, 0.5), \\ (\alpha_2^0, \beta_2^0, \delta_{21}, \gamma_{21}) = (0.2, 0.5, 0, 0), \ d = 1, \end{cases}$$

$$\text{ALTER1(B)}: \ \begin{cases} (\alpha_1^0, \beta_1^0, \delta_{12}, \gamma_{12}) = (0.2, 0.5, 0.2, 0.5), \\ (\alpha_2^0, \beta_2^0, \delta_{21}, \gamma_{21}) = (0.2, 0.5, 0, 0), \ d = 4, \end{cases}$$

$$\text{ALTER2(A)}: \ \begin{cases} (\alpha_1^0, \beta_1^0, \delta_{12}, \gamma_{12}) = (0.2, 0.5, 0.1, 0.19), \\ (\alpha_2^0, \beta_2^0, \delta_{21}, \gamma_{21}) = (0.2, 0.5, 0.1, 0.19), \ d = 1, \end{cases}$$

$$\text{ALTER2(B)}: \ \begin{cases} (\alpha_1^0, \beta_1^0, \delta_{12}, \gamma_{12}) = (0.2, 0.5, 0.1, 0.19), \\ (\alpha_2^0, \beta_2^0, \delta_{21}, \gamma_{21}) = (0.2, 0.5, 0.1, 0.19), \ d = 4. \end{cases}$$

There is no volatility spillover between Y_{1t} and Y_{2t} under NULL1(A,B). Note that h_{it}^0 is an IGARCH(1,1) process under NULL1(B). Under ALTER1(A,B), there exists volatility spillover from Y_{2t} to Y_{1t} with respect to I_{t-1} but not from

Y_{1t} to Y_{2t}, with respect to I_{t-1}. Under ALTER2(A,B), there exists volatility spillover both from Y_{1t} to Y_{2t} with respect to I_{t-1} and from Y_{2t} to Y_{1t} with respect to I_{t-1}. There is a one-period lag in volatility spillover under ALTER1(A) and ALTER2(A), and there is a four-period lag in volatility spillover under ALTER1(B) and ALTER2(B).

To investigate the conjecture that the proposed tests apply to higher order GARCH processes, we also consider the following GARCH(1,4) processes.

$$h_{it}^0 = \omega_i^0 + \sum_{j=1}^{4} \alpha_{ij}^0 \varepsilon_{it-j}^2 + \beta_i^0 h_{it-1}^0, \quad i = 1, 2,$$

with two parameter combinations, respectively:

$$\text{NULL2(A)}: \begin{cases} (\alpha_{11}^0, \alpha_{12}^0, \alpha_{13}^0, \alpha_{14}^0, \beta_1^0) = (0.1, 0.1, 0.1, 0.1, 0.3), \\ (\alpha_{21}^0, \alpha_{22}^0, \alpha_{23}^0, \alpha_{24}^0, \beta_2^0) = (0.1, 0.1, 0.1, 0.1, 0.3), \end{cases}$$

$$\text{NULL2(B)}: \begin{cases} (\alpha_{11}^0, \alpha_{12}^0, \alpha_{13}^0, \alpha_{14}^0, \beta_1^0) = (0.1, 0.1, 0.1, 0.1, 0.6), \\ (\alpha_{21}^0, \alpha_{22}^0, \alpha_{23}^0, \alpha_{24}^0, \beta_2^0) = (0.1, 0.1, 0.1, 0.1, 0.6), \end{cases}$$

where h_{it}^0 is an IGARCH(1,4) process under NULL2(B).

For all the cases, we set $\omega_i^0 = 1$; simulation shows that the test statistics are robust to the choice of ω_i^0. Three sample sizes, $T = 300, 500, 800$, are considered. For each T, we first generate $T + 1000$ observations using the GAUSS random number generator on a personal computer and then discard the first 1000 to reduce the possible effect of the start-up value h_{i0}^*. We set $h_{i0}^* = 1/(1 - \alpha_i^0 - \beta_i^0)$ for NULL1(A) and ALTER1-2, set $h_{i0}^* = 1/(1 - \sum_{j=1}^4 \alpha_{ij}^0 - \beta_i^0)$ for NULL2(A), and set $h_{i0}^* = 1000$ for NULL1(B) and NULL2(B).[10] We estimate an univariate GARCH(1,1) model for h_{it}^0 under NULL1 and ALTER1-2 respectively, and an univariate GARCH(1,4) model for h_{it}^0 under NULL2, using Berndt et al. (1974, BHHH) algorithm. The resulting squared residuals standardized by their conditional variance estimators are then used to construct the tests. We consider test statistics Q_1 in (22) and Q_2 in (31). The Q_1 test is more suitable for testing ALTER1, and the Q_2 test is more suitable for testing ALTER2. We also compute Q_1^* and Q_2^*, the modified versions of Q_1 and Q_2. Throughout, we conduct 1000 iterations for each experiment.

To examine the effect of the choice of $k(\cdot)$ on the size and power, we use four kernels: the Bartlett, Daniell, QS and truncated kernels (see Section 3 for their expressions). The first three kernels are non-uniform. To examine the effect of the choice of M, we consider $M = 10$, 20 and 30, for each T.

[10] More ideally, h_{i0}^* can be drawn from the underlying strictly stationary distribution of the conditional variance process. Nevertheless, our analysis shows that the choice of start-up values has negligible impact on the limit distribution of the test statistics, as is confirmed in the simulation study.

Table 1
Size at the 10% and 5% levels under GARCH(1,1) processes[a]

		NULL1(A)						NULL1(B)					
	M	10		20		30		10		20		30	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Q_1	BAR	10.3	6.9	10.7	6.4	11.4	6.8	10.3	7.2	10.1	6.4	11.3	6.3
	DAN	11.2	7.4	11.3	6.5	11.6	6.8	10.8	7.4	10.8	6.7	10.7	6.9
	QS	11.3	7.3	11.3	6.6	11.6	6.9	10.7	7.4	10.8	6.5	10.9	6.8
Q_1^*	BAR	10.4	6.9	10.7	6.4	11.4	6.9	10.3	7.1	10.0	6.4	11.1	6.2
	DAN	11.1	7.4	11.3	6.4	11.5	6.6	10.8	7.4	10.8	6.7	10.7	7.0
	QS	11.3	7.2	11.4	6.7	11.5	7.0	10.7	7.4	10.8	6.5	10.9	6.9
Q_1	TRUN	10.5	6.1	10.8	6.6	9.4	5.5	10.4	6.2	10.1	6.7	10.1	6.4
S_1		9.7	4.6	9.9	5.5	9.1	4.5	9.7	4.8	9.6	5.4	9.2	5.3
S_1^*		10.2	5.1	11.2	6.1	11.0	6.0	9.9	5.2	10.7	6.1	11.5	6.7
Q_2	BAR	10.5	7.2	11.2	7.3	11.3	7.4	10.8	7.3	11.7	7.7	11.6	7.2
	DAN	10.7	7.1	10.5	7.0	12.3	6.4	11.1	7.6	11.5	7.4	11.4	6.6
	QS	11.0	6.9	10.7	7.1	12.7	6.5	11.2	7.7	11.5	7.3	11.7	6.8
Q_2^*	BAR	10.4	7.1	11.2	7.2	11.3	7.2	10.8	7.2	11.4	6.9	10.3	6.9
	DAN	10.7	7.1	10.6	7.0	12.5	6.5	10.8	7.4	10.2	6.3	9.1	6.0
	QS	10.9	6.9	10.5	7.1	12.9	6.5	10.7	7.6	10.8	6.8	10.0	6.4
Q_2	TRUN	11.2	6.7	10.6	5.7	8.0	4.7	10.9	6.6	10.4	6.2	8.4	4.7
S_2		10.5	5.2	10.0	4.7	7.4	4.3	9.8	5.3	10.0	5.5	8.3	4.2
S_2^*		11.3	6.0	12.0	5.7	10.5	5.8	10.9	5.8	11.6	6.3	10.2	5.2

[a]NULL1(A): $Y_i = 1 + X_{it} + \varepsilon_{it}, \varepsilon_{it} = \xi_{it} h_{it}^{1/2},\ h_{it} = 1 + 0.2\varepsilon_{it-1}^2 + 0.5h_{it-1}$; NULL1(B): $Y_i = 1 + X_{it} + \varepsilon_{it}, \varepsilon_{it} = \xi_{it} h_{it}^{1/2},\ h_{it} = 1 + 0.2\varepsilon_{it-1}^2 + 0.8h_{it-1}$; The sample size $T = 500$, 1000 iterations; BAR, DAN, QS, TRUN = Bartlett, Daniell, Quadratic-spectral, truncated kernels.

We compare our tests with corresponding Cheung and Ng (1996) test statistics

$$S_1 = T \sum_{j=1}^{M} \hat{\rho}_{uv}^2(j), \tag{32}$$

$$S_2 = T \sum_{j=-M}^{M} \hat{\rho}_{uv}^2(j). \tag{33}$$

Under H_0, S_1 and S_2 are asymptotically distributed as χ_M^2 and χ_{2M+1}^2, respectively. We also compute S_1^* and S_2^*, the modified versions of S_1 and S_2, respectively.

To save space, we only report the results for $T = 500$ in detail. Tables 1 and 2 report the size under NULL1-2 at the 10% and 5% levels. Overall, Q_1 and Q_2 perform well at the 10% level, but tend to overreject a little at the 5% level. The non-uniform kernels perform similarly, and for each kernel, the choice of M has little impact on the size. For each i, Q_i and Q_i^* perform almost the same, while S_i and S_i^* perform a little differently.

Table 2
Size at the 10% and 5% level under GARCH(1,4) processes[a]

		NULL2(A)						NULL2(B)					
	M	10		20		30		10		20		30	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Q_1	BAR	11.2	6.4	11.3	6.6	11.2	6.6	11.0	7.3	10.9	6.5	11.7	7.4
	DAN	11.6	6.7	11.1	7.2	11.2	6.8	11.1	7.5	11.2	7.2	11.5	7.9
	QS	11.5	6.4	11.2	7.1	11.3	6.9	11.3	7.4	11.4	7.0	11.3	7.9
Q_1^*	BAR	11.2	6.4	11.3	6.6	11.1	6.7	11.0	7.2	10.8	6.6	12.0	7.4
	DAN	11.5	6.7	11.2	7.2	11.2	6.8	11.1	7.5	11.1	7.2	11.6	7.8
	QS	11.6	6.7	11.1	7.2	11.5	6.8	11.3	7.5	11.2	6.9	11.3	7.8
Q_1	TRUN	11.0	6.7	10.3	6.5	8.1	4.8	10.3	6.3	11.3	5.9	9.5	5.4
S_1		10.5	5.2	9.7	4.6	8.1	4.0	9.7	5.4	10.8	4.5	9.2	4.2
S_1^*		10.8	5.5	10.8	5.8	10.4	5.3	10.0	5.5	11.7	5.4	10.8	5.9
Q_2	BAR	10.9	6.7	11.6	6.8	10.4	6.6	11.0	7.5	11.2	7.6	11.3	6.9
	DAN	11.0	7.4	10.6	7.4	11.3	6.8	10.9	7.8	10.9	7.2	11.3	6.4
	QS	10.9	7.3	10.6	7.2	11.6	6.7	11.2	7.9	10.8	7.0	11.5	6.4
Q_2^*	BAR	10.8	6.6	11.6	6.8	10.6	6.7	10.9	7.5	11.2	7.6	11.3	6.8
	DAN	11.0	7.4	10.6	7.3	11.4	6.8	10.8	7.8	10.8	7.1	11.6	6.4
	QS	10.9	7.3	10.7	7.0	11.4	6.7	11.2	7.9	10.7	7.0	11.5	6.4
Q_2	TRUN	11.3	6.1	9.5	4.5	7.2	4.6	10.3	5.9	10.1	4.4	7.7	4.4
S_2		10.6	5.2	8.5	4.0	7.0	3.7	9.8	5.0	9.6	3.8	7.4	3.9
S_2^*		11.4	5.4	11.4	4.5	10.4	5.3	10.2	5.2	11.0	4.5	9.8	5.3

[a]NULL2(A): $Y_i = 1 + X_{it} + \varepsilon_{it}$, $\varepsilon_{it} = \xi_{it} h_{it}^{1/2}$, $h_{it} = 1 + 0.1\varepsilon_{it-1}^2 + 0.1\varepsilon_{it-2}^2 + 0.1\varepsilon_{it-3}^2 + 0.1\varepsilon_{it-4}^2 + 0.3h_{it-1}$; NULL2(B): $Y_i = 1 + X_{it} + \varepsilon_{it}, \varepsilon_{it} = \xi_{it} h_{it}^{1/2}$, $h_{it} = 1 + 0.1\varepsilon_{it-1}^2 + 0.1\varepsilon_{it-2}^2 + 0.1\varepsilon_{it-3}^2 + 0.1\varepsilon_{it-4}^2 + 0.6h_{it-1}$; The sample size $T = 500$, 1000 iterations; BAR, DAN, QS, TRUN = small Bartlett, Daniell, Quadratic-spectral, truncated kernels.

This suggests that the correction by factor $(1-j/T)^{-1}$ is not necessary for Q_1 when it has been properly standardized by $C_{1T}(k)$ and $D_{1T}(k)$. The S_i^* test does not have better sizes than the S_i test at the 5% level under NULL1, but it does under NULL2. Both S_i and S_i^* have better sizes than Q_i and Q_i^* with non-uniform kernels at the 5% level, but not at the 10% level. The truncated kernel-based test $Q_{i\text{TRUN}}$ performs similarly to S_i. There is little difference between NULL1 and NULL2, suggesting that Q_i may be applicable to higher order GARCH processes.

Table 3 reports the power under ALTER1, the one-way volatility spillover from Y_{1t} to Y_{2t} with respect to I_{t-1}. We use the empirical critical values at the 10% and 5% levels, which, obtained from the 1000 iterations under NULL1(A), give size-adjusted powers so that all the tests are compared on an equal ground. Because Q_i and Q_i^* have almost the same power, we report the power of Q_i only. We first consider ALTER1(A), which has a one-period lag in volatility spillover. Not surprisingly, one-way tests Q_1, S_1 and S_1^* have

Table 3
Size-adjusted power at the 10% and 5% level under ALTER1[a]

		ALTER1(A)						ALERT1(B)					
	M	10		20		30		10		20		30	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Q_1	BAR	73.3	64.1	74.3	65.7	71.2	63.6	42.2	27.4	61.7	50.3	62.7	54.4
	DAN	73.5	64.8	72.8	64.9	68.5	59.6	46.9	33.2	65.4	55.3	62.6	54.0
	QS	73.4	64.7	73.4	65.9	68.6	59.1	48.2	33.9	65.5	55.4	62.4	53.9
	TRUN	70.0	59.0	56.3	41.8	48.7	35.5	65.8	55.9	54.8	43.2	47.4	35.1
S_1		70.0	59.0	56.3	41.8	48.7	35.5	65.8	55.9	54.8	43.2	47.4	35.1
S_1^*		69.9	58.7	55.3	41.7	47.6	34.9	65.9	56.0	54.8	43.4	46.9	34.3
Q_2	BAR	61.2	44.7	62.8	46.9	60.2	46.1	27.7	14.7	47.7	30.0	49.4	35.8
	DAN	63.4	47.1	62.5	48.6	56.1	45.0	33.9	17.9	52.6	37.9	50.0	39.9
	QS	64.0	47.0	62.7	48.3	56.2	44.5	34.6	18.7	52.9	38.0	50.0	39.6
	TRUN	55.8	43.7	44.2	32.5	38.3	27.4	53.0	42.3	42.4	32.8	38.7	26.6
S_2		55.8	43.7	44.2	32.5	38.3	27.4	53.0	42.3	42.4	32.8	38.7	26.6
S_2^*		55.6	43.4	43.4	31.8	37.6	26.6	52.9	42.5	41.7	31.9	37.7	25.9

[a]ALTER1(A): $Y_i = 1 + X_{it} + \varepsilon_{it}$, $\varepsilon_{it} = \xi_{it} h_{it}^{1/2}$, $h_{1t} = 1 + 0.2\varepsilon_{1t-1}^2 + 0.5h_{1t-1} + 0.2\varepsilon_{2t-1}^2 + 0.5h_{2t-1}$, $h_{2t} = 1 + 0.2\varepsilon_{2t-1}^2 + 0.5h_{2t-1}$; ALTER1(B): $Y_i = 1 + X_{it} + \varepsilon_{it}$, $\varepsilon_{it} = \xi_{it} h_{it}^{1/2}$, $h_{1t} = 1 + 0.2\varepsilon_{1t-1}^2 + 0.5h_{1t-1} + 0.2\varepsilon_{2t-4}^2 + 0.5h_{2t-4}$, $h_{2t} = 1 + 0.2\varepsilon_{2t-1}^2 + 0.5h_{2t-1}$; The sample size $T =$ 500, 1000 iterations; BAR, DAN, QS, TRUN = Bartlett, Daniell, Quadratic-spectral, truncated kernels.

better power than the two-way tests Q_2, S_2 and S_2^*, respectively. The three non-uniform kernels give rather similar power. As expected, the truncated kernel delivers similar power to Cheung and Ng's tests. The non-uniform kernels give better power than the truncated kernel and Cheung and Ng's tests, especially for larger M. For the non-uniform kernels, the three different M's give similar power, but for the truncated kernel and Cheung and Ng's tests, a larger M gives smaller power. This confirms our expectation that the use of non-uniform weighting alleviates the impact of choosing too large a M because non-uniform weighting discounts higher order lags.

Next, we consider ALTER1(B), which has a four-period lag in volatility spillover. As under ALTER1(A), Q_1, S_1 and S_1^* have better power than Q_2, S_2 and S_2^*; the three non-uniform kernels give similar power; and the truncated kernel has similar power to S_1 and S_1^*. A feature different from their performances under ALTER1(A) is that now the non-uniform kernels have smaller power than the truncated kernel and Cheung and Ng's tests when $M = 10$, but they become more powerful when M becomes larger.

Table 4 reports the power at the 10% and 5% levels under ALTER2, the two-way volatility spillover between Y_{1t} and Y_{2t}. Now Q_2, S_2 and S_2^* have better power than Q_1, S_1 and S_1^* respectively. Except for this, all power patterns are similar to those under ALTER1.

Table 4
Size-adjusted power at the 10% and 5% level under ALTER2[a]

		ALTER2(A)						ALTER2(B)					
	M	10		20		30		10		20		30	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Q_1	BAR	76.4	66.2	78.2	69.4	75.7	69.2	64.1	50.2	78.5	70.1	79.6	73.5
	DAN	76.7	67.2	77.1	70.4	72.9	65.8	68.4	56.3	81.1	74.0	80.5	74.3
	QS	76.8	67.4	77.0	70.4	73.0	65.6	68.9	57.6	81.5	74.8	80.1	74.3
	TRUN	73.4	64.4	62.8	50.5	57.4	44.5	81.6	74.7	75.1	65.3	71.8	60.3
S_1		73.4	64.4	62.8	50.5	57.4	44.5	81.6	74.7	75.1	65.3	71.8	60.3
S_1^*		73.4	64.2	62.2	50.2	56.1	43.2	81.6	74.6	74.8	65.4	71.2	59.9
Q_2	BAR	95.6	87.3	96.7	92.3	95.6	91.5	79.0	63.3	93.2	86.6	95.1	89.7
	DAN	96.5	90.4	96.9	92.7	94.8	90.4	84.7	73.0	95.6	90.3	96.1	91.7
	QS	96.6	90.6	96.8	92.9	94.2	90.2	86.0	73.8	95.7	90.3	96.0	91.7
	TRUN	94.0	88.6	86.8	78.2	78.8	67.1	95.0	91.2	92.7	89.1	89.9	84.6
S_2		94.0	88.6	86.8	78.2	78.8	67.1	95.0	91.2	92.7	89.1	89.9	84.6
S_2^*		93.9	88.5	85.3	76.9	77.6	65.5	95.0	91.4	92.6	88.6	89.4	83.9

[a]ALTER2(A): $Y_i=1+X_{it}+\varepsilon_{it}$, $\varepsilon_{it}=\xi_{it}h_{it}^{1/2}$, $h_{it}=1+0.2\varepsilon_{it-1}^2+0.5h_{it-1}+0.1\varepsilon_{jt-1}^2+0.19h_{jt-2}$; ALTER2(B): $Y_i=1+X_{it}+\varepsilon_{it}$, $\varepsilon_{it}=\xi_{it}h_{it}^{1/2}$, $h_{it}=1+0.2\varepsilon_{it-1}^2+0.5h_{it-1}+0.1\varepsilon_{jt-4}^2+0.19h_{jt-4}$; The sample size T=500, 1000 iterations; BAR, DAN, QS, TRUN = Bartlett, Daniell, Quadratic-spectral, truncated kernels.

We now briefly report the results for $T = 300, 800$. The size patterns for $T = 300, 800$ are similar to those for $T = 500$. In particular, all the tests have reasonable sizes at the 10% level, but tend to overreject a little at the 5% level. The overrejections become weaker as T increases, but slowly. Both Q_i and Q_i^* continue to have almost the same sizes whether $T = 300$ or 800, but S_i^* has better sizes than S_i when $T = 300$. On the other hand, the power patterns for $T = 300, 800$ are similar to those for $T = 500$.

In summary, the new tests have reasonable sizes at the 10% level, but tend to overreject a little at the 5% level. The choice of non-uniform kernels and the lag truncation number have little impact on the sizes of the new tests. For the alternatives under study, non-uniform weighting often yields better power than uniform weighting, which delivers power similar to that of Cheung and Ng's tests. Moreover, the use of non-uniform weighting renders the power relatively robust to the choice of M.

The reason that the tests Q_1 and Q_2 tend to overreject at the 5% level under H_0 is that they behave as a standardized version of a $k^2(j/M)$-weighted sum of independent centered χ_1^2. Such a standardization converges to $N(0, 1)$ in distribution as $M \to \infty$, and is expected to work reasonably well for large M. For small and moderate M, however, such a standardized version is right-skewed in distribution, and consequently, the $N(0, 1)$ approximation will result in overrejection under H_0. To obtain more accurate finite sample approximation, higher order asymptotic approximation can be considered. This is beyond the scope of this paper and should be left for other work.

6. Application to exchange rates

Exchange rates volatility clustering has been well documented and has been a recurrent topic in the literature (e.g., Baillie and Bollerslev, 1989, 1990; Bekaert, 1995; Bollerslev, 1990; Diebold and Nerlove, 1989; Domowitz and Hakkio, 1985; Engle and Bollerslev, 1986; Engle et al., 1990; Gallant et al., 1989; Hsieh, 1988, 1989; Hodrick, 1989; West and Cho, 1995; Zhou, 1996; see also Bollerslev et al., 1992, for a detailed survey and the references therein). Investigating volatility spillover is important to understand the causal relations between exchange rates and the nature of exchange rate interaction, which are helpful for volatility prediction and forecasting (cf. Baillie and Bollerslev, 1990; Engle et al., 1990). We now apply our tests to investigate volatility spillover between two important nominal U.S. dollar exchange rates—Deutschemark (DM) and Japanese yen (YEN), which are among most active currencies traded in the foreign exchange market. We use the weekly spot rates from the first week of 1976:1 to the last week of 1995:11, with totally 1039 observations. The data are interbank closing spot rates on Wednesdays, obtained from the Bloomberg L.P. The use of Wednesday data avoids the so-called weekend effect. Also, very few holidays occur on Wednesday; for these holidays, the data on the following Thursdays are used. Both exchange rates are measured in units of local currency per dollar.

It has been well documented (e.g., Bollerslev, 1990; Diebold and Nerlove, 1989) that the weekly logarithmic exchange rates are first order homogeneous nonstationary.[11] Visual inspection suggests little serial correlation for $\Delta \ln \mathrm{DM}_t$, although it exhibits persistence in conditional variance. On the other hand, $\Delta \ln \mathrm{YEN}_t$ shows a little serial correlation with non-zero intercept, in addition to obvious volatility clustering. To account for any possible weak serial correlation, we follow Diebold and Nerlove (1989) and specify an AR(3) model with non-zero mean and GARCH(1,1) errors:

$$Y_{it} = b_{i0} + \sum_{j=1}^{3} b_{ij} Y_{it-j} + \varepsilon_{it}, \quad i = 1, 2,$$

$$\varepsilon_{it} = \xi_{it} h_{it}^{1/2},$$

$$h_{it} = \omega_i + \alpha_i \varepsilon_{it-1}^2 + \beta_i h_{it-1},$$

[11] Bollerslev (1990) and Diebold and Nerlove (1989) use the weekly spot rates on Wednesdays, with a different time period, that were obtained from the *International Monetary Markets Yearbook*.

Table 5
Quasi-maximum likelihood estimation of univariate GARCH(1, 1) models for Deutschemark and Japanese Yen[a]

	Deutschemark		Japanese Yen	
Parameter	Estimate		Estimate	
b_0	−0.073	(0.041)	−0.097	(0.042)
b_1	0.049	(0.033)	0.051	(0.034)
b_2	0.067	(0.033)	0.093	(0.034)
b_3	−0.028	(0.033)	0.066	(0.033)
ω	0.051	(0.030)	0.116	(0.068)
α	0.114	(0.027)	0.084	(0.026)
β	0.873	(0.033)	0.863	(0.055)
Sample size	1038		1038	
Log-likelihood	−1862.307		−1813.625	
Box–Pierce test				
BP(5)	1.602	[0.901]	1.924	[0.860]
BP(10)	6.610	[0.762]	7.934	[0.635]
BP(20)	10.680	[0.954]	18.349	[0.564]
$BP^2(5)$	8.407	[0.135]	1.803	[0.876]
$BP^2(10)$	15.761	[0.107]	3.255	[0.975]
$BP^2(20)$	26.228	[0.158]	6.623	[0.998]

[a]The numbers in the parentheses are standard errors for the estimates and the numbers in the square brackets are the p-values for Box–Pierce test statistics; BP(M) and $BP^2(M)$ are Box–Pierce portmanteau statistics for the first M autocrrelations of the standardized residual and squared standardized residuals, respectively.

where $Y_{1t} = 100\Delta \ln \mathrm{DM}_t$ and $Y_{2t} = 100\Delta \ln \mathrm{YEN}_t$.[12] Table 5 summarizes the QMLE results of univariate GARCH models for $\Delta \ln \mathrm{DM}_t$ and $\Delta \ln \mathrm{YEN}_t$. For $\Delta \ln \mathrm{DM}_t$, the second lag $\Delta \ln \mathrm{DM}_{t-2}$ is significant at the 5% level, but the intercept, $\Delta \ln \mathrm{DM}_{t-1}$ and $\Delta \ln \mathrm{DM}_{t-3}$ are insignificant. The GARCH parameter estimates $(\hat{\alpha}_1, \hat{\beta}_1)$ are highly significant, with $\hat{\alpha}_1 + \hat{\beta}_1 \simeq 0.99$, suggesting a nearly integrated GARCH process. Table 5 also reports some diagnostic statistics. The p-values of Box–Pierce portmanteau test statistics for autocorrelation in standardized residuals $\{\hat{\varepsilon}_{1t}/\hat{h}_{1t}^{1/2}\}$ are 0.90, 0.76 and 0.95, for $M = 5, 10$ and 20 respectively, all well above the 10% level. Similarly, the p-values of Box–Pierce tests for squared standardized residuals $\{\hat{\varepsilon}_{1t}^2/\hat{h}_{1t}\}$,

[12] Diagnostic tests, reported in Table 5, suggest that AR(3) models are adequate for both $\Delta \ln \mathrm{DM}_t$ and $\Delta \ln \mathrm{YEN}_t$. We also tried higher order AR models; the higher order terms are insignificant and the causality test statistics based on these higher order AR models remain largely the same.

Table 6
Test statistics for causality in mean between Deutschemark and Japanese Yen[a]

	M	5	10	20	30	40
Q_2	DAN	145.230	102.511	72.491	58.949	50.638
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q_2	TRUN	79.454	57.614	40.021	32.676	28.032
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
S_2		383.674	394.383	403.408	421.921	437.789
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q_1	DAN	−0.975	−1.034	−1.116	−1.385	−1.719
		(0.835)	(0.849)	(0.868)	(0.917)	(0.957)
Q_1	TRUN	−0.875	−0.909	−1.801	−1.685	−1.677
		(0.809)	(0.818)	(0.964)	(0.954)	(0.953)
S_1		2.233	5.935	8.609	16.945	25.003
		(0.816)	(0.821)	(0.987)	(0.973)	(0.969)
Q_{-1}	DAN	−0.547	−0.388	−0.060	−0.088	−0.297
		(0.708)	(0.651)	(0.524)	(0.535)	(0.617)
Q_{-1}	TRUN	−0.417	0.154	−0.468	−0.359	−0.556
		(0.662)	(0.439)	(0.680)	(0.640)	(0.711)
S_{-1}		3.681	10.688	17.039	27.216	35.025
		(0.596)	(0.382)	(0.650)	(0.612)	(0.693)

[a] Q_2 and S_2 are the two-way tests for causality in mean between $\Delta\ln(\mathrm{DM}_t)$ and $\Delta\ln(\mathrm{YEN}_t)$ with respect to I_{t-1}; Q_1 and S_1 are the one-way tests for causality in mean from $\Delta\ln(\mathrm{DM}_t)$ to $\Delta\ln(\mathrm{YEN}_t)$ with respect to I_{t-1}; Q_{-1} and S_{-1} are the one-way tests for causality in mean from $\Delta\ln(\mathrm{YEN}_t)$ to $\Delta\ln(\mathrm{DM}_t)$ with respect to I_{t-1}; DAN and TRUN are the Daniell and truncated kernels; The numbers in the parentheses are the p-values.

for $M = 5, 10$ and 20, are 0.14, 0.11 and 0.16, respectively. Again, these values are larger than the 10% level. This suggests the adequacy of the AR(3)-GARCH(1,1) model for $\Delta\ln\mathrm{DM}_t$. These findings are much in line with the previous studies on exchange rates.

In the AR(3)-GARCH(1,1) model for $\Delta\ln\mathrm{YEN}_t$, the intercept, $\Delta\ln\mathrm{YEN}_{t-2}$ and $\Delta\ln\mathrm{YEN}_{t-3}$ are all significant at the 5% level. The GARCH parameter estimates $(\hat{\alpha}_2, \hat{\beta}_2)$ are highly significant, with $\hat{\alpha}_2+\hat{\beta}_2 \simeq 0.95$, suggesting strong volatility clustering. The p-values of Box–Pierce test statistics for autocorrelation in standardized residuals $\{\hat{\varepsilon}_{2t}/\hat{h}_{2t}^{1/2}\}$ are 0.86, 0.64 and 0.56 for $M=5, 10$ and 20, respectively, all well above the 10% level. Also, the p-values of Box–Pierce tests for squared standardized residuals $\{\hat{\varepsilon}_{2t}^2/\hat{h}_{2t}\}$, for $M = 5, 10$ and 20, are 0.88, 0.98 and 1.00 respectively. These diagnostic results suggest the adequacy of the AR(3)-GARCH(1,1) model for $\Delta\ln\mathrm{YEN}_t$.

We first consider causality in mean between $\Delta\ln\mathrm{DM}_t$ and $\Delta\ln\mathrm{YEN}_t$. Table 6 reports our statistics and Cheung and Ng (1996) statistics, together with their p-values. All the statistics here are based on the sample cross-correlation function between standardized residuals $\{\hat{\varepsilon}_{1t}/\hat{h}_{1t}^{1/2}\}$ and $\{\hat{\varepsilon}_{2t}/\hat{h}_{2t}^{1/2}\}$. Because commonly used non-uniform kernels deliver similar power, we only report the Daniell kernel. The Daniell kernel-based test $Q_{2\mathrm{DAN}}$ for two-way causality in mean yields values of 145.23, 102.51, 72.49, 58.95 and 50.64, for $M = 5, 10, 20, 30$ and 40, respectively. These statistics are significant at any reasonable levels (compare to the upper-tailed $\mathrm{N}(0, 1)$ critical values), suggesting very strong causality in mean. Cheung and Ng's S_2 test is also very significant at any reasonable levels for all M. It is impossible to compare the powers of $Q_{2\mathrm{DAN}}$ and S_2 using p-values, because they are all essentially zero. However, our truncated kernel-based test $Q_{2\mathrm{TRUN}}$, which is a normalized version of S_2, has much smaller values than $Q_{2\mathrm{DAN}}$ for all M.

In order to identify the direction for causality in mean, we compute two one-way causality tests: $Q_{1\mathrm{DAN}}$ is a test for whether $\Delta \ln \mathrm{DM}_t$ Granger-causes $\Delta \ln \mathrm{YEN}_t$ in mean with respect to I_{t-1}, and $Q_{-1\mathrm{DAN}}$ is a test for whether $\Delta \ln \mathrm{YEN}_t$ Granger-causes $\Delta \ln \mathrm{DM}_t$ in mean with respect to I_{t-1}. Both tests, for all $M = 5, 10, 20, 30$ and 40, yields p-values well above the 10% level, suggesting neither $\Delta \ln \mathrm{DM}_t$ Granger-causes $\Delta \ln \mathrm{YEN}_t$ with respect to I_{t-1} nor $\Delta \ln \mathrm{YEN}_{t-1}$ Granger-causes $\Delta \ln \mathrm{DM}_t$ in mean with respect to I_{t-1}. Therefore, the significant power of $Q_{2\mathrm{DAN}}$ should have come solely from the simultaneous causality between $\Delta \ln \mathrm{DM}_t$ and $\Delta \ln \mathrm{YEN}_t$. In other words, for causality in mean, there exists only simultaneous interaction between $\Delta \ln \mathrm{DM}_t$ and $\Delta \ln \mathrm{YEN}_t$. These results are consistent with the findings by Baillie and Bollerslev (1990, Table IV) using robust LM tests and intra-day exchange rates data. To some extent, our results support Bollerslev's (1990) multivariate GARCH(1,1) model with constant conditional cross-correlation in modeling the comovement of exchange rate changes. These findings are hardly surprising, because the movements of exchange rates, to a large extent, depend on a common set of international economic variables, unobservable on a weekly basis. In particular, both exchange rates are bilateral dollar rates; new information coming to the market will affect both dollar rates. We note that Cheung and Ng's S_1 and S_{-1} tests as well as the truncated kernel-based tests $Q_{1\mathrm{TRUN}}$ and $Q_{-1\mathrm{TRUN}}$ deliver the same conclusions as $Q_{1\mathrm{DAN}}$ and $Q_{-1\mathrm{DAN}}$.

Table 7 reports tests for causality in variance. The values of the two-way test $Q_{2\mathrm{DAN}}$, for $M = 5, 10, 20, 30$ and 40, are 65.94, 46.73, 32.57, 26.10 and 22.45, respectively. These values are significant at any reasonable levels, suggesting strong causality in variance between $\Delta \ln \mathrm{DM}_t$ and $\Delta \ln \mathrm{YEN}_t$. Similarly, S_2 is significant at any reasonable level for all M. The truncated kernel-based test $Q_{2\mathrm{TRUN}}$, a normalized version of S_2, is also significant at any reasonable level for all M, but its values are much smaller than those of $Q_{2\mathrm{DAN}}$.

Again, to identify the direction for causality in variance, we compute tests for one-way volatility spillover: Q_1, Q_{-1}, S_1 and S_{-1}, where Q_1 and S_1 test for volatility spillover from past $\Delta \ln \mathrm{DM}_t$ to $\Delta \ln \mathrm{YEN}_t$, and Q_{-1} and S_{-1}

Table 7
Test statistics for causality in variance between Deutschemark and Japanese Yen[a]

	M	5	10	20	30	40
Q_2	DAN	65.941	46.728	32.572	26.100	22.447
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q_2	TRUN	36.401	25.722	17.404	14.806	12.322
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
S_2		181.738	187.670	198.600	224.537	237.839
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q_1	DAN	5.420	3.967	2.336	1.442	1.022
		(0.000)	(0.000)	(0.010)	(0.075)	(0.153)
Q_1	TRUN	3.251	1.653	0.564	0.611	−0.155
		(0.000)	(0.049)	(0.286)	(0.270)	(0.562)
S_1		15.282	17.391	23.566	34.736	38.610
		(0.009)	(0.066)	(0.262)	(0.252)	(0.533)
Q_{-1}	DAN	−0.203	−0.536	−0.797	−0.996	−0.929
		(0.580)	(0.704)	(0.788)	(0.842)	(0.824)
Q_{-1}	TRUN	−0.507	−0.616	−1.269	−0.421	−0.428
		(0.694)	(0.731)	(0.898)	(0.663)	(0.666)
S_{-1}		3.398	7.246	11.976	26.742	36.170
		(0.639)	(0.702)	(0.917)	(0.637)	(0.643)

[a] Q_2 and S_2 are the two-way tests for causality in variance between $\Delta\ln(\mathrm{DM}_t)$ and $\Delta\ln(\mathrm{YEN}_t)$ with respect to I_{t-1}; Q_1 and S_1 are the one-way tests for causality in variance from $\Delta\ln(\mathrm{DM}_t)$ to $\Delta\ln(\mathrm{YEN}_t)$ with respect to I_{t-1}; Q_{-1} and S_{-1} are the one-way tests for causality in variance from $\Delta\ln(\mathrm{YEN}_t)$ to $\Delta\ln(\mathrm{DM}_t)$ with respect to I_{t-1}; DAN and TRUN are the Daniell and truncated kernels; The numbers in the parentheses are the p-values.

test for volatility spillover from past $\Delta\ln\mathrm{YEN}_t$ to $\Delta\ln\mathrm{DM}_t$. The values of $Q_{1\mathrm{DAN}}$ are 5.42, 3.97, 2.34, 1.44, 1.02 for $M = 5, 10, 20, 30, 40$, respectively, suggesting significant volatility spillover from past $\Delta\ln\mathrm{DM}_t$ to $\Delta\ln\mathrm{YEN}_t$ at the 1% level (except for $M = 30, 40$). This finding differs from that of Baillie and Bollerslev (1990, Table VII), who find no volatility spillover from past $\Delta\ln\mathrm{DM}_t$ to $\Delta\ln\mathrm{YEN}_t$. In contrast, $Q_{1\mathrm{TRUN}}$ and S_1 are significant at the 5% level only for $M = 5, 10$, and both are less powerful than $Q_{1\mathrm{DAN}}$ in terms of p-value. Clearly, non-uniform weighting gives stronger evidence on volatility spillover. Finally, the one-directional tests $Q_{-1\mathrm{DAN}}, Q_{-1\mathrm{TRUN}}$ and S_{-1} are all insignificant at any reasonable levels for all $M = 5, 10, 20, 30$ and 40, implying that there is no volatility spillover from past $\Delta\ln\mathrm{YEN}_t$ to current $\Delta\ln\mathrm{DM}_t$.

To summarize: (1) for causality in mean, there exists only strong simultaneous interaction between $\Delta\ln\mathrm{DM}_t$ and $\Delta\ln\mathrm{YEN}_t$; (2) for causality in volatility, there exists strong simultaneous interaction between $\Delta\ln\mathrm{DM}_t$ and $\Delta\ln\mathrm{YEN}_t$. Also, $\Delta\ln\mathrm{DM}_t$ Granger-causes $\Delta\ln\mathrm{YEN}_t$ with respect to I_{t-1} but $\Delta\ln\mathrm{YEN}_t$ does not Granger-cause $\Delta\ln\mathrm{DM}_t$ with respect to I_{t-1}; (3)

non-uniform weighting is more powerful than uniform weighting in detecting volatility spillover between exchange rates.

7. Conclusions

A class of asymptotic $\mathrm{N}(0,1)$ tests for volatility spillover are proposed. The new tests are based on the sample cross-correlation function between two squared standardized residual series. We do not assume any specific innovation distribution (e.g., normality) and our tests apply to time series that exhibit conditional heteroskedasticity and may have infinite unconditional variances. We permit to use all the sample cross-correlations. This enhances power against the alternatives whose cross-correlation decays to zero slowly. We also introduce a flexible weighting scheme for the cross-correlation at each lag. In particular, we permit larger weights for lower order lags. This is expected to give good power against the alternatives with decaying cross-correlations as the lag order increases. Indeed, non-uniform weighting often delivers better power than uniform weighting, as is illustrated in a simulation study and an application to exchange rates. Cheung and Ng (1996) test and Granger (1969)-type regression-based test are equivalent to a uniform weighting based test. Simulation studies show that our tests perform reasonably well in finite samples. Finally, the new tests are applied to investigate causality between two weekly nominal dollar exchange rates—Deutschemark and Japanese yen. It is found that for causality in mean, there exists only strong simultaneous interaction between the two exchange rates. For causality in variance, there exists strong simultaneous interaction between the two exchange rates; also, a change in past Deutschemark volatility Granger-causes a change in current Japanese yen volatility, but a change in past Japanese yen volatility does not Granger-cause a change in current Deutschemark volatility. This finding differs from such studies as Baillie and Bollerslev (1990), who find no volatility spillover between Deutschemark and Japanese yen.

Like tests for causality in mean, our tests, as well as Cheung and Ng (1996) tests, will fail to detect non-linear causation patterns with zero cross-correlation between the squared innovations. However, the present approach can be extended to develop a test that has power against such non-linear alternatives. For this, we can check whether the squared standardized errors $\{u_t^0\}$ and $\{v_t^0\}$ are independent. Let $I(j)$ be a measure of dependence between u_t^0 and v_{t-j}^0 such that $I(j)=0$ if and only if u_t^0 and v_{t-j}^0 are independent. One example is the Kullback–Leibler Information Criterion

$$I(j)=\mathrm{E}[\ln\{f_{uv}(u_t^0, v_{t-j}^0)/f_u(u_t^0)f_v(v_{t-j}^0)\}],$$

where f_{uv}, f_u and f_v are the joint and marginal probability density functions of u_t^0 and v_t^0, respectively. Such a measure has been used in identifying possible lag structures of non-linear time series (e.g., Granger and Lin, 1994) and in hypothesis testing (e.g., Hong and White, 2001; Robinson, 1991). A test that will have power against alternatives with zero cross-correlation between u_t^0 and v_{t-j}^0 can be based on

$$\sum_{j=1}^{T-1} k^2(j/M)\hat{I}_T(j), \tag{34}$$

where $\hat{I}_T(j) = T^{-1}\sum_{t=j+1}^{T} \ln\{\hat{f}(\hat{u}_t, \hat{v}_{t-j})/\hat{f}_u(\hat{u}_t)\hat{f}_v(\hat{v}_{t-j})\}$, with $\hat{f}_{uv}$, $\hat{f}_u$ and $\hat{f}_v$ some non-parametric estimators for the joint and marginal density functions of the squared standardized residuals $\hat{u}_t$ and $\hat{v}_{t-j}$. The approach of this paper and that of Hong and White (2001) may be useful in obtaining a well-defined distribution of (34).

Acknowledgements

I'd like to thank three referees for their insightful reviews and comments. I'd also like to thank J. Chu, R. Engle, C. Hsiao, N. Kiefer, T. Lee, J. Lu, W. Lin, I. Perrigne, P. Rilstone, Q. Vuong, H. Wan, and seminar participants at the 1998 California Econometric Camp at Catalina Island, Los Angeles, and York University for useful comments and suggestions. F. Xu provided excellent research assistance. All remaining errors are solely attributed to the author.

Appendix A: Mathematical appendix

Throughout the appendix, $||\cdot||$ denotes the Euclidean norm of a vector or a matrix and $0<\Delta<\infty$ denotes a generic constant that may differ in different places.

Proof of Theorem 1. Recalling $u_t^0 = \xi_{1t}^2 - 1$ and $v_t^0 = \xi_{2t}^2 - 1$, we define $\hat{C}_{uv}^0(j) = T^{-1}\sum_{t=j+1}^{T} u_t^0 v_{t-j}^0$ for $j \geqslant 0$, $C_{uu}^0(0) = \mathrm{E}(u_t^0)^2$ and $C_{vv}^0(0) = \mathrm{E}(v_t^0)^2$. By (17), we can write

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/M)\hat{\rho}_{uv}^2(j)\\
&\quad=\{\hat{C}_{uu}(0)\hat{C}_{vv}(0)\}^{-1}\sum_{j=1}^{T-1} k^2(j/M)\hat{C}_{uv}^0(j)^2\\
&\qquad+\{\hat{C}_{uu}(0)\hat{C}_{vv}(0)\}^{-1}\sum_{j=1}^{T-1} k^2(j/M)\{\hat{C}_{uv}(j)^2 - \hat{C}_{uv}^0(j)^2\}
\end{aligned}$$

$$
\begin{aligned}
&= \{C_{uu}^0(0)C_{vv}^0(0)\}^{-1}\sum_{j=1}^{T-1}k^2(j/M)\hat{C}_{uv}^0(j)^2 \\
&\quad + [\{\hat{C}_{uu}(0)\hat{C}_{vv}(0)\}^{-1} - \{C_{uu}^0(0)C_{vv}^0(0)\}^{-1}]\sum_{j=1}^{T-1}k^2(j/M)\hat{C}_{uv}^0(j)^2 \\
&\quad + \{\hat{C}_{uu}(0)\hat{C}_{vv}(0)\}^{-1}\sum_{j=1}^{T-1}k^2(j/M)\{\hat{C}_{uv}(j)^2 - \hat{C}_{uv}^0(j)^2\} \qquad \text{(A.1)} \\
&= \{C_{uu}^0(0)C_{vv}^0(0)\}^{-1}\sum_{j=1}^{T-1}k^2(j/M)\hat{C}_{uv}^0(j)^2 + o_P(M^{1/2}/T). \qquad \text{(A.2)}
\end{aligned}
$$

Here, we have made use of $\sum_{j=1}^{T-1} k^2(j/M)\hat{C}_{uv}^0(j)^2 = O_P(M/T)$ by Markov's inequality and

$$
\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/M)\mathrm{E}\hat{C}_{uv}^0(j)^2 \\
&\quad = (M/T)C_{uu}^0(0)C_{vv}^0(0)\left\{M^{-1}\sum_{j=1}^{T-1}(1-j/T)k^2(j/M)\right\} = O_P(M/T),
\end{aligned}
$$

where $M^{-1}\sum_{j=1}^{T-1}(1-j/T)k^2(j/M) \to \int_0^\infty k^2(z)\,\mathrm{d}z < \infty$ by Assumptions A.4–A.5. This, together with Lemma A.1 below, implies that the second term in (A.1) is $O_P(M/T^{3/2}) = o_P(M^{1/2}/T)$ as $M/T \to 0$. That the last term in (A.1) is $o_P(M^{1/2}/T)$ follows by Lemma A.2 below. The asymptotic normality of Q_1 then follows from (A.2) and Theorem A.3 below. This completes the proof. □

Lemma A.1. Suppose Assumptions A.1–A.5 *hold under the model described by* (28)–(30)*. Then* $\hat{C}_{uu}(0) - C_{uu}^0(0) = O_P(T^{-1/2})$ *and* $\hat{C}_{vv}(0) - C_{vv}^0(0) = O_P(T^{-1/2})$.

Lemma A.2. Suppose the conditions of Theorem 1 *hold. Then*

$$\sum_{j=1}^{T-1} k^2(j/M)\{\hat{C}_{uv}(j)^2 - \hat{C}_{uv}^0(j)^2\} = o_P(M^{1/2}/T).$$

Theorem A.3. Suppose the conditions of Theorem 1 *hold. Then*

$$\left[\{C_{uu}^0(0)C_{vv}^0(0)\}^{-1}T\sum_{j=1}^{T-1}k^2(j/M)\hat{C}_{uv}^0(j)^2 - C_{1T}(k)\right]\Big/\{2D_{1T}(k)\}^{1/2} \to^d \mathrm{N}(0,1).$$

Proof of Lemma A.1. We shall focus on the proof for $\hat{C}_{uu}(0) = T^{-1}\sum_{t=1}^T \hat{u}_t^2$; the proof for $\hat{C}_{vv}(0)$ is the same. By the triangle inequality, we have

$$|\hat{C}_{uu}(0) - C_{uu}^0(0)| \leqslant |\hat{C}_{uu}(0) - \hat{C}_{uu}^0(0)| + |\hat{C}_{uu}^0(0) - C_{uu}^0(0)|,$$

where $\hat{C}^0_{uu}(0)-C^0_{uu}(0)=O_P(T^{-1/2})$ by Chebyshev's inequality and Assumption A.1. Therefore, it remains to show $\hat{C}_{uu}(0)-\hat{C}^0_{uu}(0)=O_P(T^{-1/2})$. Recall $\hat{u}_t \equiv u_t(\hat{\theta}_1)=\varepsilon^2_{1t}(\hat{\theta}_1)/h_{1t}(\hat{\theta}_1)-1$, where $\varepsilon_{1t}(\theta_1)$ and $h_{1t}(\theta_1)$ are as in (15) and (16), we have

$$\begin{aligned}\hat{C}_{uu}(0)-\hat{C}^0_{uu}(0) &= T^{-1}\sum_{t=1}^{T}\{u_t(\hat{\theta}_1)^2-(u^0_t)^2\}\\ &\leqslant T^{-1}\sum_{t=1}^{T}\{u_t(\hat{\theta}_1)-u^0_t\}^2\\ &\quad +2\left[T^{-1}\sum_{t=1}^{T}\{u_t(\hat{\theta}_1)-u^0_t\}^2\right]^{1/2}\left\{T^{-1}\sum_{t=1}^{T}(u^0_t)^2\right\}^{1/2}.\end{aligned}$$

Because $T^{-1}\sum_{t=1}^{T}(u^0_t)^2=O_P(1)$ by Markov's inequality and Assumption A.1, it suffices to show $T^{-1}\sum_{t=1}^{T}\{u_t(\hat{\theta}_1)-u^0_t\}^2=O_P(T^{-1})$. Denote $\tilde{u}_t(\theta_1)=\varepsilon^2_1(\theta_1)/\tilde{h}_{1t}(\theta_1)-1$, where

$$\begin{aligned}\tilde{h}_{1t}(\theta_1)&=\omega_1+\alpha_1\varepsilon^2_{1t-1}(\theta_1)+\beta_1\tilde{h}_{1t-1}(\theta_1)\\ &=\omega_1/(1-\beta_1)+\alpha_1\sum_{i=0}^{\infty}\beta_1^i\varepsilon^2_{1t-1-i}(\theta_1)\end{aligned}$$

is an unobservable strictly stationary process which starts from the infinite past (cf. Lee and Hansen, 1994; Lumsdaine, 1996). Note that $\tilde{h}_{1t}(\theta^0_1)=h^0_{1t}$ but $h_{1t}(\theta^0_1)\neq h^0_{1t}$ due to the start-up value h^*_{10}. Consequently, $\tilde{u}_t(\theta^0_1)=u^0_t$ but $u_t(\theta^0_1)\neq u^0_t$. Now, noting that $u_t(\hat{\theta}_1)-u^0_t=\{u_t(\hat{\theta}_1)-\tilde{u}_t(\hat{\theta}_1)\}+\{\tilde{u}_t(\hat{\theta}_1)-u^0_t\}$, we can write

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}\{u_t(\hat{\theta}_1)-u^0_t\}^2 &\leqslant 2T^{-1}\sum_{t=1}^{T}\{u_t(\hat{\theta}_1)-\tilde{u}_t(\hat{\theta}_1)\}^2\\ &\quad +2T^{-1}\sum_{t=1}^{T}\{\tilde{u}_t(\hat{\theta}_1)-u^0_t\}^2\\ &= 2\hat{A}_{1T}+2\hat{A}_{2T},\ \text{say}.\end{aligned}$$

It suffices to show $\hat{A}_{iT}=O_P(T^{-1})$, $i=1,2$. We first consider $\hat{A}_{1T}$. Noting that $\tilde{h}_{1t}(\theta_1)-h_{1t}(\theta_1)=\beta_1^t\{\tilde{h}_{10}(\theta_1)-h^*_{10}\}$, $u_t(\theta_1)=\varepsilon^2_{1t}(\theta_1)/h_{1t}(\theta_1)-1$, $h_{1t}(\theta_1)\geqslant\Delta^{-1}$, $\tilde{h}_{1t}(\theta_1)\geqslant\Delta^{-1}$, where Δ is independent of θ_1, we have

$$\begin{aligned}\hat{A}_{1T} &= T^{-1}\sum_{t=1}^{T}\varepsilon^4_{1t}(\hat{\theta}_1)[\{\tilde{h}_{1t}(\hat{\theta}_1)-h_{1t}(\hat{\theta}_1)\}/\{h_{1t}(\hat{\theta}_1)\tilde{h}_{1t}(\hat{\theta}_1)\}]^2\\ &\leqslant \Delta^{-4}T^{-1}\{\tilde{h}_{10}(\hat{\theta}_1)-h^*_{10}\}^2\sum_{t=1}^{T}\hat{\beta}_1^{2t}\varepsilon^4_{1t}(\hat{\theta}_1).\end{aligned}$$

Let Θ_1^0 be a convex compact neighborhood of θ_1^0. Because $\mathrm{E}\sup_{\theta_1\in\Theta_1^0}|\tilde{h}_{10}(\theta_1)|^{2p} \leqslant \Delta < \infty$ for $0<p<\frac{1}{2}$, as shown in Lumsdaine (1996), we have $\sup_{\theta_1\in\Theta_1^0}\tilde{h}_{10}^2(\theta_1) = O_P(1)$ by Markov's inequality. In addition, for $0<p<\frac{1}{4}$, we have

$$\begin{aligned}
\mathrm{E}\left\{\sup_{\theta_1\in\Theta_1^0}|\varepsilon_{1t}^4(\theta_1)|^p\right\} &= \mathrm{E}\sup_{\theta_1\in\Theta_1^0}|\varepsilon_{1t}(\theta_1^0)+(b_1-b_1^0)|^{4p} \\
&\leqslant 4\mathrm{E}\{\varepsilon_{1t}(\theta_1^0)\}^{4p} + 4\sup_{\theta_1\in\Theta_1^0}|b_1-b_1^0|^{4p} \\
&= 4\mathrm{E}(\xi_{1t}^{4p})\mathrm{E}(h_{1t}^0)^{4p} + 4\sup_{\theta_1\in\Theta_1^0}|b_1-b_1^0|^{4p}. \\
&\leqslant 8\Delta,
\end{aligned}$$

where $\mathrm{E}(h_{1t}^0)^{4p} \leqslant \Delta < \infty$, as shown in Nelson (1990). It follows that $\sup_{\theta_1\in\Theta_1^0}\sum_{t=1}^T \beta_1^{2t}\varepsilon_{1t}^4(\theta_1) = O_P(1)$ by Markov's inequality and $0<\beta_1\leqslant 1-\delta_1<1$ for $\theta_1\in\Theta_1^0$, where $\delta_1>0$ is some arbitrarily small constant. Therefore, $A_{1T}=O_P(T^{-1})$.

Next, by the mean value theorem and the Cauchy–Schwarz inequality, we have

$$A_{2T} \leqslant ||\hat{\theta}_1-\theta_1^0||^2\left\{T^{-1}\sum_{t=1}^T||\nabla_{\theta_1}\tilde{u}_t(\bar{\theta}_1)||^2\right\} = O_P(T^{-1})$$

given Assumption A.3, where $\bar{\theta}_1$ lies in the segment between $\hat{\theta}_1$ and θ_1^0, and ∇_{θ_1} is the gradient operator with respect to θ_1. Here, we have made use of $\sup_{\theta_1\in\Theta_1^0}T^{-1}\sum_{t=1}^T||\nabla_{\theta_1}\tilde{u}_t(\theta_1)||^2 = O_P(1)$, which follows by the weak uniform law of large numbers (e.g., Andrews, 1992, Theorem 3) and the facts that $\sup_{\theta_1\in\Theta_1^0}\mathrm{E}||\nabla_{\theta_1}\tilde{u}_t(\theta_1)||^2\leqslant\Delta$ and the elements of $\sup_{\theta_1\in\Theta_1^0}\mathrm{E}\{\nabla_{\theta_1}||\nabla_{\theta_1}\tilde{u}_t(\theta_1)||^2\}$ are bounded by Δ (cf. Lee and Hansen, 1994, p. 49, p. 52, for their proofs of Lemmas 9 and 12 in that paper). This completes the proof. □

Proof of Lemma A.2. Noting that $\hat{C}_{uv}(j)^2-\hat{C}_{uv}^0(j)^2 = \{\hat{C}_{uv}(j)-\hat{C}_{uv}^0(j)\}^2 + 2\hat{C}_{uv}^0(j)\{\hat{C}_{uv}(j)-\hat{C}_{uv}^0(j)\}$, we have

$$\begin{aligned}
&\sum_{j=1}^{T-1}k^2(j/M)\{\hat{C}_{uv}(j)^2-\hat{C}_{uv}^0(j)^2\} \\
&\quad=\sum_{j=1}^{T-1}k^2(j/M)\{\hat{C}_{uv}(j)-\hat{C}_{uv}^0(j)\}^2 \\
&\qquad+2\sum_{j=1}^{T-1}k^2(j/M)\hat{C}_{uv}^0(j)\{\hat{C}_{uv}(j)-\hat{C}_{uv}^0(j)\} \\
&\quad=\hat{A}_{3T}+2\hat{A}_{4T},\ \text{say.}
\end{aligned}\tag{A.3}$$

We shall prove that both $\hat{A}_{3T}$ and $\hat{A}_{4T}$ are $o_P(M^{1/2}/T)$. For $j \geqslant 0$, write

$$\hat{C}_{uv}(j) - \hat{C}^0_{uv}(j) = T^{-1}\sum_{t=j+1}^{T}(\hat{u}_t - u^0_t)v^0_{t-j} + T^{-1}\sum_{t=j+1}^{T}u^0_t(\hat{v}_{t-j} - v^0_{t-j}) + T^{-1}\sum_{t=j+1}^{T}(\hat{u}_t - u^0_t)(\hat{v}_{t-j} - v^0_{t-j}) = \hat{B}_{1T}(j) + \hat{B}_{2T}(j) + \hat{B}_{3T}(j),\ \text{say.}$$

It follows that

$$\hat{A}_{3T} \leqslant 8\sum_{j=1}^{T-1}k^2(j/M)\{\hat{B}^2_{1T}(j) + \hat{B}^2_{2T}(j) + \hat{B}^2_{3T}(j)\}. \tag{A.4}$$

For the last term in (A.4), we have

$$\sup_{1 \leqslant j \leqslant T-1}\hat{B}^2_{3T}(j) \leqslant \left\{T^{-1}\sum_{t=1}^{T}(\hat{u}_t - u^0_t)^2\right\}\left\{T^{-1}\sum_{t=1}^{T}(\hat{v}_t - v^0_t)^2\right\} = O_P(T^{-2}) \tag{A.5}$$

by Cauchy–Schwarz inequality, $T^{-1}\sum_{t=1}^{T}(\hat{u}_t - u^0_t)^2 = O_P(T^{-1})$ and $T^{-1}\sum_{t=1}^{T}(\hat{v}_t - v^0_t)^2 = O_P(T^{-1})$ as shown in the proof of Lemma A.1. Therefore,

$$\sum_{j=1}^{T-1}k^2(j/M)\hat{B}^2_{3T}(j) \leqslant M\sup_{1 \leqslant j \leqslant T-1}\hat{B}^2_{3T}(j)\left\{M^{-1}\sum_{j=1}^{T-1}k^2(j/M)\right\} = O_P(M/T^2). \tag{A.6}$$

Next, we consider the first term $\hat{B}_{1T}(j)$ in (A.4). Because $u_t(\hat{\theta}_1) - u^0_t = \{u_t(\hat{\theta}_1) - \tilde{u}_t(\theta_1)\} + \{\tilde{u}_t(\theta_1) - u^0_t\}$, we obtain

$$\hat{B}_{1T}(j) = T^{-1}\sum_{t=j+1}^{T}\{u_t(\hat{\theta}_1) - \tilde{u}_t(\hat{\theta}_1)\}v^0_{t-j} + T^{-1}\sum_{t=j+1}^{T}\{\tilde{u}_t(\hat{\theta}_1) - u^0_t\}v^0_{t-j} = \hat{B}_{11T}(j) + \hat{B}_{12T}(j),\ \text{say.} \tag{A.7}$$

For the first term $\hat{B}_{11T}(j)$ in (A.7), recalling that $u_t(\hat{\theta}_1) - \tilde{u}_t(\hat{\theta}_1) = \varepsilon^2_{1t}(\hat{\theta}_1)\{h^{-1}_{1t}(\hat{\theta}_1) - \tilde{h}^{-1}_{1t}(\hat{\theta}_1)\}$ and $\tilde{h}_{1t}(\hat{\theta}_1) - h_{1t}(\hat{\theta}_1) = \hat{\beta}^t_1\{\tilde{h}_{10}(\hat{\theta}_1) - h^*_{10}\}$, we have

$$\hat{B}_{11T}(j) = T^{-1}\sum_{t=j+1}^{T}\varepsilon^2_{1t}(\hat{\theta}_1)\{h^{-1}_{1t}(\hat{\theta}_1) - \tilde{h}^{-1}_{1t}(\hat{\theta}_1)\}v^0_{t-j} = T^{-1}\{\tilde{h}_{10}(\hat{\theta}_1) - h^*_{10}\}\sum_{t=j+1}^{T}\hat{\beta}^t_1\varepsilon^2_{1t}(\hat{\theta}_1)h^{-1}_{1t}(\hat{\theta}_1)\tilde{h}^{-1}_{1t}(\hat{\theta}_1)v^0_{t-j}.$$

By Cauchy–Schwarz inequality and $0<\hat{\beta}_1 \leqslant 1-\delta_1<1$ for some small $\delta_1>0$, we obtain

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/M)\hat{B}_{11T}^2(j) \\
&\quad \leqslant T^{-2}\left\{\sum_{j=1}^{T-1} k^2(j/M)\right\} \\
&\qquad \times \left[(\tilde{h}_{10}(\hat{\theta}_1)-h_{10}^*)^2 \sum_{t=1}^{T} \hat{\beta}_1^t \varepsilon_{1t}^4(\hat{\theta}_1)\{h_{1t}^{-1}(\hat{\theta}_1)\tilde{h}_{1t}^{-1}(\hat{\theta}_1)\}^2\right]\left\{\sum_{t=1}^{T}\hat{\beta}_1^t (v_t^0)^2\right\} \\
&\quad = O_P(M/T^2), \qquad\qquad (A.8)
\end{aligned}$$

where we use $\sup_{\theta_1\in\Theta_1^0}\tilde{h}_{10}^2(\theta_1)=O_P(1)$, $\sum_{t=j+1}^{T}\hat{\beta}_1^t\varepsilon_{1t}^4(\hat{\theta}_1)\{h_{1t}^{-1}(\hat{\theta}_1)\tilde{h}_{1t}^{-1}(\hat{\theta}_1)\}^2 = O_P(1)$ (as can be shown using a reasoning analogous to the proof of Lemma A.1), and $\sum_{t=1}^{T}\hat{\beta}_1^t(v_t^0)^2 \leqslant \sum_{t=1}^{T}(1-\delta_1)^t(v_t^0)^2 = O_P(1)$ by Markov's inequality and $0<\hat{\beta}_1 \leqslant 1-\delta_1<1$ for $\hat{\theta}_1\in\Theta_1^0$.

Next, for the second term in (A.7), by a two-term Taylor expansion, we obtain

$$\begin{aligned}
\hat{B}_{12T}(j) &= (\hat{\theta}_1-\theta_1^0)' T^{-1}\sum_{t=j+1}^{T}\nabla_{\theta_1}\tilde{u}_t(\theta_1^0)v_{t-j}^0 \\
&\quad + \frac{1}{2}(\hat{\theta}_1-\theta_1^0)'\left\{T^{-1}\sum_{t=j+1}^{T}\nabla_{\theta_1}^2\tilde{u}_t(\bar{\theta}_1)v_{t-j}^0\right\}(\hat{\theta}_1-\theta_1^0) \\
&= \hat{B}_{121T}(j)+\hat{B}_{122T}(j), \text{ say}, \qquad\qquad (A.9)
\end{aligned}$$

where $\bar{\theta}_1$ lies in the segment between $\hat{\theta}_1$ and θ_1^0, and $\nabla_{\theta_1}^2$ is the Hessian operator with respect to θ_1. For the first term in (A.9), we have

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/M)\hat{B}_{121T}^2(j) \\
&\quad \leqslant ||\hat{\theta}_1-\theta_1^0||^2\left\{\sum_{j=1}^{T-1}k^2(j/M)\left|\left|T^{-1}\sum_{t=j+1}^{T}\nabla_{\theta_1}\tilde{u}_t(\theta_1^0)v_{t-j}^0\right|\right|^2\right\} \\
&\quad = O_P(M/T^2) \qquad\qquad (A.10)
\end{aligned}$$

by Assumption A.3 and $\sum_{j=1}^{T-1}k^2(j/M)||T^{-1}\sum_{t=j+1}^{T}\nabla_{\theta_1}\tilde{u}_t(\theta_1^0)v_{t-j}^0||^2=O_P(M/T)$, which follows by Chebyshev's inequality and the fact that

$$\begin{aligned}
\mathrm{E}\left|\left|T^{-1}\sum_{t=j+1}^{T}\nabla_{\theta_1}\tilde{u}_t(\theta_1^0)v_{t-j}^0\right|\right|^2 &= T^{-2}\sum_{t=j+1}^{T}\mathrm{E}||\nabla_{\theta_1}\tilde{u}_t(\theta_1^0)||^2\mathrm{E}(v_{t-j}^0)^2 \\
&= O(T^{-1}),
\end{aligned}$$

where the first equality follows by independence between ξ_{1t} and ξ_{2t}.

For the second term in (A.9), we have

$$\sum_{j=1}^{T-1} k^2(j/M)\hat{B}_{122T}^2(j) \leqslant ||\hat{\theta}_1 - \theta_1^0||^4 \left\{\sum_{j=1}^{T-1} k^2(j/M)\right\} \times \left\{T^{-1}\sum_{t=1}^{T}||\nabla_{\theta_1}^2 \tilde{u}_t(\bar{\theta}_1)||^2\right\}\left\{T^{-1}\sum_{t=1}^{T}(v_t^0)^2\right\} = O_P(M/T^2), \quad (A.11)$$

where $\sup_{\theta_1\in\Theta_1^0} T^{-1}\sum_{t=1}^{T}||\nabla_{\theta_1}^2 \tilde{u}_t(\bar{\theta})||^2 = O_P(1)$, which follows by the weak law of large numbers (e.g., Andrews, 1992) and the fact that $\sup_{\theta_1\in\Theta_0^1} \mathrm{E}||\nabla_{\theta_1}^2 \tilde{u}_t(\theta_1)||^2 \leqslant \Delta$ and the elements of $\sup_{\theta_1\in\Theta_1^0} \mathrm{E}\{\nabla_{\theta_1}||\nabla_{\theta_1}^2 \tilde{u}_t(\theta_1)||^2\}$ are bounded by Δ, using a reasoning analogous to that of Lee and Hansen (1994, pp. 50–51). Combining (A.7)–(A.11), we have

$$\sum_{j=1}^{T-1} k^2(j/M)\hat{B}_{1T}^2(j) = O_P(M/T^2) = o_P(M^{1/2}/T). \quad (A.12)$$

Similarly, we also have

$$\sum_{j=1}^{T-1} k^2(j/M)\hat{B}_{2T}^2(j) = O_P(M/T^2) = o_P(M^{1/2}/T). \quad (A.13)$$

Collecting (A.4), (A.6) and (A.12)–(A.13), we have $\hat{A}_{3T} = O_P(M/T^2) = o_P(M^{1/2}/T)$.

Finally, by Cauchy–Schwarz inequality, we have

$$|\hat{A}_{4T}| \leqslant \left\{\sum_{j=1}^{T-1} k^2(j/M)\hat{C}_{uv}^0(j)^2\right\}^{1/2} (\hat{A}_{3T})^{1/2} = O_P(M/T^{3/2}) = o_P(M^{1/2}/T),$$

given $M/T \to 0$, where $\sum_{j=1}^{T-1} k^2(j/M)\hat{C}_{uv}^0(j)^2 = O_P(M/T)$ by Markov's inequality. This completes the proof. □

Proof of Theorem A.3. Put $S_T = T\sum_{j=1}^{T-1} k^2(j/M)\hat{C}_{uv}^0(j)^2$, $\sigma_u^2 = C_{uu}^0(0)$ and $\sigma_v^2 = C_{vv}^0(0)$. Then $\mathrm{E}(S_T) = \sigma_u^2\sigma_v^2 C_{1T}(k)$ and $\sigma^2(T) \equiv Var(S_T) = \sigma_u^4\sigma_v^4 2D_{1T}(k) = \sigma_u^4\sigma_v^4 2MD(k)\{1+o(1)\}$. It suffices to show $\sigma(T)^{-1}\{S_T - \sigma_u^2\sigma_v^2 C_T(k)\} \to^d \mathrm{N}(0,1)$. Then

$$S_T = T^{-1}\left\{\sum_{j=1}^{T-2} k^2(j/M)\sum_{t=j+2}^{T}\sum_{s=j+1}^{t-1}(u_t^0 v_{t-j}^0)^2\right\} + 2T^{-1}\left\{\sum_{j=1}^{T-2} k^2(j/M)\sum_{t=j+2}^{T}\sum_{s=j+1}^{t-1} u_t^0 u_s^0 v_{t-j}^0 v_{s-j}^0\right\} = \bar{H}_T + \bar{W}_T, \text{ say.}$$

We shall show (i) $\sigma^{-1}(T)\{\bar{H}_T - \sigma_u^2\sigma_v^2 C_{1T}(k)\} = o_P(1)$, and (ii) $\sigma^{-1}(T)\bar{W}_T \to^d$ N(0,1). First, we verify (i). Noting $\mathrm{E}\bar{H}_T = \sigma_u^2\sigma_v^2 C_{1T}(k)$ and using Minkowski's Inequality, we obtain

$$\begin{aligned} Var(\bar{H}_T) &\leqslant \left(T\sum_{j=1}^{T-1} k^2(j/M) \left[\mathrm{E}\left\{ T^{-1}\sum_{t=j+1}^{T} [(u_t^0 v_{t-j}^0)^2 - \sigma_u^2\sigma_v^2] \right\}^2 \right]^{1/2} \right)^2 \\ &\leqslant \mu_{4u}\mu_{4v}(M^2/T)\left\{ M^{-1}\sum_{j=1}^{T-1} k^2(j/M) \right\}^2 \\ &= O(M^2/T), \end{aligned}$$

where $\mu_{4u} \equiv \mathrm{E}(u_t^0)^4$ and $\mu_{4v} \equiv \mathrm{E}(v_t^0)^4$. It follows that $\sigma^{-1}(T)\{\bar{H}_T - \sigma_u^2\sigma_v^2 C_{1T}(k)\} = O_P(M^{1/2}/T^{1/2}) = o_P(1)$ by Chebyshev's inequality and $M/T \to 0$. This proves (i).

Next, we turn to prove (ii). Rewrite

$$\bar{W}_T = T^{-1}\sum_{t=3}^{T} 2u_t^0 \left(\sum_{s=2}^{t-1} u_s^0 \sum_{j=1}^{s-1} k^2(j/M) v_{t-j}^0 v_{s-j}^0 \right) = T^{-1}\sum_{t=3}^{T} W_{Tt}, \text{ say.}$$

Note that $(W_{Tt}, \mathscr{F}_t)$ is a martingale difference sequence because $\mathrm{E}(W_{Tt} \mid \mathscr{F}_{t-1}) = 0$ under H_0, where $(\mathscr{F}_t)$ is the sequence of sigma fields consisting of (u_s^0, v_s^0), $s \leqslant t$. We show (ii) by Brown's (1971) martingale limit theorem, which implies $\{Var(\bar{W}_T)\}^{-1/2}\bar{W}_T \to^d \mathrm{N}(0,1)$ if

$$\{Var(\bar{W}_T)\}^{-1}\sum_{t=3}^{T} T^{-2}\mathrm{E}[W_{Tt}^2 1\{T^{-1}|W_{Tt}| > \eta Var^{1/2}(\bar{W}_T)\}] \to 0 \tag{A.14}$$

for every $\eta > 0$, and

$$\{Var(\bar{W}_T)\}^{-1}\sum_{t=3}^{T} T^{-2}\hat{W}_{Tt}^2 \to^p 1, \tag{A.15}$$

where $1(\cdot)$ denotes the indicator function, and

$$\hat{W}_{Tt}^2 = \mathrm{E}\{W_{Tt}^2 | \mathscr{F}_{t-1}\} = 4\sigma_u^2 \left\{ \sum_{s=2}^{t-1} u_s^0 \sum_{j=1}^{s-1} k^2(j/M) v_{t-j}^0 v_{s-j}^0 \right\}^2.$$

We note that $Var(\bar{W}_T) = \sigma^2(T) \sim M$.

To verify condition (A.14), it suffices to show $\sigma^{-4}(T)T^{-4}\sum_{t=3}^{T}\mathrm{E}W_{Tt}^4 = o(1)$. Put

$$G_{ts}^v = \sum_{j=1}^{s-1} k^2(j/M) v_{t-j}^0 v_{s-j}^0. \tag{A.16}$$

Then $W_{Tt}=2u_t^0\sum_{s=2}^{t-1}u_s^0G_{ts}^v$. Given Assumption A.1 and independence between u_t^0 and v_s^0,

$$\begin{aligned}
\mathrm{E}W_{Tt}^4 &= 16\mu_{4u}\left(\sum_{s=2}^{t-1}u_s^0G_{ts}^v\right)^4\\
&= 16\mu_{4u}^2\sum_{s=2}^{t-1}\mathrm{E}(G_{ts}^v)^4+96\mu_{4u}\sigma_u^4\sum_{s_2=s_1+1}^{t-1}\sum_{s_1=2}^{s_2-1}\mathrm{E}\{(G_{ts_1}^v)^2(G_{ts_2}^v)^2\}\\
&\leqslant 48\mu_{4u}^2\left[\sum_{s=2}^{t-1}\{\mathrm{E}(G_{ts}^v)^4\}^{1/2}\right]^2\\
&= O(t^2M^2),
\end{aligned}$$

where for $t>s$,

$$\begin{aligned}
\mathrm{E}(G_{ts}^v)^4 &= \mathrm{E}\left\{\sum_{j=1}^{s-1}k^2(j/M)v_{t-j}^0v_{s-j}^0\right\}^4\\
&= \sum_{j=1}^{s-1}k^8(j/M)\mathrm{E}(v_{t-j}^0v_{s-j}^0)^4\\
&\quad +6\sum_{j=2}^{s-1}\sum_{i=1}^{j-1}k^4(i/M)k^4(j/M)\mathrm{E}(v_{t-i}^0v_{t-j}^0v_{s-i}^0v_{s-j}^0)^2\\
&\leqslant 3\mu_{4v}^2M^2\left\{M^{-1}\sum_{j=1}^{s-1}k^4(j/M)\right\}^2\\
&= O(M^2).
\end{aligned}$$

It follows that $\sigma^{-4}(T)T^{-4}\sum_{t=3}^{T}\mathrm{E}W_{Tt}^4=O(T^{-1})$. Hence, (A.14) holds.

We now verify (A.15) by showing $\sigma^{-4}(T)Var(T^{-2}\sum_{t=3}^{T}\hat{W}_{Tt}^2)\to 0$. By definition of $\hat{W}_{Tt}^2$, we have $\hat{W}_{Tt}^2=4\sigma_u^2\left\{\sum_{s=2}^{t-1}u_s^0G_{ts}^v\right\}^2$. Therefore, it suffices to show $M^{-2}Var(T^{-2}\sum_{t=3}^{T}\hat{W}_{Tt}^2)\to 0$. Write

$$\mathrm{E}\left(T^{-2}\sum_{t=3}^{T}\hat{W}_{Tt}^2\right)^2=T^{-4}\sum_{t=3}^{T}\mathrm{E}\hat{W}_{Tt}^4+2T^{-4}\sum_{t_2=4}^{T}\sum_{t_1=3}^{t_2-1}\mathrm{E}(\hat{W}_{Tt_2}^2\hat{W}_{Tt_1}^2). \tag{A.17}$$

Because $\mathrm{E}\hat{W}_{Tt}^4=(\mu_{4u}/\sigma_u^4)\mathrm{E}W_{Tt}^4$, the first term

$$T^{-4}\sum_{t=3}^{T}\mathrm{E}\hat{W}_{Tt}^4=(\mu_{4u}/\sigma_u^4)T^{-4}\sum_{t=3}^{T}\mathrm{E}W_{Tt}^4=O(M^2/T), \tag{A.18}$$

as shown in verifying (A.14). For the second term, given $t_2 > t_1$ and the independence between the i.i.d. sequences $\{u_t^0\}$ and $\{v_s^0\}$, we have

$$\begin{aligned}
\mathrm{E}(\hat{W}_{Tt_2}^2 \hat{W}_{t_1}^2) &= 16\sigma_u^4 \mathrm{E}\left\{\left(\sum_{s=2}^{t_2-1} u_s^0 G_{t_2 s}^v\right)^2 \left(\sum_{s=2}^{t_1-1} u_s^0 G_{t_1 s}^v\right)^2\right\} \\
&= 16\sigma_u^4 \sum_{s_2=2}^{t_2-1}\sum_{s_1=2}^{t_1-1} \mathrm{E}(u_{s_2}^0 u_{s_1}^0)^2 \mathrm{E}\{(G_{t_2 s_2}^v)^2 (G_{t_1 s_1}^v)^2\} \\
&\quad + 64\sigma_u^4 \sum_{s_2=3}^{t_1-1}\sum_{s_1=2}^{s_2-1} \mathrm{E}(u_{s_2}^0 u_{s_1}^0)^2 \mathrm{E}(G_{t_2 s_2}^v G_{t_2 s_1}^v G_{t_1 s_2}^v G_{t_1 s_1}^v) \\
&= 16(\mu_{4u} - \sigma_u^4)\sigma_u^4 \sum_{s=2}^{t_1-1} \mathrm{E}\{(G_{t_2 s}^v)^2 (G_{t_1 s}^v)^2\} \\
&\quad + 16\sigma_u^8 \sum_{s_2=2}^{t_2-1}\sum_{s_1=2}^{t_1-1} \mathrm{E}\{(G_{t_2 s_2}^v)^2 (G_{t_1 s_1}^v)^2\} \\
&\quad + 64\sigma_u^8 \sum_{s_2=3}^{t_1-1}\sum_{s_1=2}^{s_2-1} \mathrm{E}(G_{t_2 s_2}^v G_{t_2 s_1}^v G_{t_1 s_2}^v G_{t_1 s_1}^v),
\end{aligned}$$

where for the first term,

$$\sum_{s=2}^{t_1-1} \mathrm{E}\{(G_{t_2 s}^v)^2 (G_{t_1 s}^v)^2\} = O(t_1 M^2)$$

by the Cauchy–Schwarz inequality and $\mathrm{E}(G_{ts}^v)^4 = O(M^2)$, as shown in verifying (A.14). Next, substituting (A.16) and by direct but tedious algebra, we obtain that for $t_2 > t_1$, $t_2 > s_2$, $t_1 > s_1$,

$$\begin{aligned}
&\mathrm{E}\{(G_{t_2 s_2}^v)^2 (G_{t_1 s_1}^v)^2\} \\
&\quad = \sum_{i=1}^{s_2-1}\sum_{j=1}^{s_1-1} k^4(i/M) k^4(j/M) \mathrm{E}(v_{t_2-i}^0 v_{s_2-i}^0 v_{t_1-j}^0 v_{s_1-j}^0)^2 + O(M) \\
&\quad = \sigma_v^8 \sum_{i=1}^{s_2-1}\sum_{j=1}^{s_1-1} k^4(i/M) k^4(j/M) + O(M),
\end{aligned}$$

and for $t_2 > t_1 > s_2 > s_1$,

$$\mathrm{E}\{G_{t_2 s_2}^v G_{t_2 s_1}^v G_{t_1 s_2}^v G_{t_1 s_1}^v\} = \sigma_v^8 M \left\{M^{-1} \sum_{j=1}^{s_1-1} k^8(j/M)\right\} = O(M).$$

It follows that

$$\begin{aligned}
2T^{-4} \sum_{t_2=4}^{T}\sum_{t_1=3}^{t_2-1} \mathrm{E}(\hat{W}_{Tt_2}^2 \hat{W}_{Tt_1}^2) &= 4\sigma_u^4 \sigma_v^8 \left[\sum_{j=1}^{T-1} (1-j/T)\{1-(j+1)/T\} k^4(j/M)\right]^2 \\
&\quad + O(M^2/T + M) \\
&= 4\sigma_u^4 \sigma_v^8 \{D_{1T}(k)\}^2 + O(M^2/T + M). \qquad \text{(A.19)}
\end{aligned}$$

Combining (A.17)–(A.19), we obtain

$$M^{-2}\mathrm{E}\left(T^{-2}\sum_{t=3}^{T}\hat{W}_{Tt}^{2}\right)^{2} = 4\sigma_u^8\sigma_v^8\{M^{-1}D_{1T}(k)\}^2 + O(T^{-1}+M^{-1}). \tag{A.20}$$

given $M \to \infty$, $M/T \to 0$. On the other hand,

$$M^{-1}T^{-2}\sum_{t=3}^{T}\mathrm{E}\hat{W}_{Tt}^{2} = 2\sigma_u^4\sigma_v^4 M^{-1}D_{1T}(k). \tag{A.21}$$

Combining (A.20)–(A.21) yields $M^{-2}Var(T^{-2}\sum_{t=3}^{T}\hat{W}_{Tt}^{2}) = o(1)$, ensuring condition (A.15). It follows $\sigma^{-1}(T)\bar{W}_T \to^d \mathrm{N}(0,1)$ by Brown's (1971) theorem. This completes the proof. □

Proof of Theorem 2. Because $M^{-1}C_{1T}(k) \to \int_0^\infty k^2(z)\,\mathrm{d}z$, $M^{-1}D_{1T}(k) \to M \int_0^\infty k^4(Z)$ and $M/T \to 0$, we have

$$\frac{M^{1/2}}{T}Q_1 = \left\{2\int_0^\infty k^4(z)\,\mathrm{d}z\right\}^{-1/2}\left\{\sum_{j=1}^{T-1}\hat{\rho}_{uv}^2(j)\right\}\{1+o(1)\}+o(1). \tag{A.22}$$

Moreover, by (A.1), Lemma A.1, Lemmas A.5–A.6 below, and $\rho_{uv}(j) = C_{uv}^0(j)/\{C_{uu}^0(0)C_{vv}^0(0)\}^{1/2}$, we obtain

$$\sum_{j=1}^{T-1}\hat{\rho}_{uv}^2(j) \to^p \sum_{j=1}^{\infty}\rho_{uv}^2(j). \tag{A.23}$$

Combining (A.22) and (A.23) yields the desired result. □

Lemma A.5. Suppose the conditions of Theorem 2 *hold. Then*

$$\sum_{j=1}^{T-1}k^2(j/M)\{\hat{C}_{uv}(j)^2 - \hat{C}_{uv}^0(j)^2\} = O_P(M^{1/2}/T^{1/2}).$$

Lemma A.6. Suppose the conditions of Theorem 2 *hold. Then*

$$\sum_{j=1}^{T-1}k^2(j/M)\hat{C}_{uv}^0(j)^2 \to^p \sum_{j=1}^{\infty}C_{uv}^0(j)^2.$$

Proof of Lemma A.5. Putting $\hat{A}_{3T} = \sum_{j=1}^{T-1}k^2(j/M)\{\hat{C}_{uv}(j) - \hat{C}_{uv}^0(j)\}^2$ as in (A.3), we have

$$\sum_{j=1}^{T-1}k^2(j/M)\left|\hat{C}_{uv}(j)^2 - \hat{C}_{uv}^0(j)^2\right| \leqslant \hat{A}_{3T} + 2(\hat{A}_{3T})^{1/2}\left\{\sum_{j=1}^{T-1}k^2(j/M)\hat{C}_{uv}^0(j)^2\right\}^{1/2}.$$

Because $\sum_{j=1}^{T-1} k^2(j/M)\hat{C}_{uv}^0(j)^2 = O_P(1)$ by Lemma A.6 and $\sum_{j=1}^{\infty} C_{uv}^0(j)^2 < \infty$, it suffices to show $\hat{A}_{3T} = O_P(M/T)$, which we now focus on. By (A.4), it suffices to show

$$\sum_{j=1}^{T-1} k^2(j/M)\hat{B}_{iT}^2(j) = O_P(M/T), \quad i = 1,2,3, \tag{A.24}$$

where the $\hat{B}_{iT}(j)$ are defined as in (A.4). As shown in the proof of Lemma A.1, we have $T^{-1}\sum_{t=1}^{T}(\hat{u}_t - u_t^0)^2 = O_P(T^{-1})$ and $T^{-1}\sum_{t=1}^{T}(\hat{v}_t - v_t^0)^2 = O_P(T^{-1})$ under Assumptions A.1–A.5. It follows that $\sup_{1\leqslant j\leqslant T-1}\hat{B}_{iT}^2(j) = O_P(T^{-1})$ for $i = 1, 2, 3$ by the Cauchy–Schwarz inequality and Assumption A.1. This, together with $\sum_{j=1}^{T-1} k^2(j/M) = O(M)$, implies (A.24). The proof is then completed. $\square$

Proof of Lemma A.6. We first write

$$\begin{aligned}
\sum_{j=1}^{T-1} k^2(j/M)\hat{C}_{uv}(j)^2 &= \sum_{j=1}^{T-1} k^2(j/M)C_{uv}^0(j)^2 \\
&\quad + \sum_{j=1}^{T-1} k^2(j/M)\{\hat{C}_{uv}^0(j) - C_{uv}^0(j)\}^2 \\
&\quad + 2\sum_{j=1}^{T-1} k^2(j/M)\{\hat{C}_{uv}^0(j) - C_{uv}^0(j)\}C_{uv}^0(j) \\
&= \hat{A}_{4T} + \hat{A}_{5T} + 2\hat{A}_{6T}, \quad \text{say.}
\end{aligned} \tag{A.25}$$

For the first term in (A.25), we have

$$\begin{aligned}
\hat{A}_{4T} &= \sum_{j=1}^{\infty} C_{uv}^0(j)^2 + \sum_{j=1}^{T-1}\{k^2(j/M) - 1\}C_{uv}^0(j)^2 - \sum_{j=T}^{\infty} C_{uv}^0(j)^2 \\
&\to \sum_{j=1}^{\infty} C_{uv}^0(j)^2,
\end{aligned} \tag{A.26}$$

where $\sum_{j=T}^{\infty} C_{uv}^0(j)^2 \to 0$ given $\sum_{j=1}^{\infty} C_{uv}^0(j)^2 < \infty$, and $\sum_{j=1}^{T-1}\{k^2(j/M) - 1\} C_{uv}^0(j)^2 \to 0$ by dominated convergence, $k^2(j/M) - 1 \to 0$ for any given j as $M \to \infty |k^2(j/M) - 1| \leqslant 2$, and $\sum_{j=1}^{\infty} C_{uv}^0(j)^2 < \infty$.

Next, we consider the second term of (A.25). Because $\sup_{1\leqslant j\leqslant T-1} Var\{\hat{C}_{uv}^0(j)\} \leqslant \Delta T^{-1}$ given $\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|\kappa_{uvuv}(i,j,l)| < \infty$ and $\sum_{j=1}^{\infty}\rho_{uv}^2(j) < \infty$ (cf. Hannan, 1970, (3.3), p. 209), we have

$$\hat{A}_{5T} = O_P(M/T) \tag{A.27}$$

by Markov's inequality and $\sum_{j=1}^{T-1} k^2(j/M) = O(M)$. Finally, by Cauchy–Schwartz inequality and (A.26)–(A.27), we have

$$\hat{A}_{6T} = O_P(M^{1/2}/T^{1/2}). \tag{A.28}$$

Combining (A.25)–(A.28) yields the desired result. This completes the proof. □

References

Admati, A., Pfleiderer, P., 1988. A theory of intraday patterns: volume and price variability. Review of Financial Studies 1, 3–40.
Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Andrews, D.W.K., 1992. Generic uniform convergence. Econometric Theory 8, 241–257.
Baillie, R., Bollerslev, T., 1989. The message in daily exchange rates: a conditional variance tale. Journal of Business and Economics Statistics 7, 297–305.
Baillie, R., Bollerslev, T., 1990. Intra-day and inter-market volatility in foreign exchange rates. Review of Economic Studies 58, 565–585.
Bekaert, G., 1995. The time variation of expected returns and volatility in foreign-exchange markets. Journal of Business and Economic Statistics 13, 397–408.
Bera, A.K., Higgins, M.L., 1993. ARCH models: properties, estimation and testing. Journal of Economic Surveys 7, 305–366.
Bollerslev, T., 1986. A generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.
Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. Review of Economics and Statistics 72, 498–505.
Bollerslev, T., Chou, R., Kroner, K., 1992. ARCH modelling in finance: a review of the theory and empirical evidence. Journal of Econometrics 52, 5–60.
Bollerslev, T., Wooldridge, J.M., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11, 143–172.
Brown, B.M., 1971. Martingale central limit theorems. Annals of Mathematical Statistics 42, 59–66.
Cheung, Y.W., Ng, L.K., 1990. The dynamics of S&P 500 index and S&P 500 futures intraday price volatilities. Review of Futures Markets 9, 458–486.
Cheung, Y.W., Ng, L.K., 1996. A causality-in-variance test and its application to financial market prices. Journal of Econometrics 72, 33–48.
Diebold, F.X., Nerlove, M., 1989. The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. Journal of Applied Econometrics 1, 1–21.
Domowitz, I., Hakkio, C., 1985. Conditional variance and the risk premium in the foreign exchange market. Journal of International Economics 19, 47–66.
Drost, F.C., Nijman, T.E., 1993. Temporal aggregation of GARCH processes. Econometrica 61, 909–927.
Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 987–1007.
Engle, R.F., Bollerslev, T., 1986. Modelling the persistence of conditional variances. Econometric Reviews 5, 1–50.
Engle, R.F., Ito, T., Lin, W.L., 1990. Meteor shower or heat waves? heteroskedastic intra-daily volatility in the foreign exchange market. Econometrica 59, 524–542.
Engle, R.F., Susmel, R., 1993. Common volatility in international equity markets. Journal of Business and Economic Statistics 11, 167–176.

Gallant, A.R., Hsieh, D.A., Tauchen, G., 1989. On fitting a recalcitrant series: The pound/dollar exchange rate 1974–83. In: Barnett, W., Powell, J., Tauchen, T. (Eds.), Non-parametric and Semiparametric Methods in Econometrics, Cambridge University Press, Cambridge, pp. 199–240.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

Granger, C.W.J., 1980. Testing for causality: A personal view. Journal of Economic Dynamics and Control 2, 329–352.

Granger, C.W.J., Lin, J.L., 1994. Using the mutual information coefficient to identify lags in nonlinear models. Journal of Time Series Analysis 15, 371–384.

Granger, C.W.J., Robins, P.P., Engle, R.F., 1986. Wholesale and retail prices: bivariate time-series modeling with forecastable error variances. In: Belsley, D.A., Kuh, E. (Eds.), Model Reliability, MIT Press, Cambridge, pp. 1–17.

Berndt, E., Hall, B., Hall, R., Hausman, J., 1974. Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement 3, 653–665.

Hamao, Y., Masulis, R.W., Ng, V., 1990. Correlations in price changes and volatility across international stock markets. Review of Financial Studies 3, 281–307.

Hannan, E., 1970. Multiple Time Series Analysis. John Wiley, New York.

Hodrick, R.J., 1989. Risk, Uncertainty, and Exchange Rates. Journal of Monetary Economics 23, 433–459.

Hong, Y., 1996. Testing for independence between two covariance stationary time series. Biometrika 83, 615–625.

Hong, Y.M., White, H., 2001. Asymptotic theory for nonparametric entropy measures of serial dependence. Department of Economics and Department of Statistical Science, Cornell University.

Hsieh, D., 1988. The statistical properties of daily foreign exchange rates: 1974–1983. Journal of International Economics 24, 129–145.

Hsieh, D., 1989. Modeling heteroskedasticity in daily foreign exchange rates. Journal of Business and Economics Statistics 7, 307–317.

King, M., Wadhwani, S., 1990. Transmission of volatility between stock markets. Review of Financial Studies 3, 5–33.

King, M., Sentana, E., Wadhwani, S., 1994. Volatility and links between national stock markets. Econometrica 62, 901–933.

Kyle, A.A., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Lee, S.W., Hansen, B.E., 1994. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. Econometric Theory 10, 29–52.

Lin, W.L., Engle, R.F., Ito, T., 1994. Do bulls and bears move across borders? International transmission of stock returns and volatility. Review of Financial Studies 7, 507–538.

Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65 297–303.

Lumsdaine, R.L., 1996. Asymptotic properties of the quasi-maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models. Econometrica 64, 575–596.

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

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.

Nelson, D.B., Cao, C.Q., 1992. Inequality constraints in the univariate GARCH model. Journal of Business and Economic Statistics 10, 229–235.

Pierce, D., Haugh, L., 1977. Causality in temporal systems. Journal of Econometrics 5, 265–293.

Priestley, M.B., 1981. Spectral analysis and time series. Academic Press, London.

Robinson, P., 1991. Consistent nonparametric entropy-based testing. Review of Economic Studies 58, 437–453.

Ross, S.A., 1989. Information and volatility: the no-arbitrage martingale approach to timing and resolution irrelevancy. Journal of Finance XLIV, 1–17.
Shalen, R., 1993. Volume, volatility, and the dispersion of beliefs. Review of Financial Studies 6, 405–434.
Schwert, R.J., 1989. Why does stock market volatility change over time? Journal of Finance XLXV, 1115–1153.
Sentana, E., 1995. Quadratic ARCH models. Review of Economic Studies 62, 639–661.
Shiller, R.J., 1981a. Do stock prices move too much to be justified by subsequent changes in dividends. American Economic Review 75, 421–436.
Shiller, R.J., 1981b. The use of volatility measures in assessing market efficiency. Journal of Finance 17, 291–304.
West, K., Cho, D., 1995. The predictive ability of several models of exchange rate volatility. Journal of Econometrics 60, 367–391.
Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. Journal of Business and Economic Statistics 14, 45–52.