A New Class of Characteristic-Function-Based Distribution Tests and Its Application to GARCH Model

Yi-Ting Chen

Institute for Social Sciences and Philosophy

Academia Sinica

This version: March 26, 2002

*Correspondence address: Yi-Ting Chen, Institute for Social Sciences and Philosophy, Academia Sinica, Taipei 115, Taiwan. Tel:+886-2-2789-8183. Fax: +886-2-2785-4160. *E-mail*: ytchen@gate.sinica.edu.tw

** The author is very grateful for Professor Chung-Ming Kuan's constructive suggestions and comments.

Abstract

This paper proposes a new class of characteristic-function-based distribution tests for two-sample comparison, simple hypotheses, and conditional distribution assumptions of econometric models. The proposed tests are easy to compute and have $\chi^2(2)$ as the asymptotic null distribution. Comparing to the Kolmogorov-Smirnov (KS) and Cramér-von Mises (CV) tests, the proposed test can flexibly account for the difference between the true and postulated distributions at different frequencies. A Monte Carlo simulation study shows that the proposed test significantly outperforms the KS, CV, and Pearson's χ^2 tests, especially when the true and postulated distributions are both symmetric. In an empirical study of stock index returns, we apply the proposed test to check distribution assumptions of the standardized errors of GARCH(1,1) model, including the standard normal, standardized t, logistic, generalized error, and generalized lambda distributions. The proposed test accepts the standardized t distribution but rejects the standard normal distribution for all the returns considered; the appropriateness of other distribution assumptions are data-specific. This empirical study also shows that the conditional normality assumption may render the GARCH(1,1) model over-estimating the impact effect of external shocks on volatility but under-estimating the persistence of these shocks' influences, especially when the markets are volatile.

JEL classification: C12, C22, C52, G19

Keywords: distribution test, characteristic function, GARCH, stock returns.

1 Introduction

Distribution assumptions play an important role in various statistical and economic problems. In financial economics, distributions of asset returns are essential for pricing assets and selecting portfolios. Researchers often use distribution tests (the so-called goodness-of-fit tests) to choose appropriate approximations to these distributions; see e.g., Kon (1984), Hsieh (1988), and Peiró (1994). Distribution assumptions are also crucial for interpreting and estimating econometric models. For parametric survival models, the hazard functions, and hence the transition behaviors of lifetime random variables, vary with conditional distribution assumptions. For generalized autoregressive conditional heteroskedasticity (GARCH) models, various conditional distributions may yield quite different interpretations of the volatility of asset returns. It is also well known that incorrect conditional distribution assumptions may render the quasi-maximum likelihood estimator (QMLE) of econometric models inefficient or even inconsistent; see Gourieróux et al. (1984).

The discussions above reveal that testing distribution assumption is of both theoretical and practical importance. In empirical studies, distribution assumptions are often checked by the distribution tests, such as the Kolmogorov-Smirnov (KS) test, Cramér-von Mises (CV) test, and Pearson's χ^2 (PC) test. The KS and CV tests check if the empirical distribution function is sufficiently close to a postulated cumulative distribution function; the PC test checks if the numbers of observations in specific data cells are predicted by a postulated distribution. Although these tests are widely used, there are some known problems with their applicability and power performance. First, these tests may not be directly applied to distribution assumptions with unknown parameters because their asymptotic null distributions need not be distribution-free in this case; see e.g., the conditional Kolmogorov test of Andrews (1997). Second, the performance of the PC test is sensitive to the choice of data cells; see Moore (1986). In addition, the KS and CV tests can only efficiently account for distributions' Fourier coefficients at very low frequency. Thus, they may lack power to distinguish between distributions which behave similarly at lower frequency; see Durbin and Knott (1972), Eubank and La Riccia (1992), and Fan (1996).

In this paper we propose a new class of characteristic-function-based distribution tests for the two-sample comparison, simple hypotheses, and conditional distribution assumptions of econometric models. The proposed test statistic checks if a weighted discrepancy between the characteristic functions of two random variables, which correspond to the true and postulated distributions, is equal to zero. (In the one-sample case, the data of the postulated distribution are sampled by simulation.) Since the characteristic function is in the frequency domain, the proposed tests would be more flexible than conventional tests in capturing distribution information at different frequencies. The proposed tests are easy to compute and have $\chi^2(2)$ as the asymptotic null distribution. Consequently, they are very convenient in applications. A Monte Carlo simulation study shows that the KS and CV tests lack power to discriminate between symmetric distributions, such as normal and Student's t distributions, in finite samples. On the other hand, the proposed test significantly outperforms the KS, CV, and PC tests under different pairs of simple hypotheses, especially when the true and postulated distributions are both symmetric.

In the empirical study of daily returns on stock price indices, we use the proposed test to examine distribution assumptions of the standardized errors of GARCH(1,1) model, including the standard normal, standardized t, logistic, generalized error, and generalized lambda distributions. Among these distributions, the standardized t distribution is the only one which passes this test for all the returns, whereas the conditional normality assumption is strongly rejected in all cases. We also find that conditional normality does not only render less efficient QMLEs but also over-value the impact effect of external shocks on volatility and under-value the persistence of shocks' influences. The latter evidence is particularly significant when the markets are volatile, and it illustrates how different conditional distribution assumptions affect the interpretations of an econometric model.

The rest of this paper is organized as follows. We propose the new tests in Section 2 and discuss how to implement these tests in Section 3. In Section 4 we conduct a Monte Carlo simulation to compare the performance of different tests. In Section 5 we apply the proposed test to investigate conditional distribution assumptions of the GARCH model in an empirical study. Finally, we conclude this paper in the last section.

2 The Proposed Tests

Let $\{\varepsilon_t\}$ be a sequence of independently and identically distributed (i.i.d.) random variables with an unknown distribution function $\mathcal{F}$, and $\{\nu_t\}$ be another sequence of i.i.d. random variables, independent of $\{\varepsilon_t\}$, with the distribution $\mathcal{G}(\cdot\,;\alpha_o)$, where α_o represents the parameters vector of $\mathcal{G}$. The testing hypotheses are

$$\begin{aligned} H_o &: \mathcal{F} = \mathcal{G}(\cdot\,;\alpha_o), \\ H_1 &: \mathcal{F} \neq \mathcal{G}(\cdot\,;\alpha_o). \end{aligned} \tag{1}$$

The problem of checking whether two samples are drawn from the same distribution is referred to as the "two-sample case", and the problem of testing the adequacy of a postulated distribution is called the "one-sample case". In the former case, both the realizations of ε_t and ν_t are observable. In the latter case, we have only the data of ε_t, and $\mathcal{G}$ is a postulated distribution; (1) is referred to as a simple hypothesis when α_o is known. In this section we will first propose a test for the two-sample case, and then extend it to simple hypotheses and conditional distribution assumptions of econometric models.

2.1 The Two-Sample Case

The characteristic functions of ε_t and ν_t are of the forms:

$$k_f(\omega) := \mathbb{E}[\exp(i\omega\varepsilon_t)] = \mathbb{E}[\cos(\omega\varepsilon_t)] + i\mathbb{E}[\sin(\omega\varepsilon_t)],$$

and

$$k_g(\omega) := \mathbb{E}_g[\exp(i\omega\nu_t)] = \mathbb{E}_g[\cos(\omega\nu_t)] + i\mathbb{E}_g[\sin(\omega\nu_t)],$$

where $i = \sqrt{-1}$ and $\omega \in \mathbb{R}$; $\mathbb{E}$ and $\mathbb{E}_g$ are the expectation operators taken with respect to $\mathcal{F}$ and $\mathcal{G}$, respectively; k_f and k_g are in the frequency domain with ω representing a frequency. Denote

$$e(\omega) = \begin{bmatrix} \mathbb{E}[\sin(\omega\varepsilon_t)] - \mathbb{E}_g[\sin(\omega\nu_t)] \\ \mathbb{E}[\cos(\omega\varepsilon_t)] - \mathbb{E}_g[\cos(\omega\nu_t)] \end{bmatrix}.$$

From the uniqueness of the characteristic function, (1) can be equivalently expressed as

$$
\begin{aligned}
H_o \ &: \ e(\omega) = 0_{2\times 1}, \qquad \forall\, \omega \in \mathbb{R}^+, \\
H_1 \ &: \ e(\omega) \neq 0_{2\times 1}, \qquad \text{for some } \omega,
\end{aligned} \tag{2}
$$

where $0_{2\times 1}$ is the 2×1 zero vector; the range of ω changes from $\mathbb{R}$ to $\mathbb{R}^+$ because the cosine (sine) function is symmetric (odd-symmetric) about the origin. This condition allows us to check the equivalence between $\mathcal{F}$ and $\mathcal{G}$ by testing whether $e(\omega)$ is equal to $0_{2\times 1}$ for all $\omega \in \mathbb{R}^+$.

An immediate problem with testing (2) is that we must examine $e(\omega)$ for infinitely many ω. To deal with this problem, we may base the tests on the supremum of $e(\omega)$, or introduce a weighting function of ω and base the tests on the weighted result. The former method typically generates non-parametric tests, which are of some complicated forms and may not have standard distributions; see Feuerverger and Mureika (1977) for related discussions. For simplicity and convenience, in this paper we adopt the latter method. As we will see below, this method yields very simple test statistics, which are asymptotically distribution-free and have $\chi^2(2)$ as the asymptotic null distribution.

Let z be a weighting function of ω such that $z(\omega) \geq 0$ and $\int_{\mathbf{R}^+} z(\omega)\mathrm{d}\omega = 1$; that is, z is the probability density function of a lifetime distribution, such as the exponential, Weibull, and Gamma distributions. The choice of z will be discussed in Section 3. By introducing z to e and changing the order of integrations, we can define the functions:

$$
\psi_s(\epsilon) = \int_{\mathbf{R}^+} \sin(\omega\epsilon) z(\omega)\mathrm{d}\omega, \qquad \psi_c(\epsilon) = \int_{\mathbf{R}^+} \cos(\omega\epsilon) z(\omega)\mathrm{d}\omega, \tag{3}
$$

where ϵ represents a real number, and the parameters vector:

$$
\vartheta = \int_{\mathbf{R}^+} e(\omega) z(\omega)\mathrm{d}\omega = \begin{bmatrix} \mu_{fs} - \mu_{gs} \\ \mu_{fc} - \mu_{gc} \end{bmatrix},
$$

where $\mu_{fs} := \mathbb{E}[\psi_s(\varepsilon_t)]$, $\mu_{fc} := \mathbb{E}[\psi_c(\varepsilon_t)]$, $\mu_{gs} := \mathbb{E}_g[\psi_s(\nu_t)]$, and $\mu_{gc} := \mathbb{E}_g[\psi_c(\nu_t)]$.

Because ψ_s is an odd function and that ψ_c is an even function, the parameters μ_{fs} and μ_{fc} can be interpreted as indicators of the symmetry (about the origin) and the dispersion of $\mathcal{F}$. Similarly, μ_{gs} and μ_{gc} are symmetry and dispersion indicators of $\mathcal{G}$.

Clearly, under the null hypothesis $\mathcal{F}$ and $\mathcal{G}$ should have the same patterns of symmetry and dispersion, so that

$$\vartheta = 0_{2\times 1}. \tag{4}$$

This condition provides a testable implication of the null hypothesis that is free of ω. The distribution tests based on (4) are in spirit similar to the normality test of Jarque and Bera (1980), which compares the shapes of the true distribution and normal distributions by their skewness and kurtosis. However, skewness and kurtosis are obviously inappropriate for characterizing the symmetry and dispersion of the distributions without proper moments. By contrast, the parameters μ_{fs} and μ_{fc} (μ_{gs} and μ_{gc}) exist for *all* distributions.

Before discussing the proposed tests, we define the statistic:

$$\hat{\vartheta}_T = \begin{bmatrix} \hat{\mu}_{fs} - \hat{\mu}_{gs} \\ \hat{\mu}_{fc} - \hat{\mu}_{gc} \end{bmatrix},$$

where $\hat{\mu}_{fs}$, $\hat{\mu}_{fc}$, $\hat{\mu}_{gs}$, and $\hat{\mu}_{gc}$ are the sample averages of $\psi_s(\varepsilon_t)$, $\psi_c(\varepsilon_t)$, $\psi_s(\nu_t)$, and $\psi_c(\nu_t)$, respectively. We also denote the variance-covariance matrices of the vectors $(\psi_s(\varepsilon_t), \psi_c(\varepsilon_t))'$ and $(\psi_s(\nu_t), \psi_c(\nu_t))'$ by Ω_f and Ω_g, respectively. The sample counterparts of Ω_f and Ω_g are denoted by $\hat{\Omega}_f$ and $\hat{\Omega}_g$.

Since $\{\varepsilon_t\}$ and $\{\nu_t\}$ are two independent sequences of i.i.d. random variables and that all the moments of $\psi_s(\varepsilon_t)$, $\psi_c(\varepsilon_t)$, $\psi_s(\nu_t)$, and $\psi_c(\nu_t)$ exist, the law of large numbers ensures that $\hat{\vartheta}_T$, $\hat{\Omega}_f$, and $\hat{\Omega}_g$ are consistent for ϑ, Ω_f, and Ω_g, respectively. In addition, the Lindeberg-Lévy central limit theorem implies that

$$\sqrt{T}(\hat{\vartheta}_T - \vartheta) \overset{A}{\sim} N\big(0_{2\times 1},\ \Omega_f + \Omega_g\big),$$

or asymptotically equivalently,

$$T(\hat{\vartheta}_T - \vartheta)'(\hat{\Omega}_f + \hat{\Omega}_g)^{-1}(\hat{\vartheta}_T - \vartheta) \overset{A}{\sim} \chi^2(2). \tag{5}$$

As a result, under the null hypothesis, $\vartheta = 0_{2\times 1}$ and the statistic:

$$\mathcal{A}_T := T\hat{\vartheta}_T'(\hat{\Omega}_f + \hat{\Omega}_g)^{-1}\hat{\vartheta}_T \tag{6}$$

has the asymptotic null distribution $\chi^2(2)$. On the other hand, under the local alternative hypothesis: $\vartheta = \delta/\sqrt{T}$ for some $\delta \neq 0_{2\times 1}$, the statistic $\mathcal{A}_T$ has the asymptotic alternative distribution:

$$\mathcal{A}_T \overset{A}{\sim} \chi^2(2; \Delta_a), \qquad \Delta_a := \delta'(\Omega_f + \Omega_g)^{-1}\delta,$$

where $\chi^2(2; \Delta_a)$ denotes a non-central χ^2 distribution with the degrees of freedom two and the non-centrality parameter Δ_a. Therefore, the statistic $\mathcal{A}_T$ can be used to test the null hypothesis in the two-sample case, and will be referred to as the $\mathcal{A}$ test subsequently.

2.2 The One-Sample Case

In many applications, we are more interested in the one-sample case than the two-sample case. However, the proposed $\mathcal{A}$ test is not directly applicable because we do not have the data of ν_t. Nevertheless, this problem can be easily circumvented by utilizing the pseudo-random-number generator to sample realizations of ν_t from the postulated distribution $\mathcal{G}$. The use of pseudo-random number generators is quite standard in Monte Carlo simulation studies; see e.g., Davidson and MacKinnon (1994) and Gamerman (1997). In what follows, we simply assume that this sampling method can independently and identically draw data from $\mathcal{G}$.

2.2.1 Simple Hypothesis

To control the sampling variation of this method, we must consider a reasonable re-sampling number R. Let $\{\nu_t(i)\}$ be a sequence of i.i.d. random variables drawn from the distribution $\mathcal{G}(\cdot;\ \alpha_o)$ with the known parameters vector α_o; $i = 1, 2, \ldots, R$; the sequences $\{\nu_t(1)\}, \{\nu_t(2)\}, \ldots, \{\nu_t(R)\}$ are independent of each other and also independent of ε_t. By replacing ν_t with $\nu_t(i)$, we rewrite the statistics $\hat{\mu}_{gs}$, $\hat{\mu}_{gc}$, and $\hat{\Omega}_g$ as $\hat{\mu}_{gs}(i)$, $\hat{\mu}_{gc}(i)$, and $\hat{\Omega}_g(i)$, and define $\bar{\Omega}_g = R^{-1}\sum_{i=1}^{R}\hat{\Omega}_g(i)$ and

$$\bar{\vartheta}_{T,R} = \begin{bmatrix} \hat{\mu}_{fs} - \bar{\mu}_{gs} \\ \hat{\mu}_{fc} - \bar{\mu}_{gc} \end{bmatrix},$$

where $\bar{\mu}_{gs} = R^{-1}\sum_{i=1}^{R}\hat{\mu}_{gs}(i)$ and $\bar{\mu}_{gc} = R^{-1}\sum_{i=1}^{R}\hat{\mu}_{gc}(i)$.

The law of large numbers implies that $\bar{\mu}_{gs}$, $\bar{\mu}_{gc}$, and $\bar{\Omega}_g$ are consistent for μ_{gs}, μ_{gc}, and Ω_g, respectively. Moreover, the Lindeberg-Lévy central limit theorem implies

$$\sqrt{T}(\bar{\vartheta}_{T,R} - \vartheta) \overset{A}{\sim} N\big(0_{2\times 1},\ \Omega_f + R^{-1}\Omega_g\big)$$

as $T \to \infty$, or asymptotically equivalently,

$$T(\bar{\vartheta}_{T,R} - \vartheta)'(\hat{\Omega}_f + R^{-1}\bar{\Omega}_g)^{-1}(\bar{\vartheta}_{T,R} - \vartheta) \overset{A}{\sim} \chi^2(2).$$

In consequence, the statistic

$$\mathcal{B}_{T,R} := T\bar{\vartheta}'_{T,R}(\hat{\Omega}_f + R^{-1}\bar{\Omega}_g)^{-1}\bar{\vartheta}_{T,R}, \tag{7}$$

has $\chi^2(2)$ as the asymptotic null distribution. On the other hand, under the local alternative hypothesis: $\vartheta = \delta/\sqrt{T}$, the asymptotic distribution of $\mathcal{B}_{T,R}$ is

$$\mathcal{B}_{T,R} \overset{A}{\sim} \chi^2(2; \Delta_b(R)), \qquad \Delta_b(R) := \delta'(\Omega_f + R^{-1}\Omega_g)^{-1}\delta.$$

Now, the statistic $\mathcal{B}_{T,R}$ can be applied to the simple hypotheses, and will be referred to as the $\mathcal{B}$ test.

Since $\lim_{R\to\infty}(\Omega_f + R^{-1}\Omega_g) = \Omega_f$, $\operatorname{plim}_{R\to\infty}\bar{\mu}_{gs} = \mu_{gs}$, and $\operatorname{plim}_{R\to\infty}\bar{\mu}_{gc} = \mu_{gc}$, the sampling variation of $\mathcal{B}_{T,R}$ would disappear as $R \to \infty$. This also illustrates that, when the re-sampling number R is large enough, the $\mathcal{B}$ test amounts to checking if the estimators $\hat{\mu}_{fs}$ and $\hat{\mu}_{fc}$ are sufficiently close to μ_{gs} and μ_{gc}, the symmetry and dispersion indicators of $\mathcal{G}$.

2.2.2 Conditional Distribution Assumptions

In empirical studies, researchers often need to check conditional distribution assumptions of the model:

$$Y_t = m_t(X_t, \gamma) + \varepsilon_t h_t(X_t, \delta)^{1/2}, \tag{8}$$

where Y_t is the dependent variable; X_t consists of the explanatory variables and lagged dependent variables; m_t and h_t represent the conditional mean and variance of Y_t, respectively. To concentrate on investigating $\mathcal{F}$, the distribution of the standardized errors

ε_t, we assume that m_t and h_t are correctly specified in what follows. After controlling m_t and h_t, the conditional distribution of Y_t is determined by $\mathcal{F}$.

This model is often estimated by the quasi-maximum likelihood method that approximates the unknown $\mathcal{F}$ by a postulated distribution $\mathcal{G}(\cdot\,;\alpha_o)$. Under the hypothesis of $\mathcal{F} = \mathcal{G}$ and suitable regularity conditions, $\hat{\gamma}_T$, $\hat{\delta}_T$, and $\hat{\alpha}_T$ (the QMLEs of γ, δ and α_o) would have consistency, efficiency, and asymptotic normality. However, these properties do not necessarily hold when $\mathcal{F} \neq \mathcal{G}$; see e.g., Gourieroux et al. (1984) and White (1994). Therefore, it is important to check the adequacy of $\mathcal{G}$. However, the introduced $\mathcal{B}$ test is not applicable here because the standardized errors ε_t and the parameters vector α_o are unlikely to be known. To construct the modified test, we will replace the roles of α_o and ε_t in the $\mathcal{B}$ test with the estimator $\hat{\alpha}_T$ and the standardized residuals:

$$\hat{\varepsilon}_t = -\frac{m_t(X_t, \hat{\gamma}_T) - m_t(X_t, \gamma)}{h_t(X_t, \hat{\delta}_T)^{1/2}} + \frac{h_t(X_t, \delta)^{1/2}}{h_t(X_t, \hat{\delta}_T)^{1/2}}\varepsilon_t.$$

Let $\{\hat{\nu}_t(i)\}$ be a sequence of i.i.d. random variables drawn from the distribution $\mathcal{G}(\cdot\,;\hat{\alpha}_T)$, $i = 1, 2, \ldots, R$; $\{\hat{\nu}_t(1)\}, \{\hat{\nu}_t(2)\}, \ldots, \{\hat{\nu}_t(R)\}$ are independent of each other and also independent of $\{\hat{\varepsilon}_t\}$. By replacing $(\varepsilon_t, \nu_t(i))$ with $(\hat{\varepsilon}_t, \hat{\nu}_t(i))$ in the introduced statistics: $\hat{\mu}_{fs}$, $\hat{\mu}_{fc}$, $\hat{\Omega}_f$, $\hat{\mu}_{gs}(i)$, $\hat{\mu}_{gc}(i)$, and $\hat{\Omega}_g(i)$, we denote the new statistics as $\tilde{\mu}_{fs}$, $\tilde{\mu}_{fc}$, $\tilde{\Omega}_f$, $\tilde{\mu}_{gs}(i)$, $\tilde{\mu}_{gc}(i)$, and $\tilde{\Omega}_g(i)$, respectively. In addition, we also define $\ddot{\Omega}_g = R^{-1}\sum_{i=1}^{R}\tilde{\Omega}_g(i)$ and

$$\tilde{\vartheta}_{T,R} = \begin{bmatrix} \tilde{\mu}_{fs} - \ddot{\mu}_{gs} \\ \tilde{\mu}_{fc} - \ddot{\mu}_{gc} \end{bmatrix},$$

where $\ddot{\mu}_{gs} = R^{-1}\sum_{i=1}^{R}\tilde{\mu}_{gs}(i)$ and $\ddot{\mu}_{gc} = R^{-1}\sum_{i=1}^{R}\tilde{\mu}_{gc}(i)$.

From the first-order expansions of $\tilde{\mu}_{fs}$ and $\tilde{\mu}_{fc}$ around the parameters vectors γ and δ, we have

$$\begin{aligned} \sqrt{T}(\tilde{\mu}_{fs} - \mu_{fs}) &= \sqrt{T}(\hat{\mu}_{fs} - \mu_{fs}) + \eta_{s\gamma}\sqrt{T}(\hat{\gamma}_T - \gamma) + \eta_{s\delta}\sqrt{T}(\hat{\delta}_T - \delta) + o_p(1), \\ \sqrt{T}(\tilde{\mu}_{fc} - \mu_{fc}) &= \sqrt{T}(\hat{\mu}_{fc} - \mu_{fc}) + \eta_{c\gamma}\sqrt{T}(\hat{\gamma}_T - \gamma) + \eta_{c\delta}\sqrt{T}(\hat{\delta}_T - \delta) + o_p(1), \end{aligned} \tag{9}$$

where

$$\eta_{s\gamma} = -\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \nabla_\gamma m_t \left[\psi_s'(\varepsilon_t)/h_t^{1/2}\right], \qquad \eta_{s\delta} = -\operatorname*{plim}_{T\to\infty} \frac{1}{2T}\sum_{t=1}^{T} \nabla_\delta h_t \left[\psi_s'(\varepsilon_t)\varepsilon_t/h_t\right],$$

$$\eta_{c\gamma} = -\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \nabla_\gamma m_t \left[\psi_c'(\varepsilon_t)/h_t^{1/2}\right], \qquad \eta_{c\delta} = -\operatorname*{plim}_{T\to\infty} \frac{1}{2T}\sum_{t=1}^{T} \nabla_\delta h_t \left[\psi_c'(\varepsilon_t)\varepsilon_t/h_t\right].$$

Under suitable regularity conditions that ensure asymptotic normality of QMLEs, the statistics $\sqrt{T}(\hat{\mu}_{fs} - \mu_{fs})$, $\sqrt{T}(\hat{\mu}_{fc} - \mu_{fc})$, $\sqrt{T}(\hat{\gamma}_T - \gamma)$, and $\sqrt{T}(\hat{\delta}_T - \delta)$ would have an asymptotic multivariate normal distribution. We denote the variance-covariance matrix of this distribution by V. (The explicit formula of V can be easily derived. To avoid introducing too much notations, we omit it here.) In consequence,

$$\begin{bmatrix} \sqrt{T}(\tilde{\mu}_{fs} - \mu_{fs}) \\ \sqrt{T}(\tilde{\mu}_{fc} - \mu_{fc}) \end{bmatrix} \overset{A}{\sim} N\big(0_{2\times 1}, \Sigma_f\big), \tag{10}$$

where the asymptotic variance-covariance matrix Σ_f is defined in accordance with (9) and V. The consistent estimator for Σ_f, denoted by $\tilde{\Sigma}_f$, can be obtained by the kernel covariance estimators of Andrews (1991), Politis and Romano (1994), and among others, or by bootstrap. The bootstrap estimator would be especially useful for complicated models.

Let $\mathbb{E}_{\hat{g}}$ be the expectation operator taken with respect to $\mathcal{G}(\cdot\,;\hat{\alpha}_T)$. From the first order-expansion of $\mathcal{G}(\cdot\,;\hat{\alpha}_T)$ around the parameters vector α_o, it yields

$$\begin{aligned}\mathbb{E}_{\hat{g}}[\psi_s(\hat{\nu}_t(i))] &= \int_{\mathbf{R}} \psi_s(\epsilon)\mathrm{d}\mathcal{G}(\epsilon, \hat{\alpha}_T) \\ &= \int_{\mathbf{R}} \psi_s(\epsilon)\mathrm{d}\mathcal{G}(\epsilon, \alpha_o) + (\hat{\alpha}_T - \alpha_o)\int_{\mathbf{R}} \psi_s(\epsilon)\frac{\partial}{\partial\alpha}\log\left(\frac{\partial}{\partial\epsilon}\mathcal{G}(\epsilon, \alpha_o)\right)\mathrm{d}\mathcal{G}(\epsilon, \alpha_o) + o_p(1).\end{aligned}$$

Suppose that the probability density function of $\mathcal{G}$ is differentiable with respect to α_o. The above expansion implies

$$\operatorname*{plim}_{T\to\infty} \mathbb{E}_{\hat{g}}[\psi_s(\hat{\nu}_t(i))] = \int_{\mathbf{R}} \psi_s(\epsilon)\mathrm{d}\mathcal{G}(\epsilon, \alpha_o) = \mu_{gs}$$

provided that $\hat{\alpha}_T$ is a consistent estimator of α_o. Since the law of large numbers ensures that $\ddot{\mu}_{gs}$ would converge in probability to $\mathbb{E}_{\hat{g}}[\psi_s(\hat{\nu}_t(i))]$ as $R \to \infty$, the estimator $\ddot{\mu}_{gs}$

is consistent for μ_{gs} provided that R is large enough. Similarly, it can be shown that $\ddot{\mu}_{gc}$ and $\ddot{\Omega}_g$ are also consistent for μ_{gc} and Ω_g, respectively. By invoking a central limit theorem, we have

$$\begin{bmatrix} \sqrt{T}(\ddot{\mu}_{gs} - \mu_{gs}) \\ \sqrt{T}(\ddot{\mu}_{gc} - \mu_{gc}) \end{bmatrix} \overset{A}{\sim} N\big(0_{2\times 1}, R^{-1}\Omega_g\big). \tag{11}$$

Combined with the independence of $\{\hat{\varepsilon}_t\}$ and $\{\hat{\nu}_t(i)\}$, (10) and (11) imply

$$\sqrt{T}(\tilde{\vartheta}_{T,R} - \vartheta) \overset{A}{\sim} N\big(0_{2\times 1},\ \Sigma_f + R^{-1}\Omega_g\big),$$

or asymptotically equivalently,

$$T(\tilde{\vartheta}_{T,R} - \vartheta)'(\tilde{\Sigma}_f + R^{-1}\ddot{\Omega}_g)^{-1}(\tilde{\vartheta}_{T,R} - \vartheta) \overset{A}{\sim} \chi^2(2).$$

Therefore, the statistic

$$\mathcal{C}_{T,R} := T\tilde{\vartheta}'_{T,R}(\tilde{\Sigma}_f + R^{-1}\ddot{\Omega}_g)^{-1}\tilde{\vartheta}_{T,R} \tag{12}$$

has the asymptotic distribution $\chi^2(2)$ under the null hypothesis. On the other hand, under the alternative hypothesis $\vartheta = \delta/\sqrt{T}$, the asymptotic distribution of $\mathcal{C}_{T,R}$ is

$$\mathcal{C}_{T,R} \overset{A}{\sim} \chi^2(2; \Delta_c(R)), \qquad \Delta_c(R) := \delta'(\Sigma_f + R^{-1}\Omega_g)^{-1}\delta.$$

This statistic is a modified version of $\mathcal{B}_{T,R}$ for testing conditional distribution assumptions of (8). Similar to $\mathcal{B}_{T,R}$, the sampling variation of $\mathcal{C}_{T,R}$ disappears as $R \to \infty$. This modified test will be referred to as the $\mathcal{C}$ test.

We remark that, given a static transformation ϕ, the roles of (ε_t, ν_t) in the $\mathcal{A}$ test, $(\varepsilon_t, \nu_t(i))$ in the $\mathcal{B}$ test, and $(\hat{\varepsilon}_t, \hat{\nu}_t(i))$ in the $\mathcal{C}$ test may be replaced with $(\phi(\varepsilon_t), \phi(\nu_t))$, $(\phi(\varepsilon_t), \phi(\nu_t(i)))$, and $(\phi(\hat{\varepsilon}_t), \phi(\hat{\nu}_t(i)))$, respectively. This replacement would not change the asymptotic null distribution of the proposed tests because the null hypothesis is irrelevant to static transformations. This point will be further discussed in the next section.

3 Implementation of the Proposed Tests

The proposed tests are indeed a class of distribution tests that depend on the weighting function z and the transformation ϕ. In this section, we discuss how to choose these two ingredients.

3.1 The Weighting Function z

Recall that z is defined as the probability density function of a lifetime distribution. Under this definition, the functions ψ_s and ψ_c in (3) are exactly the imaginary and real parts of the characteristic function corresponding to the probability density function z. Therefore, $\psi_s(\epsilon)$ and $\psi_c(\epsilon)$ may have explicit formula when z has an analytic characteristic function. Chen et al. (2000) showed that when z is of the exponential distribution (denoted as $z = \exp$) with the parameter $\beta > 0$:

$$z(\omega) = \exp(-\omega/\beta)/\beta, \qquad \omega \geq 0, \tag{13}$$

the function ψ_s is of the form:

$$\psi_s(\epsilon) = \frac{\beta\epsilon}{1 + (\beta\epsilon)^2}. \tag{14}$$

They utilized the sample mean of this $\psi_s(\varepsilon_t)$ to construct a class of unconditional symmetry tests. Indeed, we can also show that, given $z = \exp$, the function ψ_c is

$$\psi_c(\epsilon) = \frac{1}{1 + (\beta\epsilon)^2}. \tag{15}$$

Therefore, it would be very convenient to calculate the proposed test statistics with $z = \exp$; cf. the non-parametric tests of Bowman (1992) and Fan (1998).

Ushakov (1999) showed that the Bessel, Gamma, hyper-exponential, and many other lifetime distributions also possess analytic characteristic functions. It means that we may derive other analytic $\psi_s(\epsilon)$ and $\psi_c(\epsilon)$ by setting z as one of these distributions' probability density functions. However, it should be noted that the resulting $\psi_s(\epsilon)$ and $\psi_c(\epsilon)$ would be much more complicated than (14) and (15).

The choice of $z = \exp$ does not only provide very simple ψ_s and ψ_c but also possess some desirable features. First, if $\mathcal{F}$ and $\mathcal{G}$ are absolutely continuous distributions, then their characteristic functions k_f and k_g would decay to zero as $\omega \to \infty$; see Lukacs (1970). As a result, $e(\omega) \to 0_{2\times 1}$ as $\omega \to \infty$, whether or not $\mathcal{F}$ and $\mathcal{G}$ are the same. This implies that a sensible weighting function z should not put too much weight on large ω. From (13), it is easy to see that $z = \exp$ meets this requirement.

Second, $z = \exp$ allows the proposed tests to capture distribution information at different frequency in a flexible way. By setting a smaller β, $z = \exp$ would be more concentrated on lower frequency (smaller ω); $z = \exp$ would account for more components of $e(\omega)$ at higher frequency (larger ω) as β becomes larger. On the contrary, the conventional distribution tests do not have this flexibility. Durbin and Knott (1972) and Fan (1996) showed that the KS and CV tests can only efficiently capture distribution information at very low frequency. Fan (1998) further demonstrated that the CV test statistic is indeed a weighted sum of the squared differences between the empirical and postulated characteristic functions with the weighting function ω^{-2}. Therefore, these empirical-distribution-function-based tests may not be able to tell the difference between the distributions that behave similarly around the zero frequency. By contrast, the proposed tests do not suffer from this limitation when the parameter β is properly chosen.

We note that the flexibility of the proposed tests could be further improved by replacing $z = \exp$ with other generalized density functions, such as the Weibull or Gamma distribution. However, as compared to $z = \exp$, these generalized weighting functions yield much more complicated test statistics, and selecting the parameters of these functions is also difficult. Hence, we confine ourselves to $z = \exp$ in what follows.

3.2 The Parameter $\boldsymbol{\beta}$

The proposed tests with $z = \exp$ depend on the value of β. Since the parameter ϑ is always equal to $0_{2\times 1}$ under the null hypothesis, the choice of β should have no effect on the size performance of the proposed tests. However, under the alternative hypothesis, the parameters ϑ, Ω_f, Ω_g, and Σ_f are functions of β, so that the power performance of these tests may be affected by the choice of β. To determine the appropriate β value, it is important to note that the powers of the $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ tests would increase with the the non-centrality parameters $N_a(\beta) = \vartheta'(\Omega_f + \Omega_g)^{-1}\vartheta$, $N_b(\beta) = \vartheta'\Omega_f^{-1}\vartheta$, and $N_c(\beta) = \vartheta'\Sigma_f^{-1}\vartheta$, respectively, as $T \to \infty$ and $R \to \infty$. Therefore, it is natural to choose β as the maximizer of these non-centrality parameters. The parameter β determined in this way would be optimal for the proposed tests with $z = \exp$ as $T \to \infty$ and $R \to \infty$.

However, it is hard to derive these optimal β in practice because the non-centrality parameters are unknown. Nonetheless, we may still approximate these β by estimating the non-centrality parameters. Suppose that $N_a(\beta)$, $N_b(\beta)$, and $N_c(\beta)$ are smooth with respect to β and that they can be consistently estimated by $\hat{N}_a(\beta) = \hat{\vartheta}'_T(\hat{\Omega}_f + \hat{\Omega}_g)^{-1}\hat{\vartheta}_T$, $\bar{N}_b(\beta) = \bar{\vartheta}'_{T,R}\hat{\Omega}_f^{-1}\bar{\vartheta}_{T,R}$, and $\tilde{N}_c(\beta) = \tilde{\vartheta}'_{T,R}\tilde{\Sigma}_f^{-1}\tilde{\vartheta}_{T,R}$, respectively. Then the maximizers:

$$\beta_a^* = \operatorname*{argmax}_{\beta>0} \hat{N}_a(\beta), \qquad \beta_b^* = \operatorname*{argmax}_{\beta>0} \bar{N}_b(\beta), \qquad \beta_c^* = \operatorname*{argmax}_{\beta>0} \tilde{N}_c(\beta). \tag{16}$$

would, respectively, converge in probability to the optimal β of the $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ tests with $z = \exp$ as $T \to \infty$ and $R \to \infty$. (In applications, the values of β_a^*, β_b^*, and β_c^* can be determined by trying a set of grid points of β.) This approximation allows the data to reveal the best way in accounting for $e(\omega)$ at different frequencies, and hence is a possible solution to the choice of β for the proposed tests.

3.3 The Transformation ϕ

As noted previously, we can base the proposed tests on the static transformation ϕ. In other words, we can redefine the functions ψ_s and ψ_c as

$$\psi_s(\epsilon) = \int_{\mathbf{R}^+} \sin(\omega\phi(\epsilon))z(\omega)\mathrm{d}\omega, \qquad \psi_c(\epsilon) = \int_{\mathbf{R}^+} \cos(\omega\phi(\epsilon))z(\omega)\mathrm{d}\omega, \tag{17}$$

and rewrite the test statistics accordingly. Similar to the choice of z and β, the selection of ϕ should has no effect on the size performance of the proposed tests because $\phi(\varepsilon_t)$ and $\phi(\nu_t)$ have the same distribution under the null hypothesis. On the other hand, under the alternative hypothesis, the choice of ϕ may affect the non-centrality parameters, and hence the power performance of the proposed tests.

In general, it is difficult to analyze the relationship between the power performance and the transformation ϕ. However, when $\mathcal{F}$ and $\mathcal{G}$ are both symmetric (about the origin), it is possible to improve the powers by considering ϕ that transforms $\mathcal{F}$ and $\mathcal{G}$

to be asymmetric. The intuition is that the symmetry indicators μ_{fs} and μ_{gs} are both equal to zero regardless of the null hypothesis when $\mathcal{F}$ and $\mathcal{G}$ are both symmetric; in this case, the powers of the proposed tests with $\phi(\epsilon) = \epsilon$ are simply due to the difference between the dispersion indicators μ_{fc} and μ_{gc}. On the contrary, if we base the tests on the transformation $\phi(\epsilon) = \epsilon^2$, then the powers would increase because the differences of the symmetry and dispersion indicators of the transformed distributions can be used. In the next section, this point will be verified by Monte Carlo simulation.

4 Monte Carlo Simulation

To simplify the Monte Carlo simulation, we focus on comparing finite sample performance of the $\mathcal{B}$ test with the KS, CV, and PC tests for simple hypotheses with different pairs of $(\mathcal{F}, \mathcal{G})$. In this experiment, the normal size is 5%, the sample size is $T = 50, 200$, and the number of replications is 1000. To calculate the $\mathcal{B}$ test statistic, $\mathcal{B}_{T,R}$, we set $R = 1000$ and $z = \exp$, and consider $(\phi(\epsilon), \beta) = (\epsilon, 1)$, (ϵ, β^*), $(\epsilon^2, 1)$, and (ϵ^2, β^*), where β^* is the sample average of β_b^* that appear in 1000 replications with $T = 200$. This β^* is employed to represent an approximation of the optimal β that controls the sampling variation of β_b^*. The values of β^* vary with $(\mathcal{F}, \mathcal{G})$ but are between 0.1 and 2.5.

The KS and CV test statistics are of the forms:

$$\mathrm{KS}_T = \max\left(\max_{1\le t\le T}\left(\frac{t}{T} - \mathcal{G}(\varepsilon_{(t)};\ \alpha_o)\right), \max_{1\le t\le T}\left(\mathcal{G}(\varepsilon_{(t)};\ \alpha_o) - \frac{t-1}{T}\right)\right),$$

and

$$\mathrm{CV}_T = \sum_{t=1}^{T}\left(\mathcal{G}(\varepsilon_{(t)};\ \alpha_o) - \frac{t-0.5}{T}\right)^2 + \frac{1}{12T},$$

where $\varepsilon_{(1)} \le \varepsilon_{(2)} \le \ldots \le \varepsilon_{(T)}$ are the order statistics of ε_t. The 95% critical values of $\mathrm{KS}_T(T^{1/2} + 0.12 + 0.11T^{-1/2})$ and $(\mathrm{CV}_T^2 - 0.4T^{-1} + 0.6T^{-2})(1 + T^{-1})$ are 1.358 and 0.461, respectively; see Stephens (1974). The PC test statistic is

$$\mathrm{PC}_T = \sum_{j=1}^{p}(d_j - o_j)^2/o_j, \qquad o_j = T\int_{c_{j-1}}^{c_j} \mathrm{d}\mathcal{G}(\epsilon, \alpha_o),$$

where $c_0 = -\infty$, $c_p = \infty$, $(c_{j-1}, c_j]$ denotes the j-th data cell, and d_j is the number of observations in the j-th cell. The asymptotic null distribution of PC_T is $\chi^2(p-1)$.

Following Moore (1986), we adopt the equi-probable data cells with $o_j = T/(p+1)$, and set p as the integer part of $4\left(\sqrt{2}T/1.645\right)^{2/5} - 1$.

In this simulation study, the $\mathcal{F}$ and $\mathcal{G}$ considered are as follows:

1. the standard normal distribution $N(0,1)$;

2. the Student's t distribution with the degrees of freedom $\alpha_o = 4$; denoted as $t(4)$;

3. the Cauchy distribution (or $t(1)$); denoted as CA;

4. the logistic distribution:

   $$\mathcal{G}(\epsilon, \alpha_o) = (1 + \exp(-\epsilon/\alpha_o))^{-1}$$

   with the scale parameter $\alpha_o = 1$; denoted as LG;

5. the generalized error distribution:

   $$\frac{\partial}{\partial \epsilon}\mathcal{G}(\epsilon, \alpha_o) = \alpha_o \exp\left(-\frac{1}{2}\left|\epsilon/\kappa(\alpha_o)\right|^{\alpha_o}\right)\left[\kappa(\alpha_o)\Gamma(1/\alpha_o)2^{(1+1/\alpha_o)}\right]^{-1}, \qquad \alpha_o > 0,$$

   where Γ is the gamma function, and

   $$\kappa(\alpha_o) = [2^{(-2/\alpha_o)}\Gamma(1/\alpha_o)\Gamma(3/\alpha_o)]^{1/2};$$

   see Nelson (1991); we denote this distribution with $\alpha_o = 1.5$ as GE;

6. the generalized lambda distribution of Ramberg and Schmeiser (1974):

   $$\varepsilon_t = \lambda_1 + (u_t^{\lambda_3} - (1-u_t)^{\lambda_4})/\lambda_2, \qquad u_t \sim U(0,1)$$

   with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, -0.4, -0.16, -0.16)$ and $(0, -0.4, -0.1, -0.2)$; denoted as GL_s and GL_a, respectively;

7. the standardized log-normal distribution:

   $$\varepsilon_t = \frac{\exp(\alpha_o u_t - \alpha_o^2/2) - 1}{(\exp(\alpha_o^2) - 1)^{1/2}}, \qquad u_t \sim N(0,1)$$

   with the asymmetry parameter $\alpha_o = 0.5$; denoted as LN.

These distributions are considered because they were used for approximating marginal distributions of asset returns and conditional distributions of GARCH models. Some of them will also be used in our empirical study later. Among these distributions, GL_a and LN are asymmetric, and the others are symmetric distributions with different tail and peakedness properties. The t distribution, CA, LG, GE, and GL_s have longer tails and higher peakedness than $N(0,1)$. In empirical studies, researchers are often of interest to discriminate between these leptokurtic distributions; see e.g., Hsieh (1988), Baillie and DeGennaro (1990), and Perió (1994).

In Table 1, we summarize the empirical rejection frequency of these tests with respect to $T = 50$. The case of $T = 200$ is shown in Table 2. In these two tables, the columns $\mathcal{B}_1$, $\mathcal{B}_1^*$, $\mathcal{B}_2$, and $\mathcal{B}_2^*$ represent the $\mathcal{B}$ test with $(\phi(\epsilon), \beta) = (\epsilon, 1)$, (ϵ, β^*), $(\epsilon^2, 1)$, and (ϵ^2, β^*), respectively. For the size performance (the cases of $\mathcal{F} = \mathcal{G}$), the empirical sizes of the KS, CV, PC, and $\mathcal{B}$ tests are all around with the 5% nominal level, and the choice of $(\phi(\epsilon), \beta)$ do not change the size performance of the $\mathcal{B}$ test, as noted previously.

[Tables 1 and 2 about here.]

For the power performance (the cases of $\mathcal{F} \neq \mathcal{G}$), the powers of the $\mathcal{B}$ test would change with the choice of $(\phi(\epsilon), \beta)$. Given the choice of ϕ, the test with $\beta = \beta^*$ ($\mathcal{B}_1^*$ or $\mathcal{B}_2^*$) outperforms the test with $\beta = 1$ ($\mathcal{B}_1$ or $\mathcal{B}_2$) for most $(\mathcal{F}, \mathcal{G})$ considered. More importantly, the superiority of $\beta = \beta^*$ over $\beta = 1$ could be very significant for certain distributions. For example, given $T = 50$ and $\phi(\epsilon) = \epsilon$, the change of β from 1 to β^* increases the power of the $\mathcal{B}$ test from 50.3% to 80.2% when $(\mathcal{F}, \mathcal{G}) = (CA, t(4))$; the power is even improved from 6.5% to 100% when $(\mathcal{F}, \mathcal{G}) = (LG, CA)$. This suggests that (16) provides a useful choice of β.

When the selection of β is fixed, the $\mathcal{B}$ test with $\phi(\epsilon) = \epsilon^2$ ($\mathcal{B}_2$ or $\mathcal{B}_2^*$) outperforms the test with $\phi(\epsilon) = \epsilon$ ($\mathcal{B}_1$ or $\mathcal{B}_1^*$) for most *symmetric* $\mathcal{F}$ and $\mathcal{G}$. However, the performance is reversed when one of $\mathcal{F}$ and $\mathcal{G}$ is asymmetric. For example, given $T = 200$ and $\beta = \beta^*$, the power of the $\mathcal{B}$ test is improved from 66.5% to 81.7% by changing $\phi(\epsilon)$ from ϵ to ϵ^2 as $(\mathcal{F}, \mathcal{G}) = (t(4), N(0,1))$, but this change reduces the power from 96.4% to 52.4% when $(\mathcal{F}, \mathcal{G}) = (N(0,1), LN)$. This demonstrates that the static transformation $\phi(\epsilon) = \epsilon^2$ may be suitable for symmetric $\mathcal{F}$ and $\mathcal{G}$ but not for asymmetric distributions. In consequence, a pre-test for the symmetry of $\mathcal{F}$ may be useful to choose one from $\phi(\epsilon) = \epsilon$ and $\phi(\epsilon) = \epsilon^2$.

For the comparison of the $\mathcal{B}$ test and the KS, CV, and PC tests, Tables 1 and 2 show that the $\mathcal{B}$ test is superior to the conventional tests for symmetric $\mathcal{F}$ and $\mathcal{G}$ and that the superiority of the $\mathcal{B}$ test could be quite significant. For example, given $T = 50$ (200) and $(\mathcal{F}, \mathcal{G}) = (GE, t(4))$, the powers of the KS, CV, and PC tests are 6.8%, 6.6%, and 13.1% (23.3%, 24.1%, and 38.6%). By contrast, the power of the $\mathcal{B}$ test with $(\phi(\epsilon), \beta) = (\epsilon, \beta^*)$ is 74.2% (99.8%). Similar examples, such as $(\mathcal{F}, \mathcal{G}) = (N(0, 1), t(4))$ and (CA, LG), can also be observed from these two tables. A possible interpretation to this result is that the characteristic functions of symmetric $\mathcal{F}$ and $\mathcal{G}$ are real and both of the value one at $\omega = 0$, so that the KS and CV tests cannot efficiently tell their differences; on the other hand, the $\mathcal{B}$ test can still discriminate between different symmetric distributions by utilizing the information at higher frequencies (larger ω). For asymmetric distributions, the $\mathcal{B}$ test also performs similar to or even superior to the best among the conventional tests provided that $(\phi(\epsilon), \beta)$ is properly chosen. This comparison supports that the $\mathcal{B}$ test is a useful complement to the conventional distribution tests, especially when both the true and postulated distributions are symmetric.

5 An Empirical Study

It is well documented that asset returns may be explained by the GARCH model of Bollerslev (1986). This model is typically estimated by the quasi-maximum likelihood method with the conditional normality assumption. However, some empirical studies reported that conditional distributions of GARCH models ought to be leptokurtic rather than normal; see e.g., Bollerslev (1987), Baillie and DeGennaro (1990), and Su and Fleisher (1998). If this is true, then the conditional normality assumption may lead to two undesirable results. First, the resulting QMLEs are inefficient. Second, the volatility of asset returns may be mistakenly interpreted by the estimated GARCH model because this assumption cannot fully account for the occurrence of big shocks like market crashes, as compared to leptokurtic distributions.

In this empirical study, we will apply the $\mathcal{C}$ test to check the adequacy of normal and certain leptokurtic distributions for estimating the GARCH model, and compare the influence of these distribution assumptions on estimating and interpreting GARCH models.

5.1 Data and Model

The data for our analysis are seven daily stock price indices, including the All Ordinaries index (AO) in Australia, the Toronto Stock Exchange index (TS) in Canada, the Heng Seng index (HS) in Hong Kong, the Neikkei 225 index (NK) in Japan, the Financial Times Stock Exchange 100 index (FT) in the U.K., and the Dow Jones Industrial Average index (DJ) and the Russell 2000 index (RS) in the U.S. These data are taken from Yahoo! Finance. The sample period is from January 1, 1990 through December 31, 1999. The number of observations vary with indices but are approximately 2500.

Let $r_t = 100 \times (\log P_t - \log P_{t-1})$ be the daily return of the stock price index P_t. In Table 3 we show some summary statistics of r_t. The sample kurtosis indicates that the distributions of these returns are highly leptokurtic. The Box-Pierce's $Q(30)$ statistic based on $\{r_t\}$ suggests that these returns are serially correlated. To filter out such correlations, we fit r_t by the ARMA(p,q) models with $p = 1, \ldots, 20$ and $q = 1, \ldots, 5$, and use the Akaike's information criterion to select the fitted models. The adequacy of the selected ARMA models is judged by the $Q(30)$ statistics, calculated from $\{Y_t\}$ and $\{Y_t^2\}$, as showed in Table 3, where $\{Y_t\}$ denotes the residuals of these ARMA models. These Q statistics reveal that $\{Y_t\}$ are serially uncorrelated but that $\{Y_t^2\}$ are strongly serially correlated. This result motivates us to consider the GARCH(1,1) model:

$$Y_t = \varepsilon_t h_t^{1/2}, \qquad h_t = w + \rho_1 Y_{t-1}^2 + \rho_2 h_{t-1}, \tag{18}$$

where $w, \rho_1, \rho_2 \geq 0$, $\rho_1 + \rho_2 < 1$, and h_t is the conditional variance of Y_t that measures the volatility of r_t; the standardized errors ε_t have the unknown distribution $\mathcal{F}$.

[Table 3 about here.]

We first estimate the GARCH(1,1) model by the quasi-maximum likelihood method with the conditional normality assumption: $\mathcal{F} = N(0,1)$. If h_t is correctly specified, this assumption would at least ensure consistency and asymptotic normality of the resulting QMLEs. Let $\hat{\varepsilon}_t$ be the resulting standardized residuals. In Table 4, we show the $Q(30)$ statistic based on $\{\hat{\varepsilon}_t^2\}$ and Engel's (1982) LM test statistic. Both of these statistics support that the GARCH(1,1) model successfully explains the serial correlation in $\{Y_t^2\}$. In the same table, the sample kurtosis of $\hat{\varepsilon}_t$ indicates that $\mathcal{F}$ may be leptokurtic for all the returns. This provides an evidence against the conditional normality assumption. To check the symmetry of $\mathcal{F}$, we adopt the conditional symmetry test of Chen (2001) as described in the footnote of this table. This test accepts that $\mathcal{F}$ is symmetric, except for RS. These testing results together suggest that all the returns but RS have symmetric and leptokurtic $\mathcal{F}$.

[Table 4 about here.]

Note that the observed non-normality might also be attributed to model misspecification of (8). To be sure, we adopt the BDS test of Brock et al. (1996) to check if the GARCH(1,1) model is correctly specified. The BDS test statistics in Table 4 accept that this model is correctly specified for the returns of AO, DJ, HS, NK, and TS, but reject the null for FT and RS at the 5% significance level. This indicates that the GARCH(1,1) model cannot fully explain the dependence structure of the returns of FT and RS. To concentrate on our study of $\mathcal{F}$, in what follows we will only discuss the cases of AO, DJ, HS, NK, and TS.

5.2 Conditional Distribution Assumptions

Since $\mathcal{F}$ may be symmetric and leptokurtic for the returns considered, we adopt the standardized t, logistic, and generalized error distributions as competitors to $N(0,1)$ for approximating the unknown $\mathcal{F}$. The generalized lambda distribution is also adopted because it encompasses a wide variety of distributions; see Ramberg and Schmeiser (1974) for details.

The adequacy of the above postulated distributions will be checked by the $\mathcal{C}$ test with $R = 1000$, $z = \exp$, and $(\phi(\epsilon), \beta) = (\epsilon, 1)$, (ϵ, β_c^*), $(\epsilon^2, 1)$, and (ϵ^2, β_c^*), in which the asymptotic variance-covariance estimator $\tilde{\Sigma}_f$ is computed by bootstrap with 200 re-samples. To draw the realizations of $\hat{\nu}_t(i)$ from $\mathcal{G}(\cdot;\ \hat{\alpha}_T)$, we estimate the parameters of the standardized t, logistic, and generalized error distributions by the quasi-maximum likelihood method. Because the generalized lambda distribution does not have an analytic probability density function, its parameter is estimated by the percentile method of Karian and Dudewicz (2000). The test statistic $\mathcal{C}_{T,R}$ and the value of β_c^* are shown in Table 5; the estimators $\hat{\alpha}_T$ are summarized in Table 6.

[Tables 5 and 6 about here.]

Table 5 shows that the $\mathcal{C}$ test significantly rejects the hypothesis of $\mathcal{F} = N(0,1)$ at 5% level (or even 1% level) for all $(\phi(\epsilon), \beta)$ and the returns considered. This result is consistent with the fact that $\mathcal{F}$ is leptokurtic for all the returns. By contrast, the $\mathcal{C}$ test cannot reject the standardized t distribution at any reasonable significance level for all $(\phi(\epsilon), \beta)$ and the returns considered. This may not be surprising because the standardized t distribution is leptokurtic distributions when its degrees of freedom is not too large. For other distribution assumptions, the generalized lambda distribution is accepted for the returns of AO, DJ, HS, and NK, but it is rejected for the case of TS at the 5% significance level as $(\phi(\epsilon), \beta) = (\epsilon^2, \beta_c^*)$. The logistic distribution is proper for all the returns but HS. This reflects that the logistic distribution is less flexible than the standardized t distribution because its kurtosis is a constant 4.2; see Johnson et al. (1995). The $\mathcal{C}$ test also rejects the generalized error distribution for the returns of AO, DJ, and HS at 5% level, so that it is an in-between case of $N(0,1)$ and the standardized t distribution.

To confirm the results of the $\mathcal{C}$ test, we also compare the Gaussian kernel density estimate of $\mathcal{F}$ with the probability density function of $\mathcal{G}(\cdot\ ; \hat{\alpha}_T)$. In Figure 1 we plot this estimate of $\mathcal{F}$ and the density functions of $N(0,1)$ and the fitted standardized t distribution for the returns of DJ and HS; other returns are omitted because their results are quite similar. This figure shows that the density estimate of $\mathcal{F}$ has obviously higher peakedness than the density function of $N(0,1)$, but it is very close to the density function of the standardized t distribution. This comparison supports the results in Table 5.

[Figure 1 about here.]

In summary, the standardized t distribution appears to be a good approximation to $\mathcal{F}$ for all the returns considered, while the conditional normality is not. The appropriateness of other distributions is data-specific. We also remark that the $\mathcal{C}$ test with $(\phi(\epsilon), \beta) = (\epsilon^2, \beta_c^*)$ is more powerful than the test with other $(\phi(\epsilon), \beta)$; see Table 5. This also confirms why we may want to consider the transformation of $\phi(\epsilon) = \epsilon^2$.

5.3 Models with Different Conditional Distributions

To explore the effects of conditional distribution assumptions, we estimate the GARCH(1,1) model by the quasi-maximum likelihood method with $N(0,1)$, standardized t, logistic, and generalized error distributions, as showed in Table 6. The GARCH(1,1) model accompanied with these distribution assumptions are referred to as GARCH-n, GARCH-t, GARCH-l, and GARCH-g, respectively. Since GARCH-l and GARCH-t yield quite similar QMLEs, in Table 7 we report only the estimators of GARCH-n, GARCH-g, and GARCH-t.

[Table 7 about here.]

This table shows some common features between the returns considered. First, the QMLEs of GARCH-t are more efficient than those of GARCH-n and GARCH-g because they all have smaller standard deviations. Second, the QMLEs and their standard deviations of GARCH-g are between those of GARCH-n and GARCH-t, suggesting again that the generalized error distribution is an in-between case of $N(0,1)$ and the standardized t distribution for approximating $\mathcal{F}$. Third, GARCH-n modes yield larger $\hat{w}$ and $\hat{\rho}_1$ but smaller $\hat{\rho}_2$ than GARCH-t. It means that the volatility implied by the GARCH(1,1) model may change with the conditional distribution assumptions. More specifically, this model with the conditional normality assumptions may over-value the impact effect of external shocks on volatility but under-value the shocks' persistence.

To illustrate this point, we compare the conditional variances estimated by GARCH-n and GARCH-t, which are denoted as $\hat{h}_t^n$ and $\hat{h}_t^t$, respectively. Let $\mathcal{D}\hat{h}_t = \hat{h}_t^n - \hat{h}_t^t$. In Figure 2, we plot the logarithm of HS index and the corresponding path of $\mathcal{D}\hat{h}_t$. The other indices are omitted because they have similar patterns. This figure indicates that the behavior of $\mathcal{D}\hat{h}_t$ varies with the market status. When the market is tranquil, $\mathcal{D}\hat{h}_t$ are around the origin; hence, different conditional distribution assumptions do not make a big difference in explaining the volatility. However, when the market is volatile, $\mathcal{D}\hat{h}_t$ has much larger fluctuations. It jumps synchronously with the market crashes. After the crashes, it falls rapidly to a negative value, and returns to the origin as the market becomes tranquil again. Basing on our test results and taking the GARCH-t variance estimates as a bench mark, we conclude that the conditional normality assumption tends to over-estimate the impact effect of shocks on volatility but under-estimate these shocks' persistence effect.

[Figure 2 about here.]

In Figure 2, the fluctuations of $\mathcal{D}\hat{h}_t$ are evident for the crashes of HS index in August 1990, August 1991, December 1993, January 1994, March 1996, and the Asian Financial Crisis spreading from September and October 1997. (The dotted lines in this figure denote these crashes.) In particular, the Asian Financial Crisis caused an especially volatile fluctuation of $\mathcal{D}\hat{h}_t$. This evidence suggests that the adequacy of the conditional distribution assumption of the GARCH model would be particularly important for interpreting the volatility behavior of the returns when the market is volatile.

6 Conclusions

In this paper we propose a new class of characteristic-function-based distribution tests. The proposed tests statistics are easy to compute and have the asymptotic null distribution $\chi^2(2)$. As compared to conventional distribution tests, the proposed test can flexibly account for the difference between the true and postulated distributions at different frequencies. The Monte Carlo simulation shows that the proposed test significantly outperforms the KS, CV, and Pearson's χ^2 tests, especially when the true and postulated distributions are both symmetric. In an empirical study of stock index returns, we apply the proposed test to examine conditional distribution assumptions of GARCH(1,1) model. The proposed test rejects the conditional normality assumption, but accepts the standardized Student's t distribution for the returns considered. We also find that the conditional normality assumption does not only render inefficient QMLEs of GARCH(1,1) model but also over-estimate the impact effect and under-estimate the persistence effect of shocks on volatility. The latter effects are particularly evident when the market is volatile.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, **59**, 817–858.

Andrews, D. W. K. (1997). A conditional Kolmogorov test, *Econometrica*, **65**, 1097–1128.

Baillie, R. T. and R. P. DeGennaro (1990). Stock returns and volatility, *Journal of Financial and Quantitative Analysis*, **25**, 203–214.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.

Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative sprices and rates of return, *Review of Economics and Statistics*, **69**, 542–547.

Bowman, A. W. (1992). Density based tests for goodness-of-fit, *Journal of Statistical Computation and Simulation*, **40**, 1–13.

Brock, W. A., W. D. Dechert, J. A. Scheinkman and B. LeBaron (1996). A test for independence based on the correlation dimension, *Econometric Review*, **15**, 197–235.

Chen, Y.-T. (2001). Testing conditional symmetry with an application to financial returns, Working Paper, Institute for Social Sciences and Philosophy, Academia Sinica.

Chen, Y.-T., R. Y. Chou, and C.-M. Kuan (2000). Testing time reversibility without moment restrictions, *Journal of Econometrics*, **95**, 199–218.

Davidson, R. and J. G. MacKinnon (1994). *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Durbin, J. and M. Knott (1972). Components of Cramér-von Mises statistics: I, *Journal of the Royal Statistical Association*, **B, 42**, 53–63.

EuBank, R. L. and V. N. LaRiccia (1992). Asymptotic comparison of Cramér-von Mises and non-parametric function estimation techniques for testing goodness-of-fit, *Annals of Statistics*, **20**, 2071–2086.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1006.

Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation, *Journal of the American Statistical Association*, **91**, 674–688.

Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters, *Econometric Theory*, **14**, 604-621.

Feuerverger, A. and R. A. Mureika (1977). The empirical characteristic function and its applications, *Annals of Statistics*, **5**, 88–97.

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Boca Raton: Chapman & Hall Press.

Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo-maximum likelihood methods: Theory, *Econometrica*, **52**, 681–700.

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

Jarque, C. M. and A. K. Bera (1980). Efficient tests for normality, homoscedaticity and serial independence of regression residuals, *Economics Letters*, **6**, 255–259.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions*, **2**, New York: Wiley.

Karian, Z. A. and E. J. Dudewicz (2000). *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*, Boca Raton: CRC Press.

Kon, S. J. (1984). Models of stock returns: A comparison, *Journal of Finance*, **39**, 147–165.

Lukacs, E. (1970). *Characteristic Functions*, London: Griffin.

Moore, D. S. (1986). Tests of chi-squared type, in D'Agostino, R. B. and M. A. Stephens(Ed.), *Goodness-of-Fit Techniques*, New York: M. Dekker, 63–95.

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

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap, *Journal of the American Statistical Association*, **89**, 1303–1313.

Peiró, A. (1994). The distribution of stock returns: International evidence, *Applied Financial Economics*, **4**, 431–439.

Ramberg, J. S. and B. W. Schmeiser (1974). An approximate method for generating asymmetric random variables, *Communications of the Association for Computing Machinery*, **17**, 78–82.

Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons, *Journal of the American Statistical Association*, **69**, 730–737.

Su, D. and B. M. Fleisher (1998). Risk, return and regulation in Chinese stock markets, *Journal of Economics and Business*, **50**, 239–256.

Ushakov, N. G. (1999). *Selected Topics in Characteristic Functions*, Utrecht: VNU Science Press.

White, H. (1994). *Estimation, Inference and Specification Analysis*, Cambridge University Press.

Zheng, J. X. (1998). Consistent specification testing for conditional symmetry, *Econometric Theory*, **14**, 139–149.

Table 1: Empirical sizes and powers of distributions tests with $T = 50$.

$\mathcal{G} = N(0,1)$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$	$\mathcal{G} = t(4)$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$
$\mathcal{F} = N(0,1)$	5.4	4.7	5.9	5.1	5.8	6.3	6.7	$\mathcal{F} = N(0,1)$	5.1	5.3	12.1	18.9	84.6	19.7	83.2
$t(4)$	6.2	6.3	10.2	15.3	20.2	16.2	23.8	$t(4)$	5.2	4.7	4.8	6.8	6.2	5.8	5.5
CA	47.9	44.7	60.9	79.7	97.0	95.1	98.5	CA	23.2	23.5	50.6	50.3	80.2	73.6	75.7
LG	48.8	54.4	75.3	88.9	93.4	91.0	98.7	LG	27.4	26.5	40.6	66.2	65.0	64.9	68.4
GE	4.2	4.7	5.8	10.9	9.4	16.3	15.1	GE	6.8	6.6	13.1	32.4	74.2	30.1	64.7
GL_s	8.7	6.8	14.7	33.4	33.3	52.0	53.5	GL_s	15.4	12.9	26.1	72.1	74.5	74.1	83.3
GL_a	76.7	74.0	42.0	78.0	73.1	73.6	60.0	GL_a	85.1	82.2	57.3	93.3	90.9	86.3	85.2
LN	26.6	25.0	43.0	40.7	41.9	23.4	16.7	LN	27.2	29.6	60.2	70.8	60.5	66.5	66.5
$\mathcal{G} = CA$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$	$\mathcal{G} = LG$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$
$\mathcal{F} = N(0,1)$	24.2	32.0	77.2	95.9	100	99.4	100	$\mathcal{F} = N(0,1)$	37.2	47.3	61.8	97.3	100	98.3	100
$t(4)$	10.2	10.3	40.5	65.6	100	84.2	99.7	$t(4)$	17.8	15.0	26.4	65.7	74.8	65.0	87.8
CA	5.5	6.1	5.6	6.3	6.8	6.9	6.5	CA	8.3	7.5	20.4	5.4	27.0	40.5	69.8
LG	7.8	6.3	38.8	6.5	100	43.5	84.5	LG	4.1	4.1	5.5	4.9	5.9	5.4	5.9
GE	32.5	43.1	71.8	99.0	100	99.4	100	GE	51.8	65.5	74.6	98.9	100	98.3	100
GL_s	60.5	76.0	86.0	99.9	99.9	99.9	99.9	GL_s	79.0	90.9	90.6	100	99.9	100	100
GL_a	98.6	99.7	96.7	100	100	100	99.5	GL_a	99.7	99.8	98.1	100	100	100	100
LN	98.2	79.3	98.4	99.9	89.4	100	100	LN	98.0	92.7	98.4	99.9	98.2	99.9	99.8
$\mathcal{G} = GE$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$	$\mathcal{G} = GL_s$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$
$\mathcal{F} = N(0,1)$	6.9	6.7	9.2	7.1	11.0	13.5	12.7	$\mathcal{F} = N(0,1)$	14.9	13.0	15.6	27.3	34.0	48.2	48.7
$t(4)$	10.3	10.7	13.8	27.3	27.0	22.6	29.8	$t(4)$	23.9	22.4	32.7	61.6	59.9	64.3	70.0
CA	50.0	54.1	67.8	89.8	98.4	95.0	97.9	CA	73.2	79.7	88.3	97.9	98.5	98.9	98.8
LG	62.2	69.1	84.2	95.5	95.4	95.9	98.2	LG	81.8	87.6	96.1	99.8	99.5	99.7	99.8
GE	5.4	4.3	5.6	7.0	5.5	8.3	6.3	GE	9.0	7.5	8.8	14.5	16.3	24.2	22.8
GL_s	6.0	5.7	8.7	17.5	15.2	26.0	25.9	GL_s	5.9	5.1	4.1	5.3	4.1	7.3	6.2
GL_a	63.5	64.8	31.4	68.4	75.7	49.3	37.2	GL_a	48.1	55.7	22.6	45.3	70.5	13.0	13.3
LN	29.3	28.9	41.8	31.2	40.6	14.7	17.7	LN	35.9	33.2	48.9	36.2	53.1	29.1	34.1
$\mathcal{G} = GL_a$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$	$\mathcal{G} = LN$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$
$\mathcal{F} = N(0,1)$	68.8	65.1	79.0	67.1	68.7	70.9	70.8	$\mathcal{F} = N(0,1)$	31.5	28.5	48.2	35.2	41.5	15.8	17.8
$t(4)$	76.7	71.1	86.0	88.3	85.9	79.1	82.8	$t(4)$	41.7	37.4	62.9	58.8	58.2	41.8	53.5
CA	92.8	93.5	98.5	99.4	99.4	98.7	98.6	CA	84.2	81.6	93.5	97.2	98.6	99.4	99.5
LG	97.2	97.8	99.7	100	99.9	99.9	99.9	LG	90.5	90.3	98.0	99.3	99.2	99.4	99.9
GE	60.4	60.5	66.6	52.8	52.5	41.4	41.5	GE	31.9	30.0	40.9	29.7	40.0	9.7	13.1
GL_s	48.3	56.0	41.6	39.6	56.0	10.0	10.8	GL_s	37.9	34.2	33.2	38.1	50.8	32.1	37.5
GL_a	5.8	5.4	5.3	6.9	5.8	6.9	6.6	GL_a	97.5	98.4	65.3	96.3	94.8	56.5	59.1
LN	92.3	92.6	83.9	91.9	89.8	48.4	54.1	LN	4.9	4.8	5.5	6.5	6.4	6.8	7.0

Note: The entries are rejection frequency in percentage. The columns $\mathcal{B}_1$, $\mathcal{B}_1^*$, $\mathcal{B}_2$, and $\mathcal{B}_2^*$ represent the $\mathcal{B}$ test with $(\phi(\epsilon), \beta) = (\epsilon, 1)$, (ϵ, β^*), $(\epsilon^2, 1)$, and (ϵ^2, β^*), respectively.

Table 2: Empirical sizes and powers of distribution tests with $T = 200$.

$\mathcal{G} = N(0,1)$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$	$\mathcal{G} = t(4)$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$
$\mathcal{F} = N(0,1)$	4.4	3.7	6.2	6.1	6.1	5.6	4.5	$\mathcal{F} = N(0,1)$	6.9	8.9	37.1	53.3	100	66.8	100
$t(4)$	12.5	13.7	46.9	44.2	66.5	56.5	81.7	$t(4)$	4.9	4.6	5.5	5.7	5.5	5.6	5.5
CA	99.7	99.8	99.9	100	100	100	100	CA	89.3	89.6	99.9	98.1	99.9	99.9	99.9
LG	100	100	100	100	100	100	100	LG	88.7	92.4	93.7	99.6	99.5	99.5	99.6
GE	8.8	6.5	10.8	15.9	19.1	35.4	33.7	GE	23.3	24.1	38.6	85.7	99.8	81.4	98.9
GL_s	51.4	49.2	53.3	84.2	88.7	97.5	97.7	GL_s	81.4	87.7	83.0	99.9	99.9	99.9	99.9
GL_a	100	100	99.1	100	100	99.9	99.6	GL_a	100	100	99.9	100	100	100	100
LN	86.8	86.7	98.7	91.4	95.2	60.8	53.7	LN	100	98.4	100	99.8	99.5	100	100

$\mathcal{G} = CA$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$	$\mathcal{G} = LG$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$
$\mathcal{F} = N(0,1)$	100	100	100	100	100	100	100	$\mathcal{F} = N(0,1)$	100	100	100	100	100	100	100
$t(4)$	94.4	95.8	99.5	100	100	100	100	$t(4)$	90.1	95.0	90.7	100	99.9	100	100
CA	6.0	5.5	6.4	6.2	6.0	6.6	6.5	CA	28.9	18.2	91.9	6.2	91.4	93.9	100
LG	13.8	13.9	99.2	6.2	100	96.4	100	LG	6.5	6.3	6.8	5.1	5.0	4.7	3.7
GE	100	100	100	100	100	100	100	GE	100	100	100	99.9	99.9	99.9	99.9
GL_s	100	100	100	100	100	100	100	GL_s	100	100	100	100	100	100	100
GL_a	100	100	100	100	100	100	100	GL_a	100	100	100	100	100	100	100
LN	100	100	100	100	100	100	100	LN	100	100	100	100	100	100	100

$\mathcal{G} = GE$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$	$\mathcal{G} = GL_s$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$
$\mathcal{F} = N(0,1)$	9.9	8.9	10.5	14.4	23.4	34.9	33.3	$\mathcal{F} = N(0,1)$	56.7	54.4	57.2	87.1	90.6	97.8	98.2
$t(4)$	30.1	33.7	53.0	79.0	77.0	73.8	84.5	$t(4)$	82.2	89.8	91.3	99.9	99.8	99.8	100
CA	99.8	99.9	100	100	100	100	100	CA	100	100	100	100	100	100	100
LG	100	100	100	100	100	100	100	LG	100	100	100	100	100	100	100
GE	4.6	4.8	6.3	3.6	4.2	5.7	5.6	GE	23.2	19.5	23.5	47.2	47.9	69.1	67.8
GL_s	17.8	12.5	20.2	48.2	48.7	68.4	68.1	GL_s	6.4	5.8	6.3	7.1	6.6	7.8	7.5
GL_a	100	100	95.2	100	100	95.8	89.9	GL_a	99.3	100	82.9	96.3	99.9	25.1	24.9
LN	83.4	86.0	99.0	84.8	92.5	39.4	52.4	LN	91.1	91.9	100	92.6	97.8	81.1	87.7

$\mathcal{G} = GL_a$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$	$\mathcal{G} = LN$	KS	CV	PC	$\mathcal{B}_1$	$\mathcal{B}_1^*$	$\mathcal{B}_2$	$\mathcal{B}_2^*$
$\mathcal{F} = N(0,1)$	100	99.9	99.9	100	100	99.9	99.9	$\mathcal{F} = N(0,1)$	93.1	91.6	100	94.3	96.4	49.5	52.4
$t(4)$	100	100	100	100	100	100	100	$t(4)$	99.5	98.9	100	100	99.7	99.1	99.8
CA	100	100	100	100	100	100	100	CA	100	100	100	100	100	100	100
LG	100	100	100	100	100	100	100	LG	100	100	100	100	100	100	100
GE	99.5	99.3	99.5	99.3	99.2	95.7	96.0	GE	87.2	87.0	99.1	85.7	94.9	30.9	47.0
GL_s	98.5	99.0	95.4	95.3	98.4	23.9	22.7	GL_s	93.2	92.7	98.1	92.5	98.7	83.1	88.8
GL_a	5.2	5.1	6.1	5.0	6.0	5.9	5.0	GL_a	100	100	100	100	100	97.5	98.4
LN	100	100	100	100	100	98.3	99.5	LN	5.8	5.7	5.6	5.0	5.0	6.6	5.8

Note: The entries are rejection frequency in percentage. The columns $\mathcal{B}_1$, $\mathcal{B}_1^*$, $\mathcal{B}_2$, and $\mathcal{B}_2^*$ represent the $\mathcal{B}$ test with $(\phi(\epsilon), \beta) = (\epsilon, 1)$, (ϵ, β^*), $(\epsilon^2, 1)$, and (ϵ^2, β^*), respectively.

Table 3: Summary statistics of the stock index returns.

Index	T	mean$\times 10^2$	std.$\times$10	skew.	kurt.	$Q_r(30)$	p^*	q^*	$Q_Y(30)$	$Q_{Y^2}(30)$
AO	2531	2.55	8.22	-0.26	7.43	50.40_a	9	5	14.31	509.59^b
DJ	2527	5.58	8.92	-0.41	8.20	62.94^b	6	5	34.22	481.86^b
FT	2527	4.14	9.13	0.07	5.13	72.24^b	5	3	32.85	1345.0^b
HS	2478	7.21	17.35	0.05	13.56	83.90^b	9	2	25.85	1070.2^b
NK	2465	-2.90	15.22	0.33	6.81	61.89_a	2	4	23.05	505.02^b
RS	2527	4.31	7.91	-0.92	8.67	239.56^b	9	5	23.32	1293.0^b
TS	2524	2.94	7.23	-0.82	10.81	172.23^b	14	5	19.03	1023.5^b

Note: The columns "std.", "skew.", and "kurt." denote the standard deviation, sample skewness coefficient, and kurtosis of the return r_t; p^* and q^* are the order of the selected ARMA(p,q) models. The columns $Q_r(30)$, $Q_Y(30)$, and $Q_{Y^2}(30)$ denote the Box-Pierce's $Q(30)$ statistic based on $\{r_t\}$, $\{Y_t\}$, and $\{Y_t^2\}$, respectively. The symbols ^{a} and ^{b} denote the corresponding statistic is significant at 5% and 1% level, respectively.

Table 4: Diagnostic checks of the GARCH(1,1) model with conditional normality.

Index	$Q_{\hat{\varepsilon}^2}(30)$	LM	symmetry	kurt.	$B_T(2)$	$B_T(5)$
AO	24.89	0.38	0.17	4.15	0.41	-0.31
DJ	24.99	1.20	1.00	5.81	-0.42	-0.35
FT	28.28	0.99	0.09	3.88	-3.23^b	-0.92
HS	8.00	0.27	0.37	7.99	0.50	0.81
NK	18.87	0.15	0.39	4.90	0.08	-0.30
RS	13.66	1.02	4.01^b	6.01	2.12^b	3.14^b
TS	9.54	1.54	1.49	7.13	0.41	0.09

Note: $Q_{\hat{\varepsilon}^2}(30)$ is the Box-Pierces' $Q(30)$ statistic calculated from $\{\hat{\varepsilon}_t^2\}$; "LM" denotes the ARCH-LM test statistic of Engel (1982) with five lags; "symmetry" represents the conditional symmetry test statistic of Chen (2001):

$$\mathcal{S}_T = \sqrt{T}\tilde{\mu}_{fs}/\tilde{\varpi}_{fs}, \qquad \tilde{\mu}_{fs} = \frac{1}{T}\sum_{t=1}^{T}\frac{\beta\hat{\varepsilon}_t}{1+(\beta\hat{\varepsilon}_t)^2},$$

where $\tilde{\varpi}_{fs}^2$ is a consistent estimator of the asymptotic variance of $\sqrt{T}\tilde{\mu}_{\exp}$, $\mathcal{S}_T \overset{A}{\sim} N(0,1)$. In this table, $\beta = 1$ and $\tilde{\varpi}_{fs}^2$ is obtained by bootstrap with 100 re-samples. We also tried other β for this test, and obtained the same conclusion. The column "kurt." represents the sample kurtosis of $\hat{\varepsilon}_t$; "$B_T(2)$" and "$B_T(5)$" are the BDS test statistics with the embedding dimensions are 2 and 5, in which the distance parameter is the standard deviation of $\hat{\varepsilon}_t$; see Brock et al. (1996) for the details. The symbol ^{b} denotes the corresponding statistic is significant at 1% level.

Table 5: The $\mathcal{C}$ test statistics with $z = \exp$ for different approximations of $\mathcal{F}$.

		$(\phi(\epsilon), \beta)$		$(\phi(\epsilon), \beta)$		value of β_c^*	
null hypothesis	index	$(\epsilon, 1)$	(ϵ, β_c^*)	$(\epsilon^2, 1)$	(ϵ^2, β_c^*)	$\phi(\epsilon) = \epsilon$	$\phi(\epsilon) = \epsilon^2$
	AO	6.702^a	8.615^a	9.094^a	8.825^a	2.932	0.969
	DJ	36.707^c	31.636^c	45.010^c	37.594^c	1.428	0.643
$\mathcal{F} = N(0,1)$	HS	58.705^c	56.516^c	87.121^c	84.165^c	1.986	0.796
	NK	24.301^c	22.975^c	33.652^c	30.623^c	1.791	0.646
	TS	26.109^c	34.399^c	37.580^c	40.433^c	1.746	0.798
	AO	0.024	0.840	0.086	0.894	3.430	3.002
	DJ	1.005	1.153	0.015	0.001	1.725	1.590
$\mathcal{F} =$ standardized	HS	0.289	2.193	2.266	3.419	3.334	2.288
t distribution	NK	0.100	0.478	0.015	0.282	2.002	2.252
	TS	2.131	3.531	0.914	1.006	1.996	1.326
	AO	0.943	1.520	3.461	3.923	2.596	1.136
	DJ	1.265	1.062	0.112	0.105	1.429	1.428
$\mathcal{F} =$ logistic	HS	1.966	7.094^a	8.820^a	9.323^b	3.685	1.701
distribution	NK	0.165	0.500	0.021	0.081	2.067	2.155
	TS	1.835	2.618	0.773	0.630	1.902	1.620
	AO	0.895	1.052	2.937	3.075	1.450	1.184
	DJ	0.042	0.027	0.369	0.229	2.182	1.293
$\mathcal{F} =$ generalized	HS	0.931	0.947	0.241	1.119	0.784	0.597
lambda distribution	NK	0.081	0.002	1.218	0.930	1.713	1.590
	TS	0.523	0.602	0.884	8.632^a	0.967	0.232
	AO	1.130	2.711	5.840	7.258^a	3.292	1.117
	DJ	3.098	2.003	5.061	6.234^a	1.600	1.459
$\mathcal{F} =$ generalized	HS	9.083^a	8.098^a	10.070^b	10.690^b	1.969	0.641
error distribution	NK	1.090	0.672	3.440	3.410	1.628	0.983
	TS	4.025	3.755	2.743	3.180	1.342	1.049

Note: The symbols ^{a}, ^{b}, and ^{c} denote that the corresponding statistic is significant at the 5%, 1%, and 0.5% level.

Table 6: The estimates of the postulated distributions.

				generalized lambda distribution			
Index	GE	t	LG	$\hat{\lambda}_{1T}$	$\hat{\lambda}_{2T}$	$\hat{\lambda}_{3T}$	$\hat{\lambda}_{4T}$
AO	1.514	11.20	0.558	-0.665	0.180	0.108	0.122
DJ	1.483	6.429	0.543	0.350	-0.412	-0.222	-0.211
HS	1.469	5.360	0.534	-0.577	-0.225	-0.979	-0.113
NK	1.489	6.918	0.546	0.832	0.534	0.330	0.269
TS	1.484	6.763	0.543	0.372	-0.100	-0.530	-0.507

Note: GE, t, and LG denote the generalized error distribution, the standardized t distribution, and the logistic distribution, respectively; the probability density function of GE and the distribution of LG are shown in Section 4. The entries are the estimates of the parameters of these distributions.

Table 7: The GARCH(1,1) model with different conditional distributions.

	GARCH-n			GARCH-g			GARCH-t		
Index	$\hat{w}_T$	$\hat{\rho}_{1T}$	$\hat{\rho}_{2T}$	$\hat{w}_T$	$\hat{\rho}_{1T}$	$\hat{\rho}_{2T}$	$\hat{w}_T$	$\hat{\rho}_{1T}$	$\hat{\rho}_{2T}$
AO	0.076	0.090	0.793	0.050	0.072	0.853	0.031	0.056	0.894
	(0.054)	(0.039)	(0.116)	(0.031)	(0.028)	(0.070)	(0.019)	(0.022)	(0.047)
DJ	0.007	0.047	0.944	0.006	0.042	0.949	0.005	0.040	0.953
	(0.005)	(0.021)	(0.026)	(0.004)	(0.015)	(0.019)	(0.003)	(0.012)	(0.015)
HS	0.089	0.116	0.853	0.064	0.101	0.868	0.048	0.096	0.891
	(0.039)	(0.021)	(0.027)	(0.024)	(0.020)	(0.026)	(0.018)	(0.023)	(0.026)
NK	0.074	0.104	0.865	0.057	0.096	0.878	0.045	0.095	0.889
	(0.029)	(0.020)	(0.027)	(0.020)	(0.016)	(0.021)	(0.014)	(0.014)	(0.016)
TS	0.011	0.089	0.891	0.008	0.072	0.907	0.007	0.059	0.927
	(0.005)	(0.032)	(0.037)	(0.004)	(0.024)	(0.031)	(0.004)	(0.021)	(0.028)

Note: GARCH-n, GARCH-g, and GARCH-t represent the GARCH(1,1) model with the conditional distribution assumptions: $N(0, 1)$, the generalized error distribution, and the standardized t distribution, respectively. The notations $\hat{w}_T$, $\hat{\rho}_{1T}$ and $\hat{\rho}_{2T}$ are the corresponding QMLEs of w, ρ_1 and ρ_2.

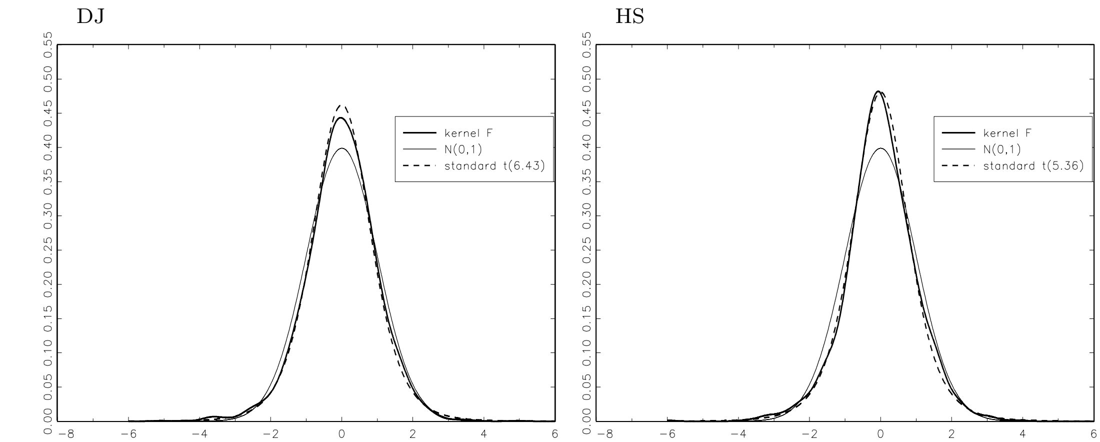

Figure 1: The Gaussian kernel density estimates and the probability density functions of $N(0,1)$ and the fitted standardized t distributions for the returns DJ and HS.

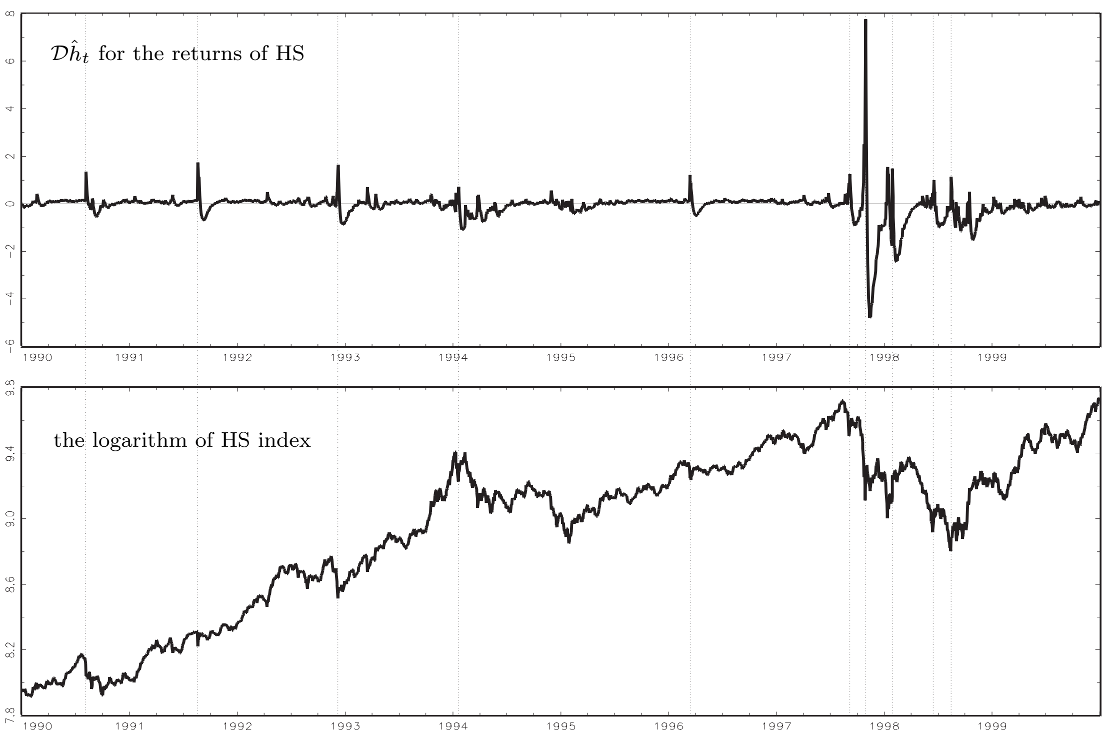

Figure 2: The path of $\mathcal{D}\hat{h}_t$ for the returns of HS and the logarithm of HS.