ELSEVIER

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

Yongmiao Hong [a,b,*], Yanhui Liu [c], Shouyang Wang [c]

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

ARTICLE INFO

ABSTRACT

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 Cochrance (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)).[1] 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 downside 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 T 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 Pierre'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 Δ_0 denote bounded constants; $\to^d$ and $\to^p$ 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 $T \to \infty$. 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 $1 - \alpha$, where $\alpha \in (0, 1)$, VaR is defined as the loss over the time horizon τ that is not exceeded with probability $1 - \alpha$. Statistically speaking, VaR, denoted by $V_t \equiv V(I_{t-1}, \alpha)$, 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:

$$P(Y_t < -V_t | I_{t-1}) = \alpha \quad \text{almost surely (a.s.)}, \tag{2.1}$$

where $I_{t-1} \equiv \{Y_{t-1}, Y_{t-2}, \ldots\}$ 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_t 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:

$$\begin{cases} Y_t = \mu_t + \sigma_t \varepsilon_t, \\ \{\varepsilon_t\} \sim \text{m.d.s. } (0, 1) \text{ with conditional CDF } F_t(\cdot), \end{cases} \tag{2.2}$$

where $\mu_t \equiv \mu_t(I_{t-1})$ and $\sigma_t^2 \equiv \sigma_t^2(I_{t-1})$ are the conditional mean and conditional variance of Y_t given I_{t-1} respectively, and $F_t(\cdot) \equiv F_t(\cdot | I_{t-1})$ is the conditional cumulative distribution function (CDF) of ε_t given I_{t-1}. By definition, the standardized innovation $\{\varepsilon_t\}$ is a conditionally homoskedastic martingale difference sequence (m.d.s.) with $E(\varepsilon_t | I_{t-1}) = 0$ a.s. and $\text{var}(\varepsilon_t | I_{t-1}) = 1$ 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 $\{\varepsilon_t\}$ follows a generalized Student-t-distribution with time-varying shape parameters.

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

$$V_t = -\mu_t + \sigma_t z_t(\alpha), \tag{2.3}$$

where $z_t(\alpha) \equiv z(I_{t-1}, \alpha)$ is the left-tailed critical value at level α of the conditional distribution $F_t(\cdot)$ of ε_t; that is, $z_t(\alpha)$ satisfies

---

[1] VaR is a measure of extreme downside risk and is similar in methodology to lower partial moments in the earlier literature (e.g., Roy (1952)).

[2] For example, Artzner (1999) defines certain properties that a good risk measure should have and shows that VaR does not satisfy all of them. Many other kinds of risk measures have been proposed but none gained the popularity as VaR.

$F_t[z_t(\alpha)] = \alpha$. Obviously, V_t depends on not only the conditional mean μ_t and conditional variance σ_t^2 of Y_t, but also its higher order conditional moments (e.g., skewness and kurtosis). There is increasing empirical evidence that the conditional skewness and kurtosis of financial time series are time-varying (e.g., Gallant et al. (1991), Hansen (1994), Harvey and Siddique (1999, 2000) and Jondeau and Rockinger (2003)). Some time series econometric models, such as those of Gallant and Tauchen (1996), Hansen (1994) and Harvey and Siddique (2000), can capture time-varying higher order conditional moments. Note that in (2.2), Y_t is not covariance-stationary if σ_t^2 follows an integrated GARCH process (Engle and Bollerslev, 1986). In this case, the unconditional variance of $\{Y_t\}$ does not exist, but its VaR is well-defined.

2.2. Granger causality in risk

To characterize whether the occurrence of a large risk in one market (Y_{1t}) can help predict the occurrence of a large risk in another market (Y_{2t}) in the spirit of Granger (1969, 1980) causality, we now define a concept of Granger causality in the tail distributions. Put $I_{t-1} \equiv (I_{1(t-1)}, I_{2(t-1)})$, where $I_{1(t-1)} = \{Y_{1(t-1)}, \ldots, Y_{11}\}$ and $I_{2(t-1)} = \{Y_{2(t-1)}, \ldots, Y_{21}\}$ are the information sets available at time $t - 1$ for two time series respectively. Suppose

$$\mathbb{H}_1^0 : P\left(Y_{1t} < -V_{1t}|I_{1(t-1)}\right) = P\left(Y_{1t} < -V_{1t}|I_{t-1}\right) \quad a.s.,$$

we say that the time series $\{Y_{2t}\}$ *does not* Granger-cause the time series $\{Y_{1t}\}$ in risk at level α with respect to information set I_{t-1}. On the other hand, if

$$\mathbb{H}_1^A : P\left(Y_{1t} < -V_{1t}|I_{1(t-1)}\right) \neq P\left(Y_{1t} < -V_{1t}|I_{t-1}\right), \tag{2.4}$$

we say that the time series $\{Y_{2t}\}$ Granger-causes the time series $\{Y_{1t}\}$ in risk at level α with respect to I_{t-1}. In this case, the information of the occurrence of a risk in $\{Y_{2t}\}$ can be used to predict the occurrence of a future risk in $\{Y_{1t}\}$. In practice, level α can be determined by regulators or practitioners, depending on their objective function or risk attitude.

In time series econometrics, the most commonly used Granger causality concept is Granger causality in mean, which was first introduced in Granger (1969). Granger et al. (1986, p. 2) also introduce a concept of Granger causality in variance, which can be used to investigate volatility spillover between financial markets (Engle et al., 1990; Cheung and Ng, 1996), or between macroeconomic time series (Granger et al., 1986). Here, the concept of Granger causality in risk focuses on the comovements between the left tails of two distributions, which is more suitable than the concept of Granger causality in variance in characterizing extreme downside risk spillover between financial markets, because as pointed out earlier volatility is a two-sided risk measure and it cannot capture heavy tails due to jumps. Granger causality in risk can arise not only from comovements in mean and in variance, but also from the comovements in higher order conditional moments (e.g., skewness and kurtosis). Therefore, it may rise even in the absence of Granger causality in mean and in variance.

Granger (1980) introduces a general Granger causality in terms of the entire conditional probability distribution $P(Y_{1t} \leq y|I_{1(t-1)}) \neq P(Y_{1t} \leq y|I_{t-1})$ for all $y \in (-\infty, \infty)$. Our concept of Granger causality in risk is more closely related to this general Granger causality, but again we only focus on left tail probabilities, which are more relevant to large downside market risks.

On the other hand, our concept of Granger causality in risk in (2.4) can be extended to define the following concept of Granger causality in expected shortfall:

$$\mathbb{H}_1^A : E\left(Y_{1t}|Y_{1t} < -V_{1t}, I_{1(t-1)}\right) \neq E\left(Y_{1t}|Y_{1t} < -V_{1t}, I_{t-1}\right). \tag{2.5}$$

The expected shortfall has been argued in the literature to be a better measure for downside market risk than VaR. It is straightforward to extend our procedures developed below to test (2.5), but for the sake of space we defer it to subsequent research.

3. Method and test statistics

To develop tests for Granger causality in risk, we first formulate our hypotheses $\mathbb{H}_1^0$ versus $\mathbb{H}_1^A$ as hypotheses on Granger causality in mean, after a proper transformation of $\{Y_{1t}, Y_{2t}\}$. Define the risk indicator

$$Z_{lt} \equiv \mathbf{1}\left(Y_{lt} < -V_{lt}\right), \quad l = 1, 2$$

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

$$\mathbb{H}_1^0 : E\left(Z_{1t}|I_{1(t-1)}\right) = E\left(Z_{1t}|I_{t-1}\right) \quad a.s. \tag{3.1}$$

versus

$$\mathbb{H}_1^A : E\left(Z_{1t}|I_{1(t-1)}\right) \neq E\left(Z_{1t}|I_{t-1}\right). \tag{3.2}$$

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_{lt} 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 $\mathbb{H}_1^0$ 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

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

where $\rho(j) \equiv \text{corr}(Z_{1t}, Z_{2(t-j)})$. 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 $\mathbb{H}_1^0$. In this paper, we use $f(\omega)$, which has a number of appealing features, as will be seen below. Under $\mathbb{H}_1^0$, $\rho(j) = 0$ for all $j > 0$ and as a consequence, $f(\omega)$ becomes

$$f_1^0(\omega) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{0} \rho(j)e^{-ij\omega}, \quad \omega \in [-\pi, \pi].$$

Thus, we can compare $f(\omega)$ and $f_1^0(\omega)$ to test $\mathbb{H}_1^0$. Any nontrivial difference between them is evidence against $\mathbb{H}_1^0$.

It may be noted that the history of the risk indicators $\{Z_{2s}, s < t\}$ 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 $\{Z_{2s}, s < t\}$ 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 $f_1^0(\omega)$ 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

$$V_{lt}(\theta_l) \equiv V_l(I_{l(t-1)}, \theta_l), \quad l = 1, 2,$$

is a parametric VaR model for V_{lt}, where θ_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 (CAViaR) models. Suppose further we have a random sample $\{Y_{1t}, Y_{2t}\}_{t=1}^T$ of size T, and an estimator $\hat{\theta}_l$. Put

$$\hat{Z}_{lt} \equiv Z_{lt}(\hat{\theta}_l), \quad l = 1, 2,$$

where $Z_{lt}(\theta_l) \equiv \mathbf{1}[Y_{lt} < -V_{lt}(\theta_l)]$. Then we can define the sample cross-covariance function between $\{\hat{Z}_{1t}\}$ and $\{\hat{Z}_{2t}\}$,

$$\hat{C}(j) \equiv \begin{cases} T^{-1} \sum_{t=1+j}^{T} (\hat{Z}_{1t} - \hat{\alpha}_1)(\hat{Z}_{2(t-j)} - \hat{\alpha}_2), & 0 \le j \le T-1, \\ T^{-1} \sum_{t=1-j}^{T} (\hat{Z}_{1(t+j)} - \hat{\alpha}_1)(\hat{Z}_{2t} - \hat{\alpha}_2), & 1 - T \le j < 0, \end{cases} \tag{3.3}$$

where $\hat{\alpha}_l \equiv T^{-1} \sum_{t=1}^T \hat{Z}_{lt}$. The sample cross-correlation function between $\{\hat{Z}_{1t}\}$ and $\{\hat{Z}_{2t}\}$ is

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

where $\hat{S}_l^2 \equiv \hat{\alpha}_l(1 - \hat{\alpha}_l)$ is the sample variance of $\{\hat{Z}_{lt}\}$.[3] With $\{\hat{\rho}(l)\}$, the kernel estimators for the cross-spectral densities $f(\omega)$ and $f_1^0(\omega)$ can be given as follows:

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

$$\hat{f}_1^0(\omega) \equiv \frac{1}{2\pi} \sum_{j=1-T}^{0} k(j/M)\hat{\rho}(j)e^{-ij\omega}, \tag{3.5}$$

To compare $\hat{f}(\omega)$ and $\hat{f}_1^0(\omega)$, we use the quadratic form

$$L^2\left(\hat{f}, \hat{f}_1^0\right) \equiv 2\pi \int_{-\pi}^{\pi} \left|\hat{f}(\omega) - \hat{f}_1^0(\omega)\right|^2 d\omega = \sum_{j=1}^{T-1} k^2(j/M)\hat{\rho}^2(j), \tag{3.6}$$

where the second equality follows by Paserval's identity. We need not calculate numerical integrations over frequency ω. Our test statistic for $\mathbb{H}_1^0$ versus $\mathbb{H}_1^A$ is a standardized version of the quadratic form:

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

where the centering and standardization constants are

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

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

The factors $(1 - j/T)$ and $[1 - (j + 1)/T]$ are finite sample corrections. They could be replaced by 1. Both $C_{1T}(M)$ and $D_{1T}(M)$ are approximately the mean and variance of the quadratic form $TL^2(\hat{f}, \hat{f}_0)$. 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

$$k_T(z) = \mathbf{1}(|z| \le 1), \tag{3.8}$$

where $\mathbf{1}(\cdot)$ is the indicator function. This yields the following test statistic[4]

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

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

$$\hat{Z}_{1t} = \alpha_0 + \sum_{j=1}^{M} \alpha_j \hat{Z}_{2(t-j)} + u_t, \tag{3.10}$$

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 $\hat{Z}_{1t}$ because $\{Z_{1t}\}$ is a sequence of i.i.d. Bernoulli random variables under the null hypothesis. For the estimated $\hat{Z}_{1t}$, (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

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

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 $\mathbb{H}_1^0$. This procedure is simple and intuitive. It could be shown that $Q_{1REG}(M)$ is asymptotically equivalent to $Q_{1TRUN}(M)$ under $\mathbb{H}_1^0$. When M is large, however, both $Q_{1TRUN}(M)$ and $Q_{1REG}(M)$ 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_{1t} on the $Z_{2(t-j)}$ 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.[5] In contrast, the $Q_{1TRUN}(M)$ and $Q_{1REG}(M)$ tests are not fully efficient when M is large. See Sections 4 and 6 for more discussion and simulation studies.

---

[3] We can replace $\hat{\alpha}_l$ with α. This does not affect the asymptotic distribution of the proposed test statistic under $\mathbb{H}_1^0$.

[4] For $k(z) = \mathbf{1}(|z| \le 1)$, we have $C_{1T}(M) = M[1 - (M + 1)/(2T)]$ and $D_{1T}(M) = 2M[1 - (M+2)/T + (M+1)(M+2)/(3T^2)]$. Under suitable conditions on M, we can conveniently approximate $C_{1T}(M)$ and $D_{1T}(M)$ 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 orders (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 spillover at the stop-loss level.

4. Asymptotic theory

We now derive the limit distribution of the $Q_1(M)$ test under $\mathbb{H}_1^0$. Its derivation is complicated by the fact that we do not observe the true parameter values $\{\theta_l^0\}$ and have to estimate them. Parameter estimation uncertainty in $\{\hat{\theta}_l\}$ 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 $\hat{\theta}_l$ asymptotically. Intuitively, $\hat{\theta}_l$ converges to θ_l^0 faster than the nonparametric estimators $\hat{f}(\omega)$ and $\hat{f}_1^0(\omega)$ to $f(\omega)$ and $f_1^0(\omega)$ respectively. As a consequence, the limit distribution of $Q_1(M)$ is solely determined by the kernel estimators $\hat{f}(\omega)$ and $\hat{f}_1^0(\omega)$. One can proceed as if θ_l^0 were known and equal to $\hat{\theta}_l$. Thus, replacing θ_l^0 with $\hat{\theta}_l$ 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 $\{\hat{\theta}_l\}$ 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 $\{Y_{lt}\}$, the VaR models $V_{lt}(\theta_l)$, the parameter estimators $\hat{\theta}_l$, and the kernel function $k(\cdot)$.

Assumption 1. For $l = 1, 2$, $\{Y_{lt}\}$ is a stochastic time series process with unknown twice continuously differentiable conditional distribution function $F_{lt}(y) \equiv P(Y_{lt} \leq y | I_{t-1})$, where $y \in \mathbb{R}$ and I_{t-1} is the information available at time $t - 1$.

Assumption 2. For $\theta_l \in \Theta_l \subset R^{d_l}$, where d_l is a positive integer, $l = 1, 2$, $V_{lt}(\theta_l) \equiv V_l(I_{l(t-1)}, \theta_l)$ is a VaR model at level $\alpha \in (0, 1)$ such that (i) for each $\theta_l \in \Theta_l$, $V_{lt}(\theta_l)$ is a measurable function of $I_{l(t-1)}$; (ii) with probability 1, $V_{lt}(\cdot)$ is twice continuously differentiable with respect to $\theta_l \in \Theta_l$, with $\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E \sup_{\theta_l \in \Theta_l} \| \frac{\partial}{\partial \theta} F_{lt}[-V_{lt}(\theta_l)] \|^4 < \infty$ and $\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E \sup_{\theta_l \in \Theta_l} \| \frac{\partial^2}{\partial \theta \partial \theta'} F_{lt}[-V_{lt}(\theta_l)] \|^2 < \infty$.

Assumption 3. For $l = 1, 2$, there exists some $\theta_l^0 \in \Theta_l$ such that (i) $P[Y_{lt} < -V_{lt}(\theta_l^0) | I_{l(t-1)}] = \alpha$ a.s.; (ii) the risk indicator $Z_{2t}(\theta_2^0) \equiv \mathbf{1}[Y_{2t} < -V_{2t}(\theta_2^0)]$ depends on an arbitrarily long but finite length of the current and past history of $\{Z_{1s}(\theta_1^0) \equiv \mathbf{1}[Y_{1s} < -V_{1s}(\theta_1^0)], s \leq t\}$.

Assumption 4. $T^{\frac{1}{2}}(\hat{\theta}_l - \theta_l^*) = O_P(1)$ for $l = 1, 2$, where $\theta_l^* \equiv p \lim \hat{\theta}_l$ and $\theta_l^* = \theta_l^0$ under the null hypothesis of interest.

Assumption 5. Put $S_t(\theta) \equiv [S_{1t}(\theta_1)', S_{2t}(\theta_2)']'$ and $Z_t(\theta) \equiv [Z_{1t}(\theta_1), Z_{2t}(\theta_2)]'$, where $S_{lt}(\theta_l) \equiv \frac{\partial}{\partial \theta_l} F_{lt}[-V_{lt}(\theta_l)]$. Then $\{S_t(\theta^*)', Z_t(\theta^*)'\}'$ is a fourth order stationary process such that

(i) $\sum_{j=0}^{\infty} \| \Gamma(j) \| \leq \Delta$, where $\Gamma(j) \equiv \text{cov}[S_t(\theta^*), Z_{t-j}(\theta^*)']$;
(ii) $\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \| \kappa_0(j, k, l) \| \leq \Delta$, where $\kappa_0(j, k, l)$ is the fourth order cumulant of the joint distribution of $\{S_t(\theta^*)' - ES_t(\theta^*)', Z_{t-j}(\theta^*)' - EZ_{t-j}(\theta^*)', S_{t-k}(\theta^*)' - ES_{t-k}(\theta^*)', Z_{t-l}(\theta^*)' - EZ_{t-l}(\theta^*)'\}'$.

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 $\int_{-\infty}^{\infty} k^2(z) dz < \infty$.

Assumption 1 is a standard regularity condition on the bivariate data generating process for $\{Y_{1t}, Y_{2t}\}$. We allow for some covariance-nonstationary processes $\{Y_{lt}\}$. 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 $V_{lt}(\theta_l)$. 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_{lt} 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 $\mathbb{H}_1^0$. 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 $\mathbb{H}_1^0$, although $\{Z_{2s}(\theta_2^0), s < t\}$ does not affect $Z_{1t}(\theta_1^0)$, $Z_{2t}(\theta_2^0)$ may depend on the current and past history of $\{Z_{1s}(\theta_1^0), s \leq t\}$. In other words, there may exist instantaneous Granger causality between $Z_{1t}(\theta_1^0)$ and $Z_{2t}(\theta_2^0)$ and/or Granger causality from $\{Z_{1s}(\theta_1^0), s < t\}$ to $Z_{2t}(\theta_2^0)$ under $\mathbb{H}_0^1$. For simplicity, Assumption 3(ii) assumes that $Z_{2t}(\theta_2^0)$ depends on an arbitrarily long but finite history of $\{Z_{1s}(\theta_1^0), s \leq t\}$. It is possible to allow $Z_{2t}(\theta_2^0)$ to depend on the entire past history of $\{Z_{1s}(\theta_1^0), s \leq t\}$, with a suitable rate condition on the dependence of $Z_{2t}(\theta_2^0)$ on the history of $\{Z_{1s}(\theta_1^0), s \leq t\}$, 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, $\hat{\theta}_l$ need not be asymptotically most efficient; any $\sqrt{T}$-consistent estimator of θ_l^0 suffices under $\mathbb{H}_0^1$. An example is Engle and Manganelli's (2004) regression quantile estimator. We do not require parameter estimation consistency under the alternative $\mathbb{H}_A^1$. Thus, the probability limit θ_l^* may not coincide with θ_l^0 under $\mathbb{H}_A^1$. Moreover, we need not know the asymptotic expansion of $\{\hat{\theta}_l\}$. 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 $\{S_t(\theta^*)', Z_t(\theta^*)'\}'$. Under $\mathbb{H}_0$, we have $Z_{lt}(\theta^*) = Z_{lt}(\theta_l^0) = Z_{lt}$. Thus, $\{Z_{1t}(\theta_1^*)\}$ is an i.i.d. Bernoulli(α) sequence and $Z_{1t}(\theta_1^*)$ is independent of $\{Z_{2s}(\theta_2^*), s < t\}$. However, the derivative $\frac{\partial}{\partial \theta_1} E[Z_{1t}(\theta_1^*) | I_{t-1}] = S_{1t}(\theta_1^*)$ generally depends on I_{t-1} even under $\mathbb{H}_1^0$. 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 $\hat{f}_1^0(\omega)$ 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 $Q_1(M)$ under $\mathbb{H}_1^0$.

Theorem 1. *Suppose Assumptions 1–6 hold and $M = cT^\nu$, where $0 < c < \infty, 0 < \nu < \frac{1}{2}, \nu < \min(\frac{2}{d-2}, \frac{3}{d-1})$ if $d \equiv \max(d_1, d_2) > 2$, and d_l is the dimension of θ_l. Then $Q_1(M) \to^d N(0, 1)$ under $\mathbb{H}_1^0$.*

The condition that $\nu < \min(\frac{2}{d-2}, \frac{3}{d-1})$ 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 $\{\hat{\theta}_l\}$. 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 $Z_{lt}(\theta_l)$ is not differentiable with respect to θ_l. From the proof of Theorem 1 (see Theorem A.1 in the Appendix), we find that parameter estimation uncertainty in $\hat{\theta}_l$ has no impact on the limit distribution of $Q_1(M)$. This occurs because $\hat{\theta}_l$ converges to θ_l^0 faster than the kernel cross-spectral estimators $\hat{f}(\omega)$ and $\hat{f}_1^0(\omega)$ to $f(\omega)$ and $f_1^0(\omega)$ respectively.

To understand the intuition why $Q_1(M)$ is asymptotically $N(0, 1)$, we consider the $Q_{1\text{TRUN}}(M)$ test that is based on the truncated kernel in (3.8). First, suppose θ_l^0 were known. Then as $T \to \infty$, we have $\sqrt{T}\hat{\rho}(j) \to^d N(0, 1)$ at each lag $j > 0$ and $\text{cov}[\sqrt{T}\hat{\rho}(j_1), \sqrt{T}\hat{\rho}(j_2)] \to 0$ for any $j_1 \neq j_2$ under $\mathbb{H}_0^1$. Consequently, $\sum_{j=1}^M T\hat{\rho}^2(j)$, being the sum of M asymptotically independent χ_1^2 random variables, are asymptotically distributed as χ_M^2. By the well-known normal approximation of χ_M^2 when M is large, we obtain the asymptotic normality of $Q_{1\text{TRUN}}(M)$. The impact of parameter estimation uncertainty in $\hat{\theta}_l$ 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 $Q_1(M)$ test under $\mathbb{H}_1^A$, we impose a condition on the cross-correlation $\rho(j)$ and a fourth order cumulant condition.

Assumption 7. *Let $\rho(j) = \text{cov}[Z_{1t}(\theta^*), Z_{2(t-j)}(\theta^*)]$. Then (i) $\sum_{j=1}^\infty \rho^2(j) < \infty$; (ii) $\sum_{j=-\infty}^\infty \sum_{k=-\infty}^\infty \sum_{l=-\infty}^\infty |\kappa_1(j, k, l)| < \infty$, where $\kappa_1(j, k, l)$ is the fourth order cumulant of the joint distribution of $\{Z_{1t}(\theta_1^*) - EZ_{1t}(\theta_1^*), Z_{2(t-j)}(\theta_1^*) - EZ_{2(t-j)}(\theta_2^*), Z_{1(t-k)}(\theta_1^*) - EZ_{1(t-k)}(\theta_1^*), Z_{2(t-l)}(\theta_2^*) - EZ_{2(t-l)}(\theta_2^*)\}$.*

Assumption 7(i) implies that the dependence of $Z_{1t}(\theta_1^*)$ on $\{Z_{2s}(\theta_2^*), s < t\}$ 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 slow hyperbolic rate. We do not impose any condition on the dependence of $Z_{2t}(\theta_2^*)$ on $\{Z_{1s}(\theta_1^*), s \leq t\}$, because we only check the one-way Granger causality from $\{Z_{2t}(\theta_2^*)\}$ to $\{Z_{1t}(\theta_1^*)\}$ with respect to I_{t-1}.

Theorem 2. *Suppose Assumptions 1–7 hold and $M = cT^\nu$ for $0 < c < \infty$ and $0 < \nu < 1$. Then $(M^{\frac{1}{2}}/T)Q_1(M) \to^p [2\int_0^\infty k^4(z) dz]^{-\frac{1}{2}} \sum_{j=1}^\infty \rho^2(j)$ under $\mathbb{H}_1^A$.*

Thus, for any sequence of constants, $K_T = o(T^{1-\frac{\nu}{2}})$, we have $P[Q_1(M) > K_T] \to 1$ whenever $\rho(j) \neq 0$ for some $j > 0$. In other words, the $Q_1(M)$ test has asymptotic unit power at any given significance level whenever $\rho(j) \neq 0$ for some $j > 0$. Because $Q_1(M) \to +\infty$ whenever $\rho(j) \neq 0$ 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 $\mathbb{H}_1^A$ is weaker than that under $\mathbb{H}_1^0$.

Hong (1996) shows that over a class of kernels

$$\mathbb{K}(\tau) = \left\{ k(\cdot) : k(0) = 1, \int_{-\infty}^\infty k(z) dz < \infty, \right.$$
$$\left. k_2 \equiv \lim_{z \to 0} \frac{1 - k(z)}{z^2} \in (0, \infty) \right\}$$

which includes the Parzen and Quadratic-Spectral kernels (but not the Bartlett kernel), the Daniell kernel $k_D(z) = \sin(\pi z)/\pi z$ minimizes $\int_0^\infty k^4(z) dz$. Using this result, it can be shown that the Daniell kernel $k_D(\cdot)$ maximizes the asymptotic power of $Q_1(M)$ in terms of Bahadur's (1960) asymptotic efficiency criterion. Of course, the relative efficiency of nonuniform kernels in $\mathbb{K}(\tau)$ 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 $|\rho(j)|$ 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 $\{Y_{1t}\}$ nor $\{Y_{2t}\}$ Granger-causes each other in risk at level α with respect to I_{t-1}. The hypotheses of interest are

$$\mathbb{H}_2^0 : P\left(Y_{lt} < -V_{lt}|I_{l(t-1)}\right)$$
$$= P\left(Y_{lt} < -V_{lt}|I_{t-1}\right) a.s. \quad \text{for both } l = 1, 2 \quad (5.1)$$

versus

$$\mathbb{H}_2^A : P\left(Y_{lt} < -V_{lt}|I_{l(t-1)}\right)$$
$$\neq P\left(Y_{lt} < -V_{lt}|I_{t-1}\right) \quad \text{for at least one } l. \quad (5.2)$$

Using the risk indicator $\{Z_{lt}\}$, we can write these hypotheses as

$$\mathbb{H}_2^0 : E\left(Z_{lt}|I_{l(t-1)}\right) = E\left(Z_{lt}|I_{t-1}\right) a.s. \quad \text{for both } l = 1, 2.$$

versus

$$\mathbb{H}_2^A : E\left(Z_{lt}|I_{l(t-1)}\right) \neq E\left(Z_{lt}|I_{t-1}\right)$$

for at least one l, where $l = 1, 2$.

Under $\mathbb{H}_2^0$, 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:

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

where $\rho(0)$ is nonzero when there exists instantaneous causality between Z_{1t} and Z_{2t}. A consistent estimator for $f_2^0(\omega)$ is

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

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

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

where the centering and scaling factors are

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

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

Note that $D_{2T}(M)$ involves the cross-correlation estimator $\hat{\rho}(0)$, which has taken into account the possible instantaneous correlation between Z_{1t} and Z_{2t} under $\mathbb{H}_2^0$.

The $Q_2(M)$ statistic is asymptotically $N(0, 1)$ under $\mathbb{H}_2^0$, as is stated below.

Theorem 3. *Suppose Assumptions 1–3(i) and 4–6 hold, and $M = cT^\nu$, where $0 < \nu < \frac{1}{2}$, $\nu < \max(\frac{2}{d-2}, \frac{3}{d-1})$ if $d \equiv \max(d_1, d_2) > 2$, and d_l is the dimension of θ_l, $l = 1, 2$. Then $Q_2(M) \to^d N(0, 1)$ under $\mathbb{H}_2^0$.*

We do not need Assumption 3(ii) here, because under $\mathbb{H}_2^0$, neither $\{Z_{1t}\}$ nor $\{Z_{2t}\}$ Granger-causes each other with respect to I_{t-1}. Nevertheless, we allow for instantaneous Granger causality between Z_{1t} and Z_{2t} under $\mathbb{H}_2^0$.

To study the asymptotic behavior of the $Q_2(M)$ test under $\mathbb{H}_2^A$, we strengthen Assumption 7 slightly to cover two-way cross-correlations.

Assumption 8. (i) $\sum_{j=-\infty}^{\infty} \rho^2(j) < \infty$ and (ii) $\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\kappa_1(j, k, l)| < \infty$, where $\rho(j)$ and $\kappa_1(j, k, l)$ are as in Assumption 7.

Theorem 4. *Suppose Assumptions 1, 2, 4–6 and 8 hold, and $M = cT^\nu$ for $0 < c < \infty$ and $0 < \nu < 1$. Then $\{\sqrt{1 + \rho^4(0)} M^{\frac{1}{2}}/T\} Q_2(M) \to^p [2 \int_{-\infty}^{\infty} k^4(z) dz]^{-\frac{1}{2}} \sum_{|j|=1}^{\infty} \rho^2(j)$ under $\mathbb{H}_2^A$.*

Thus, whenever there exists Granger causality in risk between $\{Y_{1t}\}$ and $\{Y_{2t}\}$ with respect to I_{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 $\mathbb{H}_2^0$.

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 expected to perform similarly.

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

$$\begin{cases} Y_{lt} = \beta_{l1} Y_{1t-1} + \beta_{l2} Y_{2t-1} + u_{lt}, & l = 1, 2, \\ u_{lt} = \sigma_{lt} \varepsilon_{lt}, \\ \sigma_{lt}^2 = \gamma_{l0} + \gamma_{l1} \sigma_{lt-1}^2 + \gamma_{l2} u_{1t-1}^2 + \gamma_{l3} u_{2t-1}^2, \\ \varepsilon_{lt} \sim m.d.s. (0, 1). \end{cases} \quad (6.1)$$

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

NULL [No Granger Causality in Risk]:

$$\begin{cases} (\beta_{11}, \beta_{12}, \gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{13}) = (0.5, 0, 0.1, 0.6, 0.2, 0), \\ (\beta_{21}, \beta_{22}, \gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{23}) = (0, 0.5, 0.1, 0.6, 0, 0.2). \end{cases}$$

ALTER1 [Granger Causality in Risk from Mean]:

$$\begin{cases} (\beta_{11}, \beta_{12}, \gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{13}) = (0.5, 0.2, 0.1, 0.6, 0.2, 0), \\ (\beta_{21}, \beta_{22}, \gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{23}) = (0, 0.5, 0.1, 0.6, 0, 0.2). \end{cases}$$

ALTER2 [Granger Causality in Risk from Variance]:

$$\begin{cases} (\beta_{11}, \beta_{12}, \gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{13}) = (0.5, 0, 0.1, 0.5, 0.2, 0.7), \\ (\beta_{21}, \beta_{22}, \gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{23}) = (0, 0.5, 0.1, 0.5, 0.2, 0). \end{cases}$$

Under NULL, there is no Granger causality in risk between $\{Y_{1t}\}$ and $\{Y_{2t}\}$ with respect to I_{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 $\{\varepsilon_{lt}\}$ following a generalized Student-t-distribution with time-varying shape parameters. More specifically, Hansen's (1994) generalized t density is given by

$$g_\varepsilon(\varepsilon|\lambda, \eta) = \begin{cases} bc \left[1 + \frac{1}{\eta - 2} \left(\frac{b\xi + a}{1 - \lambda} \right)^2 \right]^{-(\eta+1)/2} \\ \quad \text{if } \varepsilon < -a/b, \\ bc \left[1 + \frac{1}{\eta - 2} \left(\frac{b\xi + a}{1 + \lambda} \right)^2 \right]^{-(\eta+1)/2} \\ \quad \text{if } \varepsilon \geq -a/b, \end{cases}$$

where

$$a \equiv 4\lambda c \frac{\eta - 2}{\eta - 1}, \qquad b^2 \equiv 1 + 3\lambda^2 - a^2,$$

$$c \equiv \frac{\Gamma((\eta + 1)/2)}{\sqrt{\pi (\eta - 2)} \Gamma(\eta/2)}.$$

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

$$\begin{cases} \lambda_{lt} = \frac{1 - \exp(\tilde{\lambda}_{it})}{1 + \exp(\tilde{\lambda}_{it})}, \\ \eta_{lt} = 4 + \frac{4}{1 + \exp(\tilde{\eta}_{it})}, \end{cases}$$

where

$$\begin{cases} \tilde{\lambda}_{lt} = \delta_{l0} + \delta_{l1} u_{1t-1} + \delta_{l2} u_{2t-1} + \delta_{l3} \tilde{\lambda}_{1t-1} + \delta_{l4} \tilde{\lambda}_{2t-1}, \\ \tilde{\eta}_{lt} = \tau_{l0} + \tau_{l1} u_{1t-1} + \tau_{l2} u_{2t-1} + \tau_{l3} \tilde{\eta}_{1t-1} + \tau_{l4} \tilde{\eta}_{2t-1}. \end{cases}$$

We use the following parameter combinations:

ALTER3 [Granger Causalities in Risk from Skewness and Kurtosis]:

$$\begin{cases} (\beta_{11}, \beta_{12}, \gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{13}) = (0.3, 0, 0.1, 0.5, 0.2, 0), \\ (\beta_{21}, \beta_{22}, \gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{23}) = (0, 0.5, 0.1, 0.6, 0, 0.2), \\ (\delta_{10}, \delta_{11}, \delta_{12}\delta_{13}\delta_{14}, \tau_{10}, \tau_{11}, \tau_{12}, \tau_{13}, \tau_{14}) \\ \quad = (-0.2, 1, -5, 0, -0.9, -0.2, 1, -5, 0, -0.9), \\ (\delta_{20}, \delta_{21}, \delta_{22}\delta_{23}\delta_{24}, \tau_{20}, \tau_{21}, \tau_{22}, \tau_{23}, \tau_{24}) \\ \quad = (-0.2, 0, 1, 0, 0, -0.2, 0, 1, 0, 0). \end{cases}$$

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 I_{t-1}. To our knowledge, the financial econometric literature has been focusing on spillover in mean and in variance; no study on

Table 1
Size at the 10% and 5% significant levels.

	M	5		10		15		20		25		30	
	T	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Q_{1DAN}^{10}	500	9.7	6.9	10.5	6.4	11.6	7.2	12.3	6.6	11.3	6.9	12.2	6.9
	1000	9.6	6.7	10.4	6.5	10.8	6.5	10.9	6.1	10.6	5.8	10.2	5.6
	2000	9.6	6.7	10.3	6.3	10.1	6.8	10.9	6.9	10.5	7.0	11.6	7.3
Q_{1DAN}^{5}	500	9.4	6.8	10.2	7.3	10.2	7.0	10.2	6.4	10.2	7.0	10.3	7.1
	1000	7.5	6.4	10.2	6.9	11.6	7.7	11.7	7.2	11.5	6.9	11.5	7.3
	2000	8.6	5.8	10.3	7.0	9.9	7.2	11.0	7.7	11.1	7.2	11.1	6.8
Q_{1TRUN}^{10}	500	12.2	7.1	10.5	6.5	11.3	6.5	9.8	6.0	10.0	6.4	10.9	7.1
	1000	10.5	7.1	12.3	7.4	11.9	7.3	11.4	6.7	11.3	6.0	11.6	6.1
	2000	11.1	6.7	10.0	6.5	10.7	6.8	12	6.5	12.0	5.7	11.4	5.7
Q_{1TRUN}^{5}	500	11.5	7.6	12.4	7.9	12.2	8.4	11.8	8.3	12.7	8.5	11.7	7.4
	1000	9.2	6.6	11.1	6.9	12.1	7.6	11.0	6.7	11.6	7.8	11.3	6.8
	2000	10.3	6.4	9.9	6.8	10.7	7.5	10.5	6.3	10.9	6.4	10.3	5.9
Q_{1REG}^{10}	500	11.5	7.3	10.8	7.3	10.7	5.7	10.8	5.3	10.2	5.7	11.6	6.7
	1000	11.2	6.9	11.8	7.2	11.7	6.4	11.3	6.7	11.2	6.1	9.9	5.8
	2000	10.0	7.0	10.5	6.9	8.8	5.3	8.0	4.4	8.3	4.8	9.7	5.0
Q_{1REG}^{5}	500	11.3	7.6	11.3	7.3	10.0	6.4	9.3	6.6	10.4	7.0	10.4	6.6
	1000	11.1	7.4	11.0	7.8	11.5	6.8	10.2	5.6	10.0	5.9	9.9	5.9
	2000	9.6	6.3	10.4	7.0	10.5	6.2	9.5	5.3	8.9	4.3	9.1	5.3

NULL: $Y_{it} = 0.5Y_{it-1} + u_{it}$, $u_{it} = \sqrt{h_{it}}\varepsilon_{it}$, $h_{it} = 0.1 + 0.6h_{it-1} + 0.2u_{it-1}^2$, $\varepsilon_{it} \sim m.d.s.\ N(0, 1)$, $i = 1, 2$; the sample size $T = 500, 1000$ and 2000; Q_{1DAN}^{10}, Q_{1TRUN}^{10}, Q_{1REG}^{10} and Q_{1DAN}^{5}, Q_{1TRUN}^{5}, Q_{1REG}^{5} represent one-way tests for Granger causality in risk from Y_{2t} to Y_{1t} at the 10% and 5% risk levels respectively, where the subscripts DAN, TRUN and REG denote the Daniell kernel, the truncated kernel, and the regression-based tests.

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 correspond to two to eight years of daily financial data. These sample sizes may be still relatively small in view of estimating parameters involved in conditional variance and higher order conditional moments (skewness and kurtosis). 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 $(h_{l0}^*, \tilde{\lambda}_{l0}^*, \tilde{\eta}_{l0}^*) = (1/(1 - \gamma_{21} - \gamma_{23}), -0.2, 4.1)$. 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 $Q_{1TRUN}(M)$ in (3.7) and the Granger-type regression test (1969) $Q_{1REG}(M)$ 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:

$$\begin{cases} Y_{lt} = \beta_{l1}Y_{lt-1} + u_{lt}, & l = 1, 2, \\ u_{lt} = \sigma_{lt}\varepsilon_{lt}, \\ \sigma_{lt}^2 = \gamma_{l0} + \gamma_{l1}\sigma_{lt-1}^2 + \gamma_{l2}u_{lt-1}^2, \\ \varepsilon_{lt} \sim i.i.d.\ N(0, 1). \end{cases} \quad (6.2)$$

The BHHH 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 $Q_{1DAN}(M)$ in (3.7) with Daniell kernel, $Q_{1TRUN}(M)$ in (3.9) and $Q_{1REG}(M)$ tests in (3.11) at the 10% and 5% significance levels under NULL.[7] Overall, the

$Q_{1DAN}(M)$ 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 $Q_{1DAN}(M)$ test. The truncated kernel-based test $Q_{1TRUN}(M)$ performs similarly to $Q_{1DAN}(M)$ at the 10% significant level. The regression procedure $Q_{1REG}(M)$, on the contrary, tends to a bit overreject the null hypothesis at the 10% significant level. Both $Q_{1TRUN}(M)$ and $Q_{1REG}(M)$ 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 $Q_{1DAN}(M)$ test has good power against ALTER1 and it becomes more powerful as T increases. Given each sample size T; the power of $Q_{1DAN}(M)$ declines as the lag order M increases, but not dramatically (which is apparently due to the downward weighting of $k^2(\cdot)$). The $Q_{1DAN}(M)$ test with the 10% shortfall probability (or risk level) is more powerful than the $Q_{1DAN}(M)$ 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, $Q_{1TRUN}(M)$ and $Q_{1REG}(M)$ 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 $Q_{1TRUN}(M)$ 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 $Q_{1DAN}(M)$ test has good power against ALTER2. As under ALTER1, the power of $Q_{1DAN}(M)$ declines as M increases, but not dramatically. Again, $Q_{1TRUN}(M)$ and $Q_{1REG}(M)$ 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 $Q_{1DAN}(M)$ 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.

	M	5		10		15		20		25		30	
	T	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Q_{1DAN}^{10}	500	58.0	52.1	52.6	47.8	50.2	44.2	47.5	40.6	44.7	37.6	43.2	35.7
	1000	80.2	76.0	74.9	69.2	70.5	63.5	66.8	58.8	63.1	54.5	60.0	51.3
	2000	96.3	94.6	94.4	92.8	92.7	89.9	90.7	87.3	88.5	84.0	86.6	82.2
Q_{1DAN}^{5}	500	43.5	38.6	41.2	35.4	39.5	32.8	37.7	30.3	35.6	27.6	33.3	25.7
	1000	56.9	52.7	54.5	48.4	51.7	44.7	48.2	41.6	46.1	38.6	45.0	36.1
	2000	77.9	73.7	75.0	69.0	70.1	64.1	67.7	60.3	64.9	57.7	61.8	53.9
Q_{1TRUN}^{10}	500	48.6	42.4	41.6	33.8	37.5	28.2	33.7	25.2	32.6	23.2	30.9	21.9
	1000	69.7	62.6	58.5	51.5	52.4	43.9	49.0	39.1	46.7	35.1	41.9	31.0
	2000	92.3	89.8	86.6	81.7	81.8	75.1	77.2	70.8	75.7	67.2	73.1	63.9
Q_{1TRUN}^{5}	500	38.7	33.0	34.4	25.2	28.5	22.2	25.5	18.4	23.8	18.8	23.7	17.1
	1000	50.9	44.7	43.5	35.0	38.9	31.2	35.8	28.9	35.6	25.7	32.4	24.5
	2000	70.8	63.2	60.4	54.7	55.0	46.9	52.4	42.3	48.7	39.5	46.9	38.2
Q_{1REG}^{10}	500	48.1	37.5	40.8	31.7	34.7	27.3	30.9	24.0	27.7	21.9	27.2	21.5
	1000	72.8	65.9	62.7	53.7	55.2	46.1	49.7	40.0	44.4	34.9	41.6	32.8
	2000	93.2	90.8	87.1	81.3	82.1	75.7	77.7	69.8	73.1	65.2	69.5	59.8
Q_{1REG}^{5}	500	42.6	31.4	32.1	25.2	27.3	21.0	22.6	17.0	19.8	16.2	18.3	14.6
	1000	49.6	43.4	43.7	36.3	37.5	30.4	34.5	27.2	32.2	24.1	30.0	22.7
	2000	72.1	65.4	61.9	55.5	55.4	47.5	50.8	42.3	48.5	39.9	46.9	37.0

ALTER1: $Y_{1t} = 0.5Y_{1t-1} + 0.2Y_{2t-1} + u_{1t}$, $Y_{2t} = 0.5Y_{2t-1} + u_{2t}$, $u_{it} = \sqrt{h_{it}}\varepsilon_{it}$, $h_{it} = 0.1 + 0.6h_{it-1} + 0.2u_{it-1}^2$, $\varepsilon_{it} \sim m.d.s. N(0, 1)$, $i = 1, 2$; the sample size $T = 500$, 1000 and 2000; Q_{1DAN}^{10}, Q_{1TRUN}^{10}, Q_{1REG}^{10} and Q_{1DAN}^{5}, Q_{1TRUN}^{5}, Q_{1REG}^{5} represent one-way tests for Granger causality in risk from Y_{2t} to Y_{1t} at the 10% and 5% risk levels respectively, where the subscripts DAN, TRUN and REG denote the Daniell kernel, the truncated kernel, and the regression-based tests.

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

	M	5		10		15		20		25		30	
	T	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Q_{1DAN}^{10}	500	42.8	36.9	43.1	36.4	40.1	33.4	38.3	30.3	35.5	27.7	33.4	26.1
	1000	59.3	53.5	59.9	52.9	56.5	49.0	52.3	44.4	47.9	41.7	46.2	40.0
	2000	86.6	83.0	84.8	80.4	82.4	77.3	79.3	73.6	76.6	71.5	74.7	68.4
Q_{1DAN}^{5}	500	51.4	46.0	53.3	46.9	52.0	45.1	48.6	42.0	46.3	38.0	44.4	35.9
	1000	67.2	62.2	65.6	59.4	62.2	56.3	59.9	52.8	57.3	50.3	55.3	48.2
	2000	84.5	85.4	82.2	83.5	80.6	80.2	78.3	77.3	75.9	74.0	73.7	72.0
Q_{1TRUN}^{10}	500	39.8	33.0	33.8	25.0	30.2	22.6	27.2	19.0	22.9	16.5	23.6	16.9
	1000	55.3	47.8	47.1	38.2	42.5	32.6	37.3	27.5	34.7	25.5	32.0	23.4
	2000	81.4	76.2	75.5	68.2	68.4	60.5	64.2	53.5	59.3	48.5	55.5	43.6
Q_{1TRUN}^{5}	500	50.3	44.5	44.7	36.3	38.9	30.5	34.5	26.9	32.4	26.6	33.2	25.1
	1000	62.4	56.3	55.1	48.5	49.7	42.3	46.8	37.0	43.8	34.7	41.5	32.0
	2000	85.0	80.2	77.0	71.7	72.3	66.0	68.6	60.0	64.6	55.9	60.8	51.9
Q_{1REG}^{10}	500	40.0	33.5	33.3	25.2	28.7	20.5	24.4	16.6	20.7	14.3	20.6	13.7
	1000	57.4	50.8	50.6	42.1	42.7	33.8	37.9	28.5	34.7	27.1	32.3	24.5
	2000	82.7	77.3	73.6	67.0	67.1	57.2	60.3	50.2	55.7	44.6	52.2	40.8
Q_{1REG}^{5}	500	50.1	44.3	42.9	34.1	37.0	29.4	32.2	24.4	30.5	24.0	29.8	22.5
	1000	60.9	55.7	57.6	48.8	50.5	42.9	47.0	38.3	44.4	33.8	41.0	31.7
	2000	84.7	80.1	76.3	70.7	72.4	65.8	69.0	60.8	64.1	56.7	60.8	53.5

ALTER2: $Y_{it} = 0.5Y_{it-1} + u_{it}$, $u_{it} = \sqrt{h_{it}}\varepsilon_{it}$, , $\varepsilon_{it} \sim m.d.s. N(0, 1)$, $i = 1, 2$, $h_{1t} = 0.1 + 0.5h_{1t-1} + 0.2u_{1t-1}^2 + 0.7u_{2t-1}^2$, $h_{2t} = 0.1 + 0.5h_{2t-1} + 0.2u_{2t-1}^2$; the sample size $T = 500$, 1000 and 2000; Q_{1DAN}^{10}, Q_{1TRUN}^{10}, Q_{1REG}^{10} and Q_{1DAN}^{5}, Q_{1TRUN}^{5}, Q_{1REG}^{5} represent one-way tests for Granger causality in risk from Y_{2t} to Y_{1t} at the 10% and 5% risk levels respectively, where the subscripts DAN, TRUN and REG denote the Daniell kernel, the truncated kernel, and the regression-based tests.

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 $Q_{1DAN}(M)$ test with the 10% shortfall probability is more powerful than the $Q_{1DAN}(M)$ test with the 5% shortfall probability. With $M = 5$ and $T = 2000$, for example, the rejection rates of $Q_{1DAN}(M)$ at the 5% significance level are 86.5% and 39.5% for the 10% and 5% shortfall probabilities respectively. Again this may be due to the possibility that spillover mainly comes from skewness which occurs in the main body of the distribution. All power patterns of $Q_{1TRUN}(M)$ and $Q_{1REG}(M)$ are similar to those of $Q_{1TRUN}(M)$ and $Q_{1REG}(M)$ 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):

$$\begin{cases} Y_{lt} = \beta_{l1}Y_{1t-1} + \beta_{l2}Y_{2t-1} + u_{lt} + \sum_{1}^{n_{lt}} v_{lt}, & l = 1, 2, \\ u_{lt} = \sigma_{lt}\varepsilon_{lt}, \\ \sigma_{lt}^2 = \gamma_{l0} + \gamma_{l1}\sigma_{lt-1}^2 + \gamma_{l2}u_{1t-1}^2 + \gamma_{l3}u_{2t-1}^2, \\ \varepsilon_{lt} \sim m.d.s. (0, 1), \\ v_{lt} \sim i.i.d. N(0, \delta_l^2), \quad n_{lt} \sim Poisson(\lambda_l). \end{cases}$$

Table 4
Power at the 10% and 5% significant levels under ALTER3.

	M	5		10		15		20		25		30	
	T	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Q^{10}_{1DAN}	500	47.5	41.1	43.0	36.6	39.4	32.3	36.9	30.2	35.7	27.3	34.3	25.7
	1000	67.0	62.1	60.7	53.8	54.8	47.8	50.8	43.7	48.0	40.1	44.9	37.1
	2000	88.8	86.5	84.7	81.6	82.1	76.5	77.8	71.9	74.9	68.1	71.5	64.7
Q^{5}_{1DAN}	500	25.8	22.9	25.4	21.9	25.7	20.6	24.5	19.6	24.2	18.2	22.6	17.5
	1000	31.5	26.9	28.7	23.8	27.1	21.2	25.4	19.4	24.5	18.2	23.3	17.4
	2000	45.7	39.5	40.0	33.8	35.1	29.1	32.6	25.8	31.1	22.6	29.9	21.7
Q^{10}_{1TRUN}	500	38.5	30.9	32.6	26.3	29.7	22.0	28.4	19.9	26.9	17.7	25.0	16.8
	1000	55.6	46.9	44.4	36.6	38.3	31.4	36.7	28.4	34.1	24.4	33.4	23.3
	2000	81.3	76.1	72.0	64.3	65.4	55.6	58.7	49.6	55.5	46.0	52.7	43.0
Q^{5}_{1TRUN}	500	24.7	20.7	23.3	17.8	20.5	15.7	19.0	14.7	19.3	14.5	19.6	14.2
	1000	26.3	20.2	21.9	16.7	19.8	14.9	20.2	14.0	19.9	13.7	18.9	12.8
	2000	35.6	29.4	28.6	22.3	24.9	18.2	23.6	16.0	22.6	16.0	20.6	13.9
Q^{10}_{1REG}	500	38.0	30.9	32.8	25.8	27.6	20.8	25.8	17.2	23.6	16.3	23.1	14.8
	1000	57.0	29.1	48.2	22.9	44.5	22.6	40.9	20.9	39.3	20.8	36.8	19.0
	2000	80.4	75.4	71.7	65.3	65.8	56.5	62.0	53.2	59.0	48.1	54.8	44.1
Q^{5}_{1REG}	500	23.7	19.8	22.6	17.2	21.3	15.7	18.0	12.5	17.5	13.1	17.9	13.2
	1000	50.1	23.0	40.3	16.9	35.4	16.1	31.3	14.6	29.1	14.3	25.4	13.8
	2000	35.7	28.7	29.1	22.3	27.7	21.3	25.6	18.9	24.4	16.8	22.5	15.8

ALTER3: $Y_{it} = 0.3Y_{it-1} + u_{it}$, $i = 1, 2$, $u_{it} = \sqrt{h_{it}}\varepsilon_{it}$, $h_{it} = 0.1 + 0.5h_{it-1} + 0.2u^2_{it-1}$, $\varepsilon_{it} \sim GT(\varepsilon_t|\lambda_t, \eta_t)$, $\lambda_{it} = (1 - \exp(\tilde{\lambda}_{it}))/(1 + \exp(\tilde{\lambda}_{it}))$, $\eta_{it} = 4/(1 + \exp(\tilde{\eta}_{it})) + 4$, where $\tilde{\lambda}_{1t} = -0.2 + \varepsilon_{1t-1} - 5\varepsilon_{2t-1} - 0.9\tilde{\lambda}_{2t-1}$, $\tilde{\lambda}_{2t} = -0.2 + \varepsilon_{2t-1}$, $\tilde{\eta}_{1t} = -0.2 + \varepsilon_{1t-1} - 5\varepsilon_{2t-1} - 0.9\tilde{\eta}_{2t-1}$, $\tilde{\eta}_{2t} = -0.2 + \varepsilon_{2t-1}$; the sample size $T = 500, 1000$ and 2000; $Q^{10}_{1DAN}, Q^{10}_{1TRUN}, Q^{10}_{1REG}$ and $Q^{5}_{1DAN}, Q^{5}_{1TRUN}, Q^{5}_{1REG}$ represent one-way tests for Granger causality in risk from Y_{2t} to Y_{1t} at the 10% and 5% risk levels respectively, where the subscripts DAN, TRUN and REG denote the Daniell kernel, the truncated kernel, and the regression-based tests.

Particularly, we test the alternative of Granger Causality in Risk from Mean, however with jump added to the process[8]:

$$\begin{cases} (\beta_{11}, \beta_{12}, \gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{13}, \delta_1, \lambda_1) \\ = (0.5, 0.2, 0.1, 0.6, 0.2, 0, 1, 0.2), \\ (\beta_{21}, \beta_{22}, \gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{23}, \delta_2, \lambda_2) \\ = (0, 0.5, 0.1, 0.6, 0, 0.2, 1, 0.2). \end{cases}$$

We find that the $Q_{1DAN}(M)$ test still has good power and it becomes more powerful as T increases. For 10% risk level, $Q_{1DAN}(5)$ is 51.3%, 72.7% and 92.9% when $M = 500, 1000$ and 2000 respectively. And for the 5% risk level, the $Q_{1DAN}(5)$ test increases from 45.2% to 91.1%.[9]

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 $Q_{1TRUN}(M)$ and the Granger-type regression procedure $Q_{1REG}(M)$ 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., Baillie 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 risk 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 $Q_{1DAN}(M)$ 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.

Table 5
Estimate and relevant statistics for asymmetric slope CAViaR specification.

Parameter	1% VaR				5% VaR			
	Euro/Dollar		Yen/Dollar		Euro/Dollar		Yen/Dollar	
θ_0	0.015	[0.032]	2.215	[0.788]	0.639	[0.646]	0.842	[0.865]
θ_1	0.984	[0.011]	0.588	[0.122]	0.741	[0.217]	0.766	[0.245]
θ_2	−0.053	[0.037]	0.029	[0.257]	0.066	[0.067]	−0.008	[0.074]
θ_3	0.157	[0.086]	0.872	[0.295]	0.178	[0.063]	0.108	[0.127]
Sample size	2390		2390		2390		2390	
Diagnostic DQ test statistics								
(p-values)	0.597		0.542		0.994		0.506	

The specified asymmetric slope VaR model for Euro/Dollar and Japanese Yen/Dollar Y_{lt} is: $V_{lt} = \theta_0 + \theta_1 V_{lt-1} + \theta_2 Y_{lt-1}^+ + \theta_3 Y_{lt-1}^-$, where $l = 1, 2$, $Y_{lt}^+ = \max(Y_{lt}, 0)$, $Y_{lt}^- = -\min(Y_{lt}, 0)$. The numbers in the square brackets are the standard deviation for the estimated parameters. Engle and Manganelli's DQ test are applied to check the adequacy of the specified VaR model, where the first four lagged hits are used as instruments.

Engle and Manganelli's (2004) Conditional Autoregressive Value at Risk (CAViaR) model for our VaR calculation:

$$V_{lt}(\theta) = \theta_{l0} + \sum_{j=1}^{q} \theta_{lj} V_{lt-j}(\theta) + \sum_{j=1}^{r} \theta_{l(q+j)} L(Y_{lt-j}), \quad l = 1, 2.$$

The autoregressive components $\{\theta_{lj} V_{lt-j}\}$ 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 time-varying error distributions, or scenarios with both time-varying error densities and volatilities. They are thus more general than those GARCH models. See Engle and Manganelli (2004) for more discussion.

Specifically, we use Engle and Manganelli's (2004) asymmetric slope model for the VaRs of Euro/Dollar and Yen/Dollar[11]:

$$V_{lt}(\theta_l) = \theta_{l0} + \theta_{l1} V_{lt-1} + \theta_{l2} Y_{lt-1}^+ + \theta_{l3} Y_{lt-1}^-, \quad l = 1, 2,$$

where $Y_{lt}^+ = \max(Y_{lt}, 0)$, $Y_{lt}^- = -\min(Y_{lt}, 0)$. The asymmetric slope quantile specification is correctly specified if an exchange rate change series follows a GARCH model in which the conditional standard deviation, rather than the conditional variance, is modelled asymmetrically with i.i.d. innovations. The unknown parameters are estimated using Engle and Manganelli's (2004) method, which is based on the regression quantile loss function, first introduced by Koenker and Bassett (1978). We also use Engle and Manganelli's (2004) dynamic quantile (DQ) test to check the adequacy of the estimated CAViaR models. For all subsequent discussions, we use the 5% significance level as the cutoff level.

Table 5 presents the parameter estimates and their standard deviations, as well as the p-values of the DQ test statistics. We find that the autoregressive coefficient (θ_{l1}) estimates are significant for both exchange rate change series, confirming the clustering of the tail of the exchange rate change distribution. The p-values of the DQ test statistics are 0.597, 0.542, 0.994 and 0.506 for Euro/Dollar and Yen/Dollar at the 1% and 5% risk levels respectively, all well above the 5% significance level. These results suggest the adequacy of the estimated asymmetric slope model for the VaRs of Euro/Dollar and Yen/Dollar.

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 one-way Granger causality in risk. The one-way test $Q_{1DAN}(M)$ in (3.7) with 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_{-1DAN}, 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 $Q_{-1DAN}(M)$ 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 $Q_{1REG}(M)$ and $Q_{-1REG}(M)$ and the truncated kernel-based tests $Q_{1TRUN}(M)$ and $Q_{-1TRUN}(M)$ deliver similar conclusions to $Q_{1DAN}(M)$ and $Q_{-1DAN}(M)$, but with weaker evidence, particularly for larger M. Their statistic values are much smaller than those of $Q_{1DAN}(M)$ and $Q_{-1DAN}(M)$. Furthermore, at the 1% risk level, $Q_{1TRUN}(M)$ is significant only for $M = 5$. It becomes insignificant for larger M, with p-values well above 10%. And at the 5% risk level, $Q_{-1REG}(M)$ is significant only for $M = 5$ and 10.

To sum up, we find that large price falls of Euro/Dollar always have a significant predictive power for large price falls of Yen/Dollar. However, risk causality from Yen/Dollar to Euro/Dollar is only significant at the 1% risk level, not at the 5% risk level. For both exchange rate changes series, the risk spillover is much stronger 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 comovements 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

---

[11] Several other CAViaR models are also considered. The asymmetric slope model is found to be adequate and has the best performance in most cases.

Table 6
Risk spillover between Euro/Dollar and Yen/Dollar.

	M	1% VaR					5% VaR				
		5	10	20	30	40	5	10	20	30	40
Eur⇐Yen	Q_{1DAN}	3.394	2.909	1.342	0.478	−0.063	−0.313	−0.465	−0.880	−1.128	−1.229
(p-values)		0.000	0.002	0.090	0.316	0.525	0.623	0.679	0.811	0.870	0.890
Eur⇒Yen	Q_{-1DAN}	16.083	11.156	7.506	5.890	4.873	4.570	3.069	2.450	1.983	1.715
(p-values)		0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.024	0.043
Eur⇐Yen	Q_{1TRUN}	2.495	0.950	−0.475	−1.094	−1.342	−0.455	−0.856	−1.566	−0.515	−0.967
(p-values)		0.006	0.171	0.683	0.863	0.910	0.675	0.804	0.941	0.697	0.833
Eur⇒Yen	Q_{-1TRUN}	8.621	5.703	3.775	2.865	1.862	2.319	2.380	1.469	1.045	1.260
(p-values)		0.000	0.000	0.000	0.002	0.031	0.010	0.009	0.071	0.148	0.104
Eur⇐Yen	Q_{1REG}	2.487	0.906	−0.461	−0.934	−1.120	−0.459	−0.825	−1.561	−0.696	−1.288
(p-values)		0.006	0.182	0.678	0.825	0.869	0.677	0.795	0.941	0.757	0.901
Eur⇒Yen	Q_{-1REG}	8.393	5.439	3.614	0.765	−0.037	2.347	2.476	1.517	0.981	1.544
(p-values)		0.000	0.000	0.000	0.222	0.515	0.009	0.007	0.065	0.163	0.061

"⇐" and "⇒" represent one-way causality in risk from the latter to the former and the former to the latter with respect to I_{t-1} respectively. Q_{1DAN}, Q_{1TRUN}, Q_{1REG} and Q_{-1DAN}, Q_{-1TRUN}, Q_{-1REG} represent one-way tests for Granger causality in risk from Japanese Yen/Dollar to Euro/Dollar and from Euro/Dollar to Japanese Yen/Dollar respectively, where the subscripts DAN, TRUN and REG denote the Daniell kernel, the truncated kernel, and the regression-based tests.

the use of a downward weighting kernel function. This downward weighting is consistent with the stylized fact that today's financial markets are more influenced by more recent events than by remote past events, thus enhancing the power of the proposed tests. This is expected to give good power against the alternatives with decaying cross-correlations as the lag order increases. Indeed, nonuniform weighting often delivers better power than uniform weighting, as is illustrated in a simulation study and an application to exchange rates. A Granger-type regression-based test (1969) is equivalent to the uniform-weighting-based test. Simulation studies show that the procedures have reasonable size and good power against a number of empirically plausible alternatives in finite samples, no matter whether risk spillover arises from spillover in mean, in variance, or in skewness and kurtosis. These procedures are therefore useful for investigating the comovements between large market changes such as financial contagions. We use the new tests to investigate intraday risk spillover causality between Euro/Dollar and Yen/Dollar. It is found that a past large downside movement of Euro Granger-causes a future large price fall of Japanese Yen, and the evidence is stronger for larger changes. On the other hand, only an extremely large past downside movement of Japanese Yen Granger-causes a future large price fall of Euro in risk.

Acknowledgements

Appendix. Mathematical proof

Proof of Theorem 1. Let $\tilde{f}(\omega)$ and $\tilde{f}_1^0(\omega)$ be defined in the same way as $\hat{f}(\omega)$ and $\hat{f}_1^0(\omega)$ in (3.4) and (3.5) with $\hat{\theta} \equiv (\hat{\theta}_1, \hat{\theta}_2)'$ replaced by $\theta_0 \equiv (\theta_1^0, \theta_2^0)'$. Given $D_{1T}(M) = M \int_0^\infty k^4(z)dz[1 + o(1)]$ 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 $\hat{\theta}$ has no impact on the limit distribution of $Q_1(M)$. The main technical challenge for the proof of Theorem A.1 is that the risk indicator $Z_{lt}(\theta_l) \equiv \mathbf{1}(Y_{lt} < -V_{lt})$ is not differentiable with respect to parameter θ_l.

Theorem A.1. $M^{-\frac{1}{2}}T[L^2(\hat{f}, \hat{f}_1^0) - L^2(\tilde{f}, \tilde{f}_1^0)] \to^p 0$, where $L_2(\cdot)$ is defined as in (3.6).

Theorem A.2. $[TL^2(\tilde{f}, \tilde{f}_1^0) - C_{1T}(M)]/[2D_{1T}(M)]^{\frac{1}{2}} \to^d N(0, 1)$.

Proof of Theorem A.1. Throughout, let $\tilde{C}(j)$ be defined as $\hat{C}(j)$ in (3.3) with $\hat{\theta}$ replaced by θ_0. We further replace the sample proportions $\hat{\alpha}_l$ and $\tilde{\alpha}_l$ in $\hat{C}(j)$ and $\tilde{C}(j)$ with α. Such a replacement does not affect the asymptotic distribution of $Q_1(M)$. Putting $\sigma^2 \equiv \alpha(1 - \alpha)$, we have

$$T\left[L^2(\hat{f}, \hat{f}_1^0) - L^2(\tilde{f}, \tilde{f}_1^0)\right] = \sigma^{-4}T\sum_{j=1}^{T-1} k^2(j/M)\left[\hat{C}(j) - \tilde{C}(j)\right]^2$$
$$+ 2\sigma^{-4}T\sum_{j=1}^{T-1} k^2(j/M)[\hat{C}(j) - \tilde{C}(j)]\tilde{C}(j)$$
$$\equiv \sigma^{-4}T\hat{Q}_1 + 2\sigma^{-4}T\hat{Q}_2, \quad \text{say.} \quad \text{(A.1)}$$

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

Proposition A.1. $M^{-\frac{1}{2}}T\hat{Q}_1 \to^p 0$.

Proposition A.2. $M^{-\frac{1}{2}}T\hat{Q}_2 \to^p 0$.

Proof of Proposition A.1. By straightforward algebra, we have for $j > 0$,

$$\hat{C}(j) - \tilde{C}(j) = \hat{M}_1(j, \hat{\theta}_1) + \hat{M}_2(j, \hat{\theta}_2) + \hat{M}_3(j, \hat{\theta}_1, \hat{\theta}_2), \quad \text{(A.2)}$$

where

$$\hat{M}_1(j, \theta_1) \equiv T^{-1}\sum_{t=j+1}^{T} [Z_{1t}(\theta_1) - Z_{1t}(\theta_1^0)][Z_{2(t-j)}(\theta_2^0) - \alpha],$$

$$\hat{M}_2(j, \theta_2) \equiv T^{-1}\sum_{t=j+1}^{T} [Z_{1t}(\theta_1^0) - \alpha][Z_{2(t-j)}(\theta_2) - Z_{2(t-j)}(\theta_2^0)],$$

$$\hat{M}_3(j, \theta_1, \theta_2)$$
$$\equiv T^{-1}\sum_{t=j+1}^{T} [Z_{1t}(\theta_1) - Z_{1t}(\theta_1^0)][Z_{2(t-j)}(\theta_2) - Z_{2(t-j)}(\theta_2^0)].$$

By the definition of $\hat{Q}_1$ in (A.1), we have

$$\hat{Q}_1 \le 3\left[\hat{Q}_{11}(\hat{\theta}_1) + \hat{Q}_{12}(\hat{\theta}_2) + \hat{Q}_{13}(\hat{\theta}_1, \hat{\theta}_2)\right], \quad \text{(A.3)}$$

where

$$\hat{Q}_{11}(\theta_1) \equiv \sum_{j=1}^{T-1} k^2(j/M)\hat{M}_1^2(j, \theta_1),$$

$$\hat{Q}_{12}(\theta_2) \equiv \sum_{j=1}^{T-1} k^2(j/M)\hat{M}_2^2(j, \theta_2),$$

$$\hat{Q}_{13}(\theta_1, \theta_2) \equiv \sum_{j=1}^{T-1} k^2(j/M)\hat{M}_3^2(j, \theta_1, \theta_2).$$

Because the risk indicator $Z_{lt}(\theta_l)$ is not differentiable with respect to θ_l, we shall use the uniform convergence argument to show that $\hat{Q}_{11}(\hat{\theta}_1)$, $\hat{Q}_{12}(\hat{\theta}_2)$ and $\hat{Q}_{13}(\hat{\theta}_1, \hat{\theta}_2)$ 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 $P(|\hat{\theta}_l - \theta_l^0| > \Delta_0 T^{-\frac{1}{2}}) < \varepsilon$ for T sufficiently large. Hence, it suffices to show Lemmas A.1–A.3.

Lemma A.1. *Put* $\Theta_l^0 \equiv \{\theta_l \in \Theta_l : |\theta_l - \theta_l^0| \leq \Delta_0 T^{-\frac{1}{2}}\}$ *for* $0 < \Delta_0 < \infty, l = 1, 2$. *Then for any given constant* $\Delta_0 > 0$, $\sup_{\theta_1 \in \Theta_1^0} |T\hat{Q}_{11}(\theta_1)| = O_P(1)$ *and* $\sup_{\theta_1 \in \Theta_1^0} |M^{-\frac{1}{2}}T\hat{Q}_{11}(\theta_1)| \to^p 0$.

Proof of Lemma A.1. This is one of the most involved proofs, due to the fact that the risk indicator $Z_{lt}(\theta_l)$ is not differentiable with respect to θ_l. Recalling $Z_{lt}(\theta_l) \equiv \mathbf{1}[Y_{lt} < -V_{lt}(\theta_l)]$, we put

$$W_{lt}(\theta_l) \equiv Z_{lt}(\theta_l) - Z_{lt}(\theta_l^0) - E[Z_{lt}(\theta_l)|I_{t-1}] + E[Z_{lt}(\theta_l^0)|I_{t-1}]$$
$$= Z_{lt}(\theta_l) - Z_{lt}(\theta_l^0) - F_{lt}[-V_{lt}(\theta_l)] + F_{lt}[-V_{lt}(\theta_l^0)], \quad (A.4)$$

where $F_{lt}(y|I_{t-1})$ is the conditional CDF of Y_{lt} given I_{t-1}. Given each $\theta_l \in \Theta_l$, we have $E[W_{lt}(\theta_l)|I_{t-1}] = 0$ a.s., i.e., $\{W_{lt}(\theta_l)\}$ is a m.d.s. with respect to I_{t-1} for each given θ_l. Noting $Z_{lt}(\theta_l^0) = Z_{lt} \equiv \mathbf{1}(Y_{lt} \leq -V_{lt})$ a.s. under Assumption 3(i) and recalling the definition of $\hat{M}_1(j, \theta_1)$ in (A.2), we have for $j > 0$,

$$\hat{M}_1(j, \theta_1) = T^{-1} \sum_{t=j+1}^{T} [Z_{1t}(\theta_1) - Z_{1t}(\theta_1^0)][Z_{2(t-j)}(\theta_2^0) - \alpha]$$

$$= T^{-1} \sum_{t=j+1}^{T} W_{1t}(\theta_1)[Z_{2(t-j)} - \alpha]$$

$$+ T^{-1} \sum_{t=j+1}^{T} \{F_{1t}[-V_{1t}(\theta_1)]$$

$$- F_{1t}[-V_{1t}(\theta_1^0)]\}[Z_{2(t-j)} - \alpha]$$

$$\equiv \hat{M}_{11}(j, \theta_1) + \hat{M}_{12}(j, \theta_1), \quad \text{say.} \quad (A.5)$$

We first consider $\hat{M}_{12}(j, \theta_1)$ in (A.5). By a second order Taylor series expansion, we have

$$\hat{M}_{12}(j, \theta_1) = (\theta_1 - \theta_1^0)'T^{-1} \sum_{t=j+1}^{T} \frac{\partial F_{1t}[-V_{1t}(\theta_1^0)]}{\partial \theta_1}[Z_{2(t-j)} - \alpha]$$

$$+ \frac{1}{2}(\theta_1 - \theta_1^0)'T^{-1} \sum_{t=j+1}^{T} \frac{\partial^2 F_{1t}[-V_{1t}(\bar{\theta}_1)]}{\partial \theta_1 \partial \theta_1'}$$

$$\times [Z_{2(t-j)} - \alpha](\hat{\theta}_1 - \theta_1^0)$$

$$\equiv (\theta_1 - \theta_1^0)'\hat{M}_{121}(j, \theta_1) + \frac{1}{2}(\theta_1 - \theta_1^0)'\hat{M}_{122}(j, \bar{\theta}_1)(\theta_1 - \theta_1^0),$$

$$\text{say,} \quad (A.6)$$

where $\bar{\theta}_1$ lies between θ_1 and θ_1^0. Put $\tilde{\Gamma}(j) \equiv T^{-1} \sum_{t=j+1}^{T} S_{1t}(\theta_1^0)[Z_{2(t-j)} - \alpha]$, where $S_{1t}(\theta_1)$ is defined in Assumption 5. Then

$$\sup_{\theta_1 \in \Theta_1^0} T \sum_{j=1}^{T-1} k^2(j/M)[(\theta_1 - \theta_1^0)'\hat{M}_{121}(j, \theta_1)]^2$$

$$\leq 2T\|\theta_1 - \theta_1^0\|^2 \sum_{j=1}^{T-1} k^2(j/M) \|\Gamma(j)\|^2$$

$$+ 2T\|\theta_1 - \theta_1^0\|^2 \sum_{j=1}^{T-1} k^2(j/M)\|\tilde{\Gamma}(j) - \Gamma(j)\|^2$$

$$= O_P(1) + O_P(M/T), \quad (A.7)$$

where $\sup_{\theta_1 \in \Theta_1^0} \|\theta_1 - \theta_1^0\| \leq \Delta_0 T^{-\frac{1}{2}}$, $\sum_{j=1}^{T-1} k^2(j/M) \|\Gamma(j)\|^2 \to \sum_{h=1}^{\infty} \|\Gamma(h)\|^2 = O(1)$ given Assumptions 5(i) and 6, $M \to \infty, M/T \to 0$, and

$$\sum_{j=1}^{T-1} k^2(j/M)\|\tilde{\Gamma}(j) - \Gamma(j)\|^2 = O_P(M/T)$$

by Markov's inequality and $\sup_{0<j<T} E\|\tilde{\Gamma}(j) - \Gamma(j)\|^2 \leq \Delta T^{-1}$. Note that $\sup_{0<j<T} E\|\hat{\Gamma}(j) - \Gamma(j)\|^2 \leq \Delta T^{-1}$ follows from Assumption 5 (cf. Hannan (1970, p. 209)).

For the second term in (A.6), we have

$$\sup_{\theta_1 \in \Theta_1^0} T \sum_{j=1}^{T-1} k^2(j/M)|(\theta_1 - \theta_1^0)'\hat{M}_{122}(j, \theta_1)(\theta_1 - \theta_1^0)|^2$$

$$\leq \sup_{\theta_1 \in \Theta_1^0} \|\theta_1 - \theta_1^0\|^4 T \sum_{j=1}^{T-1} k^2(j/M)\|\hat{M}_{122}(j, \theta_1)\|^2 = O_P(M/T)$$

$$(A.8)$$

given $\sup_{\theta_1 \in \Theta_1^0} \|\theta_1 - \theta_1^0\| \leq \Delta_0 T^{-\frac{1}{2}}$, Assumption 2, and Markov's inequality. It follows from (A.6)–(A.8), $M \to \infty, M/T \to 0$ that

$$\sup_{\theta_1 \in \Theta_1^0} T \sum_{j=1}^{T-1} k^2(j/M)\|\hat{M}_{12}(j, \theta_1)\|^2 = O_P(1). \quad (A.9)$$

We now consider the first term $\hat{M}_{11}(j, \theta_1)$ in (A.5). Divide the cube Θ_1^0, which is centered at θ_1^0 with size $2\Delta_0 T^{-\frac{1}{2}}$, into approximately $L_T \equiv (2\Delta_0/\varepsilon_T)^{d_1}$ cubes $\{\Theta_1^+(l), l = 1, \ldots, L_T\}$ of size $\varepsilon_T/T^{\frac{1}{2}}$, where $\varepsilon_T \equiv M^{-\frac{1}{2}}/\ln(T) \to 0$ and d_1 is the dimension of θ_1. For $1 \leq l \leq L_T$, put $\theta_{1t}^a(l) \equiv \inf_{\theta_1 \in \Theta_1^+(l)} F_{1t}[-V_{1t}(\theta_1)]$ and $\theta_{1t}^b(l) \equiv \sup_{\theta_1 \in \Theta_1^+(l)} F_{1t}[-V_{1t}(\theta_1)]$. Note that $\theta_{1t}^a(l)$ and $\theta_{1t}^b(l)$ are measurable functions of $I_{1(t-1)}$ because $V_{1t}(\theta_1)$ is a measurable function of I_{1t-1}. Then, for any $\theta_1 \in \Theta_1^+(l)$, we write

$$\hat{M}_{11}(j, \theta_1) = T^{-1} \sum_{t=j+1}^{T} \{Z_{1t}(\theta_1) - Z_{1t}(\theta_1^0) - F_{1t}[-V_{1t}(\theta_1)]$$

$$+ F_{1t}[-V_{1t}(\theta_1^0)]\}Z_{2(t-j)} - \alpha T^{-1} \sum_{t=j+1}^{T} \{Z_{1t}(\theta_1) - Z_{1t}(\theta_1^0)$$

$$- F_{1t}[-V_{1t}(\theta_1)] + F_{1t}[-V_{1t}(\theta_1^0)]\}$$

$$\leq T^{-1} \sum_{t=1}^{T} \{Z_{1t}[\theta_{1t}^b(l)] - Z_{1t}(\theta_1^0) - F_{1t}[-V_{1t}(\theta_{1t}^a(l))]$$

$$+ F_{1t}[-V_{1t}(\theta_1^0)]\} - T^{-1} \sum_{t=1}^{T} \{Z_{1t}[\theta_{1t}^a(l)]$$

$$- Z_{1t}(\theta_1^0) - F_{1t}[-V_{1t}(\theta_{1t}^b(l))] + F_{1t}[-V_{1t}(\theta_1^0)]\}$$

$$= T^{-1} \sum_{t=1}^{T} \{W_{1t}[\theta_{1t}^b(l)] - W_{1t}[\theta_{1t}^a(l)]\}$$

$$+ 2T^{-1} \sum_{t=1}^{T} \{F_{1t}[-V_{1t}(\theta_{1t}^b(l))] - F_{1t}[-V_{1t}(\theta_{1t}^a(l))]\}.$$

Similarly, we can obtain

$$\hat{M}_{11}(j, \theta_1) \geq -T^{-1} \sum_{t=1}^{T} \{W_{1t}[\theta_{1t}^b(l)] - W_{1t}[\theta_{1t}^a(l)]\}$$

$$- 2T^{-1} \sum_{t=1}^{T} \{F_{1t}[-V_{1t}(\theta_{1t}^b(l))] - F_{1t}[-V_{1t}(\theta_{1t}^a(l))]\}.$$

It follows that

$$\max_{0<j<T} \sup_{\theta_1 \in \Theta_1^+(l)} \left| \hat{M}_{11}(j, \theta_1) \right| \leq \left| T^{-1} \sum_{t=1}^{T} \{W_{1t}[\theta_{1t}^b(l)] - W_{1t}[\theta_{1t}^a(l)]\} \right|$$

$$+ 2\Delta_0 \varepsilon_T T^{-\frac{1}{2}} T^{-1} \sum_{t=1}^{T} \sup_{\theta_1 \in \Theta_1^0} \left\| \frac{\partial F_{1t}[-V_{1t}(\theta_1)]}{\partial \theta_1} \right\| \tag{A.10}$$

given $\|\theta_{1t}^b(l) - \theta_{1t}^a(l)\| \leq \Delta_0 \varepsilon_T / T^{\frac{1}{2}}$. Note that the second term in (A.10) does not depend on l.

Therefore, we have

$$\sup_{\theta_1 \in \Theta_1^0} T \sum_{j=1}^{T-1} k^2(j/M) \left| \hat{M}_{11}(j, \theta_1) \right|^2$$

$$= \max_{1 \leq l \leq L_T} \sup_{\theta_1 \in \Theta_1^+(l)} T \sum_{j=1}^{T-1} k^2(j/M) \left| \hat{M}_{11}(j, \theta_1) \right|^2$$

$$\leq 2 \max_{1 \leq l \leq L_T} T \left| T^{-1} \sum_{t=1}^{T} \{W_{1t}[\theta_{1t}^b(l)] - W_{1t}[\theta_{1t}^a(l)]\} \right|^2 \sum_{j=1}^{T-1} k^2(j/M)$$

$$+ 4T(\Delta_0^2 \varepsilon_T^2 / T) \left[T^{-1} \sum_{t=1}^{T} \sup_{\theta_1 \in \Theta_1^0} \left\| \frac{\partial F_{1t}[-V_{1t}(\theta_1)]}{\partial \theta_1} \right\| \right]^2$$

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

$$= O_P(M\varepsilon_T^2) = o_P(1) \tag{A.11}$$

given $\varepsilon_T = M^{-\frac{1}{2}} / \ln(T)$, where we have made use of the fact that

$$P\left(\max_{1 \leq l \leq L_T} T \left| T^{-1} \sum_{t=1}^{T} \{W_{1t}[\theta_{1t}^b(l)] - W_{1t}[\theta_{1t}^a(l)]\} \right|^2 > \Delta_0^2 \varepsilon_T^2 \right)$$

$$\leq \sum_{l=1}^{L_T} P\left(\left| T^{-1} \sum_{t=1}^{T} \{W_{1t}[\theta_{1t}^b(l)] - W_{1t}[\theta_{1t}^a(l)]\} \right|^2 > \Delta_0^2 \varepsilon_T^2 / T \right)$$

$$\leq \sum_{l=1}^{L_T} (\Delta_0^2 \varepsilon_T^2 / T)^{-2} E \left| T^{-1} \sum_{t=1}^{T} \{W_{1t}[\theta_{1t}^b(l)] - W_{1t}[\theta_{1t}^a(l)]\} \right|^4$$

$$\leq (2\Delta_0 / \varepsilon_T)^{d_1} (\Delta_0^2 \varepsilon_T^2 / T)^{-2} \left[T^{-2} (\Delta_0 \varepsilon_T / T^{\frac{1}{2}})^2 + T^{-3} (\Delta_0 \varepsilon_T / T^{\frac{1}{2}}) \right]$$

$$= O(\varepsilon_T^{-(d_1+2)} / T + \varepsilon_T^{-(d_1+3)} / T^{\frac{3}{2}}) \to 0$$

given $M = cT^\nu, \varepsilon_T = M^{-\frac{1}{2}} / \ln(T), \nu < \frac{2}{d_1+2}$, where the third inequality follows from

$$E \left| T^{-1} \sum_{t=1}^{T} \{W_{1t}[\theta_{1t}^b(l)] - W_{1t}[\theta_{1t}^a(l)]\} \right|^4$$

$$\leq T^{-2} (\Delta_0 \varepsilon_T / T^{\frac{1}{2}})^2 + T^{-3} (\Delta_0 \varepsilon_T / T^{\frac{1}{2}})$$

by Rosenthal's inequality (e.g., Hall and Heyde (1980, p. 23)), the fact that $\{W_{1t}[\theta_{1t}^b(l)] - W_{1t}[\theta_{1t}^a(l)], \mathcal{F}_{t-1}\}$ is a m.d.s. and the fact that

$$E|W_{1t}[\theta_{1t}^b(l)] - W_{1t}[\theta_{1t}^a(l)]|^m \leq \Delta_0 \varepsilon_T / T^{\frac{1}{2}} \quad \text{for any } m \geq 1$$

by the law of iterated expectation and $|\theta_{1t}^b(l) - \theta_{1t}^a(l)| \leq \Delta_0 \varepsilon_T / T^{\frac{1}{2}}$. Note that a larger m does not imply a faster convergence rate due to the very nature of the indicator function. The desired result then follows from (A.5), (A.9) and (A.11). □

Lemma A.2. *For any given constant* $\Delta_0 > 0$, *we have* $\sup_{\theta_2 \in \Theta_2^0} |T\hat{Q}_{12}(\theta_2)| = O_P(1)$ *and* $\sup_{\theta_2 \in \Theta_2^0} \left| M^{-\frac{1}{2}} T\hat{Q}_{12}(\theta_2) \right| \to^p 0$.

Proof of Lemma A.2. Similar to the proof of Lemma A.1. □

Lemma A.3. *Put* $\Theta_0 \equiv \Theta_1^0 \otimes \Theta_2^0$ *and* $\theta \equiv (\theta_1, \theta_2)'$. *Then for any given constant* $\Delta_0 > 0$, $\sup_{\theta \in \Theta_0} \left| T\hat{Q}_{13}(\theta_1, \theta_2) \right| \to^p 0$ *and* $\sup_{\theta \in \Theta_0} \left| M^{-\frac{1}{2}} T\hat{Q}_{13}(\theta_1, \theta_2) \right| \to^p 0$.

Proof of Lemma A.3. Recalling the definition of $\hat{M}_3(j, \theta_1, \theta_2)$ as in (A.2) and $Z_{lt}(\theta_l^0) = Z_{lt}$, we write

$$\hat{M}_3(j, \theta_1, \theta_2) = T^{-1} \sum_{t=j+1}^{T} [Z_{1t}(\theta_1) - Z_{1t}(\theta_1^0)]$$

$$\times [Z_{2(t-j)}(\theta_2) - Z_{2(t-j)}(\theta_2^0)]$$

$$= T^{-1} \sum_{t=j+1}^{T} W_{1t}(\theta_1)[Z_{2(t-j)}(\theta_2) - Z_{2(t-j)}]$$

$$+ T^{-1} \sum_{t=j+1}^{T} \{F_{1t}[-V_{1t}(\theta_1)] - F_{1t}[-V_{1t}(\theta_1^0)]\}$$

$$\times [Z_{2(t-j)}(\theta_2) - Z_{2(t-j)}]$$

$$\equiv \hat{M}_{31}(j, \theta_1, \theta_2) + \hat{M}_{32}(j, \theta_1, \theta_2). \tag{A.12}$$

For $\hat{M}_{31}(j, \theta_1, \theta_2)$ in (A.12), following reasoning analogous to that for $\hat{M}_{11}(j, \theta_1)$ in the proof of Lemma A.1, we can obtain

$$\sup_{\theta \in \Theta_0} T \sum_{j=1}^{T-1} k^2(j/M) \left| \hat{M}_{31}(j, \theta_1, \theta_2) \right|^2 \to^p 0. \tag{A.13}$$

For $\hat{M}_{32}(j, \theta_1, \theta_2)$ in (A.12), by the mean value theorem, we have

$$\hat{M}_{32}(j, \theta_1, \theta_2) = (\theta_1 - \theta_1^0)' T^{-1} \sum_{t=j+1}^{T} \frac{\partial F_{1t}[-V_{1t}(\bar{\theta}_1)]}{\partial \theta_1}$$

$$\times [Z_{2(t-j)}(\theta_2) - Z_{2(t-j)}(\theta_2^0)] = O_P(T^{-3/4}) \tag{A.14}$$

uniformly in $(\theta_1, \theta_2) \in \Theta_0$. Here, we have used the facts that $\|\theta_1 - \theta_1^0\| \leq \Delta_0 T^{-\frac{1}{2}}$ and

$$E \sup_{\theta \in \Theta_0} \left\| T^{-1} \sum_{t=j+1}^{T} \frac{\partial F_{1t}[-V_{1t}(\theta_1)]}{\partial \theta_1} [Z_{2(t-j)}(\theta_2) - Z_{2(t-j)}(\theta_2^0)] \right\|^2$$

$$\leq \left\{ T^{-1} \sum_{t=1}^{T} E \sup_{\theta_1 \in \Theta_1^0} \left\| \frac{\partial F_{1t}[-V_{1t}(\theta_1)]}{\partial \theta_1} \right\|^2 \right\}$$

$$\times \left\{ T^{-1} \sum_{t=1}^{T} E \sup_{\theta_2 \in \Theta_2^0} \left[Z_{2t}(\theta_2) - Z_{2t}(\theta_2^0) \right]^2 \right\} \leq CT^{-\frac{1}{2}},$$

given Assumption 2 and

$$E \sup_{\theta_2 \in \Theta_2^0} \left[Z_{2t}(\theta_2) - Z_{2t}(\theta_2^0) \right]^2$$

$$\leq E \left\{ \mathbf{1}[-V_{2t}(\theta_{2t}^a) < Y_{2t} < -V_{2t}(\theta_{2t}^b)] \right\}$$

$$= E\{F_{2t}[-V_{2t}(\theta_{2t}^b)] - F_{2t}[-V_{2t}(\theta_{2t}^a)]\} \leq \Delta_0 T^{-\frac{1}{2}}$$

given $\|\theta_{2t}^a - \theta_{2t}^b\| \leq \Delta_0 T^{-\frac{1}{2}}$, where $\theta_{2t}^a \equiv \arg\inf_{\theta_2 \in \Theta_2^0} F_{2t}[-V_{2t}(\theta_2)]$ and $\theta_{2t}^b \equiv \arg\sup_{\theta_2 \in \Theta_2^0} F_{2t}[-V_{2t}(\theta_2)]$, and the equality follows by the law of iterated expectations and the fact that θ_{2t}^a and θ_{2t}^b are measurable functions of I_{2t-1}. It follows that

$$\sup_{\theta \in \Theta_2^0} T \sum_{j=1}^{T-1} k^2(j/M) \left| \hat{M}_{32}(j, \theta_1, \theta_2) \right|^2 = O_P(M/T^{\frac{1}{2}}) = o_P(1). \quad \text{(A.15)}$$

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

Proof of Proposition A.2. Recalling the definition of $\hat{Q}_2$ in (A.1) and using (A.2), we can write

$$\hat{Q}_2 = \hat{Q}_{21}(\hat{\theta}_1) + \hat{Q}_{22}(\hat{\theta}_2) + \hat{Q}_{23}(\hat{\theta}_1, \hat{\theta}_2), \quad \text{(A.16)}$$

where

$$\hat{Q}_{21}(\theta_1) \equiv \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_1(j, \theta_1) \tilde{C}(j),$$

$$\hat{Q}_{22}(\theta_2) \equiv \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_2(j, \theta_2) \tilde{C}(j),$$

$$\hat{Q}_{23}(\theta_1, \theta_2) \equiv \sum_{j=0}^{T-1} k^2(j/M) \hat{M}_3(j, \theta_1, \theta_2) \tilde{C}(j).$$

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 $\Delta_0 > 0$, $\sup_{\theta_1 \in \Theta_1^0} |M^{-\frac{1}{2}} T \hat{Q}_{21}(\theta_1)| \to^p 0$.

Proof of Lemma A.4. Recalling $\hat{M}_1(j, \theta_1) = \hat{M}_{11}(j, \theta_1) + \hat{M}_{12}(j, \theta_1)$ in (A.5), we have

$$\hat{Q}_{21}(\hat{\theta}_1) = \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_{11}(j, \hat{\theta}_1) \tilde{C}(j)$$

$$\quad + \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_{12}(j, \hat{\theta}_1) \tilde{C}(j)$$

$$\equiv \hat{Q}_{211}(\hat{\theta}_1) + \hat{Q}_{212}(\hat{\theta}_1). \quad \text{(A.17)}$$

For the first term in (A.17), we have

$$\sup_{\theta_1 \in \Theta_1^0} M^{-\frac{1}{2}} T \left| \hat{Q}_{211}(\theta_1) \right|$$

$$\leq \left[T \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_{11}^2(j, \theta_1) \right]^{\frac{1}{2}} \left[M^{-1} T \sum_{j=1}^{T-1} k^2(j/M) \tilde{C}^2(j) \right]^{\frac{1}{2}}$$

$$= o_P(1) O_P(1) = o_P(1) \quad \text{(A.18)}$$

by the Cauchy–Schwarz inequality, (A.11), and the fact that

$$M^{-1} T \sum_{j=1}^{T-1} k^2(j/M) \tilde{C}^2(j) = O_P(1), \quad \text{(A.19)}$$

which follows by Markov's inequality, and $E|\tilde{C}(j)|^2 \leq T^{-1}$ under $\mathbb{H}_1^0$.

For the second term in (A.17), recalling that $\hat{M}_{12}(j, \theta_1)$ can be decomposed as in (A.6), we have

$$T\hat{Q}_{212}(\theta_1) = (\theta_1 - \theta_1^0)' T \sum_{j=1}^{T-1} k^2(j/M) \Gamma(j) \tilde{C}(j)$$

$$\quad + (\theta_1 - \theta_1^0)' T \sum_{j=1}^{T-1} k^2(j/M) [\tilde{\Gamma}(j) - \Gamma(j)] \tilde{C}(j)$$

$$\quad + \frac{1}{2}(\theta_1 - \theta_1^0)' T \sum_{j=1}^{T-1} k^2(j/M) \hat{M}_{122}(j, \bar{\theta}_1) \tilde{C}(j)$$

$$\equiv \sqrt{T}(\theta_1 - \theta_1^0)' \hat{D}_1(\theta_1^0) + \sqrt{T}(\theta_1 - \theta_1^0)' \hat{D}_2(\theta_1^0)$$

$$\quad + \frac{1}{2}\sqrt{T}(\theta_1 - \theta_1^0)' \hat{D}_3(\bar{\theta}_1) \sqrt{T}(\theta_1 - \theta_1^0), \quad \text{(A.20)}$$

where $\bar{\theta}$ lies between θ_1 and θ_1^0. For the first term in (A.20), we have

$$\left\| \hat{D}_1(\theta_1^0) \right\| \leq \sqrt{T} \sum_{j=1}^{T-1} k^2(j/M) \|\Gamma(j)\| \left| \tilde{C}(j) \right| = O_P(1) \quad \text{(A.21)}$$

by Markov's inequality, Assumptions 5 and 6 and $E|\tilde{C}(j)|^2 \leq T^{-1}$ under $\mathbb{H}_1^0$.

For the second term in (A.20), we have

$$\left\| \hat{D}_2(\theta_1^0) \right\| \leq \sqrt{T} \sum_{j=1}^{T-1} k^2(j/M) \|\tilde{\Gamma}(j) - \Gamma(j)\| \left| \tilde{C}(j) \right|$$

$$= O_P(M/T^{\frac{1}{2}}) \quad \text{(A.22)}$$

by Markov's inequality, $E\|\tilde{\Gamma}(j) - \Gamma(j)\|^2 \leq \Delta T^{-1}$ given Assumption 5, and $E|\tilde{C}(j)|^2 \leq T^{-1}$ under $\mathbb{H}_1^0$. Similarly, we have

$$\left\| \hat{D}_3(\bar{\theta}_1) \right\| \leq T^{-1} \sum_{t=1}^{T} \sup_{\theta_1 \in \Theta_1^0} \left\| \frac{\partial^2 F_{1t}[-V_{1t}(\theta_1)]}{\partial \theta_1 \partial \theta_1'} \right\| \sum_{j=1}^{T-1} k^2(j/M) \left| \tilde{C}(j) \right|$$

$$= O_P(M/T^{\frac{1}{2}}) \quad \text{(A.23)}$$

given Assumption 6, and $E|\tilde{C}(j)|^2 \leq T^{-1}$. Collecting (A.20)–(A.23) and $M/T \to 0$, we obtain $M^{-\frac{1}{2}} \sup_{\theta_1 \in \Theta_1^0} |T\hat{Q}_{212}(\theta_1)| \to^p 0$. This completes the proof for Lemma A.4. □

Lemma A.5. For any given constant $\Delta_0 > 0$, $\sup_{\theta_2 \in \Theta_2^0} |M^{-\frac{1}{2}} T \hat{Q}_{22}(\theta_2)| \to^p 0$.

Proof of Lemma A.5. Similar to the proof of Lemma A.4. □

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). □

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 $\mathbb{H}_1^0$. 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 $Q_1(M)$ holds under if $\{u_t\}$ and $\{v_s\}$ were mutually independent. □

Proof of Theorem 2. Recall that $C_{1T}(M) = O(M)$ and $D_{1T}(M) = 2M \int k^4(z)dz[1 + o(1)]$ as $M \to \infty$ and $M/T \to 0$, we have $(M^{1/2}/T)Q_1(M) = \{2 \int k^4(z)dz\}^{-1/2} L^2(\hat{f}, \hat{f}_1^0)[1 + o(1)] + o(1)$. Thus, it suffices to show Theorems A.3 and A.4.

Theorem A.3. $L^2(\hat{f}, \hat{f}_1^0) - L^2(\tilde{f}, \tilde{f}_1^0) \to^p 0$.

Theorem A.4. $L^2(\tilde{f}, \tilde{f}_1^0) - L^2(f, f_1^0) \to^p 0$.

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

Next, recall $\hat{Q}_1 \leq 3[\hat{Q}_{11}(\hat{\theta}_1) + \hat{Q}_{12}(\hat{\theta}_2) + \hat{Q}_{13}(\hat{\theta}_1, \hat{\theta}_2)]$, where $\hat{Q}_{11}(\hat{\theta}_1)$, $\hat{Q}_{12}(\hat{\theta}_2)$ and $\hat{Q}_{13}(\hat{\theta}_1, \hat{\theta}_2)$ are defined as in (A.3). We shall show that these three terms all vanish in probability under $\mathbb{H}_1^A$.

We first consider $\hat{Q}_{11}(\hat{\theta}_1)$. Given Assumption 4, we have that for any given constant $\varepsilon > 0$, there exists $\Delta_0 \equiv \Delta_0(\varepsilon)$ such that $P(|\hat{\theta}_l - \theta_l^0| > \Delta_0 T^{-\frac{1}{2}}) < \varepsilon$ for all T sufficiently large. Thus, it suffices to show $\hat{Q}_{11}(\theta_1) \to^p 0$ uniformly in $\theta_1 \in \Theta_1^0$, where Θ_1^0 is as in Lemma A.1. By the definition of $\hat{Q}_{11}(\hat{\theta}_1)$ in (A.3), we have

$$\sup_{\theta_1 \in \Theta_1^0} |\hat{Q}_{11}(\theta_1)| \leq \max_{0 < j < T} \sup_{\theta_1 \in \Theta_1^0} |\hat{M}_1(j, \theta_1)| \sum_{j=1}^{T-1} k^2(j/M), \qquad (A.24)$$

where $\hat{M}_1(j, \theta_1)$ is defined in (A.2). Put $\theta_{1t}^a \equiv \arg\inf_{\theta_1 \in \Theta_1^0} F_{1t}[-V_{1t}(\theta_1)]$ and $\theta_{1t}^b \equiv \arg\sup_{\theta_1 \in \Theta_1^0} F_{1t}[-V_{1t}(\theta_1)]$. Note that θ_{1t}^a and θ_{1t}^b are measurable functions of $I_{1(t-1)}$, because $V_{1t}(\theta_1)$ depends on $I_{1(t-1)}$ and θ_1. Then

$$\max_{0 < j < T} \sup_{\theta_1 \in \Theta_1^0} |\hat{M}_1(j, \theta_1)| \leq T^{-1} \sum_{t=1}^T \sup_{\theta_1 \in \Theta_1^0} |Z_{1t}(\theta_1) - Z_{1t}(\theta_1^0)|$$

$$\leq T^{-1} \sum_{t=1}^T [Z_{1t}(\theta_{1t}^b) - Z_{1t}(\theta_{1t}^a)] = O_P(T^{-\frac{1}{2}}), \qquad (A.25)$$

where the $O_P(T^{-\frac{1}{2}})$ term follows from Markov's inequality and the fact that

$$E|Z_{1t}(\theta_{1t}^b) - Z_{1t}(\theta_{1t}^a)| = E\{F_{1t}[-V_{1t}(\theta_{1t}^b)] - F_{1t}[-V_t^{(1)}(\theta_{1t}^a)]\}$$

$$\leq \Delta_0 T^{-\frac{1}{2}} E \sup_{\theta_1 \in \Theta_1^0} \left\| \frac{\partial}{\partial \theta} F_{1t}[-V_{1t}(\theta_1)] \right\|$$

by the law of iterated expectations and $\|\theta_{1t}^b - \theta_{1t}^a\| \leq \Delta_0 T^{-\frac{1}{2}}$. It follows from (A.24) and (A.25), $M^2/T \to 0$ that $\sup_{\theta_1 \in \Theta_1^0} |\hat{Q}_{11}(\theta_1)| = O_P(M/T^{\frac{1}{2}}) = o_P(1)$. Similarly, we have $\sup_{\theta_2 \in \Theta_2^0} |\hat{Q}_{12}(\theta_2)| \to^p 0$ and $\sup_{\theta \in \Theta_0} |\hat{Q}_{13}(\theta)| \to^p 0$. This completes the proof. $\square$

Proof of Theorem A.4. See Hong (2001, proof of Theorem 2). $\square$

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 $\mathbb{H}_2^0$, $u_t \equiv Z_{1t} - \alpha$ is independent of $\{v_s = Z_{2t} - \alpha, s \leq t - 1\}$ and v_t is independent of $\{u_s, s \leq t - 1\}$. This is the reason why we do not need Assumption 3, which is required in Theorem 1. $\square$

Proof of Theorem 4. The proof is similar to that of Theorem 2. $\square$

References

Anderson, T., Bollerslev, T., Diebold, F., Labys, P., 2000. Exchange rate returns standardized by realized volatility are (nearly) Gaussian. Multinational Finance Journal 4, 159–179.

Artzner, P., 1999. Application of coherent risk measures to capital requirements in insurance. North American Actuarial Journal 3, 11–25.

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.

Bae, Kee-Hong, Andrew Karolyi, G., Stulz, René M., 2003. A new approach to measuring financial contagion. Review of Financial Studies 16, 717–763.

Bahadur, R., 1960. Stochastic comparison of tests. Annals of Mathematical Statistics 31, 276–295.

Bali, T.G., 2000. Testing the empirical performance of stochastic volatility models of the short term interest rate. Journal of Financial and Quantitative Analysis 35, 191–215.

Basel Committee on Banking Supervision, 1996. Amendment to the capital accord to incorporate market risks. Bank for International Settlements. Basel.

Basel Committee on Banking Supervision, 2001. The new Basel capital accord. Bank for International Settlements. Basel.

Bollerslev, T., 2001. Financial econometrics: Past developments and future challenges. Journal of Econometrics 100, 41–51.

Bollerslev, T., Wooldridge, J., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11 (2), 143–172.

Box, G., Pierre, D., 1970. Distribution of residual autocorrelations in autoregressive integrated moving average time series models. Journal of American Statistical Association 65, 1509–1526.

Campbell, J.Y., Cochrane, J.H., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.

Chernozhukov, V., Umantsev, L., 2001. Conditional Value-at-Risk: Aspects of modelling and estimation. Empirical Economics 26, 271–292.

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.

Christoffersen, P., Hahn, J., Inoue, A., 2001. Testing and comparing Value-at-Risk measures. Journal of Empirical Finance 8, 325–342.

Danielsson, J., de Vries, C.G., 2000. Value-at-Risk and extreme returns. Annales d' Economie et de Statistics 60, 239–270.

Diebold, F.X., Hahn, J., Tay, A.S., 1999. Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns of foreign exchange. Review of Economics and Statistics 81, 661–673.

Duffie, D., Pan, J., 1997. An overview of the Value at Risk. Journal of Derivatives 4 (3), 7–49.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modeling Extremal Events for Insurance and Finance. Springer, New York.

Engle, R.F., 1982. Auto-regressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 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., 1990. Meteor shower or heat wave? Heteroskedastic intra-daily volatility in the foreign exchange market. Econometrica 59, 524–542.

Engle, R.F., Manganelli, S., 2004. CAViaR: Conditional autoregressive Value at Risk by regression quantiles. Journal of Business and Economic Statistics 22, 367–381.

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., Tauchen, G., 1991. On fitting a recalcitrant series: The pound/dollar exchange rate, 1974–83. In: Barnett, W., Powell, J., Tauchen, G. (Eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics. Cambridge University Press, Cambridge, pp. 199–240.

Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.

Gourieroux, C., Jasiak, J., 2001. Financial Econometrics. Princeton University Press, Princeton, NJ.

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., 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.

Hall, P., Heyde, C.C., 1980. Martingale limit theory and its application. Academic Press, New York.

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. Wiley, New York.

Hansen, B., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.

Harvey, C.R., Siddique, A., 1999. Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis 34 (4), 465–487.

Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55 (3), 1263–1295.

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

Hong, Y., 2001. A test for volatility spillover with applications to exchange rates. Journal of Econometrics 103, 183–224.

Hong, Y., Li, H., Zhao, F., 2004. Out-of-sample performance of discrete-time short-term interest models. Journal of Business and Economic Statistics 22, 457–473.

Hong, Y., Li, H., Zhao, F., 2007. Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates. Journal of Econometrics 141, 736–776.

Jondeau, E., Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. Journal of Economic Dynamics and Control 27, 1699–1737.

Jorion, P., 2000. Value at Risk, 2nd edition. Academic Press, San Diego.

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

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

Koenker, Roger, Bassett Jr., Gilbert, 1978. Regression quantiles. Econometrica 46, 33–50.

Lee, S., Hansen, B., 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.

Longin, F.M., 2000. From Value at Risk to stress testing: The extreme value approach. Journal of Banking and Finance 24, 1097–1130.

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

Lumsdaine, R., 1996. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. Econometrica 64, 575–596.

Maheu, J.M., McCurdy, T.H., 2004. News arrival, jump dynamics and volatility components for individual stock returns. Journal of Finance 59, 755–793.

Morgan, J.P., 1996. Risk Metrics—Technical Document, 4th edition. Morgan Guaranty Trust Company, New York.

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

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.

Roy, A., 1952. Safety first and the holding of assets. Econometrica 20, 431–449.