Granger causality in risk and detection of extreme risk spillover between financial markets

Yongmiao Hong\textsuperscript{a,b,*}, Yanhui Liu\textsuperscript{c}, Shouyang Wang\textsuperscript{c}

\textsuperscript{a} Department of Economics & Department of Statistical Science, Cornell University, 424 Uris Hall, Ithaca, NY 14853-7601, USA
\textsuperscript{b} Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian, China
\textsuperscript{c} Academy of Mathematics and System Science, Chinese Academy of Sciences, Beijing, China

\textbf{A R T I C L E  I N F O}

Article history:
Available online 7 January 2009

\textbf{JEL classification:}
C14
G10

\textbf{Keywords:}
Cross-spectrum
Extreme downside risk
Financial contagion
Granger causality in risk
Nonlinear time series
Risk management
Value at Risk

\textbf{A B S T R A C T}

Controlling and monitoring extreme downside market risk are important for financial risk management and portfolio/investment diversification. In this paper, we introduce a new concept of Granger causality in risk and propose a class of kernel-based tests to detect extreme downside risk spillover between financial markets, where risk is measured by the left tail of the distribution or equivalently by the Value at Risk (VaR). The proposed tests have a convenient asymptotic standard normal distribution under the null hypothesis of no Granger causality in risk. They check a large number of lags and thus can detect risk spillover that occurs with a time lag or that has weak spillover at each lag but carries over a very long distributional lag. Usually, tests using a large number of lags may have low power against alternatives of practical importance, due to the loss of a large number of degrees of freedom. Such power loss is fortunately alleviated for our tests because our kernel approach naturally discounts higher order lags, which is consistent with the stylized fact that today’s financial markets are often more influenced by the recent events than the remote past events. A simulation study shows that the proposed tests have reasonable size and power against a variety of empirically plausible alternatives in finite samples, including the spillover from the dynamics in mean, variance, skewness and kurtosis respectively. In particular, nonuniform weighting delivers better power than uniform weighting and a Granger-type regression procedure. The proposed tests are useful in investigating large comovements between financial markets such as financial contagions. An application to the Eurodollar and Japanese Yen highlights the merits of our approach.

© 2009 Published by Elsevier B.V.

1. Introduction

Controlling and monitoring financial risk have recently received increasing attention from business practitioners, policy makers and academic researchers. For financial risk management and investment/portfolio diversification, it is important to understand the mechanism of how risk spillover occurs across different markets. When monitoring financial risk, the probability of a large adverse market movement is always of greater concern to practitioners (e.g., Bollerslev (2001)). When they occur, extreme market movements imply change hands of a huge amount of capital among market participants, unavoidably leading to bankruptcies due to various downside constraints. Market participants have been always aware of painful experiences when extreme adverse market movements occur, and their aversion to insolvency-type extreme risk is usually very high (e.g., Campbell and Cochrane (1999)). Large market movements have become commonplace nowadays. Examples include the 1994 Mexico Peso Crisis, the 1994 US bond debacle, the 1997–1998 Asian financial crisis, as well as the bankruptcies of the Long Term Capital Management, Enron, and Worldcom.

Most of the existing literature uses volatility to measure risk and focuses on volatility spillover (e.g., 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) and Hong (2001)). Volatility is an important instrument in finance and macroeconomics. However, it can only adequately represent small risks in practice (e.g., Gourieroux and Jasiak (2001, p. 427)). Volatility alone cannot satisfactorily capture risk in scenarios of occasionally occurring extreme market movements. For example, Longin (2000) and Bali (2000) point out that volatility measures based on asset return distributions cannot produce accurate estimates of market risks during volatile periods. Hong et al. (2004, 2007) also find that the innovation distributions have heavier tails when the interest rate market and the foreign...
exchange market have higher volatilities. Moreover, volatility includes both gains and losses in a symmetric way, whereas financial risk is obviously associated with losses but not profits. Also, practical downside constraints often require asymmetric treatment between potential upside and downside risk. Therefore, a more sensible measure of risk should be associated with large losses, or large adverse market movements.

In econometrics and statistics, left tail probabilities are closely related to the likelihoods of extreme downward market movements (e.g., Embrechts et al., 1997). Although not a perfect measure of extreme market risk, Value at Risk (VaR), originally proposed by J.P. Morgan in 1994, has become a standard synthetic measure of extreme market risk (e.g., Duffie and Pan (1997) and Engle and Manganelli (2004)). It measures how much a portfolio can lose within a given time period, with a prespecified probability. It has become an essential part of financial regulations for setting risk capital requirements so as to ensure that financial institutions can survive after a catastrophic event (e.g., the Basel Committee on Banking Supervision (1996); Basel Committee on Banking Supervision (2001)).

Intuitively, VaR measures the total risk in a portfolio of financial assets by summarizing many complex undesired outcomes in a single monetary number. It naturally represents a compromise between the needs of different users. The conceptual simplicity and compromise have made VaR the most popular measure of risk among practitioners in spite of its weakness.2

In this paper, we will develop econometric tools for investigating comovements of large changes between two time series. A leading motivating example is the spillover of extreme down- side movements between financial markets when markets are integrated and suffer from the same global shock, or due to “market contagion”. Using VaR as a measure of extreme downside market risk, we first introduce a new concept of Granger causality in risk, where a large risk is said to have occurred at a prespecified level if actual loss exceeds VaR at the given level. As is well-known, Granger causality (Granger, 1969, 1980) is not a relationship between “causes” and “effects”. Instead, it is defined in terms of incremental predictive ability. This concept is suitable for the purpose of predicting and monitoring risk spillover and can provide valuable information for investment decisions, risk capital allocation and external regulation. We then propose a class of econometric procedures to detect Granger causality in risk between financial markets. Utilizing the most concerned information, it checks whether the past history of the occurrences of large risks in one market has predictive ability for the future occurrences of large risks in another market. We emphasize that the scope of the applicability of the concept of Granger causality in risk is not limited to financial markets and financial positions (e.g., investment portfolios). For example, it can also be used to investigate the spillover of international business cycles, where the understanding of the mechanism of how a large negative shock transmits across different economies is vital to international policy coordination to alleviate its adverse impact on the world economy.

Our proposed procedure has a number of appealing features. First, it checks an increasing number of lags as the sample size grows. This ensures power against a wide range of alternatives of extreme downside risk spillover. Secondly, our frequency domain kernel-based approach naturally discounting higher order lags alleviates the loss of a large number of degrees of freedom and thus enhances good power of the test, which many chi-square tests with a large number of lags (e.g., Box and Pierce’s (1970) portmanteau test) suffer. Downward weighting for higher order lags is consistent with the stylized fact that today’s financial markets are often more influenced by the recent events than by the remote past events. Indeed, simulation shows that nonuniform weighting is more powerful than uniform weighting and a Granger-type regression-based procedure. Finally, our procedure is easy to implement, particularly since the VaR calculation has been available in the standard toolbox of risk managers’ desk.

In Section 2, we describe the concept of Granger causality in risk and discuss its differences from the concepts of Granger causality in mean (Granger, 1969), Granger causality in variance (Granger et al., 1986) and general Granger causality (Granger, 1980). In Section 3, we use a cross-spectral approach to test one-way Granger causality in risk. The kernel method is used. Section 4 develops the asymptotic theory, and Section 5 considers extensions to bilateral Granger causalities in risk. In Section 6, a simulation study examines the finite sample performance of the proposed procedures. Section 7 presents an empirical application to the Eurodollar and the Japanese Yen. It is found that a large downward movement in the Eurodollar Granger-causes a large downward movement in the Japanese Yen, and the causality is stronger for larger movements. On the other hand, Granger causality in risk from the Japanese Yen to the Eurodollar is much weaker or nonexistent. Section 8 concludes the paper. All mathematical proofs are collected in the Appendix. Throughout, Δ and Δ0 denote bounded constants; → almost surely; → P and → F convergences in distribution and in probability respectively; and ||A|| the usual Euclidean norm of A. Unless indicated, all limits are taken as the sample size T → ∞. A GAUSS code for implementing the proposed procedures is available from the authors.

2. Granger causality in risk

2.1. Extreme downside market risk and Value at Risk

For a given time horizon τ and confidence level 1 − α, where α ∈ (0, 1), VaR is defined as the loss over the time horizon τ that is not exceeded with probability 1 − α. Statistically speaking, VaR, denoted by \( V \), is the negative \( α \)-quantile of the conditional probability distribution of a time series \( Y_t \) (e.g., portfolio return), which satisfies the following equation:

\[
P ( Y_t < - V | I_{t-1} ) = \alpha \quad \text{almost surely (a.s.)},
\]

where \( I_{t-1} \) is the information set available at time \( t-1 \). In financial risk management, the left tail probability in (2.1) is usually called the shortfall probability. For notational simplicity, we have suppressed the dependence of \( V \) on level \( α \). In practice, commonly used levels for \( α \) are 10%, 5% or 1%.

To gain insight into VaR from a statistical perspective, we write the time series \( Y_t \) as follows:

\[
Y_t = \mu_t + \sigma_t \epsilon_t, \quad \{ \epsilon_t \} \sim \text{m.d.s.} (0, 1) \quad \text{with conditional CDF } F_{\epsilon_t}(\cdot), \quad (2.2)
\]

where \( \mu_t = \mu_t(I_{t-1}) \) and \( \sigma_t^2 = \sigma_t^2(I_{t-1}) \) are the conditional mean and conditional variance of \( Y_t \) given \( I_{t-1} \) respectively, and \( F_{\epsilon_t}(\cdot) = F_{\epsilon_t}(I_{t-1}) \) is the conditional cumulative distribution function (CDF) of \( \epsilon_t \) given \( I_{t-1} \). By definition, the standardized innovation \( \{ \epsilon_t \} \) is a conditionally homoskedastic martingale difference sequence (m.d.s.) with \( E(\epsilon_t | I_{t-1}) = 0 \) a.s. and \( \text{var}(\epsilon_t | I_{t-1}) = 1 \) a.s., but its higher order conditional moments, such as skewness and kurtosis, may be time-varying. An example is Hansen’s (1994) autoregressive conditional density model where \( \{ \epsilon_t \} \) follows a generalized Student-t distribution with time-varying shape parameters.

From (2.1) and (2.2), we obtain the VaR

\[
V_t = - \mu_t + \sigma_t z_\alpha(\alpha), \quad (2.3)
\]

where \( z_\alpha(\alpha) \equiv z(\alpha-1, \alpha) \) is the left-tailed critical value at level \( \alpha \) of the conditional distribution \( F_{\epsilon_t}(\cdot) \). That is, \( z_\alpha(\alpha) \) satisfies
To develop tests for Granger causality in risk, we first formulate our hypotheses $H^0_1$ versus $H^1_1$ as hypotheses on Granger causality in mean, after a proper transformation of $\{Y_{1t}, Y_{2t}\}$. Define the risk indicator

$$Z_t = 1(Y_{1t} < -V_0), \quad l = 1, 2$$

where $1(\cdot)$ is the indicator function. The indicator $Z_t$ takes value 1 when actual loss exceeds VaR and takes value 0 otherwise. Then $H^0_1$ and $H^1_1$ can be equivalently stated as

$$H^0_1: E(\{Z_{1t}|I_{1(t-1)}\}) = E(\{Z_{1t}|I_{1(t-1)}\}), \quad \text{a.s.}$$

versus

$$H^1_1: E(\{Z_{1t}|I_{1(t-1)}\}) \neq E(\{Z_{1t}|I_{1(t-1)}\}).$$

Thus, Granger causality in risk between $\{Y_{1t}\}$ and $\{Y_{2t}\}$ can be viewed as Granger causality in mean between $\{Z_{1t}\}$ and $\{Z_{2t}\}$. We emphasize that this does not imply that the popular regression-based test proposed by Granger (1969) can be used here, because the risk indicator $Z_t$ has to be estimated, and parameter estimation uncertainty has a nontrivial impact and should be taken care of properly. However, the formulation in (3.1) motivates us to use the cross-spectrum of $\{Z_{1t}, Z_{2t}\}$ below, which is used in Granger (1969) to define the concept of Granger causality in mean.

The cross-spectrum is a natural and powerful tool to investigate Granger causality in mean between two time series (Granger, 1969). To see the implications of $H^0_1$ on the cross-spectrum between $\{Z_{1t}\}$ and $\{Z_{2t}\}$, we first note that for a bivariate covariance-stationary process $\{Z_{1t}, Z_{2t}\}$, the normalized cross-spectral density is

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho(j)e^{-ij\omega}, \quad \omega \in [-\pi, \pi], \quad i = \sqrt{-1},$$

where $\rho(j) = \text{corr}(Z_{1t}, Z_{2t-j})$. Because $\rho(j) \neq \rho(-j), f(\omega)$ is generally complex-valued.

The patterns of $\rho(j)$ and $f(\omega)$ contain valuable information on Granger causality in risk between $\{Y_{1t}\}$ and $\{Y_{2t}\}$. Because $\rho(j)$ and $f(\omega)$ are Fourier transforms of each other, they contain the same information about cross-correlation between $\{Z_{1t}\}$ and $\{Z_{2t}\}$. One could use either $\rho(j)$ or $f(\omega)$ to test $H^0_1$. In this paper, we use $f(\omega)$, which has a number of appealing features, as will be seen below. Under $H^0_1$, $\rho(j) = 0$ for all $j > 0$ and as a consequence, $f(\omega)$ becomes

$$f^0_1(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{0} \rho(j)e^{-ij\omega}, \quad \omega \in [-\pi, \pi].$$

Thus, we can compare $f(\omega)$ and $f^0_1(\omega)$ to test $H^0_1$. Any nontrivial difference between them is evidence against $H^1_1$.

It may be noted that the history of the risk indicators $\{Z_{1s}, s < t\}$ is only a subset of $I_{2(t-1)}$. One can also use other information in $I_{2(t-1)}$ to predict Granger causality in risk. However, the use of the risk indicators $\{Z_{2t}, s < t\}$ is suitable when one is interested in the comovements of extreme changes between two markets. Moreover, a large change in one market may be induced by a change in another market only when the latter exceeds a certain threshold. We note that Bae et al. (2003) also consider the coincidence of extreme return shocks across countries, but their approach is based on the marginal distributions of asset returns. In practice, the spillover in the tails of the conditional distributions could be more relevant and important. A risk manager, for example, may be concerned with whether an incurred loss for a portfolio will exceed a certain prespecified value given that a large loss in another market or another portfolio has occurred.
Both $f(\omega)$ and $f_0^P(\omega)$ are unknown, but they can be estimated consistently by nonparametric methods. The kernel method is the most commonly used in nonparametric spectral estimation. In this paper, we will use the kernel method, which has simple and intuitive appeal in the present context. Most importantly, it naturally provides a flexible downward weighting for higher lag orders, which is consistent with the stylized fact that today’s financial markets are more affected by the recent events than by the remote events and is expected to enhance the power of the proposed procedure.

Suppose

$$V_h(\theta_l) = V_l(l_{h-1} - \theta_l), \quad l = 1, 2,$$

is a parametric VaR model for $V_h$, where $\theta_l$ is an unknown finite-dimensional parameter. There have been many methods to estimate VaR (e.g., Chernozhukov and Umantsev (2001), Engle and Manganelli (2004) and Jorion (2000)). Examples are historical simulation methods, Hansen’s (1994) autoregressive conditional density model, Morgan’s (1996) RiskMetrics, and Engle and Manganelli’s (2004) conditional autoregressive VaR (CVAiR) models. Suppose further we have a random sample $\{Y_t, Y_{a+1}, \ldots, Y_t\}$ of size $T$, and an estimator $\hat{\theta}_l$. Put

$$\hat{Z}_d = Z_{d}(\hat{\theta}_l), \quad l = 1, 2,$$

where $Z_{d}(\theta_l) = 1|Y_t < -V_h(\theta_l)|$. Then we can define the sample cross-covariance function between $\{\hat{Z}_1\}$ and $\{\hat{Z}_2\}$,

$$\hat{C}(j) = \left\{ \begin{array}{ll}
T^{-1} \sum_{t=T-j}^{T} (\hat{Z}_{1t} - \hat{\alpha}_1)(\hat{Z}_{2t-j} - \hat{\alpha}_2), & 0 \leq j \leq T - 1, \\
T^{-1} \sum_{t = j}^{T-j} (\hat{Z}_{1t+j} - \hat{\alpha}_1)(\hat{Z}_{2t} - \hat{\alpha}_2), & 1 - T < j < 0,
\end{array} \right. \tag{3.3}$$

where $\hat{\alpha}_1 = T^{-1} \sum_{t=1}^{T} \hat{Z}_{1t}$. The sample cross-correlation function between $\{\hat{Z}_1\}$ and $\{\hat{Z}_2\}$ is

$$\hat{\rho}(j) = \hat{C}(j)/\hat{S}_1 \hat{S}_2, \quad j = 0, \pm 1, \ldots, \pm (T - 1),$$

where $\hat{S}_j = \hat{\alpha}_1(1 - \hat{\alpha}_1)$ is the sample variance of $\{\hat{Z}_1\}$. The kernel estimators for the cross-spectral densities $f(\omega)$ and $f_0^P(\omega)$ can be given as follows:

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{t=1}^{T-j} k(j/M) \hat{\rho}(j) e^{-ij\omega}, \tag{3.4}$$

$$\hat{f}_0^P(\omega) = \frac{1}{2\pi} \sum_{j=1}^{0} k(j/M) \hat{\rho}(j) e^{-ij\omega}. \tag{3.5}$$

To compare $\hat{f}(\omega)$ and $\hat{f}_0^P(\omega)$, we use the quadratic form

$$l^2(\hat{f}, \hat{f}_0^P) = 2\pi \int_{-\pi}^{\pi} \left| \hat{f}(\omega) - \hat{f}_0^P(\omega) \right|^2 d\omega = \sum_{j=1}^{T-j} k^2(j/M) \hat{\rho}^2(j). \tag{3.6}$$

where the second equality follows by Passerval’s identity. We need not calculate numerical integrations over frequency $\omega$. Our test statistic for $l^2_{20}$ versus $l^2_1$ is a standardized version of the quadratic form:

$$Q_1(M) = \left[ \sum_{j=1}^{T-j} k^2(j/M) \hat{\rho}^2(j) - C_{11}(M) \right]/D_{11}(M) \frac{1}{2}, \tag{3.7}$$

where the centering and standardization constants are

$$C_{11}(M) = \sum_{j=1}^{T-j} (1 - j/T) k^2(j/M),$$

$$D_{11}(M) = \sum_{j=1}^{T-j} (1 - j/T) (1 - (j + 1)/T) k^4(j/M).$$

The factors $(1 - j/T)$ and $(1 - (j + 1)/T)$ are finite sample corrections. They could be replaced by 1. Both $C_{11}(M)$ and $D_{11}(M)$ are approximately the mean and variance of the quadratic form $\sum_{j=1}^{T-j} \hat{S}_j \hat{S}_j$. What $Q_1(M)$ checks here is not the original hypothesis but only its necessary condition. However, it captures the most important information to deliver a feasible test.

To compute $Q_1(M)$, one can use the truncated kernel

$$k_T(\omega) = \mathbf{1}(|\omega| \leq 1), \tag{3.8}$$

where $\mathbf{1}(\cdot)$ is the indicator function. This yields the following test statistic

$$Q_{\text{STRM}}(M) = \left[ \sum_{j=1}^{M} \hat{\rho}^2(j) - M \right] / (2M)^{1/2}. \tag{3.9}$$

This test gives an equal weight to each of the $M$ lags. It is essentially equivalent to a Granger-type procedure based on the following auxiliary regression

$$\hat{Z}_d = \alpha_0 + \sum_{j=0}^{M} \alpha_j \hat{Z}_{d-j} + u_1, \tag{3.10}$$

which checks whether the coefficients $\{\alpha_j\}_{j=1}^{M}$ are jointly zero. This is similar to Pierce and Haugh’s (1977) residual-based test for Granger causality in mean. Here we need not include the lagged variables of $\hat{Z}_1$ because $\{\hat{Z}_1\}$ is a sequence of i.i.d. Bernoulli random variables under the null hypothesis. For the estimated $\hat{Z}_1$, (3.10) holds asymptotically, which is almost satisfied in practical applications where usually large samples are used to estimate the parameters in VaR model. Granger (1969) proposes a popular test for causality in mean based on a regression similar to (3.10), with a fixed but arbitrarily large $M$. To ensure that the regression test has power against a large class of alternatives, we let $M$ grow with the sample size $T$ properly. This delivers a $R^2$-based test statistic

$$Q_{\text{REG}} = (R^2 - M)/(2M^{1/2}), \tag{3.11}$$

where $R^2$ is the centered squared multi-correlation coefficient from the regression in (3.10). We may view this test as a generalized version of Granger’s (1969) test for $H_0^0$. This procedure is simple and intuitive. It could be shown that $Q_{\text{REG}}(M)$ is asymptotically equivalent to $Q_{\text{STRM}}(M)$ under $H_0^0$. When $M$ is large, however, both $Q_{\text{STRM}}(M)$ and $Q_{\text{REG}}(M)$ may not deliver good power against alternatives of practical importance. As a stylized fact, today’s financial markets are often more influenced by the recent events than by the remote events, which implies that the dependence of $Z_1$ on the $Z_{2t-j}$ will eventually diminish as lag order $j$ increases. Consequently, it is more efficient to discount higher order lags. The most commonly used kernels are downward weighting for higher order lags. Examples are the Bartlett, Daniell, Parzen, and Quadratic-Spectral kernels. In contrast, the $Q_{\text{STRM}}(M)$ and $Q_{\text{REG}}(M)$ tests are not fully efficient when $M$ is large. See Sections 4 and 6 for more discussion and simulation studies.

---

3 We can replace $\hat{\alpha}_1$ with $\alpha$. This does not affect the asymptotic distribution of the proposed test statistic under $H_0^0$.

4 For $k(z) = 1(|z| \leq 1), \mathbf{H}_1(\mathbf{M}) = (M + 1)/2T$ and $D_{11}(M) = 2[M(1 + (M + 1)/2)]T/(M + 1)/2T = c_1^2$. Under suitable conditions on $M$, we can conveniently approximate $C_{11}(M)$ and $D_{11}(M)$ by $M$ and $2M$ respectively.

5 Engle (1982), in the context of testing the existence of ARCH effects, also considers linearly declining weighting for lag order (which is equivalent to the Bartlett kernel) to increase the power of his LM test. Here, we allow for a more general flexible weighting and allow $M$ to grow with $T$. 

---

The key step in implementing our procedure lies in the VaR estimation. This is relatively simple for practitioners in the real financial industry, because VaR can be easily calculated by most standard risk management softwares. Furthermore, VaR can be set not only at the commonly used 1% or 5% level, but also at any level which the investors or risk managers may be interested in. For example, investors often impose the stop-loss rule for their portfolio investments. Our procedure can be applied to investigate risk slippover at the stop-loss level.

4. Asymptotic theory

We now derive the limit distribution of the $Q_1(M)$ test under $H_0^1$. Its derivation is complicated by the fact that we do not observe the true parameter values ($\theta^0$) and have to estimate them. Parameter estimation uncertainty in $[\hat{\theta}]$ has to be dealt with properly, as is encountered by Engle and Manganelli (2004), where the interest is in testing the adequacy of an univariate VaR model, and parameter estimation uncertainty has a nontrivial impact on the limit distribution of the test statistic, which complicates the construction of their test statistic. In particular, it involves nonparametric estimation of the conditional probability density of the underlying process.

Our nonparametric cross-spectral approach fortunately enables us to get rid of the impact of $[\hat{\theta}]$ asymptotically. Intuitively, $[\hat{\theta}]$ converges to $\theta^0$ faster than the nonparametric estimators $f(\omega)$ and $f_1^0(\omega)$ respectively. As a consequence, the limit distribution of $Q_1(M)$ is solely determined by the kernel estimators $f(\omega)$ and $f_1^0(\omega)$. One can proceed as if $\theta^0$ were known and equal to $\hat{\theta}$. Thus, replacing $\theta^0$ with $\hat{\theta}$ has no impact on the limit distribution of $Q_1(M)$. This greatly simplifies the construction and implementation of our test because we need not know the asymptotic expansion of $[\hat{\theta}]$ and can choose any convenient $\sqrt{T}$-consistent estimator.

To justify the above heuristics, we impose a set of regularity conditions on the data generating process ($X_t$), the VaR models $V_t(\theta)$, the parameter estimators $\hat{\theta}$ and the kernel function $k(\cdot)$.

**Assumption 1.** For $t = 1, 2, \ldots$, $\mathbb{E}[X_t] = \rho_t$ where $\rho_t$ is a positive integer, $t = 1, 2, \ldots$, $V_t(\theta) \equiv V_{t-1}(\theta)$ is a VaR model at level $\alpha$ in $(0,1)$ such that (i) for each $\theta \in \Theta$, $V_t(\theta)$ is a measurable function of $X_{t-1}$; (ii) with probability 1, $V_t(\cdot)$ is twice continuously differentiable with respect to $\theta_t \in \Theta$, with

$$\lim_{t \to \infty} \sum_{i=1}^{T-1} E \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta} F_t \left[ -V_t(\theta) \right] \right\|^2 < \infty$$

and

$$\lim_{t \to \infty} \sum_{i=1}^{T-1} E \sup_{\theta \in \Theta} \left\| \frac{\partial^2}{\partial \theta \partial \theta} F_t \left[ -V_t(\theta) \right] \right\|^2 < \infty$$

**Assumption 2.** For $t = 1, 2$, there exists some $\theta^0 \in \Theta$ such that (i) $P[X_0 < -V_t(\theta^0)]_{[t-1]} = \alpha$ a.s.; (ii) the risk indicator $Z_{t}(\theta^0) = 1[X_0 < -V_t(\theta^0)]$, $s < t$ depends on an arbitrarily long but finite length of the current and past history of $[Z_{t-1}(\theta^0)] = 1[X_0 < -V_t(\theta^0)]$.

**Assumption 3.** For $t = 1, 2$, there exists some $\theta_t^0 \in \Theta$ such that (i) $\eta_t = p \lim_{t \to \infty} \eta_t^0$ under the null hypothesis of interest.

**Assumption 4.** $T^2(\hat{\theta} - \theta^*) = O_p(1)$ for $l = 1, 2$, where $\eta_t^0 = \theta_t^0$ under the null hypothesis of interest.

**Assumption 5.** Put $S_t(\theta) = [S_0(\theta), S_2(\theta)]'$ and $Z_t(\theta) = [Z_0(\theta), Z_2(\theta)]'$, where $S_0(\theta) = \frac{\partial}{\partial \theta} F_t \left[ -V_t(\theta) \right]$. Then $[S_t(\theta^0), Z_t(\theta^0)]'$ is a fourth order stationary process such that

(i) $\sum_{k=0}^{\infty} \| F(j) \| \leq \Delta$, where $F(j) \equiv \text{cov}([S_t(\theta^0), Z_t-j(\theta^0)])$;

(ii) $\sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \| \lambda_{k,m,l} \| \leq \Delta$, where $\lambda_{k,m,l}$ is the fourth order cumulant of the joint distribution of $[S_t(\theta^0)' - ES_t(\theta^0)', Z_{t-j}(\theta^0)' - EZ_{t-j}(\theta^0)', S_{t-k}(\theta^0)' - ES_{t-k}(\theta^0)', Z_{t-k}(\theta^0)' - EZ_{t-k}(\theta^0)'].$

**Assumption 6.** $k: \mathbb{R} \to [-1,1]$ is a symmetric function that is continuous at 0 and all points except a finite number of points on $\mathbb{R}$, with $k(0) = 1$ and $\int_{-\infty}^{\infty} k^2(\omega) d\omega < \infty$.

**Assumption 1** is a standard regularity condition on the bivariate data generating process for $[Y_t, Y_{t-1}]$. We allow for some covariance non-stationary processes $[Y_t]$. An example is the integrated GARCH process (Engle and Bollerslev, 1986), which is strictly stationary but not covariance-stationary (Nelson, 1991).

**Assumption 2** provides regularity smoothness and moment conditions on the VaR models $V_t(\theta)$. There are various VaR models in the literature (e.g., Chernozhukov and Umantsev (2001), Duffie and Pan (1997), Engle and Manganelli (2004) and Jorion (2000)). Some of them essentially specify the whole conditional distribution of $Y_t$ while others only specify the left tail of the conditional distribution. Examples of the former include Morgan’s (1996) RiskMetrics, GARCH models with i.i.d. innovations, Hansen’s (1994) autoregressive conditional density model with a generalized Student-t-distribution, and examples of the latter include Engle and Manganelli’s (2004) CAViaR models.

**Assumption 3** imposes some conditions on the VaR models which will be required only under $H_0^1$. **Assumption 3(i) is the condition on the adequacy of VaR models, which can be checked using the methods of Chernozhukov and Umantsev (2001), Christoffersen et al. (2001) and Engle and Manganelli (2004).** **Assumption 3(ii) allows for the possibility that under $H_0^1$, although $[Z_{t-1}(\theta_t^0)] \leq \alpha$ does not affect $[Z_{t-1}(\theta_t^0)]$ and $[Z_{t-1}(\theta_t^0)]$ may depend on the current and past history of $[Z_{t-1}(\theta_t^0)]$, $s < t$. In other words, there may exist instantaneous Granger causality between $Z_{t-1}(\theta_t^0)$ and $Z_{t-1}(\theta_t^0)$, and/or Granger causality from $[Z_{t-1}(\theta_t^0)]$ to $[Z_{t-1}(\theta_t^0)]$ under $H_0^1$. For simplicity, **Assumption 3(iii) assumes that $[Z_{t-1}(\theta_t^0)]$ depends on an arbitrarily long but finite history of $[Z_{t-1}(\theta_t^0)]$, $s < t$. It is possible to allow $Z_{t-1}(\theta_t^0)$ to depend on the entire past history of $[Z_{t-1}(\theta_t^0)]$, $s < t$, with a suitable rate condition on the dependence of $Z_{t-1}(\theta_t^0)$ on the history of $[Z_{t-1}(\theta_t^0)]$, $s < t$, and the test statistic and its limit distribution remain unchanged. However, we do not consider this possibility here for simplicity.**

**Assumption 4** does not require any specific estimation method. In particular, $\hat{\theta}$ need not be asymptotically most efficient; any $\sqrt{T}$-consistent estimator of $\theta^0$ suffices under $H_0^1$. An example is Engle and Manganelli’s (2004) regression quantile estimator. We do not require parameter estimation consistency under the alternative $H_1^1$. Thus, the probability limit $\theta^0$ may not coincide with $\theta^0$ under $H_0^1$. Moreover, we need not know the asymptotic expansion of $[\hat{\theta}]$. These features greatly simplify the construction and implementation of the proposed tests.

**Assumption 5** is a regularity condition on the serial dependence of the process $[S_t(\theta^0), Z_t(\theta^0)]$. Under $H_0^1$, we have $[Z_{t-1}(\theta_t^0)] = [Z_{t-1}(\theta_t^0)]$. Thus, $[Z_{t-1}(\theta_t^0)]$ is an i.i.d. Bernoulli(α) sequence and $[Z_{t-1}(\theta_t^0)]$ is independent of $[Z_{t-1}(\theta_t^0)]$, $s < t$. However, the derivative $\frac{\partial}{\partial \theta} [E[Z_{t-1}(\theta_t^0)]]_{[t-1]} = S_{t-1}(\theta_t^0)$ generally depends on $t_1$ even under $H_0^1$. We note that the fourth order cumulant condition in **Assumption 5**(ii) is a standard assumption in time series analysis (e.g., HANNAN (1970)).

Finally, **Assumption 6** is a standard regularity condition on the kernel $k(\cdot)$. Among other things, the condition that $k(0) = 1$ ensures that the asymptotic biases of the kernel-based cross-spectral density estimators $\hat{f}(\omega)$ and $\hat{f}_1(\omega)$ vanish as sample size
$T \to \infty$. Most commonly used kernels satisfy Assumption 6 (see, e.g., Priestley [1981]).

We now state the asymptotic normality of $Q_t(M)$ under $\mathbb{H}_0^1$.

**Theorem 1.** Suppose Assumptions 1–6 hold and $M = cT^{\alpha}$, where $0 < c < \infty$, $0 < v < \frac{1}{2}$, $v < \min\left(\frac{3}{2}, \frac{3}{2}\right)$ if $d \equiv \max(d_1, d_2) > 2$, and $d_t$ is the dimension of $\theta_t$. Then $Q_t(M) \to^d N(0, 1)$ under $\mathbb{H}_0^1$.

The condition that $v < \min\left(\frac{3}{2}, \frac{3}{2}\right)$ if $d > 2$ is sufficient but may not be necessary. This is imposed to simplify the treatment of the impact of parameter estimation uncertainty in $\hat{\theta}_t$. It could be weakened at a cost of a more tedious proof. In the present context, the technical treatment of parameter estimation uncertainty is not trivial, because the risk indicator $Z_0(\hat{\theta}_t)$ is not differentiable with respect to $\theta_t$. From the proof of Theorem 1 (see Theorem A.1 in the Appendix), we find that parameter estimation uncertainty in $\hat{\theta}_t$ has no impact on the limit distribution of $Q_t(M)$. This occurs because $\hat{\theta}_t$ converges to $\theta_0^1$ faster than the kernel cross-spectral estimators $\hat{f}(\omega)$ and $\hat{f}_2(\omega)$ to $f(\omega)$ and $f_2(\omega)$ respectively.

To understand the intuition why $Q_t(M)$ is asymptotically $N(0, 1)$, we consider the $Q_{\text{TRUN}}(M)$ test that is based on the truncated kernel in (3.8). First, suppose $\theta_0^1$ were known. Then as $T \to \infty$, we have $\sqrt{T}\hat{\rho}(j) \to^d N(0, 1)$ for each lag $j > 0$ and $\cos(\sqrt{T}\hat{\rho}(j)) \to 0$ for any $j \neq j_1$ under $\mathbb{H}_0^1$. Consequently, $\sum_{j=1}^M T\hat{\rho}^2(j)$, being the sum of $M$ asymptotically independent $\chi^2_d$ random variables, are asymptotically distributed as $\chi^2_d$. By the well-known approximate distribution of $\chi^2_d$ when $M$ is large, we obtain the asymptotic normality of $Q_{\text{TRUN}}(M)$. The impact of parameter estimation uncertainty in $\hat{\theta}_t$ is at most an finite adjustment, which is asymptotically negligible as $M$ becomes large. This intuition remains valid for nonuniform kernels.

To investigate the asymptotic behavior of the $Q_t(M)$ test under $\mathbb{H}_1^1$, we impose a condition on the cross-correlation $\rho(\cdot)$ and a fourth order cumulant condition.

**Assumption 7.** Let $\rho(\cdot) = \text{cov}(Z_t(\theta_0^1), Z_{t-l}(\theta_0^1))$. Then (i) $\sum_{j=1}^M \rho^2(j) < \infty$; (ii) $\sum_{j=1}^M \sum_{k=1}^M \sum_{l=1}^M |\kappa_1(j, k, l)| < \infty$, where $\kappa_1(j, k, l)$ is the fourth order cumulant of the joint distribution of $Z_t(\theta_0^1), Z_{t-l}(\theta_0^1), Z_{t-l-1}(\theta_0^1), Z_{t-l-2}(\theta_0^1)$.

Assumption 7(i) implies that the dependence of $Z_t(\theta_0^1)$ on $Z_{t-l}(\theta_0^1)$, $s < t$ decays to zero at a suitable rate, but it still allows for certain strongly cross-dependent processes whose cross-correlation decays to zero at a slower hyperbolic rate. We do not impose any condition on the dependence of $Z_{t-l}(\theta_0^1)$ on $Z_t(\theta_0^1)$, $s \leq t$, because we only check the one-way Granger causality from $Z_t(\theta_0^1)$ to $Z_{t-l}(\theta_0^1)$ with respect to $l_{t-l}$.

**Theorem 2.** Suppose Assumptions 1–7 hold and $M = cT^{\alpha}$ for $0 < c \leq \infty$ and $0 < v < 1$. Then $(M^2/T)Q_t(M) \to^d N(0, 1)$ under $\mathbb{H}_1^1$.

Thus, for any sequence of constants, $K_t = o(T^{1-\frac{1}{2}})$, we have $P(Q_t(M) > K_t) \to 1$ whenever $\rho(\cdot) \neq 0$ for some $j > 0$. In other words, the $Q_t(M)$ test has asymptotic unit power at any given significance level whenever $\rho(\cdot) \neq 0$ for some $j > 0$. Because $Q_t(M) \to \pm \infty$ whenever $\rho(\cdot) \neq 0$ for some $j > 0$, upper-tailed $N(0, 1)$ critical values are appropriate. For example, the critical value at the 5% significance level is 1.65. We note also that the condition on $M$ under $\mathbb{H}_1^1$ is weaker than that under $\mathbb{H}_0^1$.

Hong (1996) shows that over a class of kernels

$$K(\tau) = \left\{ k(\cdot) : k(0) = 1, \int_{-\infty}^{\infty} k(z)dz < \infty, \right\}$$

$$k_2 = \lim_{z \to 0} \frac{1 - k(z)}{z^2} \in (0, \infty)$$

which includes the Parzen and Quadratic-Spectral kernels (but not the Bartlett kernel), the Daniell kernel $k_2(\cdot) = \sin(\pi z)/\pi z$ minimizes $\int_{-\infty}^{\infty} k^2(z)dz$. Using this result, it can be shown that the Daniell kernel $k_2(\cdot)$ maximizes the asymptotic power of $Q_t(M)$ in terms of Bahadur’s (1960) asymptotic efficiency criterion. Of course, the relative efficiency of nonuniform kernels in $K(\tau)$ is very close to each other. This implies that the choice of kernel $k(\cdot)$ is not important, provided the truncated (i.e., uniform) kernel in (3.8) is not used. Intuitively, the cross-dependence $|\rho(\cdot)|$ decays to zero as $j \to \infty$ under Assumption 7, so it is more efficient to discount higher order lags than to put an equal weight for each lag.

5. Bilateral Granger causality in risk

We now extend our analysis to two-way Granger causalities in risk. We consider the hypothesis that neither $\{Y_{t+1}\}$ nor $\{Y_{t+1}\}$ Granger-causes each other in risk at level $\alpha$ with respect to $k_{l-1}$. The hypotheses of interest are

$$\mathbb{H}_0^2 : P \{ Y_{t+1} < -V_t | l_{t+1-1} \}$$

$$= P \{ Y_{t+1} < -V_t | l_{t+1-1} \} \text{ a.s. for both } l = 1, 2$$

versus

$$\mathbb{H}_1^2 : P \{ Y_{t+1} < -V_t | l_{t+1-1} \}$$

$$\neq P \{ Y_{t+1} < -V_t | l_{t+1-1} \} \text{ for at least one } l.$$ (5.2)

Using the risk indicator $Z_{t+1}$, we can write these hypotheses as

$$\mathbb{H}_0^2 : E \{ Z_{t+1} | l_{t+1-1} \} = E \{ Z_{t+1} | l_{t+1-1} \} \text{ a.s. for both } l = 1, 2$$

versus

$$\mathbb{H}_1^2 : E \{ Z_{t+1} | l_{t+1-1} \} \neq E \{ Z_{t+1} | l_{t+1-1} \} \text{ for at least one } l,$$ where $l = 1, 2$.

Under $\mathbb{H}_0^2$, the past information of one series is not useful for predicting the risk of the other series, and their cross-spectral density $f(\omega)$ becomes a flat spectrum:

$$f_2^0(\omega) = \frac{1}{2\pi} \rho(0), \quad \omega \in [-\pi, \pi]$$

where $\rho(0)$ is nonzero when there exists instantaneous causality between $Z_{t+1}$ and $Z_{t+1}$. A consistent estimator for $f_2^0(\omega)$ is

$$\hat{f}_2^0(\omega) = \frac{1}{2\pi} \hat{\rho}(0), \quad \omega \in [-\pi, \pi].$$

Our test statistic for $\mathbb{H}_0^2$ versus $\mathbb{H}_1^2$ is a properly standardized version of a quadratic form between $\hat{f}(\omega)$ and $\hat{f}_2^0(\omega)$:

$$Q_2(M) = \left[ T \sum_{j=1}^M \hat{k}(j) \hat{\rho}(j) - C_{2T}(M) \right]/\{D_{2T}(M)\}^2,$$ (5.3)

where the centering and scaling factors are

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

$$D_{2T}(M) = 2 \left[ 1 + \frac{\hat{\rho}^4(0)}{M} \sum_{j=1}^M \{1 - (j/T)\} \{1 - (j + 1)/T\} \hat{k}(j) \right].$$

Note that $D_{2T}(M)$ involves the cross-correlation estimator $\hat{\rho}(0)$, which has taken into account the possible instantaneous correlation between $Z_{t+1}$ and $Z_{t+1}$ under $\mathbb{H}_0^2$. The $Q_2(M)$ statistic is asymptotically $N(0, 1)$ under $\mathbb{H}_0^2$, as is stated below.
Theorem 3. Suppose Assumptions 1–3(i) and 4–6 hold, and $M = cT^ν$, where $0 < ν < \frac{1}{2}$, $ν < \max\left(\frac{2}{2-ν}, \frac{2}{2-ν-2}\right)$ if $d = \max(d_1, d_2) > 2$, and $d_i$ is the dimension of $θ_i$, $i = 1, 2$. Then $Q_2(M) \rightarrow d N(0, 1)$ under $ι^2$.

We do not need Assumption 3(ii) here, because under $ι^3$, neither $Z_1t$ nor $Z_2t$ Granger-causes each other with respect to $l_{t-1}$. Nevertheless, we allow for instantaneous Granger causality between $Z_1t$ and $Z_2t$ under $ι^3$.

To study the asymptotic behavior of the $Q_2(M)$ test under $ι^4$, we strengthen Assumption 7 slightly to cover two-way cross-correlations.

Assumption 8. (i) $\sum_{j=-\infty}^{\infty} \rho^2(j) < \infty$ and (ii) $\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |k_j(k, j, l)| < \infty$, where $\rho(j)$ and $k_j(k, j, l)$ are as in Assumption 7.

Theorem 4. Suppose Assumptions 1, 2, 4–6 and 8 hold, and $M = cT^ν$ for $0 < c < \infty$ and $0 < ν < 1$. Then $\left\{\sqrt{\frac{1 + \rho^2(0)T^{-\frac{1}{2}}}{2}} Q_2(M) \rightarrow \left\{2 \sum_{j=-\infty}^{\infty} k^4(z)dJ(z)\right\} \sum_{j=1}^{\infty} \rho^2(j) \right\}$ under $ι^5$.

Thus, whenever there exists Granger causality in risk between $\{Y_{1t}\}$ and $\{Y_{2t}\}$ with respect to $l_{t-1}$ such that $\rho(j) \neq 0$ for some $j \neq 0$, the $Q_2(M)$ test will have asymptotic unit power at any given significance level. Note that the asymptotic variance depends on $\rho(0)$, which arises due to the presence of instantaneous risk spillover under $ι^5$.

6. Finite sample performance

We now examine the finite sample performance of the proposed tests via simulation. For the sake of space, we focus on the $Q_1(M)$ test in (3.7); $Q_2(M)$ in (5.3) is performed to expect similar results.

Throughout this section, we work with the following data generating process (DGP):

$$ Y_t = β_1 Y_{1t-1} + β_2 Y_{2t-1} + u_t, \quad l = 1, 2, $$

$$ u_t = σ_t ε_t, $$

$$ σ_t^2 = σ_0^2 + γ_1 σ_{t-1}^2 + γ_2 u_{t-1}^2 + γ_3 u_{2t-1}^2, \quad ε_t \sim \text{N}(0, 1). $$

To investigate both the size and power of our test, we consider the following cases under (6.1):

**NULL** [No Granger Causality in Risk]:

$$ \{β_{11}, β_{12}, γ_{10}, γ_{11}, γ_{12}, γ_{13}\} = (0.5, 0, 0.1, 0.6, 0.2, 0), $$

$$ \{β_{21}, β_{22}, γ_{20}, γ_{21}, γ_{22}, γ_{23}\} = (0.5, 0.1, 0.6, 0, 0.2). $$

**ALTER1** [Granger Causality in Risk from Mean]:

$$ \{β_{11}, β_{12}, γ_{10}, γ_{11}, γ_{12}, γ_{13}\} = (0.5, 0.2, 0.1, 0.6, 0.2, 0), $$

$$ \{β_{21}, β_{22}, γ_{20}, γ_{21}, γ_{22}, γ_{23}\} = (0.5, 0.1, 0.6, 0.2, 0). $$

**ALTER2** [Granger Causality in Risk from Variance]:

$$ \{β_{11}, β_{12}, γ_{10}, γ_{11}, γ_{12}, γ_{13}\} = (0.5, 0.1, 0.6, 0.2, 0.2, 0), $$

$$ \{β_{21}, β_{22}, γ_{20}, γ_{21}, γ_{22}, γ_{23}\} = (0.5, 0, 0.1, 0.5, 0.2, 0.2). $$

Under NULL, there is no Granger causality in risk between $\{Y_{1t}\}$ and $\{Y_{2t}\}$ with respect to $l_{t-1}$. This allows us to examine the size of the $Q_1(M)$ test in finite samples. On the other hand, there exists Granger causality in risk under both ALTER1 and ALTER2, but with different sources of spillover. Under ALTER1, there exists Granger causality in mean but not in any higher order conditional moments. Under ALTER2, there exists Granger causality in variance, but not in mean and other higher order conditional moments. Spillovers in mean and in variance are most commonly studied in the literature; ALTER1 and ALTER2 allow us to investigate how well our test can detect risk spillover from these sources.

Recent empirical studies (e.g., Hansen (1994), Harvey and Siddique (1999, 2000), Jondeau and Rockinger (2003) and Hong et al. (2004, 2007)) find evidence of time-varying skewness and kurtosis for various financial time series. It is therefore conceivable that Granger causality in risk between financial markets may be caused by comovements in conditional skewness or in conditional kurtosis. Indeed, skewness and kurtosis are closely related to the left tail of the innovation distribution, or extreme downside risk. To investigate Granger causality in risk from higher order conditional moments, we generate data using Hansen’s (1994) autoregressive conditional density model, which is embedded in (6.1) with the innovations $\{ε_{kt}\}$ following a generalized Student-t-distribution with time-varying shape parameters. More specifically, Hansen’s (1994) generalized t density is given by

$$ g_ε(ε | λ, ν) = \frac{bc \left\{1 + \frac{1}{\eta - 2} \left(\frac{b^ε + a}{1 - λ}\right)\right\}^{-\frac{1}{2}} \Gamma((ν + 1)/2)}{2 \sqrt{π(ν - 2)\Gamma(ν/2)}} $$

where

$$ a = 4λc(η - 2)/(η - 1), \quad b^2 = 1 + 3λ^2 - a^2, $$

$$ c = \frac{Γ((ν + 1)/2)}{\sqrt{π(ν - 2)\Gamma(ν/2)}} $$

Here, $λ$ measures skewness and $ν$ is the degree of freedom parameter. They characterize asymmetry and fat-tailedness of $ε_{kt}$ respectively. This density is well-defined for $−1 < λ < 1$ and $2 < ν < ∞$ and encompasses a variety of popular densities. For instance, if $λ = 0$, the generalized Student-t-distribution reduces to the standard Student-t-distribution. If in addition $ν = ∞$, it further reduces to a normal density. We specify that $|ε_{kt}|$ follows a generalized Student-t-distribution with time-varying parameters $λ_{kt}$ and $ν_{kt}$, where the dynamics of $λ_{kt}$ and $ν_{kt}$ follow the specification of Jondeau and Rockinger (2003):

$$ \lambda_{kt} = \frac{1 - \exp(\tilde{λ}_{kt})}{1 + \exp(\tilde{λ}_{kt})}, $$

$$ \eta_{kt} = \frac{4 + 1 + \exp(\tilde{η}_{kt})}{4 + 1 + \exp(\tilde{η}_{kt})}, $$

where

$$ \tilde{λ}_{kt} = δ_0 + δ_1 u_{t-1} + δ_2 u_{2t-1} + δ_3 \tilde{λ}_{1t-1} + δ_4 \tilde{λ}_{2t-1}, $$

$$ \tilde{η}_{kt} = η_0 + η_1 u_{t-1} + η_2 u_{2t-1} + η_3 η_{1t-1} + η_4 η_{2t-1}. $$

We use the following parameter combinations:

**ALTER3** [Granger Causality in Risk from Skewness and Kurtosis]:

$$ \{β_{11}, β_{12}, γ_{10}, γ_{11}, γ_{12}, γ_{13}\} = (0.3, 0.1, 0.6, 0.2, 0), $$

$$ \{β_{21}, β_{22}, γ_{20}, γ_{21}, γ_{22}, γ_{23}\} = (0.5, 0.1, 0.6, 0.2, 0), $$

$$ \{δ_{10}, δ_{11}, δ_{12}, δ_{13}, δ_{14}, η_0, η_1, η_2, η_3, η_4\} = (-0.2, 1, -0.9, -0.2, 1, -5, 0, -0.9), $$

$$ \{δ_{20}, δ_{21}, δ_{22}, δ_{23}, η_0, η_1, η_2, η_3, η_4\} = (-0.2, 0.1, 0.0, -0.2, 0.1, 0.0). $$

Under ALTER3, there exists no Granger causality in mean nor in variance, but there exists Granger causality in risk, due to the causality in skewness and in kurtosis from $Y_{2t}$ to $Y_{1t}$ with respect to $l_{t-1}$. To our knowledge, the financial econometric literature has been focusing on spillover in mean and in variance; no study on
spillover in skewness and kurtosis was not previously available in the literature.

For all data generating processes, we consider three sample sizes: $T = 500, 1000, 2000$, which covers a rather wider range of lag orders for the sample sizes considered here. For each $T$, we first generate $T + 500$ observations using the GAUSS Window Version random number generator on a personal computer and then discard the first $500$ to reduce the possible effect of the chosen starting values $(h_0, \lambda_{10}^*, \gamma_{10}^*) = (1/(1 - \gamma_{21} - \gamma_{22}), 0, 2, 4, 1)$. We choose two shortfall probabilities or risk levels: $\alpha = 10\%$ and $5\%$. To compute our test statistics, we use the Daniell kernel $k(z) = \sin \pi z/\pi z, z \in (-\infty, \infty)$, which enjoys some optimal power property (see Section 4).\(^6\) For comparison, we also consider the truncated kernel-based test $Q_{\text{TRUN}}(M)$ in (3.7) and the Granger-type regression test (1969) $Q_{\text{REG}}(M)$ in (3.11). To examine the impact of the choice of the lag order $M$, we consider $M = 5, 10, 15, 20, 25$ and $30$, which covers a rather wide range of lag orders for the sample sizes considered here. For data generated from each of the DGPs 1–4, we use the Quasi-MLE to estimate the unknown parameters in each individual null model:

\[ Y_{t} = \beta_{0}Y_{t-1} + \varepsilon_{t}, \quad I = 1, 2, \]
\[ u_{t} = \sigma_{\theta}e_{t}, \]
\[ \sigma_{u}^{2} = \gamma_{0} + \gamma_{1}\sigma_{u}^{2}h_{t} + \gamma_{2}\varepsilon_{t-1}^{2}, \]
\[ \varepsilon_{t} \sim \text{i.i.d. N}(0, 1). \]

The BHNN algorithm is used. This delivers $\sqrt{T}$-consistent estimators under NULL (Bollerslev and Wooldridge, 1992; Lee and Hansen, 1994; Lumsdaine, 1996)).

Table 1 reports the rejection rates of the $Q_{\text{IDAN}}(M)$ in (3.7) with Daniell kernel, $Q_{\text{TRUN}}(M)$ in (3.9) and $Q_{\text{REG}}(M)$ tests in (3.11) at the 10% and 5% significance levels under NULL.\(^7\) Overall, the $Q_{\text{IDAN}}(M)$ test, which is based on the Daniell kernel, has reasonable sizes for all three sample sizes. It tends to overreject a little at the 5% significance level, but not excessively. For each shortfall probability ($\alpha = 10\%$ or $5\%$) and each sample size $T$, the choice of $M$ has little impact on the size of the $Q_{\text{IDAN}}(M)$ test. The truncated kernel-based test $Q_{\text{TRUN}}(M)$ performs similarly to $Q_{\text{IDAN}}(M)$ at the 10% significant level. The regression procedure $Q_{\text{REG}}(M)$, on the contrary, tends to a bit overreject the null hypothesis at the 10% significant level. Both $Q_{\text{TRUN}}(M)$ and $Q_{\text{REG}}(M)$ have better sizes at the $5\%$ significant level.

Table 2 reports the power of the tests under ALTER1, where there exists Granger causality in mean from $Y_{2t}$ to $Y_{1t}$ with respect to $I_{t-1}$. The $Q_{\text{IDAN}}(M)$ test has good power against ALTER1 and it becomes more powerful as $T$ increases. Given each sample size $T$; the power of $Q_{\text{IDAN}}(M)$ declines as the lag order $M$ increases, but not dramatically (which is apparently due to the downward weighting of $k^{T}(\cdot)$). The $Q_{\text{IDAN}}(M)$ test with the 10% shortfall probability (or risk level) is more powerful than the $Q_{\text{IDAN}}(M)$ test with the 5% shortfall probability (or risk level). This is possibly because spillover in mean occurs in the main body of the distribution. On the other hand, $Q_{\text{TRUN}}(M)$ and $Q_{\text{REG}}(M)$ perform similarly, and both have relatively low power. Furthermore, a larger $M$ gives substantially smaller power. For example, at the 5% risk level, the rejection rates of $Q_{\text{TRUN}}(M)$ decrease from 70.8% and 72.1% to 38.2% and 37.0% respectively even when $T = 2000$. These results confirm our expectation that nonuniform weighting alleviates the impact of choosing too large a $M$ because nonuniform weighting discounts higher order lags.

Table 3 reports the power of the tests under ALTER2, where there exists Granger causality in variance from $Y_{2t}$ to $Y_{1t}$ with respect to $I_{t-1}$. The $Q_{\text{IDAN}}(M)$ test has good power against ALTER2. As under ALTER1, the power of $Q_{\text{IDAN}}(M)$ declines as $M$ increases, but not dramatically. Again, $Q_{\text{TRUN}}(M)$ and $Q_{\text{REG}}(M)$ perform similarly and they have relatively low power. In contrast to ALTER1, all tests have better power at the 5% risk level than at the 10% risk level under ALTER2. This is perhaps because under ALTER2, spillover in variance occurs mainly in the tails rather than the centers of the conditional distributions.

Table 4 reports the power of the tests under ALTER3, where there exists Granger causality in skewness and kurtosis from $Y_{2t}$ to $Y_{1t}$ with respect to $I_{t-1}$. As expected, the $Q_{\text{IDAN}}(M)$ test has

---

\(^6\) We have also considered the Bartlett kernel, which is outside the class of kernels over which the Daniell kernel has the optimal power; the results are similar to those based on the Daniell kernel.

\(^7\) We emphasize that the significance level of the tests is different from the risk level or shortfall probability level in the definition of Granger causality in risk.
Table 2
Power at the 10% and 5% significant levels under ALTER1.

<table>
<thead>
<tr>
<th>M</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>Q_{DAN}</td>
<td>500</td>
<td>58.0</td>
<td>52.1</td>
<td>52.6</td>
<td>47.8</td>
<td>50.2</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>80.2</td>
<td>76.0</td>
<td>74.9</td>
<td>69.2</td>
<td>70.5</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>96.3</td>
<td>94.6</td>
<td>94.4</td>
<td>92.8</td>
<td>92.7</td>
</tr>
<tr>
<td>Q_{TRUN}</td>
<td>500</td>
<td>43.5</td>
<td>38.6</td>
<td>41.2</td>
<td>35.4</td>
<td>39.5</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>56.9</td>
<td>52.7</td>
<td>54.5</td>
<td>48.4</td>
<td>51.7</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>77.9</td>
<td>73.7</td>
<td>75.0</td>
<td>69.0</td>
<td>70.1</td>
</tr>
<tr>
<td>Q_{REC}</td>
<td>500</td>
<td>48.6</td>
<td>42.4</td>
<td>41.6</td>
<td>33.8</td>
<td>37.5</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>69.7</td>
<td>62.6</td>
<td>58.5</td>
<td>51.5</td>
<td>52.4</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>92.3</td>
<td>89.8</td>
<td>86.6</td>
<td>81.7</td>
<td>81.8</td>
</tr>
</tbody>
</table>

Table 3
Power at the 10% and 5% significant levels under ALTER2.

<table>
<thead>
<tr>
<th>M</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>Q_{DAN}</td>
<td>500</td>
<td>42.8</td>
<td>36.9</td>
<td>43.1</td>
<td>36.4</td>
<td>40.1</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>59.3</td>
<td>53.5</td>
<td>59.9</td>
<td>52.9</td>
<td>56.5</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>86.6</td>
<td>83.0</td>
<td>84.8</td>
<td>80.4</td>
<td>82.4</td>
</tr>
<tr>
<td>Q_{TRUN}</td>
<td>500</td>
<td>51.4</td>
<td>46.0</td>
<td>53.3</td>
<td>46.9</td>
<td>52.0</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>67.2</td>
<td>62.2</td>
<td>65.6</td>
<td>59.4</td>
<td>62.2</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>84.5</td>
<td>85.4</td>
<td>82.2</td>
<td>83.5</td>
<td>80.6</td>
</tr>
<tr>
<td>Q_{REC}</td>
<td>500</td>
<td>39.8</td>
<td>33.0</td>
<td>33.8</td>
<td>25.0</td>
<td>30.2</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>55.3</td>
<td>47.8</td>
<td>47.1</td>
<td>38.2</td>
<td>42.5</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>81.4</td>
<td>76.2</td>
<td>75.5</td>
<td>68.2</td>
<td>68.4</td>
</tr>
<tr>
<td>Q_{TRUN}</td>
<td>500</td>
<td>50.3</td>
<td>44.5</td>
<td>44.7</td>
<td>36.3</td>
<td>38.9</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>62.4</td>
<td>56.3</td>
<td>55.1</td>
<td>48.5</td>
<td>49.7</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>85.0</td>
<td>80.2</td>
<td>77.0</td>
<td>71.7</td>
<td>72.3</td>
</tr>
<tr>
<td>Q_{REC}</td>
<td>500</td>
<td>40.0</td>
<td>33.5</td>
<td>33.3</td>
<td>25.2</td>
<td>28.7</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>57.4</td>
<td>50.8</td>
<td>50.6</td>
<td>42.1</td>
<td>42.7</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>82.7</td>
<td>77.3</td>
<td>73.6</td>
<td>67.0</td>
<td>67.1</td>
</tr>
<tr>
<td>Q_{TRUN}</td>
<td>500</td>
<td>50.1</td>
<td>44.3</td>
<td>42.9</td>
<td>34.1</td>
<td>37.0</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>60.9</td>
<td>55.7</td>
<td>57.6</td>
<td>48.8</td>
<td>50.5</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>84.7</td>
<td>80.1</td>
<td>76.3</td>
<td>70.7</td>
<td>72.4</td>
</tr>
</tbody>
</table>

ALTERT1: $Y_{t}\sim\text{N}(0,1)$, $\sigma_{Y_{t}}=\text{m.d.s.}$, $\Gamma_{Y_{t-1}}=\text{m.d.s.}$, $\epsilon_{t}\sim\text{N}(0,1)$, $\lambda_{t}\sim\text{Poission}(\lambda)$.

ALTERT2: $Y_{t}\sim\text{N}(0,1)$, $\sigma_{Y_{t}}=\text{m.d.s.}$, $\Gamma_{Y_{t-1}}=\text{m.d.s.}$, $\epsilon_{t}\sim\text{N}(0,1)$, $\lambda_{t}\sim\text{Poission}(\lambda)$.

good power against ALTER3. All power patterns are similar to those under ALTER1, where there exists Granger causality in mean. Furthermore, as under ALTER1, the $Q_{DAN}(M)$ test with the 10% shortfall probability is more powerful than the $Q_{TRUN}(M)$ test with the 5% shortfall probability. With $M = 5$ and $T = 2000$, for example, the rejection rates of $Q_{DAN}(M)$ at the 5% significance level are 86.5% and 39.5% for the 10% and 5% shortfall probabilities respectively. Again this may due to the possibility that spillover mainly comes from skewness which occurs in the main body of the distribution. All power patterns of $Q_{TRUN}(M)$ and $Q_{REC}(M)$ are similar to those of $Q_{TRUN}(M)$ and $Q_{REC}(M)$ under ALTER1.

To better capture the empirical distribution of returns, it is often proposed in the existing literature to add jumps to return and volatility process (see, e.g., Hong et al. (2007) and Maheu and McCurdy (2004)). To investigate the power of our proposed test on detecting the risk spillover based on jump processes, we further consider the Poisson jump model following Hong et al. (2007):
particularly, we test the alternative of Granger causality in risk from mean, however with jump added to the process\(^8\):

\[
\begin{align*}
(\beta_1^1, \beta_{12}^1, \gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{13}, \delta_1, \lambda_1) = & (0.5, 0.2, 0.1, 0.6, 0.2, 0.1, 0.2, 0.1) \\
(\beta_1^2, \beta_{22}^2, \gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{23}, \delta_2, \lambda_2) = & (0.5, 0.1, 0.6, 0.2, 0.1, 0.2, 0.1) \\
\end{align*}
\]

We find that the \(Q_{\text{DAN}}(M)\) test still has good power and it becomes more powerful as \(T\) increases. For 10% risk level, \(Q_{\text{DAN}}(5)\) is 51.3%, 72.7% and 92.9% when \(M = 500, 1000\) and 2000 respectively. and for the 5% risk level, the \(Q_{\text{DAN}}(5)\) test increases from 45.2% to 91.1%.

In summary, the proposed tests with the Daniell kernel have reasonable size and power against a variety of empirically plausible alternatives in finite samples. The truncated kernel-based test \(Q_{\text{TIRUN}}(M)\) and the Granger-type regression procedure \(Q_{\text{REC}}(M)\) also have reasonable sizes for all sample sizes. However, for the alternatives under study, they often yield lower power than nonuniform weighting, especially for a larger lag order \(M\). In contrast, the use of nonuniform weighting makes the power relatively robust to the choice of \(M\). This suggests that our test with nonuniform weighting is a useful tool in investigating extreme risk spillover across financial markets.

### 7. Application to exchange rates

To illustrate our procedures, we now apply them to foreign exchange rates. The foreign exchange market is one of the most important financial markets in the world, where trading takes place 24 h a day around the globe and trillions of dollars of different currencies are transacted each day. Understanding the mechanism of risk spillover between exchange rates is important for many outstanding issues in international economics and finance. The previous literature has focused on volatility spillover (e.g., Baille and Bollerslev (1989, 1990), Engle et al. (1990), Cheung and Ng (1996) and Hong (2001)). Nevertheless, extreme downside risk spillover is important because market participants are increasingly concerned with their exposure to large exchange rate fluctuations, and financial regulators are keen to measure the exchange rate exposures of the financial institutions they supervise. In this section, we use our tests to investigate intraday extreme downside risk spillover between two foreign exchange rates—Euro/Dollar and Yen/Dollar, which are among most active currencies traded in the foreign exchange market.

The data, obtained from Olsen & Associates, are indicative bid and ask quotes posted by banks from July 1, 2000 to September 8, 2000, with a total of 10 weeks. We choose the starting time from July 1, 2000 to wait for the market to stabilize after the introduction of Euro as a new currency in January 1, 1999. Similarly to Diebold et al. (1999), we sample data over a grid of half-hour intervals, i.e., we obtain the quotes nearest the half-hour time stamps. Although foreign exchange trading occurs around the clock during weekdays, trading is very thin during weekends. Following Diebold et al. (1999), we eliminate the observations from Friday 21:30 GMT to Sunday 21:00 GMT, and consequently get a total of 2400 observations. Exchange rate changes are calculated in the same way as in Anderson et al. (2000) and Diebold et al. (1999).

We first calculate the average log bid and log ask prices to get a “log price”, then calculate changes as the differences between log prices at consecutive time points. The intraday calendar effects are also removed following Diebold et al. (1999).\(^10\)

It has been well argued that the mechanism governing the behavior of the tails may be different from that of the rest of the distribution (e.g., Chernozhukov and Umantsev (2001), Danielsson and de Vries (2000) and Engle and Manganelli (2004)). Thus, instead of attempting to model the whole distribution, we use

\(^8\) We are grateful to an anonymous referee for very helpful suggestions on investigating our proposed test on detecting the risk spillover due to jump processes.

\(^9\) The \(Q_{\text{DAN}}(M)\) results for \(M = 10, 15, 20, 25\) and 30, are available from the authors upon request.

\(^10\) Unlike Diebold et al. (1999), we further add a Monday dummy variable to account for the possible weekend effect on the data.
Engle and Manganelli’s (2004) Conditional Autoregressive Value at Risk (CAViaR) model for our VaR calculation:

\[ V_{lt}(\theta) = \theta_0 + \sum_{j=1}^{q} \theta_j Y_{lt-j} + \sum_{j=1}^{r} \theta_{0+j} L(Y_{lt-j}), \quad l = 1, 2. \]

The autoregressive components \( \{\theta_j Y_{lt-j}\} \) ensure that VaR changes slowly over time. The rationale is to capture volatility clustering which is typical of financial time series. The function \( L(\cdot) \) depends on a finite number of lagged values of observable variables that belong to the information set available at time \( t - 1 \). It provides a link between these predetermined variables and VaR. CAViaR models can be used for scenarios with constant volatilities but an link between these predetermined variables and VaR. CAViaR models can be used for scenarios with constant volatilities but avoid suffering from severe loss of power due to the loss of a large number of degrees of freedom, thanks to one-way Granger causality in risk. The one-way test \( Q_{1DAN}(M) \) of Daniell kernel, which checks risk causality from Yen/Dollar to Euro/Dollar, yields statistic values of 3.394, 2.909, 1.342, 0.478 and 0.063 for \( M = 5, 10, 20, 30 \) and 40 respectively at the 1% risk level. It is significant only for \( M = 5 \) and 10, with \( p \)-values below 5%, suggesting that there may only exist weak extreme risk spillover from Yen/Dollar to Euro/Dollar. At the 5% risk level, \( Q_{1DAN} \) yields \( p \)-values well above 10% for all \( M \), suggesting that there is no risk spillover from past Yen/Dollar to Euro/Dollar. In contrast, at both the 1% and 5% risk levels, the \( p \)-values of the one-way test \( Q_{1TRUN} \), which checks causality from Euro/Dollar to Yen/Dollar, are well below the 5% significance level for all \( M \), suggesting significant extreme risk spillover from Euro/Dollar to Yen/Dollar. A further comparison reveals that one-way risk spillover is stronger at the 1% risk level than at the 5% risk level, because both \( Q_{1DAN}(M) \) and \( Q_{1TRUN}(M) \) give larger statistic values at the 1% risk level than at the 5% risk level for each \( M \). This is consistent with most of the empirical findings in the literature that the codependency may be stronger in larger downside market movements between financial markets. We note that the Granger-type regression tests \( Q_{1REC}(M) \) and \( Q_{1REC}(M) \) at the 1% risk level than at the 5% risk level, which is consistent with most of the empirical findings in the literature that the degree of correlation between financial assets or markets often becomes stronger in large downside market movements (e.g., Longin and Solnik (2001)). Finally, nonuniform weighting is more powerful than uniform weighting and the Granger-type regression-based test in detecting risk spillover between exchange rates. This highlights the practical merit of the proposed tests.

8. Conclusion

Based on a new concept of Granger causality in risk which focuses on the movements between the tails of the two distributions, a class of kernel-based tests are proposed to test whether a large downside risk in one market will Granger-cause a large downside risk in another market. The proposed tests check a large number of lags but avoid suffering from severe loss of power due to the loss of a large number of degrees of freedom, thanks to

---

**Table 5**

Estimate and relevant statistics for asymmetric slope CAViaR specification.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>1% VaR Euro/Dollar</th>
<th>2390</th>
<th>0.015 [0.032]</th>
<th>2.215 [0.788]</th>
<th>0.639 [0.646]</th>
<th>0.842 [0.865]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_0 )</td>
<td>0.015 [0.032]</td>
<td>2.215 [0.788]</td>
<td>0.639 [0.646]</td>
<td>0.842 [0.865]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>0.084 [0.011]</td>
<td>0.588 [0.122]</td>
<td>0.741 [0.217]</td>
<td>0.766 [0.245]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>-0.053 [0.037]</td>
<td>0.029 [0.257]</td>
<td>0.066 [0.067]</td>
<td>-0.008 [0.074]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_3 )</td>
<td>0.157 [0.086]</td>
<td>0.872 [0.295]</td>
<td>0.178 [0.063]</td>
<td>0.108 [0.127]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Table 6**

We now consider extreme risk spillover between Euro/Dollar and Yen/Dollar. Table 6 reports our test statistics and the Granger-type regression test statistics at the 1% and 5% risk levels, together with their \( p \)-values. As in the simulation study, we use the Daniell kernel and the truncated kernel for our test. To identify the direction of risk spillover, here we consider two directional tests for extreme risk spillover from Euro/Dollar to Yen/Dollar. Table 6 reports our test statistics and the Granger-type regression test statistics at the 1% and 5% risk levels, together with their \( p \)-values. As in the simulation study, we use the Daniell kernel and the truncated kernel for our test. To identify the direction of risk spillover, here we consider two directional tests.
Table 6  
Risk spillover between Euro/Dollar and Yen/Dollar.

<table>
<thead>
<tr>
<th></th>
<th>1% VaR</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>Eur⇒Yen</td>
<td>3.394</td>
<td>2.909</td>
<td>1.342</td>
<td>0.478</td>
<td>−0.063</td>
<td>−0.313</td>
</tr>
<tr>
<td>(p-values)</td>
<td>0.000</td>
<td>0.002</td>
<td>0.090</td>
<td>0.316</td>
<td>0.525</td>
<td>0.623</td>
</tr>
<tr>
<td>(p-values)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>Eur⇒Yen</td>
<td>2.495</td>
<td>0.950</td>
<td>−0.475</td>
<td>−1.094</td>
<td>−1.342</td>
<td>−0.455</td>
</tr>
<tr>
<td>(p-values)</td>
<td>0.006</td>
<td>0.171</td>
<td>0.683</td>
<td>0.863</td>
<td>0.910</td>
<td>0.675</td>
</tr>
<tr>
<td>Eur⇒Yen</td>
<td>8.621</td>
<td>5.703</td>
<td>3.775</td>
<td>2.865</td>
<td>1.862</td>
<td>2.319</td>
</tr>
<tr>
<td>(p-values)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>0.031</td>
<td>0.010</td>
</tr>
<tr>
<td>Eur⇒Yen</td>
<td>2.487</td>
<td>0.906</td>
<td>−0.461</td>
<td>−0.934</td>
<td>−1.120</td>
<td>−0.459</td>
</tr>
<tr>
<td>(p-values)</td>
<td>0.006</td>
<td>0.182</td>
<td>0.678</td>
<td>0.825</td>
<td>0.869</td>
<td>0.677</td>
</tr>
<tr>
<td>Eur⇒Yen</td>
<td>8.393</td>
<td>5.439</td>
<td>3.614</td>
<td>0.765</td>
<td>−0.037</td>
<td>2.347</td>
</tr>
<tr>
<td>(p-values)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.222</td>
<td>0.515</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Assumptions 6. Let \( R(t) \) and \( R(t)^0 \) be defined in the same way as \( f(\omega) \) and \( f(\omega)^0 \) in (3.4) and (3.5) with \( \theta \equiv (\theta_1, \theta_2) \) replaced by \( \theta_0 \equiv (\theta_0^1, \theta_0^2) \). Given \( D_1[(\omega)] = M \int_{-\infty}^{\infty} k^1(\omega) d\omega = 1 + o(1) \) as \( M \to \infty \) under Assumption 6, it suffices to show Theorems A.1 and A.2 under the conditions of Theorem 1. Theorem A.1 implies that parameter estimation uncertainty in \( \theta \) has no impact on the limit distribution of \( Q_t(M) \). The main technical challenge for the proof of Theorem A.1 is that the risk indicator \( Z_t(\theta) \equiv 1 (Y_t < -V_t) \) is not differentiable with respect to parameter \( \theta_t \).

Theorem A.1. \( M^{-1/2} T [L^2(\tilde{f}, \tilde{f}_t^0) - L^2(\tilde{f}, \tilde{f}_t^0)] \to \mathcal{P} 0 \), where \( L_t(\cdot) \) is defined as in (3.6).

Theorem A.2. \( [T L^2(\tilde{f}, \tilde{f}_t^0) - C_{1T}(M)]/[2D_{1T}(M)]^{1/2} \to N(0, 1) \).

Proof of Theorem A.1. Throughout, let \( \hat{c}(j) \) be defined as in (3.3) with \( \hat{\theta} \) replaced by \( \theta_0 \). We further replace the sample proportions \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) in \( \hat{c}(j) \) with \( \alpha \). Such a replacement does not affect the asymptotic distribution of \( Q_t(M) \). Putting \( \sigma^2 = \alpha (1 - \alpha) \), we have

\[
T \left[ L^2(\tilde{f}, \tilde{f}_t^0) - L^2(\tilde{f}, \tilde{f}_t^0) \right] = \sigma^4 T \sum_{j=1}^{T-1} k^2(j/M) \left[ \hat{c}(j) - \hat{c}(j) \right]^2 + 2 \sigma^4 T \sum_{j=1}^{T-1} k^2(j/M) \hat{c}(j) - \hat{c}(j) \hat{c}(j) \\
\equiv \sigma^4 T \hat{Q}_t + 2 \sigma^4 \hat{T}_Q, \quad \text{say.} \quad (A.1)
\]

We shall prove Theorem A.1 by showing Propositions A.1 and A.2.

**Proposition A.1.** \( M^{-1/2} \hat{Q}_t \to \mathcal{P} 0 \).

**Proposition A.2.** \( M^{-1/2} \hat{T}_Q \to \mathcal{P} 0 \).

**Proof of Proposition A.1.** By straightforward algebra, we have for \( j > 0 \).

\[
\hat{c}(j) - \hat{c}(j) = \hat{M}_1(j, \hat{\theta}_1) + \hat{M}_2(j, \hat{\theta}_2) + \hat{M}_3(j, \hat{\theta}_1, \hat{\theta}_2), \quad (A.2)
\]

where

\[
\hat{M}_1(j, \hat{\theta}_1) = \frac{1}{T} \sum_{t=j+1}^{T} \left[ Z_{1t} \theta_1 - Z_{1t} \theta_1^0 \right] Z_{2t-j} \theta_2^0 - \alpha, \quad \hat{M}_2(j, \hat{\theta}_2) = \frac{1}{T} \sum_{t=j+1}^{T} \left[ Z_{1t} \theta_2^0 - \alpha \right] Z_{2t-j} \theta_2 - Z_{2t-j} \theta_2^0, \\
\hat{M}_3(j, \hat{\theta}_1, \hat{\theta}_2) = \frac{1}{T} \sum_{t=j+1}^{T} \left[ Z_{1t} \theta_1 - Z_{1t} \theta_1^0 \right] Z_{2t-j} \theta_2 - Z_{2t-j} \theta_2^0.
\]

By the definition of \( \hat{Q}_t \) in (A.1), we have

\[
\hat{Q}_t \leq 3 \left[ \hat{Q}_{1t} \hat{\theta}_1 + \hat{Q}_{1t} \hat{\theta}_2 + \hat{Q}_{1t} \hat{\theta}_2 \right], \quad (A.3)
\]
where
\[
\hat{Q}_{11}(\theta_t) = \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_t^2(j, \theta_t),
\]
\[
\hat{Q}_{12}(\theta_t) = \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_t^2(j, \theta_t),
\]
\[
\hat{Q}_{13}(\theta_t, \theta_{t+1}) = \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_t^2(j, \theta_t).
\]

Because the risk indicator \(Z_t(\theta_t)\) is not differentiable with respect to \(\theta_t\), we shall use the uniform convergence argument to show that \(\hat{Q}_{11}(\theta_t), \hat{Q}_{12}(\theta_t)\) and \(\hat{Q}_{13}(\theta_t, \theta_{t+1})\) vanish in probability with suitable rates. Given Assumption 4, we have that for any given constant \(\varepsilon > 0\), there exists \(\Delta_0 \equiv \Delta_0(\varepsilon) < \infty\) such that \(P(\hat{\alpha} - \theta_0^0 > \Delta_0 T^{-\frac{1}{2}}) < \varepsilon\) for \(T\) sufficiently large. Hence, it suffices to show Lemmas 1A–1.3.

**Lemma A.1.** Put \(\theta_0^0 = \{\theta_t : |\theta_t - \theta_0^0| \leq \Delta_0 T^{-\frac{1}{2}}\}\) for \(0 < \Delta_0 < \infty, l = 1, 2\). Then for any given constant \(\Delta_0 > 0\), sup_{\theta_t \in \theta_0^0} ||F(\hat{\theta}_t) - F(\theta_0^0)|| \leq \Delta_0 T^{-\frac{1}{2}}, \sum_{j=1}^{T-1} k^2(j/M)|\Gamma(j)|^2 \rightarrow \sum_{j=0}^{\infty} ||\Gamma(j)||^2 = O(1) \) given Assumptions 5(i) and 6, \(M \rightarrow \infty, M/T \rightarrow 0\), and

\[
\sup_{\theta_t \in \theta_0^0} \sum_{j=1}^{T-1} k^2(j/M)||\hat{F}(j) - \Gamma(j)||^2 = O_p(1) + O_p(M/T).
\]

by Markov’s inequality and sup_{\theta_t \in \theta_0^0} ||F(\hat{\theta}_t) - F(\theta_0^0)|| \leq \Delta_0 T^{-\frac{1}{2}}, Assumption 2, and Markov’s inequality. It follows from (A.6)–(A.8), \(M \rightarrow \infty, M/T \rightarrow 0\) that

\[
\sup_{\theta_t \in \theta_0^0} \sum_{j=1}^{T-1} k^2(j/M)||\hat{M}(j, \theta_t)||^2 = O_p(1).
\]

We now consider the first term \(\hat{M}(j, \theta_t)\) in (A.5). Divide the cube \(\theta_0^0\), which is centered at \(\theta_0^0\) with size \(2\Delta_0 T^{-\frac{1}{2}}\), into approximately \(l_T \equiv (2\Delta_0/\varepsilon_T)^d\) cubes \(\{\theta_t^0(I), l = 1, \ldots, l_T\}\) of size \(\varepsilon_T/T^{\frac{1}{2}}\), where \(\varepsilon_T \equiv M^{-\frac{1}{2}} / (\ln(T) \rightarrow 0\) and \(d_t^0\) is the dimension of \(\theta_t\). For \(1 \leq l \leq l_t\), put \(\theta_t^0(l) = \inf_{\theta_t \in \theta_t^0(I)} \sup_{\theta_t \in \theta_t^0(I)} F_{\theta_t}(\theta_t^0(l))\) and \(\theta_t^0(l) = \sup_{\theta_t \in \theta_t^0(I)} \inf_{\theta_t \in \theta_t^0(I)} F_{\theta_t}(\theta_t^0(l))\). Note that \(\theta_t^0(l)\) and \(\theta_t^0(l)\) are measurable functions of \(I_{t-1}\) because \(V_{\theta_t}(\theta_t)\) is a measurable function of \(I_{t-1}\). Then, for any \(\theta_t \in \theta_t^0(l)\), we write

\[
\hat{M}(j, \theta_t) = \hat{M}_1(j, \theta_t) + \hat{M}_2(j, \theta_t),
\]

say. (A.5)

We first consider \(\hat{M}_1(j, \theta_t)\) in (A.5). By a second order Taylor series expansion, we have

\[
\hat{M}_1(j, \theta_t) = (\theta_t - \theta_0^0)^T \hat{M}_1(j, \theta_t) = \sum_{t=1}^{T-1} \frac{\partial F_{\theta_t}(\theta_t^0(l))}{\partial \theta_t} [Z_{t+1} - \alpha] + \frac{1}{2} (\theta_t - \theta_0^0)^T \hat{M}_1(j, \theta_t) + \frac{1}{2} (\theta_t - \theta_0^0) \hat{M}_2(j, \theta_t)
\]

say, (A.6)
\[ = T^{-1} \sum_{t=1}^{T} [W_t[\theta_0^b(l)] - W_t[\theta_0^a(l)]] \]
\[ + 2T^{-1} \sum_{t=1}^{T} [F_t[\theta_t^a(l)] - F_t[\theta_t^a(l)]] \]
\[
\text{Similarly, we can obtain}
\]
\[ \hat{M}_1(j, \theta_1) \geq -T^{-1} \sum_{t=1}^{T} [W_t[\theta_0^b(l)] - W_t[\theta_0^a(l)]] \]
\[ - 2T^{-1} \sum_{t=1}^{T} [F_t[\theta_t^a(l)] - F_t[\theta_t^a(l)]] \]
\[ \text{It follows that} \]
\[ \max_{0 < \epsilon < T} \sup_{\theta \in \Theta_0^b} \left| \hat{M}_1(j, \theta_1) \right| \leq T^{-1} \sum_{t=1}^{T} [W_t[\theta_0^b(l)] - W_t[\theta_0^a(l)]] \]
\[ + 2 \Delta_0 \epsilon T^{-1} \sum_{t=1}^{T} \sup_{\theta \in \Theta_0^b} \left| \frac{\partial F_t[\theta_t^a(l)]}{\partial \theta_1} \right| \quad (A.10) \]
\[ \text{given } \|\theta_0^b(l) - \theta_0^a(l)\| \leq \Delta_0 \epsilon T^{\frac{1}{2}}. \text{ Note that the second term in} \]
\[ \text{(A.10) does not depend on } I. \]
\[ \text{Therefore, we have} \]
\[ \sup_{\theta \in \Theta_0^b} \left| \frac{T^{-1} \sum_{j=1}^{T-1} k_j(j) \epsilon}{M_1(j, \theta_1)} \right|^2 \]
\[ = \max_{0 < \epsilon < T} \sup_{\theta \in \Theta_0^b} \left| \frac{T^{-1} \sum_{j=1}^{T-1} k_j(j) \epsilon}{M_1(j, \theta_1)} \right|^2 \]
\[ \leq 2 \max_{0 < \epsilon < T} \left| T^{-1} \sum_{j=1}^{T-1} [W_t[\theta_0^b(l)] - W_t[\theta_0^a(l)]] \right| \left| \frac{T^{-1} \sum_{j=1}^{T-1} k_j(j) \epsilon}{M_1(j, \theta_1)} \right|^2 \]
\[ + 4T(\Delta_0 \epsilon T^{1/2}) \left[ T^{-1} \sum_{j=1}^{T-1} \left| \frac{\partial F_t[\theta_t^a(l)]}{\partial \theta_1} \right| \right|^2 \]
\[ \times T^{-1} \sum_{j=1}^{T-1} k_j(j) \epsilon \]
\[ = O(\epsilon^2 M_1^2) = O(1) \quad (A.11) \]
\[ \text{given } \epsilon = M^{-\frac{1}{2}} / \ln(T), \text{ where we have made use of the fact that} \]
\[ P \left( \max_{0 < \epsilon < T} \left| T^{-1} \sum_{j=1}^{T-1} [W_t[\theta_0^b(l)] - W_t[\theta_0^a(l)]] \right|^2 > \Delta_0 \epsilon^2 \right) \]
\[ \leq \frac{1}{2} \left( T^{-1} \sum_{j=1}^{T-1} [W_t[\theta_0^b(l)] - W_t[\theta_0^a(l)]] \right)^2 \left( \Delta_0 \epsilon^2 / T \right) \]
\[ \leq \left( \Delta_0 \epsilon^2 / T \right) \epsilon \left[ T^{-1} \sum_{j=1}^{T-1} [W_t[\theta_0^b(l)] - W_t[\theta_0^a(l)]] \right] \]
\[ \leq \frac{1}{2} \left( \Delta_0 \epsilon^2 / T \right)^2 \left( T^{-2} (\Delta_0 \epsilon T^2 / 2)^2 + T^{-3} (\Delta_0 \epsilon / T) \right) \]
\[ = O(\epsilon^2 T^{3/2} / \ln(T)) \rightarrow 0 \]
\[ \text{given } M = c T^{\nu}, \epsilon = M^{-\frac{1}{2}} / \ln(T), \nu < \frac{3}{4}, \text{ where the third inequality follows from} \]
\[ E \left| T^{-1} \sum_{j=1}^{T-1} [W_t[\theta_0^b(l)] - W_t[\theta_0^a(l)]] \right|^4 \]
\[ \leq T^{-2} (\Delta_0 \epsilon T^2 / 2)^2 + T^{-3} (\Delta_0 \epsilon T^2 / 2) \]
\[ \text{by Rosenthal's inequality (e.g., Hall and Heyde (1980, p. 23)), the fact that } [W_t[\theta_0^b(l)] - W_t[\theta_0^a(l)], F_{t-1}^l] \text{ is a m.s. and the fact that} \]
\[ E[W_t[\theta_0^b(l)] - W_t[\theta_0^a(l)] \leq \Delta_0 \epsilon T^2 \quad \text{for any } m \geq 1 \]
\[ \text{by the law of iterated expectation and } |\theta_0^b(l) - \theta_0^a(l)| \leq \Delta_0 \epsilon T^{1/2}. \text{ Note that a larger } m \text{ does not imply a faster convergence rate due to} \]
\[ \text{the very nature of the indicator function. The desired result then follows from} (A.5), (A.9) \text{ and} (A.11). \]
\[ \text{□} \]
\[ \text{Lemma A.2. For any given constant } \Delta_0 > 0, \text{ we have sup}_{\theta \in \Theta_0^b} \|T \tilde{Q}_1(\theta_2)\| = O(1) \\text{and sup}_{\theta \in \Theta_0^b} \left| T \tilde{Q}_1(\theta_1, \theta_2) \right| \rightarrow^p 0. \]
\[ \text{Proof of Lemma A.2. Similar to the proof of Lemma A.1.} \]
\[ \text{Lemma A.3. Put } \theta_0 = \theta_0^1 \otimes \theta_2^0 \text{ and } \theta = (\theta_1, \theta_2). \text{ Then for any given constant } \Delta_0 > 0, \text{ sup}_{\theta \in \Theta_0^b} \|T \tilde{Q}_1(\theta_1, \theta_2)\| \rightarrow^p 0 \text{ and sup}_{\theta \in \Theta_0^b} \left| T \tilde{Q}_1(\theta_1, \theta_2) \right| \rightarrow^p 0. \]
\[ \text{Proof of Lemma A.3. Recalling the definition of } \hat{M}_1(j, \theta_1, \theta_2) \text{ as in} (A.2) \text{ and } Z_\theta(\theta_0^b) = Z_\theta, \text{ we write} \]
\[ \hat{M}_3(j, \theta_1, \theta_2) = T^{-1} \sum_{t=j+1}^{T} [Z_\theta(\theta_0^b) - Z_\theta(\theta_1^a)] \]
\[ \times [Z_\theta(\theta_0^b) - Z_\theta(\theta_2^0)] \]
\[ = T^{-1} \sum_{t=j+1}^{T} W_t[\theta_0^b(l)] \sum_{t=j+1}^{T} W_t[\theta_0^a(l)] \]
\[ + T^{-1} \sum_{t=j+1}^{T} \left[ F_t[\theta_t^a(l)] - F_t[\theta_t^a(l)] \right] \]
\[ \times \left[ Z_\theta(\theta_0^b) - Z_\theta(\theta_2^0) \right] \]
\[ \hat{M}_3(j, \theta_1, \theta_2) \text{ in } (A.12), \text{ following reasoning analogous to that for } \hat{M}_1(j, \theta_1) \text{ in the proof of Lemma A.1, we can obtain} \]
\[ \sup_{\theta \in \Theta_0^b} \left| T \sum_{j=1}^{T-1} k_j(j) \epsilon \right|^2 \rightarrow^p 0. \]
\[ \text{For } \hat{M}_3(j, \theta_1, \theta_2) \text{ in } (A.12), \text{ by the mean value theorem, we have} \]
\[ \hat{M}_3(j, \theta_1, \theta_2) = (\theta_1 - \theta_0^1)^T \frac{T^{-1} \sum_{t=j+1}^{T} \frac{\partial F_t[\theta_t^a(l)]}{\partial \theta_1}}{\theta_1 - \theta_0^1} \times [Z_\theta(\theta_0^b) - Z_\theta(\theta_2^0)] = O(T^{-3/4}) \quad (A.14) \]
\[ \text{uniformly in } (\theta_1, \theta_2) \in \Theta_0. \text{ Here, we have used the facts that} \]
\[ \|\theta_1 - \theta_0^1\| \leq \Delta_0 T^{-1/2} \text{ and} \]
\[ E \sup_{\theta \in \Theta_0^b} \left| T^{-1} \sum_{t=j+1}^{T} \frac{\partial F_t[\theta_t^a(l)]}{\partial \theta_1} \right|^2 \leq \left\{ T^{-1} \sum_{t=j+1}^{T} E \sup_{\theta \in \Theta_0^b} \left| Z_\theta(\theta_0^b) - Z_\theta(\theta_0^b) \right|^2 \right\} \]
\[ \leq c T^{-1 / 2} \]
given Assumption 2 and
\[ E \sup_{\theta_2 \in \Theta^0} \left[ Z_{21}(\theta_2) - Z_{21}(\theta_2^0) \right]^2 \]
\[ \leq E \left\{ \left[ (1 - V_{21}(\theta_2^0)) < V_{21} < (1 - V_{21}(\theta_2^0)) \right] \right\} \]
\[ = E[F_{21}[1 - V_{21}(\theta_2^0)] - F_{21}[1 - V_{21}(\theta_2^0)]] \leq \Delta_0 T^{-\frac{1}{2}} \]
given \[ \| \theta_2 - \theta_2^0 \| \leq \Delta_0 T^{-\frac{1}{2}} \], where \( \theta_2^0 \equiv \arg \inf_{\theta_2 \in \Theta^0} F_{21}[1 - V_{21}(\theta_2)] \)
and \( \theta_2^0 \equiv \arg \sup_{\theta_2 \in \Theta^0} F_{21}[1 - V_{21}(\theta_2)] \), and the equality follows by
the law of iterated expectations and the fact that \( \theta_2^0 \) and \( \theta_2^0 \) are measurable functions of \( I_{21-1} \). It follows that
\[ \sup_{\theta_2 \in \Theta^0} \sum_{j=1}^{T-1} k^2(j/M) \left| \hat{M}_{21}(j, \theta_1, \theta_2) \right|^2 = O_p(M/T^{\frac{1}{2}}) = o_p(1). \] (A.15)

Combining (A.12), (A.13) and (A.15) yields the desired result. \( \Box \)

**Proof of Proposition A.2.** Recalling the definition of \( \hat{Q}_2 \) in (A.1) and using (A.2), we can write
\[ \hat{Q}_2 = \hat{Q}_{12}(\hat{\theta}_1) + \hat{Q}_{22}(\hat{\theta}_2) + \hat{Q}_{23}(\hat{\theta}_1, \hat{\theta}_2), \] (A.16)
where
\[ \hat{Q}_{12}(\theta_1) \equiv \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_{11}(j, \theta_1) \hat{C}(j), \]
\[ \hat{Q}_{22}(\theta_2) \equiv \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_{12}(j, \theta_2) \hat{C}(j), \]
\[ \hat{Q}_{23}(\theta_1, \theta_2) \equiv \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_{12}(j, \theta_1, \theta_2) \hat{C}(j). \]

Following reasoning analogous to that of Proposition A.1, it suffices to show Lemmas A.4–A.6:

**Lemma A.4.** For any given constant \( \Delta_0 > 0 \), \( \sup_{\theta_1 \in \Theta_0^1} |M^{-\frac{1}{2}} T \hat{Q}_{21}(\theta_1)| \to p 0 \).

**Proof of Lemma A.4.** Recalling \( \hat{M}_{11}(j, \theta_1) = \hat{M}_{11}(j, \theta_1) + \hat{M}_{12}(j, \theta_1) \) in (A.5), we have
\[ \hat{Q}_{21}(\hat{\theta}_1) \equiv \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_{11}(j, \hat{\theta}_1) \hat{C}(j) \]
\[ + \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_{12}(j, \hat{\theta}_1) \hat{C}(j) \]
\[ = \hat{Q}_{211}(\hat{\theta}_1) + \hat{Q}_{212}(\hat{\theta}_1). \] (A.17)

For the first term in (A.17), we have
\[ \sup_{\theta_1 \in \Theta_0^1} M^{-\frac{1}{2}} T \left| \hat{Q}_{211}(\theta_1) \right| \leq \left( \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_{11}(j, \theta_1) \right) \left( \sum_{j=1}^{T-1} k^2(j/M) \hat{C}(j) \right) \]
\[ \leq O_p(1) = O_p(1) \] (A.18)
by the Cauchy–Schwarz inequality, (A.11), and the fact that
\[ M^{-\frac{1}{2}} T \sum_{j=1}^{T-1} k^2(j/M) \hat{C}(j) = O_p(1). \] (A.19)

which follows by Markov’s inequality, and \( E[\hat{C}(j)]^2 \leq T^{-1} \) under \( H_0^1 \).

For the second term in (A.17), recalling that \( \hat{M}_{12}(j, \theta_1) \) can be decomposed as in (A.6), we have
\[ T \hat{Q}_{122}(\theta_1) = (\theta_1 - \theta_1^0) T \sum_{j=1}^{T-1} k^2(j/M) \Gamma(j) \hat{C}(j) \]
\[ + (\theta_1 - \theta_1^0) \hat{D}_1(\theta_1) \hat{C}(j) \]
\[ + \frac{1}{2} (\theta_1 - \theta_1^0) \hat{D}_1(\theta_1) \hat{C}(j) \]
\[ + \frac{1}{2} \sqrt{T} (\theta_1 - \theta_1^0) \hat{D}_2(\theta_1) \hat{C}(j), \] (A.20)
where \( \hat{\theta} \) lies between \( \theta_1 \) and \( \theta_1^0 \). For the first term in (A.20), we have
\[ \left| \hat{D}_1(\theta_1^0) \right| \leq \sqrt{T} \sum_{j=1}^{T-1} k^2(j/M) \||\Gamma(j)||\hat{C}(j)\| \]
\[ = O_p(1) \] (A.21)
by Markov’s inequality, Assumptions 5 and 6 and \( E[\hat{C}(j)]^2 \leq T^{-1} \) under \( H_0^1 \).

For the second term in (A.20), we have
\[ \left| \hat{D}_2(\theta_1^0) \right| \leq T \sum_{j=1}^{T-1} k^2(j/M) \||\hat{C}(j)||\hat{C}(j)\| \]
\[ = O_p(M/T^{\frac{1}{2}}) \] (A.22)
by Markov’s inequality, \( E[\hat{C}(j)]^2 \leq T^{-1} \) given Assumption 5, and \( E[\hat{C}(j)]^2 \leq T^{-1} \) under \( H_0^1 \). Similarly, we have
\[ \left| \hat{D}_3(\theta_1) \right| \leq T \sum_{j=1}^{T-1} \sup_{\theta_1 \in \Theta_0^1} \left| \frac{\partial^2 F_{11}[1 - V_{11}(\theta_1)]}{\partial \theta_1 \theta_1^0} \right| \sum_{j=1}^{T-1} k^2(j/M) \hat{C}(j) \]
\[ = O_p(M/T^{\frac{1}{2}}) \] (A.23)
given Assumption 6, and \( E[\hat{C}(j)]^2 \leq T^{-1} \). Collecting (A.20)–(A.23) and \( M/T \to 0 \), we obtain \( M^{-\frac{1}{2}} T \sup_{\theta_1 \in \Theta_0^1} |T \hat{Q}_{212}(\theta_1) | \to p 0 \). This completes the proof for Lemma A.4. \( \Box \)

**Lemma A.5.** For any given constant \( \Delta_0 > 0 \), \( \sup_{\theta_1, \theta_2 \in \Theta_0^1} |M^{-\frac{1}{2}} T \hat{Q}_{22}(\theta_2)| \to p 0 \).

**Proof of Lemma A.5.** Similar to the proof of Lemma A.4. \( \Box \)

**Lemma A.6.** For any given constant \( \Delta_0 > 0 \), \( \sup_{\theta_1, \theta_2 \in \Theta_0} |M^{-\frac{1}{2}} T \hat{Q}_{23}(\theta_1, \theta_2) | \to p 0 \).

**Proof of Lemma A.6.** The result follows by the Cauchy–Schwarz inequality, Lemma A.3 and (A.19). \( \Box \)

**Proof of Theorem A.2.** The desired result follows from a modification of the proof of Hong (2001, Theorem 1) by putting \( u_t \equiv Z_{1t} - \alpha \) and \( v_t \equiv Z_{2t} - \alpha \). Note that \( \{u_t\} \) is an i.i.d. sequence and \( u_t \) is independent of \( \{v_s, s < t\} \) under \( H_0^1 \). The difference between Theorem 1 of Hong (2001) and the present case is that in the former, \( \{u_t\} \) and \( \{v_t\} \) are mutually independent, while in the present case, we should allow for the possibility that \( v_t \) may depend on \( \{u_s, s < t\} \). Given Assumption 3(ii), however, by going through all steps in the proof of Hong (2001, Theorem 1), we can show that this does affect the asymptotic normality result of the proposed test statistic. In other words, the asymptotic normality of \( Q_t(M) \) holds under if \( \{u_t\} \) and \( \{v_t\} \) were mutually independent. \( \Box \)
Proof of Theorem 2. Recall that \( C_1T(M) = O(M) \) and \( D_1T(M) = 2M \int k^4(z)dz\{1 + o(1)\} = M \rightarrow \infty \) and \( M/T \rightarrow 0 \), we have 
\[
(1/M^{1/2})Q_1(M) = \{2 \int k^4(z)dz\}^{-1/2}L^2(\hat{f},f_0^2)\{1 + o(1)\} + o(1).
\]
Thus, it suffices to prove Theorems A.3 and A.4.

Theorem A.3. \( L^2(\hat{f},f_0^2) - L^2(\tilde{f},f_0^2) \rightarrow p 0 \).

Theorem A.4. \( L^2(\hat{f},f_0^2) - L^2(\tilde{f},f_0^2) \rightarrow p 0 \).

Proof of Theorem A.3. Recall \( L^2(\hat{f},f_0^2) = \sigma^4 \hat{Q}_1 + 2\sigma^2 \hat{Q}_2 \), where \( \hat{Q}_1 \) and \( \hat{Q}_2 \) are as in (A.1). It suffices to show \( \hat{Q}_0 \rightarrow 0 \). The second term \( \hat{Q}_2 \) will also vanish in probability by the Cauchy–Schwarz inequality and Theorem A.4, which implies \( L^2(\hat{f},f_0^2) = O_p(1) \) given Assumption 7.

Next, recall \( \hat{Q}_1 \leq 3(\hat{Q}_{11}(\theta_1) + \hat{Q}_{12}(\theta_2) + \hat{Q}_{13}(\theta_1, \theta_2)) \), where \( \hat{Q}_{11}(\theta_1) \), \( \hat{Q}_{12}(\theta_2) \) and \( \hat{Q}_{13}(\theta_1, \theta_2) \) are defined as in [A.3]. We shall show that these three terms all vanish in probability under \( A_1^0 \).

We first consider \( \hat{Q}_{11}(\theta_1) \), given Assumption 4, we have that for any given constant \( \epsilon > 0 \), there exists \( \Delta_0 = \Delta_0(\epsilon) \) such that 
\[
P(\hat{M}_t(\theta_1) \geq \epsilon) \approx \inf_{\theta_1 \in \Theta_1} E(F_t - V_{t1}(\theta_1))
\]
\[
\Delta_0 T^{-1/2} < \epsilon \text{ for all } T \text{ sufficiently large.}
\]
Thus, it suffices to show \( \hat{Q}_{11}(\theta_1) \rightarrow 0 \) uniformly in \( \theta_1 \in \Theta_1 \), where \( \Theta_1 \) is as in Lemma A.1. By the definition of \( \hat{Q}_{11}(\theta_1) \) in (A.3), we have 
\[
\sup_{\theta_1 \in \Theta_1} \hat{Q}_{11}(\theta_1) = \sup_{\theta_1 \in \Theta_1} \sup_{j \geq 1-t} \left\{ \sum_{t=1}^{T-1} \left[ M_t(j, \theta_1) - Z_t(\theta_1^*) \right] \right\},
\]
where \( M_t(j, \theta_1) \) is defined in (A.2). Put \( \theta_1^* = \arg \sup_{\theta_1 \in \Theta_1} E(F_t - V_{t1}(\theta_1)) \) and \( \theta_1^* = \arg \sup_{\theta_1 \in \Theta_1} E(F_t - V_{t1}(\theta_1)). \) Note that \( \theta_1^* \) and \( \theta_1^* \) are measurable functions of \( I_{t-1-t} \), because \( V_{t1}(\theta_1) \) depends on \( I_{t-1-t} \) and \( \theta_1 \). Then 
\[
\max_{\theta_1 \in \Theta_1} \sup_{t \geq 1-t} \left\{ \sum_{t=1}^{T-1} \left[ M_t(j, \theta_1) - Z_t(\theta_1^*) \right] \right\} \leq \Delta_0 T^{-1/2} E \sup_{\theta_1 \in \Theta_1} \left\| \frac{\partial}{\partial \theta_1} F_t - V_{t1}(\theta_1) \right\|
\]
by the law of iterated expectations and 
\[
\left\| \theta_1^* - \theta_1^* \right\| \leq \Delta_0 T^{-1/2}.
\]
It follows from (A.24) and (A.25), \( M^2/T \rightarrow 0 \) that \( \sup_{\theta_1 \in \Theta_1} [\hat{Q}_{12}(\theta_1)] = O_p(M/T^{1/2}) \rightarrow 0 \). Similarly, we have \( \sup_{\theta_1 \in \Theta_1} [\hat{Q}_{13}(\theta_1)] = O_p(1) \). This completes the proof.


Proof of Theorem 3. The proof is similar to that of Theorem 1, so we omit it here. Part of the proof is more tedious than the proof of Theorem 1 because both positive and negative j’s should be considered, but the other part is simpler because under \( \|z_{t+2}\|^2 \), \( u_t = Z_{t+2} - \alpha \) is independent of \( \{v_t = \tilde{z}_{t+2} - \alpha, s \leq t \} \) and \( v_t \) is independent of \( \{u_t \leq s - 1 \} \). This is the reason why we do not need Assumption 3, which is required in Theorem 1.

Proof of Theorem 4. The proof is similar to that of Theorem 2.
Lumsdaine, R., 1996. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in GARCH(1,1) and covariance stationary GARCH(1,1) models. Econometrica 64, 575–596.