

A Unified Approach to Validating Univariate and Multivariate Conditional Distribution Models in Time Series

Bin Chen

Department of Economics

University of Rochester

and

Yongmiao Hong

Department of Economics and Statistical Science

Cornell University,

Wang Yanan Institute for Studies in Economics (WISE)

Xiamen University

November, 2009

We would like to thank Anil Bera, Bruce Hansen, Oliver Linton, Peter Robinson and seminar participants at LSE, the Far Eastern Econometric Society Meeting in Taipei, and the International Symposium on Recent Developments of Time Series Econometrics in Xiamen for their comments and discussions. Any remaining errors are solely ours. Hong acknowledges support from the National Science Foundation of China and the Cheung Kong Visiting Scholarship by the Chinese Ministry of Education and Xiamen University, China. Correspondences: Bin Chen, Department of Economics, University of Rochester, Rochester, NY 14627; Email: bchen8@mail.rochester.edu; Yongmiao Hong, Department of Economics & Department of Statistical Science, Cornell University, Ithaca, NY 14850, USA, and Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Xiamen 361005, China; Email: yh20@cornell.edu.

A Unified Approach to Validating Univariate and Multivariate Conditional Distribution Models in Time Series

Abstract:

Modeling conditional distributions in time series has attracted increasing attention in economics and finance. We develop a new class of generalized Cramer-von Mises (GCM) specification tests for time series conditional distribution models using a novel approach, which embeds the empirical distribution function in a spectral framework. Our tests check a large number of lags and are therefore expected to be powerful against neglected dynamics at higher order lags, which is particularly useful for non-Markovian processes. Despite using a large number of lags, our tests do not suffer much from loss of a large number of degrees of freedom, because our approach naturally downweights higher order lags, which is consistent with the stylized fact that economic or financial markets are more affected by recent past events than by remote past events. Unlike the existing methods in the literature, the proposed GCM tests cover both univariate and multivariate conditional distribution models in a unified framework. They exploit the information in the joint conditional distribution of underlying economic processes. Moreover, a class of easy-to-interpret diagnostic procedures are supplemented to gauge possible sources of model misspecifications. Distinct from conventional CM and Kolmogorov-Smirnov (KS) tests, which are also based on the empirical distribution function, our GCM test statistics follow a convenient asymptotic $N(0, 1)$ distribution and enjoy the appealing "nuisance parameter free" property that parameter estimation uncertainty has no impact on the asymptotic distribution of the test statistics. Simulation studies show that the tests provide reliable inference for sample sizes often encountered in economics and finance.

Key words: Diagnostic procedure, Empirical distribution function, Frequency domain, Generalized Cramer-von Mises test, Kernel method, Non-Markovian process, Time series conditional distribution model

JEL Classifications: C4, G0.

1. INTRODUCTION

The modeling of conditional distributions in time series has been advancing rapidly, with wide applications in economics and finance (e.g., Duffie and Pan 1997, Granger 1999, Corradi and Swanson 2006). Enormous empirical evidences document that economic and financial variables are typically nonlinear and nonnormally distributed, and have asymmetric comovements.¹ Consequently, one has to go beyond the conditional mean and conditional variance to obtain a complete picture for the dynamics of time series of interest. The conditional distribution characterizes the full dynamics of economic variables. As pointed out by Granger (2003), the knowledge of the conditional distribution is essential in performing various economic policy evaluations, financial forecasts, derivative pricing and risk management.²

In time series analysis, the most popular models are autoregressive moving average (ARMA) models for conditional mean and the generalized autoregressive conditional heteroskedasticity (GARCH) models for conditional variance. However, as Hansen (1994) points out "there is no reason to assume, in general, that the only features of the conditional distribution which depend upon the conditional information are the mean and variance." There has been an interest to go beyond the first two moments in modeling the dynamics of economic time series. Although still in an early stage, some time series models have been developed to study skewness, kurtosis and even the entire distribution. Hansen (1994) develops a general model for autoregressive conditional density (ARCD), which allows for time-varying first four conditional moments via a generalized skewed t -distribution. Harvey and Siddique (1999) propose a generalized autoregressive conditional skewness model (GARCS) in a conditional non-central t -distribution framework by explicitly modeling the conditional second and third moments jointly. Brooks, Burke and Persaud (2005) develop a generalized autoregressive conditional heteroscedasticity and kurtosis (GARCHK) model via a central t distribution with time-varying degrees of freedom. Other examples of distribution models include Engle and Russell's (1998a, 1998b and 2005) autoregressive conditional duration (ACD) and autoregressive conditional multinomial (ACM) models, Bowsher's (2007) vector conditional intensity model, Hamilton's (1989,1990) Markov regime switching models and Geweke and Amisano's (2007) compound Markov mixture models.

In economics and econometrics, effort has been devoted to using higher moments and the entire distribution. Rothschild and Stiglitz's (1971,1972) seminar works have demonstrated that the risk or uncertainty should be characterized by the distribution function, rather than the first two moments. In particular, the ranking of the cumulative distribution function (CDF) by certain rules always coincides with that of the risk-avertter's preference,³ while the mean-variance analysis is only applicable to the restricted family of utility functions or distribution functions. Granger (1999), in a model evaluation context, suggests that the predictive conditional distribution should be provided, since forecasts based on conditional means are optimal only for a very limited class of loss functions.⁴

¹Empirical evidences against normality can be dated back to Mills (1927) and continue through today, see, e.g., Ang and Chen (2002), Bollerslev (1986), Longin and Solnik (2001).

²A prominent example is in the option pricing context, where the price is determined by not just the conditional mean and variance, but functions of conditional distribution. Another example is to calculate value-at-risk (VaR), where the key step is to accurately estimate the conditional distribution of asset returns and the preassumed normal distribution can significantly underestimate the downward risk.

³A closely related concept is second-order stochastic dominance, which ranks any pair of distributions with the same mean in terms of comparative risk.

⁴See also Christoffersen and Diebold (1997) for more discussion.

In asset pricing, the classical Capital Asset Pricing Model (CAPM) is based on the assumption that asset returns are normally distributed and hence mean and variance can fully characterize the whole distribution. However, the inadequacy of these traditional hypotheses has led researchers to explore the impact of higher moments of asset returns (see, e.g., Harvey and Siddique 2000, Krauss and Litzenberger 1976). The other strand involves making assumptions on the representative agent's utility function, such as a quadratic or logarithmic form, which guarantees the linear form of the stochastic discounted factor. However, outside the family of mean-variance preferences, optimal portfolio choices generally depend on the entire return distribution. For example, a loss-averse investor, who realizes a greater incremental utility penalty for a loss than for an equally large gain, may be more concerned about the left tail of the return distribution.⁵

In option pricing, Black and Scholes' (1973) model is a cornerstone but empirical studies document that there exists severe mispricing for the deep out-of-the-money and deep in-the-money options. This has been attributed to the unrealistic assumption of a normally distributed continuous rate of return. More flexible models have been proposed. For example, multi-moment approximate option pricing models, initiated by Jarrow and Rudd (1982) and developed by Corrado and Su (1996, 1997) and Rubinstein (1998), approximate the risk-neutral density by a series expansion, which incorporates the third and fourth moments of the underlying asset.

In risk management, value at risk (VaR) has become a standard quantitative measure of market risk. VaR is the loss in market value over a given time horizon, for a given confidence level. Usually the normality assumption is made but the empirical evidences show its inadequacy (e.g., McNeil and Frey 2000). As emphasized by Engle (2002b), when computing VaR, "GARCH methods proved successful but suffered if errors were assumed to be Gaussian." On the other hand, hedging is another key concern in risk management. Optimal hedging analysis is based on the expected utility maximization paradigm, which generally requires characterization of the returns' joint distributions, rather than the first two moments only (e.g., Lien and Tse 2000).

In time series econometrics, there has been an important literature on density forecast evaluation, see, e.g., Diebold, Gunther and Tay (1998), Granger and Pesaran (2000) and Corradi and Swanson (2006). They show that when a forecast density coincides with the true data generating process (DGP), the forecast density will be preferred by all forecast users regardless of their loss function. This highlights the importance of correctly modeling the distribution. On the other hand, even if the first two moments are of major interest, correct specification of the entire distribution is essential for efficient estimation. For example, as Engle and Gonzalez-Rivera (1991) point out, the quasi-maximum likelihood estimator for ARCH models is inefficient, with the degree of inefficiency increasing with the degree of departure from the assumed normality.

In addition to the univariate time series distribution modeling, the recent literature has documented a rapid growth of multivariate conditional distribution models, due to an increasing need to capture the joint dynamics of multivariate processes, such as in macroeconomic control, pricing, hedging and risk management.⁶ For example, CAPM studies the relationship between individual asset returns and the

⁵Loss aversion is a special case of prospect theory, where utility is defined over gains and losses relative to a reference point rather than over the level of wealth as in expected utility theory. See (e.g.) Barberis and Huang (2001) and Barberis, Huang and Santos (2001) for financial applications of prospect theory.

⁶Geweke and Amisano (2001) argue that "while univariate models are a first step, there is an urgent need to move on

market return, which has motivated the development of multivariate GARCH models (e.g., Bollerslev, Engle and Wooldridge 1988, Engle 2002). Among multivariate distribution models, copula-based models have become increasingly popular in characterizing the comovement between markets, risk factors and other relevant variables (e.g., Patton 2004, Hu 2006, Lee and Long 2009). Another example is the extension of Markov regime switching models to a multivariate framework (e.g., Diebold and Rudebusch 1996, Krolzig 1997). Markov regime switching models can capture the asymmetry, nonlinearity and persistence of extreme observations of time series.

Efficient parameter estimation, optimal distribution forecast, valid hypothesis testing and economic interpretation all require correct model specification. The work on testing distributional assumptions at least date back to the KS test. One undesired feature of this test is that it is not distribution free when parameters are estimated. Andrews (1997) extends the KS test to conditional distribution models for independent observations, where a bootstrap procedure is used to obtain critical values. Meanwhile, Zheng (2000) proposes a nonparametric test for conditional distribution functions based on the Kullback-Leibler information criterion and the kernel estimation of the underlying distributions. Fan, Li and Min (2006) extend Zheng's (2000) test to allow for discrete dependent variables and for mixed discrete and continuous conditional variables. However, a limitation of the above tests is that the data must be independently and identically distributed, therefore ruling out time series applications especially when the underlying time series is non-Markovian.

Observing the fact that when a dynamic distribution model is correctly specified, the probability integral transform of observed data via the model-implied conditional density is *i.i.d.* $U[0,1]$, Bai (2003) proposes a KS type test with Khmaladze's (1981) martingale transformation, whose asymptotic distribution is free of impact of parameter estimation. However, Bai's (2003) test only checks uniformity rather than the joint *i.i.d.* $U[0,1]$ hypothesis. It will have no asymptotic unit power if the transformed data is uniform but not *i.i.d.*. Moreover, in a multivariate context, the probability integral transform of data with respect to a model-implied multivariate conditional density is no longer *i.i.d.* $U[0,1]$, even if the model is correctly specified. Bai and Chen (2008) evaluate the marginal distribution of both independent and serially dependent multivariate data by using the probability integral transform for each individual component. This test is legitimate, but it may miss important information on the joint distribution of a multivariate model. In particular, when applied to each component of multivariate time series data, Bai and Chen's (2008) test may fail to detect misspecification in the joint dynamics. For example, the test may easily overlook misspecification in the conditional correlations between individual time series.

Corradi and Swanson (2006) propose bootstrap conditional distribution tests in the presence of dynamic misspecifications. However, they consider a finite dimensional information set and thus may not have good power against non-Markovian models. Their tests are designed for univariate time series. When extended to multivariate time series, their tests are not consistent against all alternatives to the null. Moreover, their critical values are data dependent and cannot be tabulated.

In a continuous-time diffusion framework, Ait-Sahalia, Fan and Peng (2009) and Li and Tkacz (2006) propose tests by comparing the model-implied distribution function with its nonparametric counterpart. Both tests maintain the Markov assumption for the DGP, and only check one lag dependence, therefore

to multivariate modeling of the time-varying distribution of asset returns".

are not suitable for non-Markovian models like GARCH or MA type models. Another undesired feature of these tests is that they have severe size distortion in finite samples and bootstrap must be used to approximate the distribution of the test statistics. Bhardwaj, Corradi and Swanson (2008) consider a simulation-based test, which is an extension of Andrews' (1997) conditional KS test, for multivariate diffusion models. The limit distribution of their test is not nuisance parameter free and asymptotic critical values must be obtained via a block bootstrap.

In this paper, we shall propose a new class of generalized Cramer-von Mises (GCM) tests of the adequacy of univariate and multivariate conditional distribution models, without requiring prior knowledge of possible alternatives (including both functional forms and lag structures). Compared with the existing tests for conditional distribution models in the literature, our approach has several main advantages.

First, our GCM tests are constructed using a new approach, which embeds the empirical distribution function in a spectral framework. Thus it can detect misspecification in both marginal distribution and dynamics of a time series. Thanks to the use of the empirical distribution function, our approach can detect a variety of linear and nonlinear functional form misspecifications. Our frequency domain approach can check a growing number of lags as the sample size increases without suffering from the curse of dimensionality. This is particularly useful for conditional distribution models in time series since the conditioning information set may depend on the entire history of the data. Indeed, most time series distribution models in the literature are non-Markov. Moreover, our approach employs a kernel function and it naturally discounts higher order lags. This is expected to enhance power because it is consistent with the stylized fact that economic and financial variables are usually more influenced by recent events than by remote past events. Unlike the traditional CM and KS tests, which also use the empirical distribution function but have nonstandard distributions contaminated by parameter estimation uncertainty, our tests have a convenient null asymptotic $N(0,1)$ distribution.

Second, by using the conditional distribution of a multivariate time series vector directly, our tests exploit the information in the joint conditional dynamics of the time series vector rather than only in the conditional distribution of individual components. Thus, they can detect misspecifications in the joint conditional distribution even if the conditional distribution of each individual series is correctly specified. Our tests are applicable to both continuous and discrete distributions. Moreover, because we impose regularity conditions directly on the conditional distribution function of a discrete sample, our tests are also applicable to multivariate continuous-time models with discretely observed samples.

Third, besides the GCM test, we propose a class of diagnostic tests. These tests can evaluate how well a time series conditional distribution model captures various specific aspects of the joint dynamics, and are easy to interpret. In particular, these tests can provide valuable information about neglected dynamics in conditional means, conditional variances, conditional correlations, conditional skewness and conditional kurtosis, respectively. Thus, they complement the popular conditional moment tests in the literature. All our GCM test and diagnostic tests are derived from a unified framework.

Fourth, we do not require a particular estimation method. Any \sqrt{T} -consistent parametric estimators can be used. Unlike tests based on the distributional function, such as the conventional CM and KS tests, parameter estimation uncertainty does not affect the asymptotic distribution of our test statistic. One can proceed as if the true model parameters were known and equal to parameter estimates. This makes our tests easy to implement. The only inputs needed to calculate the test statistics are the

original data and the model-implied CDF.

In Section 2, we introduce the framework, state the hypotheses, and characterize the correct specification of a conditional distribution model that can be either univariate or multivariate. In Section 3, we propose an empirical distribution function-based test embedded a frequency domain approach. In Section 4, we derive its asymptotic null distribution, and discuss its asymptotic power property. In Section 5, we develop a class of diagnostic tests that focus on various specific aspects of a time series conditional distribution model. In Section 6, we assess the reliability of the asymptotic theory in finite samples via simulation. Section 7 concludes. All mathematical proofs are collected in the appendix. A GAUSS code to implement our tests is available from the authors upon request. Throughout, we will use C to denote a generic bounded constant, $\|\cdot\|$ for the Euclidean norm.

2. HYPOTHESES OF INTEREST

Suppose $\{\mathbf{X}_t\}$ is a $d \times 1$ strictly stationary time series process with unknown conditional CDF $P_0(\mathbf{x}|\mathcal{I}_{t-1})$, where the dimension $d \geq 1$, and \mathcal{I}_{t-1} is the information set available at time $t-1$. We allow but do not require \mathbf{X}_t to be Markov. As a leading example, we consider a time series model

$$\mathbf{X}_t = \boldsymbol{\mu}(\mathcal{I}_{t-1}, \boldsymbol{\theta}) + \mathbf{h}^{1/2}(\mathcal{I}_{t-1}, \boldsymbol{\theta}) \boldsymbol{\varepsilon}_t, \quad (2.1)$$

where $\boldsymbol{\mu}(\mathcal{I}_{t-1}, \boldsymbol{\theta})$ is a parametric model for $E(\mathbf{X}_t|\mathcal{I}_{t-1})$, $\mathbf{h}(\mathcal{I}_{t-1}, \boldsymbol{\theta})$ is a parametric model for $var(\mathbf{X}_t|\mathcal{I}_{t-1})$, $\boldsymbol{\varepsilon}_t$ has the conditional CDF $P_\varepsilon(\boldsymbol{\varepsilon}|\mathcal{I}_{t-1}, \boldsymbol{\theta})$, and $\boldsymbol{\theta} \in \Theta$ is a finite-dimension parameter. In time series modeling, \mathcal{I}_{t-1} is possibly infinite-dimensional, as in the case of non-Markovian processes. Given $P_\varepsilon(\boldsymbol{\varepsilon}|\mathcal{I}_{t-1}, \boldsymbol{\theta})$, it is straightforward to calculate the conditional CDF of \mathbf{X}_t

$$P_{\mathbf{x}}(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta}) = P_\varepsilon \left[\frac{\mathbf{x} - \boldsymbol{\mu}(\mathcal{I}_{t-1}, \boldsymbol{\theta})}{\mathbf{h}^{1/2}(\mathcal{I}_{t-1}, \boldsymbol{\theta})} \middle| \mathcal{I}_{t-1}, \boldsymbol{\theta} \right].$$

The setup (2.1) is a general specification that nests most popular time series conditional distribution models in the literature. For example, suppose we assume that $\boldsymbol{\varepsilon}_t$ has a continuous distribution with the conditional PDF

$$p_\varepsilon(\boldsymbol{\varepsilon}|\mathcal{I}_{t-1}, \boldsymbol{\theta}) = p_\varepsilon[\boldsymbol{\varepsilon}|\boldsymbol{\alpha}(\mathcal{I}_{t-1}, \boldsymbol{\theta})],$$

where $\boldsymbol{\alpha}(\mathcal{I}_{t-1}, \boldsymbol{\theta}) = [\boldsymbol{\mu}(\mathcal{I}_{t-1}, \boldsymbol{\theta}), \mathbf{h}(\mathcal{I}_{t-1}, \boldsymbol{\theta}), \boldsymbol{\lambda}(\mathcal{I}_{t-1}, \boldsymbol{\theta}), \boldsymbol{\nu}(\mathcal{I}_{t-1}, \boldsymbol{\theta})]'$ is a low dimensional time-varying function that can effectively summarize the available information \mathcal{I}_{t-1} , and $\boldsymbol{\lambda}(\cdot)$ and $\boldsymbol{\nu}(\cdot)$ are so called time-varying shape parameters, which control serial dependence in higher order conditional moments. Then we obtain Hansen's (1994) univariate ARCD model. Specifically, Hansen (1994) considers a skewed student's t distribution with

$$p_\varepsilon(\varepsilon|\nu, \lambda) = \begin{cases} \frac{bc}{\left[1 + \frac{1}{\nu-2} \left(\frac{b\varepsilon+a}{1-\lambda}\right)^2\right]^{(\nu+1)/2}} & \text{if } \varepsilon < -\frac{a}{b}, \\ \frac{bc}{\left[1 + \frac{1}{\nu-2} \left(\frac{b\varepsilon+a}{1+\lambda}\right)^2\right]^{(\nu+1)/2}} & \text{if } \varepsilon \geq -\frac{a}{b}, \end{cases} \quad (2.2)$$

where $0 < \nu < \infty$, $-1 < \lambda < 1$, $a = 4\lambda c \frac{\nu-2}{\nu-1}$, $b^2 = 1 + 3\lambda^2 - a^2$, $c = \frac{\Gamma[(\nu+1)/2]}{[\pi(\nu-2)]^{1/2}\Gamma(\nu/2)}$.

Another example is Harvey and Siddique's (1999) GARCHS model. For a univariate GARCHS(1,1,1) model, the conditional variance $\mathbf{h}_t \equiv \mathbf{h}(\mathcal{I}_{t-1}, \boldsymbol{\theta})$ and conditional skewness $\mathbf{S}_t \equiv \mathbf{S}(\mathcal{I}_{t-1}, \boldsymbol{\theta})$ are specified

as

$$\begin{aligned}\mathbf{h}_t &= \beta_0 + \beta_1 \mathbf{h}_{t-1} + \beta_2 u_{t-1}^2 \\ \mathbf{S}_t &= \gamma_0 + \gamma_1 \mathbf{S}_{t-1} + \gamma_2 u_{t-1}^3,\end{aligned}$$

where $u_t \equiv \mathbf{h}_t^{1/2} \boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_t$ has a conditional noncentral t distribution with the degrees of freedom ν_t and the noncentrality parameter δ_t .

A third example is the copula-based multivariate GARCH model considered by Lee and Long (2009). They assume that $\boldsymbol{\mu}(\mathcal{I}_{t-1}, \boldsymbol{\theta}) = 0$, $\mathbf{h}(\mathcal{I}_{t-1}, \boldsymbol{\theta})$ adopts the forms from Engle and Kroner's (1995) BEKK model, Engle's (2002a) dynamic conditional correlation (DCC) model and Tse and Tsui's (2002) varying correlation model, and

$$\begin{aligned}\boldsymbol{\varepsilon}_t &= \Sigma_t^{-1/2} \eta_t, \\ \eta_t | \mathcal{I}_{t-1} &\sim C(F_t(\cdot), G_t(\cdot); \alpha_t),\end{aligned}\tag{2.3}$$

where $C(\cdot, \cdot; \cdot)$ is the conditional copula function, such as the Gumbel copula with $C(u, v; \alpha) = \exp\{-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{1/\alpha}\}$, $F_t(\cdot)$, $G_t(\cdot)$ are marginal CDFs.

In our setup, \mathbf{X}_t need not have a continuous distribution. An example of a conditional discrete distribution is Russell and Engle's (2005) ACM-ACD model. They assume $\mathbf{X}_t = (y_t, \tau_t)'$, where y_t is the discrete price change and τ_t is the duration between transactions. The joint conditional distribution of y_t and τ_t can be decomposed into the product of the conditional distribution of the price change and the conditional distribution of the arrival times, namely,

$$P_{\mathbf{x}}(\mathbf{x} | \mathcal{I}_{t-1}, \boldsymbol{\theta}) = P_y(y | \mathcal{I}_{y,t-1}, \mathcal{I}_{\tau,t}, \boldsymbol{\theta}) P_{\tau}(\tau | \mathcal{I}_{t-1}, \boldsymbol{\theta}),$$

where $\mathcal{I}_{y,t-1} = (y_{t-1}, y_{t-2}, \dots, y_1)$ and $\mathcal{I}_{\tau,t-1} = (\tau_{t-1}, \tau_{t-2}, \dots, \tau_1)$. The duration τ_t is assumed to follow an ACD model and its conditional density is given as

$$p_{\tau}(\tau | \mathcal{I}_{t-1}, \boldsymbol{\theta}) = \frac{1}{\psi_t} \exp\left(-\frac{\tau_t}{\psi_t}\right),$$

where $\psi_t = E(\tau_t | \mathcal{I}_{\tau,t-1})$. The price change y_t has a multinomial distribution, namely,

$$p_y(y | \mathcal{I}_{y,t-1}, \mathcal{I}_{\tau,t}, \boldsymbol{\theta}) = \sum_{j=1}^s \pi_{tj}^{\tilde{y}_{tj}},$$

where s is the number of states, \tilde{y}_t takes the j th column of the $s \times s$ identity matrix if the j th state occurs in y_t and π_j denotes the $s \times 1$ vector of conditional probabilities associated with the states.

We say that the model (2.1) is correctly specified if there exists some parameter value $\boldsymbol{\theta}_0 \in \Theta$ such that

$$\mathbb{H}_0 : P(\mathbf{x} | \mathcal{I}_{t-1}, \boldsymbol{\theta}_0) = P_0(\mathbf{x} | \mathcal{I}_{t-1}) \quad \text{almost surely (a.s.) and for all } \mathbf{x} \text{ and } t.\tag{2.4}$$

Alternatively, if for all $\boldsymbol{\theta} \in \Theta$, we have

$$\mathbb{H}_A : P(\mathbf{x} | \mathcal{I}_{t-1}, \boldsymbol{\theta}) \neq P_0(\mathbf{x} | \mathcal{I}_{t-1}) \text{ with positive probability measure,}\tag{2.5}$$

then model (2.1) is misspecified.

The empirical distribution function has been used to test correct specification of a conditional distribution model. Observing that when $d = 1$, the probability integral transform $U_t(\boldsymbol{\theta}_0) \equiv P_t(\mathbf{X}_t|\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)$ is an *i.i.d.* uniform[0,1] random variable, Bai (2003) compares the empirical distribution function of $U_t(\hat{\boldsymbol{\theta}})$ with a uniform CDF. Bai (2003) uses Khmaladze’s (1981) martingale transformation to remove the impact of parameter estimation uncertainty and his test statistic converges to a standard Brownian motion. An undesired feature of this test is that it only checks the marginal distribution of U_t and has no power against the alternatives where the independence property is violated but the marginal uniformity holds. Moreover, the probability integral transform is not applicable to the multivariate joint conditional density directly, because when $d > 1$, $U_t(\boldsymbol{\theta}_0)$ is no longer *i.i.d.* $U[0, 1]$. Bai and Chen (2008) extend it to the multivariate setup by considering the particular sequence $U_{t1}(\boldsymbol{\theta}_0) \equiv P_t(X_{t1}|\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)$, $U_{t2}(\boldsymbol{\theta}_0) \equiv P_t(X_{t2}|X_{t1}, \mathcal{I}_{t-1}, \boldsymbol{\theta}_0), \dots, U_{td}(\boldsymbol{\theta}_0) \equiv P_t(X_{td}|X_{t1}, \dots, X_{td-1}, \mathcal{I}_{t-1}, \boldsymbol{\theta}_0)$. This is legitimate, but it does not make full use of the information contained in the joint distribution of \mathbf{X}_t . In particular, it may miss important model misspecification in the joint dynamics of \mathbf{X}_t . For example, consider the DGP $\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$, where $\{\boldsymbol{\varepsilon}_t\}$ is *i.i.d.* $N(0, \Sigma)$ and Σ is a $d \times d$ ($d > 1$) constant upper-triangular matrix. Suppose one fits the data by a VAR(1) model with $\tilde{\boldsymbol{\varepsilon}}_t \sim i.i.d.N(0, \tilde{\Sigma})$, where $\tilde{\Sigma}$ is a diagonal matrix. Then this model is misspecified yet their test has no power.

To develop a test for \mathbb{H}_0 , we define a generalized model residual

$$Z_t(\mathbf{x}, \boldsymbol{\theta}) \equiv 1(\mathbf{X}_t \leq \mathbf{x}) - P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta}), \quad \mathbf{x} \in \mathbb{R}^d. \quad (2.6)$$

Then \mathbb{H}_0 is equivalent to the following MDS characterization for $Z_t(\mathbf{x}, \boldsymbol{\theta})$:

$$E[Z_t(\mathbf{x}, \boldsymbol{\theta}_0)|\mathcal{I}_{t-1}] = 0 \text{ for all } \mathbf{x} \in \mathbb{R}^d \text{ and some } \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}, \text{ a.s.} \quad (2.7)$$

It is not a trivial task to check (2.7). First, the MDS property in (2.7) must hold for all $\mathbf{x} \in \mathbb{R}^d$, not just a finite number of grid points of \mathbf{x} . This is an example of the well-known nuisance parameter problem encountered in the literature (e.g., Davies 1977, 1987 and Hansen 1996). Second, the conditioning information set \mathcal{I}_{t-1} in (2.7) has an infinite dimension as $t \rightarrow \infty$, so there is a ‘‘curse of dimensionality’’ difficulty associated with testing the model specification. Finally, $\{Z_t(\mathbf{x}, \boldsymbol{\theta}_0)\}$ may display serial dependence in its higher order conditional moments. Any test for (2.7) should be robust to time-varying conditional heteroskedasticity and higher order moments of unknown form in $\{Z_t(\mathbf{x}, \boldsymbol{\theta}_0)\}$.

There has been a large literature on empirical distribution function-based tests; see, e.g., Hoeffding (1948), Andrews (1997), Linton and Gozalo (1997), and Hong (1998). However, most tests are designed for *i.i.d.* observations. The CDF approach is particularly appealing in checking conditional distribution models because the conditional PDF usually has a simple closed form and the conditional CDF can be obtained via analytic or numerical integration. Moreover, there is a natural link between the distribution function and moments, which can be exploited to construct a class of diagnostic procedures for different specific aspects of $P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta})$ in Section 5.

So far we have assumed that all components of \mathbf{X}_t are observable. However, there are time series models with unobservable components in the literature. For example, the state-space models have been widely used in macroeconomics and finance. The simplest state-space representation is given by the

following system of equations:

$$\begin{cases} \mathbf{Y}_t = \mathbf{A}'\mathbf{Y}_{t-1} + \mathbf{H}'\boldsymbol{\xi}_t + \mathbf{w}_t, \\ \boldsymbol{\xi}_t = \mathbf{F}'\boldsymbol{\xi}_{t-1} + \mathbf{v}_t, \end{cases} \quad (2.8)$$

where \mathbf{A} , \mathbf{F} and \mathbf{H} are matrices of parameters, \mathbf{w}_t and \mathbf{v}_t are vector white noise, $\boldsymbol{\xi}_t$ is the possibly unobserved state vector, and \mathbf{Y}_t is observable. The system in (2.8) is known as the observation equation and the state equation respectively (see, e.g., Hamilton 1994 and DeJong and Dave 2007). Another example is the class of stochastic volatility (SV) models for equity returns and interest rates, see (e.g.) Shephard (2005), Anderson and Lund (1997) and Gallant, Hsieh and Tauchen (1997). With a latent volatility state variable, SV models can capture salient properties of volatility such as randomness and persistence. A first order SV model (Taylor 1986) assumes:

$$\begin{cases} S_t = V_t \varepsilon_t, \\ \ln V_t^2 = \gamma_0 + \gamma_1 \ln V_{t-1}^2 + u_t, \end{cases} \quad (2.9)$$

where V_t is the latent volatility and S_t is the asset return, γ_0 and γ_1 are both scalar parameters, and ε_t and u_t are mutually independent innovations.

To test time series models with unobservable components, we need to modify the MDS characterization (2.7) to make it operational. For this purpose, we partition $\mathbf{X}_t = (\mathbf{X}'_{1,t}, \mathbf{X}'_{2,t})'$, where $\mathbf{X}_{1,t} \subset \mathbb{R}^{d_1}$ denotes the observable components, $\mathbf{X}_{2,t} \subset \mathbb{R}^{d_2}$ denotes the unobservable components, and $d_1 + d_2 = d$. Also, partition \mathbf{x} conformably as $\mathbf{x} = (\mathbf{x}'_1, \mathbf{x}'_2)'$. Let

$$P(\mathbf{x}_1 | \mathcal{I}_{1,t-1}, \boldsymbol{\theta}) \equiv E_{\boldsymbol{\theta}}[1(\mathbf{X}_{1,t} \leq \mathbf{x}_1) | \mathcal{I}_{1,t-1}] = E_{\boldsymbol{\theta}}\{P[(\mathbf{x}'_1, \boldsymbol{\theta}')' | \mathcal{I}_{t-1}, \boldsymbol{\theta}] | \mathcal{I}_{1,t-1}\},$$

where $\mathcal{I}_{1,t-1} = \{\mathbf{X}_{1,t-1}, \mathbf{X}_{1,t-2}, \dots, \mathbf{X}_{1,1}\}$ is the information set on the observables that is available at time $t - 1$ and the second equality follows by the law of iterated expectations. Then we define

$$Z_{1,t}(\mathbf{x}_1, \boldsymbol{\theta}) \equiv 1(\mathbf{X}_{1,t} \leq \mathbf{x}_1) - P(\mathbf{x}_1 | \mathcal{I}_{1,t-1}, \boldsymbol{\theta}).$$

Under \mathbb{H}_0 , we have

$$E[Z_{1,t}(\mathbf{u}_1, \boldsymbol{\theta}_0) | \mathcal{I}_{1,t-1}] = 0 \text{ a.s. for all } \mathbf{x}_1 \in \mathbb{R}^{d_1} \text{ and some } \boldsymbol{\theta}_0 \in \Theta. \quad (2.10)$$

This provides a basis for constructing operational tests for time series models with partially observable variables. Without loss of generality, we will focus on conditional distribution models with fully observable variables for the rest of the paper.

3. GENERALIZED DYNAMIC CRAMER-VON MISES TEST

We now propose a new class of GCM tests for the adequacy of a dynamic conditional distribution model by exploiting the characterization in (2.7). To check the MDS property of $Z_t(\mathbf{x}, \boldsymbol{\theta})$, we take a frequency domain approach in combination with the empirical distribution function. It can capture both linear and nonlinear dynamics while maintaining the nice features of spectral analysis, particularly its appealing property to accommodate all lags information. In the present context, it can check departures of correct model specification over many lags in a pairwise manner, avoiding the "curse of

dimensionality" difficulty. This is not attained by many existing tests in the literature which only check a fixed lag order. The empirical distribution function is rather natural in testing conditional distribution models. Most time series conditional distributional models have closed-form PDFs.

Define a generalized covariance function

$$\Gamma_j(\mathbf{x}, \mathbf{y}) = \text{cov} [Z_t(\mathbf{x}, \boldsymbol{\theta}), \mathbf{1}(\mathbf{X}_{t-|j|} \leq \mathbf{y})], \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \quad (3.1)$$

where j is a lag or lead number. We also define the Fourier transform

$$F(\omega, \mathbf{x}, \mathbf{y}) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma_j(\mathbf{x}, \mathbf{y}) \exp(-ij\omega), \quad \omega \in [-\pi, \pi], \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \quad (3.2)$$

where ω is the frequency. The function $F(\omega, \mathbf{x}, \mathbf{y})$ may be called the distribution function-based generalized spectral density of $\{\mathbf{X}_t\}$. It contains the same information on serial dependence of $\{\mathbf{X}_t\}$ as the generalized covariance function $\Gamma_j(\mathbf{x}, \mathbf{y})$. An advantage of frequency domain analysis is that it can capture cyclical patterns caused by both linear and nonlinear serial dependence. Examples include volatility spillover, the comovements of tail distribution clustering between economic variables, and asymmetric spillover of business cycles cross different sectors or countries. Another attractive feature of $F(\omega, \mathbf{x}, \mathbf{y})$ is that it does not require the existence of any moment condition on \mathbf{X}_t due to the use of the distribution function. This is appealing for time series data with heavy tail distributions.

Under \mathbb{H}_0 , we have $\Gamma_j(\mathbf{x}, \mathbf{y}) = 0$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and all $j \neq 0$. Consequently, the generalized spectral density $F(\omega, \mathbf{x}, \mathbf{y})$ becomes a "flat" spectrum (i.e., a constant function of frequency ω):

$$F(\omega, \mathbf{x}, \mathbf{y}) = F_0(\omega, \mathbf{x}, \mathbf{y}) \equiv \frac{1}{2\pi} \Gamma_0(\mathbf{x}, \mathbf{y}), \quad \omega \in [-\pi, \pi], \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \quad (3.3)$$

Thus, we can test \mathbb{H}_0 by checking whether a consistent estimator for $F(\omega, \mathbf{x}, \mathbf{y})$ is flat with respect to frequency ω . Any significant deviation from a flat generalized spectrum is evidence of model misspecification.

Suppose we have a random sample $\{\mathbf{X}_t\}_{t=1}^T$ of size T . Then we can estimate the generalized covariance $\Gamma_j(\mathbf{x}, \mathbf{y})$ by its sample analogue

$$\hat{\Gamma}_j(\mathbf{x}, \mathbf{y}) = \frac{1}{T - |j|} \sum_{t=|j|+1}^T Z_t(\mathbf{x}, \hat{\boldsymbol{\theta}}) \left[\mathbf{1}(\mathbf{X}_{t-|j|} \leq \mathbf{y}) - \hat{P}(\mathbf{y}) \right], \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \quad (3.4)$$

where $\hat{\boldsymbol{\theta}}$ is a \sqrt{T} -consistent estimator for $\boldsymbol{\theta}_0$ and $\hat{P}(\mathbf{y}) = T^{-1} \sum_{t=1}^T \mathbf{1}(\mathbf{X}_t \leq \mathbf{y})$ is the empirical distribution function.

Then a consistent estimator for $F_0(\omega, \mathbf{x}, \mathbf{y})$ is

$$\hat{F}_0(\omega, \mathbf{x}, \mathbf{y}) = \frac{1}{2\pi} \hat{\Gamma}_0(\mathbf{x}, \mathbf{y}), \quad \omega \in [-\pi, \pi], \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \quad (3.5)$$

Consistent estimation for $F(\omega, \mathbf{x}, \mathbf{y})$ is more challenging. We use a smoothed kernel estimator

$$\hat{F}(\omega, \mathbf{x}, \mathbf{y}) = \frac{1}{2\pi} \sum_{j=1-T}^{T-1} (1 - |j|/T)^{1/2} k(j/p) \hat{\Gamma}_j(\mathbf{x}, \mathbf{y}) e^{-ij\omega}, \quad \omega \in [-\pi, \pi], \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \quad (3.6)$$

where $p \equiv p(T) \rightarrow \infty$ is a bandwidth or an effective lag order, and $k : \mathbb{R} \rightarrow [-1, 1]$ is a kernel function, assigning weights to various lags. Examples of $k(\cdot)$ include the Bartlett kernel

$$k(z) = \begin{cases} 1 - |z|, & |z| \leq 1, \\ 0, & \text{otherwise,} \end{cases} \quad (3.7)$$

the Parzen kernel

$$k(z) = \begin{cases} 1 - 6z^2 + 6|z|^3, & |z| \leq 0.5, \\ 2(1 - |z|)^3, & 0.5 < |z| \leq 1, \\ 0, & \text{otherwise,} \end{cases} \quad (3.8)$$

and the Quadratic-Spectral kernel

$$k(z) = \frac{3}{(\pi z)^2} \left[\frac{\sin(\pi z)}{\pi z} - \cos(\pi z) \right], \quad z \in \mathbb{R}. \quad (3.9)$$

In (3.6), the factor $(1 - |j|/T)^{1/2}$ is a finite-sample correction. It could be replaced by unity. Under suitable regularity conditions, $\hat{F}(\omega, \mathbf{x}, \mathbf{y})$ and $\hat{F}_0(\omega, \mathbf{x}, \mathbf{y})$ are consistent for $F(\omega, \mathbf{x}, \mathbf{y})$ and $F_0(\omega, \mathbf{x}, \mathbf{y})$ respectively. These estimators converge to the same limit under \mathbb{H}_0 but they generally converge to different limits under \mathbb{H}_A , giving the power of the test.

We can construct a test via the L_2 -norm

$$\begin{aligned} \hat{L}^2 &= \frac{\pi T}{2} \int_{-\pi}^{\pi} \iint \left| \hat{F}(\omega, \mathbf{x}, \mathbf{y}) - \hat{F}_0(\omega, \mathbf{x}, \mathbf{y}) \right|^2 d\omega dW(\mathbf{x}) dW(\mathbf{y}) \\ &= \sum_{j=1}^{T-1} k^2(j/p) (T - j) \iint \hat{\Gamma}_j^2(\mathbf{x}, \mathbf{y}) dW(\mathbf{x}) dW(\mathbf{y}), \end{aligned} \quad (3.10)$$

where the equality follows by Parseval's identity, $W : \mathbb{R}^d \rightarrow \mathbb{R}^+$ is a nondecreasing right-continuous weighting function that weighs the sets symmetric about the origin equally, and the unspecified integrals are all taken over the support of $W(\cdot)$. An example of $W(\cdot)$ is the CDF of $N(\mathbf{0}, \mathbf{I}_d)$, where \mathbf{I}_d is a $d \times d$ identity matrix. The function $W(\cdot)$ can also be a step function, analogous to the CDF of a discrete random vector.

Our GCM test statistic for \mathbb{H}_0 against \mathbb{H}_A is a standardized version of (3.10):

$$\hat{Q}_1 = \left[\sum_{j=1}^{T-1} k^2(j/p) (T - j) \int \int \hat{\Gamma}_j^2(\mathbf{x}, \mathbf{y}) dW(\mathbf{x}) dW(\mathbf{y}) - \hat{C}_1 \right] / \sqrt{\hat{D}_1}, \quad (3.11)$$

where the centering and scaling factors

$$\begin{aligned}\hat{C}_1 &= \sum_{j=1}^{T-1} k^2(j/p)(T-j)^{-1} \sum_{t=j+1}^T \int Z_t^2(\mathbf{x}, \hat{\boldsymbol{\theta}}) dW(\mathbf{x}) \int \hat{\psi}_{t-j}^2(\mathbf{y}) dW(\mathbf{y}), \\ \hat{D}_1 &= 2 \sum_{j=1}^{T-2} \sum_{l=1}^{T-2} k^2(j/p)k^2(l/p) \int \int \int \int dW(\mathbf{x}_1) dW(\mathbf{y}_1) dW(\mathbf{x}_2) dW(\mathbf{y}_2) \\ &\quad \times \left\{ [T - \max(j, l)]^{-1} \sum_{t=\max(j, l)+1}^T Z_t(\mathbf{x}_1, \hat{\boldsymbol{\theta}}) Z_t(\mathbf{x}_2, \hat{\boldsymbol{\theta}}) \hat{\psi}_{t-j}(\mathbf{y}_1) \hat{\psi}_{t-l}(\mathbf{y}_2) \right\}^2,\end{aligned}$$

where $\hat{\psi}_t(\mathbf{y}) = \mathbf{1}(\mathbf{X}_t \leq \mathbf{y}) - \hat{P}(\mathbf{y})$, and as before, $\hat{P}(\mathbf{y}) = T^{-1} \sum_{t=1}^T \mathbf{1}(\mathbf{X}_t \leq \mathbf{y})$. The factors \hat{C}_1 and \hat{D}_1 are the approximately mean and variance of the quadratic form in (3.10).

When $W(\cdot)$ is continuous, \hat{Q}_1 can be calculated by numerical integration or simulation. Alternatively, the empirical distribution function also provides a natural way of choosing a data-dependent weighting function $W(\mathbf{x}) = \hat{P}(\mathbf{x})$, where $\hat{P}(\mathbf{x})$ is the empirical CDF of \mathbf{X}_t . Then a feasible test statistic is

$$\hat{Q}_2 = \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \hat{\Gamma}_j^2(\mathbf{X}_t, \mathbf{X}_s) - \hat{C}_2 \right] / \sqrt{\hat{D}_2}, \quad (3.12)$$

where the centering and scaling factors

$$\begin{aligned}\hat{C}_2 &= \sum_{j=1}^{T-1} k^2(j/p)(T-j)^{-1} \frac{1}{T^2} \sum_{m=j+1}^T \sum_{t=1}^T Z_m^2(\mathbf{X}_t, \hat{\boldsymbol{\theta}}) \sum_{s=1}^T \hat{\psi}_{m-j}^2(\mathbf{X}_s), \\ \hat{D}_2 &= 2 \sum_{j=1}^{T-2} \sum_{l=1}^{T-2} k^2(j/p)k^2(l/p) \frac{1}{T^4} \sum_{t_1=1}^T \sum_{t_2=1}^T \sum_{s_1=1}^T \sum_{s_2=1}^T \\ &\quad \times \left[\frac{1}{T - \max(j, l)} \sum_{m=\max(j, l)+1}^T Z_m(\mathbf{X}_{t_1}, \hat{\boldsymbol{\theta}}) Z_m(\mathbf{X}_{t_2}, \hat{\boldsymbol{\theta}}) \hat{\psi}_{m-j}(\mathbf{X}_{s_1}) \hat{\psi}_{m-l}(\mathbf{X}_{s_2}) \right]^2.\end{aligned}$$

Here, no numerical integration is needed. Depending on the sample size, the computational cost of \hat{Q}_2 may or may not be higher than that of \hat{Q}_1 .

Alternatively, we can define the generalized covariance function as the autocovariance of the generalized residuals

$$\tilde{\Gamma}_j(\mathbf{x}, \mathbf{y}) = \text{cov} [Z_t(\mathbf{x}, \boldsymbol{\theta}), Z_{t-|j|}(\mathbf{y}, \boldsymbol{\theta})], \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d,$$

and estimate it by its sample analogue

$$\hat{\tilde{\Gamma}}_j(\mathbf{x}, \mathbf{y}) = \frac{1}{T - |j|} \sum_{t=|j|+1}^T Z_t(\mathbf{x}, \hat{\boldsymbol{\theta}}) Z_{t-|j|}(\mathbf{y}, \hat{\boldsymbol{\theta}}), \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.$$

Following similar derivations, we can obtain two new test statistics, corresponding to (3.11) and (3.12)

respectively:

$$\tilde{Q}_1 = \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int \int \hat{\Gamma}_j^2(\mathbf{x}, \mathbf{y}) dW(\mathbf{x}) dW(\mathbf{y}) - \tilde{C}_1 \right] / \sqrt{\tilde{D}_1}, \quad (3.13)$$

where the centering and scaling factors

$$\begin{aligned} \tilde{C}_1 &= \sum_{j=1}^{T-1} k^2(j/p)(T-j)^{-1} \sum_{t=j+1}^T \int P(\mathbf{x}|\mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}}) [1 - P(\mathbf{x}|\mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}})] dW(\mathbf{x}) \int Z_{t-j}^2(\mathbf{y}, \hat{\boldsymbol{\theta}}) dW(\mathbf{y}), \\ \tilde{D}_1 &= 2 \sum_{j=1}^{T-2} k^4(j/p) \left\{ \int \left\{ T^{-1} \sum_{t=1}^T [P(\mathbf{x} \wedge \mathbf{y}|\mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}}) - P(\mathbf{x}|\mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}})P(\mathbf{y}|\mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}})] \right\}^2 dW(\mathbf{x}) dW(\mathbf{y}) \right\}^2, \end{aligned}$$

where $\mathbf{x} \wedge \mathbf{y} \equiv \min(\mathbf{x}, \mathbf{y})$. We note that \tilde{Q}_1 is computationally simpler than \hat{Q}_1 . In particular, the integration for \tilde{D}_1 is reduced from $4d$ dimensions to $2d$ dimensions. The key difference between \tilde{Q}_1 and \hat{Q}_1 is the use of different conditioning variables. The generalized residuals $\{Z_t(\mathbf{x}, \boldsymbol{\theta}_0)\}$ are MDS under the null and we expect that \tilde{Q}_1 might have better power than \hat{Q}_1 . We will further examine the finite sample performance of \tilde{Q}_1 and \hat{Q}_1 in Section 6.

A counterpart of (3.12) is

$$\tilde{Q}_2 = \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j)T^{-2} \sum_{t=1}^T \sum_{s=1}^T \hat{\Gamma}_j^2(\mathbf{X}_t, \mathbf{X}_s) - \tilde{C}_2 \right] / \sqrt{\tilde{D}_2}, \quad (3.14)$$

where the centering and scaling factors

$$\begin{aligned} \tilde{C}_2 &= \sum_{j=1}^{T-1} k^2(j/p)(T-j)^{-1}T^{-2} \sum_{m=j+1}^T \sum_{t=1}^T P(\mathbf{X}_t|\mathcal{I}_{m-1}, \hat{\boldsymbol{\theta}}) [1 - P(\mathbf{X}_t|\mathcal{I}_{m-1}, \hat{\boldsymbol{\theta}})] \sum_{s=1}^T Z_{m-j}^2(\mathbf{X}_s, \hat{\boldsymbol{\theta}}), \\ \tilde{D}_2 &= 2 \sum_{j=1}^{T-2} k^4(j/p) \left\{ T^{-2} \sum_{t=1}^T \sum_{s=1}^T \left\{ T^{-1} \sum_{m=1}^T [P(\mathbf{X}_t \wedge \mathbf{X}_s|\mathcal{I}_{m-1}, \hat{\boldsymbol{\theta}}) - P(\mathbf{X}_t|\mathcal{I}_{m-1}, \hat{\boldsymbol{\theta}})P(\mathbf{X}_s|\mathcal{I}_{m-1}, \hat{\boldsymbol{\theta}})] \right\}^2 \right\}^2. \end{aligned}$$

Again, the computational cost is lowered by replacing the quadruple sum with the double sum.

One could also consider a test based on the supremum norm

$$\hat{S} = \sup_{-\pi \leq \omega \leq \pi} \sup_{\mathbf{x}, \mathbf{y} \in \mathbb{R}^d} \left| \hat{F}(\omega, \mathbf{x}, \mathbf{y}) - \hat{F}_o(\omega, \mathbf{x}, \mathbf{y}) \right|.$$

This delivers a generalized KS test for dynamic conditional distribution models. In this paper, we focus on the test based on (3.10). The test based on \hat{S} requires a different treatment and it is expected to follow a nonstandard asymptotic distribution.

4. ASYMPTOTIC THEORY

To derive the null asymptotic distribution of the test statistics \hat{Q}_1 , \hat{Q}_2 , \tilde{Q}_1 , \tilde{Q}_2 and investigate their asymptotic power property, we impose following regularity conditions.

Assumption A.1: $\{\mathbf{X}_t, t \in N\}$ is a d -dimensional strictly stationary time series process with unknown

CDF $P_0(\mathbf{x}|\mathcal{I}_{t-1})$, where $\mathcal{I}_{t-1} \equiv \{\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots, \mathbf{X}_1\}$ and $d \geq 1$.

Assumption A.2: Let $P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta})$ be the CDF of \mathbf{X}_t given \mathcal{I}_{t-1} for a parametric model for \mathbf{X}_t . (i) For each $\boldsymbol{\theta} \in \Theta$, each $\mathbf{x} \in \mathbb{R}^d$, and each t , $P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta})$ is measurable with respect to \mathcal{I}_{t-1} ; (ii) for each $\boldsymbol{\theta} \in \Theta$, each $\mathbf{x} \in \mathbb{R}^d$, and each t , $P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta} \in \Theta$ with probability one; (iii) $\sup_{\mathbf{x} \in \mathbb{R}^d} \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E[\sup_{\boldsymbol{\theta} \in \Theta} \|\frac{\partial}{\partial \boldsymbol{\theta}} P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta})\|^2] \leq C$ and $\sup_{\mathbf{x} \in \mathbb{R}^d} \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E[\sup_{\boldsymbol{\theta} \in \Theta} \|\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta})\|] \leq C$.

Assumption A.3: $\hat{\boldsymbol{\theta}}$ is a parameter estimator such that $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = O_P(1)$, where $\boldsymbol{\theta}^* \equiv p \lim_{T \rightarrow \infty} \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ under \mathbb{H}_0 .

Assumption A.4: For each $\mathbf{x} \in \mathbb{R}^d$, $\{\mathbf{X}_t, \frac{\partial}{\partial \boldsymbol{\theta}} P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta}^*)\}$ is a strictly stationary α -mixing process with mixing coefficient satisfying $\sum_{j=0}^{\infty} \alpha(j)^{(\nu-1)/\nu} \leq C$ for some constant $\nu > 1$.

Assumption A.5: $k : \mathbb{R} \rightarrow [-1, 1]$ is a symmetric function that is continuous at zero and all points in \mathbb{R} except for a finite number of points, with $k(0) = 1$ and $k(z) \leq C|z|^{-b}$ for some $b > \frac{1}{2}$ as $z \rightarrow \infty$.

Assumption A.6: $W : \mathbb{R}^d \rightarrow \mathbb{R}^+$ is a nondecreasing right-continuous function that weighs sets symmetric about the origin equally, with $\int_{\mathbb{R}^d} dW(\mathbf{x}) < \infty$ and $\int_{\mathbb{R}^d} \|\mathbf{x}\|^4 dW(\mathbf{x}) < \infty$.

Assumption A.7: For each sufficiently large integer q , there exists a stationary process $\{\mathbf{X}_{q,t}\}$ such that $\{\mathbf{X}_{q,t}\}$ is independent of \mathcal{I}_{t-q} for q sufficiently large, and $E\|\mathbf{X}_t - \mathbf{X}_{q,t}\|^2 \leq Cq^{-\delta}$ for some constant $\delta \geq 1$ and all large q .

Assumption A.1 imposes some regularity conditions on the DGP. Both univariate and multivariate time series processes are covered, and we allow but do not require \mathbf{X}_t to be Markov. It is important to allow the DGP to be non-Markov, because many popular time series models such as GARCH, ACM and MA models are not Markov.

Assumption A.2 provides standard regularity conditions on the conditional CDF $P(\mathbf{x}|\mathcal{I}_{t-1})$ of \mathbf{X}_t . The assumption that the conditional CDF is twice continuously differentiable with respect to $\boldsymbol{\theta}$ is weaker than the requirement that the conditional parametric density be twice continuously differentiable in $\boldsymbol{\theta}$, since the integration is a smoothing operation. Bai (2003) imposes similar regularity conditions. We allow $P(\mathbf{x}|\mathcal{I}_{t-1})$ to depend on the entire past history \mathcal{I}_{t-1} , rather than finitely many lags only. Assumption A.3 requires a \sqrt{T} -consistent estimator $\hat{\boldsymbol{\theta}}$ under \mathbb{H}_0 , which need not be asymptotically most efficient. The quasi-maximum likelihood estimator can be used. Assumption A.4 is a regularity condition on the temporal dependence of the process $\{\mathbf{X}_t, \frac{\partial}{\partial \boldsymbol{\theta}} P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta}^*)\}$. Assumption A.5 is the regularity condition on the kernel function $k(\cdot)$. The continuity of $k(\cdot)$ at 0 and the unity of $k(0)$ ensure that the bias of the generalized spectral estimator $\hat{F}(\omega, \mathbf{x}, \mathbf{y})$ vanishes to zero asymptotically as $T \rightarrow \infty$. The condition on the tail behavior of $k(\cdot)$ ensures that higher order lags have asymptotically negligible impact on the statistical properties of $\hat{F}(\omega, \mathbf{x}, \mathbf{y})$. Assumption A.5 covers most commonly used kernels. For kernels with bounded support, such as the Bartlett and Parzen kernels in (3.7) and (3.8), we have $b = \infty$. For kernels with unbounded support, b is some finite positive real number. For example, we have $b = 2$ for the Quadratic-Spectral kernel in (3.9). Assumption A.6 imposes mild conditions on the weighting function $W(\cdot)$. Any CDF with finite fourth moments satisfies Assumption A.6. Note that $W(\cdot)$ can be a step function. This provides a convenient way to implement our tests, because we can avoid high dimensional numerical integrations by using a finite number of grid points for \mathbf{x} and \mathbf{y} . This is equivalent to using the CDF of a discrete random vector.

Assumption A.7 is needed only under \mathbb{H}_0 . It assumes that $\{\mathbf{X}_t\}$ can be approximated by a q -dependent process $\{\mathbf{X}_{q,t}\}$ arbitrarily well when q is sufficiently large. Intuitively, this condition implies that the serial dependence of \mathbf{X}_t on its remote past history decays to zero sufficiently fast so that it is asymptotically negligible. It holds trivially when $\{\mathbf{X}_t\}$ is a q -dependent process with an arbitrarily large but finite order q . It also covers many non-Markov processes.

We now state the asymptotic distribution of the GCM test \hat{Q}_1 under \mathbb{H}_0 . All other tests \hat{Q}_2 , \tilde{Q}_1 and \tilde{Q}_2 follow the same asymptotic $N(0, 1)$ distribution under \mathbb{H}_0 .

Theorem 1: *Suppose Assumptions A.1–A.7 hold, and $p = cT^\lambda$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$ and $0 < c < \infty$. Then $\hat{Q}_1 \xrightarrow{d} N(0, 1)$ under \mathbb{H}_0 as $T \rightarrow \infty$.*

The asymptotic normality of our GCM test statistic \hat{Q}_1 differs sharply from the nonstandard distribution of the CM test statistic in the literature. It offers a rather convenient inference procedure. For example, the asymptotic $N(0, 1)$ critical value at the 5% significance level is 1.65. The appealing asymptotic normality is made possible due to our spectral approach. To gain intuition, we consider the case when the kernel function $k(\cdot)$ has bounded support, *i.e.*, $k(z) = 0$ if $|z| > 1$. Then \hat{Q}_1 is a weighted sum of p random variables $\left\{ \int \int \hat{\Gamma}_j^2(\mathbf{x}, \mathbf{y}) dW(\mathbf{x}) dW(\mathbf{y}) \right\}_{j=1}^p$, which are approximately independent under \mathbb{H}_0 when $p \rightarrow \infty$. This statistic thus converges to $N(0, 1)$ by CLT after appropriate centering and scaling. Of course, our formal proof does not rely on this simplistic intuition. Another important feature of \hat{Q}_1 that differs from the classical CM tests is that the use of the estimated generalized residuals $\{Z_t(\mathbf{x}, \hat{\boldsymbol{\theta}})\}$ in place of the unobservable generalized residuals $\{Z_t(\mathbf{x}, \boldsymbol{\theta}_0)\}$ has no impact on the limiting distribution of \hat{Q}_1 . One can proceed as if the true parameter value $\boldsymbol{\theta}_0$ were known and equal to $\hat{\boldsymbol{\theta}}$. Intuitively, the parametric estimator $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_0$ faster than the nonparametric estimator $\hat{F}(\omega, \mathbf{x}, \mathbf{y})$ converges to $F(\omega, \mathbf{x}, \mathbf{y})$ as $T \rightarrow \infty$. Consequently, the limiting distribution of \hat{Q}_1 is solely determined by $\hat{F}(\omega, \mathbf{x}, \mathbf{y})$, and replacing $\boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\theta}}$ has no impact asymptotically. This delivers a convenient procedure, because any \sqrt{T} -consistent estimator can be used.

Next, we investigate the asymptotic power of \hat{Q}_1 under \mathbb{H}_A .

Theorem 2: *Suppose Assumption A.1–A.7 hold, and $p = cT^\lambda$ for $0 < \lambda < \frac{1}{2}$ and $0 < c < \infty$. Then as $T \rightarrow \infty$,*

$$\begin{aligned} \frac{p^{\frac{1}{2}}}{T} \hat{Q}_1 &\xrightarrow{p} \frac{1}{\sqrt{D}} \sum_{j=1}^{\infty} \int \int \Gamma_j^2(\mathbf{x}, \mathbf{y}) dW(\mathbf{x}) dW(\mathbf{y}) \\ &= \frac{\pi}{2\sqrt{D}} \int \int \int_{-\pi}^{\pi} [F(\omega, \mathbf{x}, \mathbf{y}) - F_0(\omega, \mathbf{x}, \mathbf{y})]^2 d\omega dW(\mathbf{x}) dW(\mathbf{y}), \end{aligned}$$

where

$$D = 2 \int_0^{\infty} k^4(z) dz \int \int \left| \tilde{\Gamma}_0(\mathbf{x}_1, \mathbf{x}_2) \right|^2 dW(\mathbf{x}_1) dW(\mathbf{x}_2) \sum_{j=-\infty}^{\infty} \int \int |\Omega_j(\mathbf{y}_1, \mathbf{y}_2)|^2 dW(\mathbf{y}_1) dW(\mathbf{y}_2),$$

and $\tilde{\Gamma}_0(\mathbf{x}, \mathbf{y}) = \text{cov}[Z_t(\mathbf{x}, \boldsymbol{\theta}^*), Z_t(\mathbf{y}, \boldsymbol{\theta}^*)]$ and $\Omega_j(\mathbf{x}, \mathbf{y}) = \text{cov}[\mathbf{1}(\mathbf{X}_t \leq \mathbf{x}), \mathbf{1}(\mathbf{X}_{t-|j|} \leq \mathbf{y})]$.

The function $\Omega_j(\mathbf{x}, \mathbf{y})$ can be viewed as the indicator function-based autocovariance function of $\{\mathbf{X}_t\}$. It captures temporal dependence in $\{\mathbf{X}_t\}$. The dependence of the constant D on $\Omega_j(\mathbf{x}, \mathbf{y})$ is due to the fact that the conditioning variable $\mathbf{1}(\mathbf{X}_{t-|j|} \leq \mathbf{y})$ is a time series process.

Following Stinchcombe and White (1998), we have that for $j > 0$, $\Gamma_j(\mathbf{x}, \mathbf{y}) = 0$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ if and only if $E[Z_t(\mathbf{x}, \boldsymbol{\theta}^*)|\mathbf{X}_{t-j}] = 0$ *a.s.* for all $\mathbf{x} \in \mathbb{R}^d$. Suppose $E[Z_t(\mathbf{x}, \boldsymbol{\theta}^*)|\mathbf{X}_{t-j}] \neq 0$ at some lag $j > 0$ under \mathbb{H}_A . Then we have $\int \int |\Gamma_j(\mathbf{x}, \mathbf{y})|^2 dW(\mathbf{x}) dW(\mathbf{y}) \geq C > 0$ for any weighting function $W(\cdot)$ that is positive, monotonically increasing and continuous, with unbounded support on \mathbb{R} . As a result, $P[\hat{Q}_1 > C(T)] \rightarrow 1$ for any sequence of constants $\{C(T) = o(T/p^{1/2})\}$. Thus \hat{Q}_1 has asymptotic unit power at any given significance level $\alpha \in (0, 1)$, whenever $E[Z_t(\mathbf{x}, \boldsymbol{\theta}^*)|\mathbf{X}_{t-j}]$ is nonzero at some lag $j > 0$ under \mathbb{H}_A . Note that for a Markov process \mathbf{X}_t , we always have $E[Z_t(\mathbf{x}, \boldsymbol{\theta}^*)|\mathbf{X}_{t-j}] \neq 0$ at least for some $j > 0$ under \mathbb{H}_A . Hence, \hat{Q}_1 is consistent against \mathbb{H}_A when \mathbf{X}_t is Markov.

For a non-Markovian process \mathbf{X}_t , the hypothesis that $E[Z_t(\mathbf{x}, \boldsymbol{\theta}_0)|\mathbf{X}_{t-j}] = 0$ *a.s.* for all $\mathbf{x} \in \mathbb{R}^d$ and some $\boldsymbol{\theta}_0 \in \Theta$ and all $j > 0$ is not equivalent to the hypothesis that $E[Z_t(\mathbf{x}, \boldsymbol{\theta}_0)|\mathcal{I}_{t-1}] = 0$ *a.s.* for all $\mathbf{x} \in \mathbb{R}^d$ and some $\boldsymbol{\theta}_0 \in \Theta$. The latter implies the former but not vice versa. This is the price we have to pay for dealing with the difficulty of "the curse of dimensionality". Nevertheless, our GCM test is expected to have power against a wide range of non-Markovian processes, since we check many lag orders. The use of a large number of lags might cause the loss of power, due to the loss of a large number of degree of freedom. Fortunately, such power loss is substantially alleviated for \hat{Q}_1 , thanks to the downward weighting by $k^2(\cdot)$ for higher order lags. Generally speaking, \mathbf{X}_t is more affected by the recent events than the remote past events. In such scenarios, equal weighting to each lag is not expected to be powerful. Instead, downward weighting is expected to enhance better power because it discounts remote past information. Thus, we expect that the power of our test is not so sensitive to the choice of the lag order. This is confirmed by our simulation study below.

The asymptotic powers of \hat{Q}_2 , \tilde{Q}_1 and \tilde{Q}_2 under \mathbb{H}_A can be derived in a similar manner.

5. DIAGNOSTIC PROCEDURES

When a conditional distribution model is rejected by the GCM test \hat{Q}_1 , say, it would be interesting to explore possible sources of the rejection. For example, one may like to know whether misspecification comes from the conditional mean, conditional variance, conditional skewness or conditional kurtosis. In economic and financial applications, for example, the first four conditional moments are closely related to the return, volatility, asymmetry and fat-tail, respectively. Such information, if any, will be valuable in reconstructing the model and studying different aspects of the dynamics of economic and financial time series.

The CDF is a convenient and useful tool to gauge possible sources of model misspecification, because it can be integrated to obtain conditional moments. We now develop a class of diagnostic tests by integrating the CDF-based generalized spectral density $F(\omega, \mathbf{x}, \mathbf{y})$. This class of diagnostic tests can provide useful information about how well a conditional distribution model can capture the dynamics of various conditional moments.

To gain insight, we first consider the univariate case ($d = 1$). By straightforward derivation, we can obtain the integral of the generalized covariance function

$$\begin{aligned} \Gamma_j^m(y) &= \text{cov} \left[\int mx^{m-1} Z_t(x, \boldsymbol{\theta}) dx, \mathbf{1}(X_{t-|j|} \leq y) \right] \\ &= -\text{cov}[X_t^m - E_{\boldsymbol{\theta}}(X_t^m|\mathcal{I}_{t-1}), \mathbf{1}(X_{t-|j|} \leq y)]. \end{aligned}$$

Here, as before, $E_{\boldsymbol{\theta}}(\cdot|\mathcal{I}_{t-1})$ is the conditional expectation under the CDF model $P(x|\mathcal{I}_{t-1}, \boldsymbol{\theta})$.

For example, when $m = 1, 2, 3, 4$, we have

$$\begin{aligned}\Gamma_j^1(y) &= -\text{cov}[X_t - E_{\theta}(X_t|\mathcal{I}_{t-1}), \mathbf{1}(X_{t-|j|} \leq y)], \\ \Gamma_j^2(y) &= -\text{cov}[X_t^2 - E_{\theta}(X_t^2|\mathcal{I}_{t-1}), \mathbf{1}(X_{t-|j|} \leq y)], \\ \Gamma_j^3(y) &= -\text{cov}[X_t^3 - E_{\theta}(X_t^3|\mathcal{I}_{t-1}), \mathbf{1}(X_{t-|j|} \leq y)], \\ \Gamma_j^4(y) &= -\text{cov}[X_t^4 - E_{\theta}(X_t^4|\mathcal{I}_{t-1}), \mathbf{1}(X_{t-|j|} \leq y)],\end{aligned}$$

which can be used to check misspecifications in the first four conditional moments respectively.

Alternatively, we can define the standardized innovation

$$\varepsilon_t = [X_t - \mu(\mathcal{I}_{t-1}, \theta)]/h^{1/2}(\mathcal{I}_{t-1}, \theta)$$

and test the moment conditions based on ε_t . If $\{\varepsilon_t\}$ is *i.i.d.* $N(0, 1)$ under \mathbb{H}_0 , we have

$$\begin{aligned}\Gamma_j'^1(y) &= -\text{cov}[\varepsilon_t, \mathbf{1}(\varepsilon_{t-|j|} \leq y)], \\ \Gamma_j'^2(y) &= -\text{cov}[\varepsilon_t^2 - 1, \mathbf{1}(\varepsilon_{t-|j|} \leq y)], \\ \Gamma_j'^3(y) &= -\text{cov}[\varepsilon_t^3, \mathbf{1}(\varepsilon_{t-|j|} \leq y)], \\ \Gamma_j'^4(y) &= -\text{cov}[\varepsilon_t^4 - 3, \mathbf{1}(\varepsilon_{t-|j|} \leq y)],\end{aligned}$$

which can be used to check misspecifications in conditional mean, conditional variance, conditional skewness and conditional kurtosis respectively. This alternative has more natural economic interpretation since it checks conditional centered moments rather than uncentered moments. Nevertheless, this procedure is more like a joint test, in the sense that the test of higher conditional moments is based on the assumption that lower conditional moments are correctly specified.⁷

This set of diagnostic tests is similar to the moment-based tests used in Brooks *et al.* (2005) and Harvey and Siddique (1999). Brooks *et al.* (2005) check the temporal dependence of the standardized residuals via the following orthogonality conditions:

$$\begin{aligned}E(\varepsilon_t z_t) &= 0, \\ E(\varepsilon_t \cdot \varepsilon_{t-j}) &= 0 \text{ for } j = 1, 2, 3, 4, \\ E\left[\left(\varepsilon_t^2 - \frac{\nu_t}{\nu_t - 2}\right)\left(\varepsilon_{t-j}^2 - \frac{\nu_{t-j}}{\nu_{t-j} - 2}\right)\right] &= 0 \text{ for } j = 1, 2, 3, 4, \\ E(\varepsilon_t^3 \cdot \varepsilon_{t-j}^3) &= 0 \text{ for } j = 1, 2, 3, 4, \\ E\left\{\left[\varepsilon_t^4 - \frac{3\nu_t^2}{(\nu_t - 2)(\nu_t - 4)}\right]\left[\varepsilon_{t-j}^4 - \frac{3\nu_{t-j}^2}{(\nu_{t-j} - 2)(\nu_{t-j} - 4)}\right]\right\} &= 0 \text{ for } j = 1, 2, 3, 4,\end{aligned}$$

where ε_t is the standardized residual and ν_t is the degrees of freedom of the innovation.⁸ Compared with these conditional moment tests, our tests have several advantages: first, our GCM test \hat{Q}_1 essentially

⁷For example, the test of the conditional variance is based on the assumption that the conditional mean is correctly specified. But it coincides with the empirical convention that the specification tests usually are carried out from lower moments to higher moments.

⁸Harvey and Siddique (1999) construct a similar set of 16 orthogonality conditions.

checks every moment, which is not obtainable by the Chi-square test; second, because we employ a frequency domain approach, we check a growing number of lags as the sample size increases, while they use an arbitrary and fixed lag order; third, our GCM test and diagnostic procedures are derived in a unified framework.

Next, we consider the multivariate case ($d > 1$). We define the integral of the generalized cross-covariance function as

$$\Gamma_j^{\mathbf{m}}(\mathbf{y}) = -\text{cov} \left\{ \prod_{m_c \neq 0} [X_{ct}^{m_c} - E_{\boldsymbol{\theta}}(X_{ct}^{m_c} | \mathcal{I}_{t-1})] - E_{\boldsymbol{\theta}} \left\{ \prod_{m_c \neq 0} [X_{ct}^{m_c} - E_{\boldsymbol{\theta}}(X_{ct}^{m_c} | \mathcal{I}_{t-1})] \right\}, \mathbf{1}(\mathbf{X}_{t-|j|} \leq \mathbf{y}) \right\},$$

where $\mathbf{m} = (m_1, m_2, \dots, m_d)'$, $m_c \geq 0$ for all $1 \leq c \leq d$. We put $|\mathbf{m}| = \sum_{c=1}^d m_c$.

For illustration, consider a bivariate process $\mathbf{X}_t = (X_{1t}, X_{2t})'$ and examine the cases of $|\mathbf{m}| = 1$ and $|\mathbf{m}| = 2$ respectively:

- Case 1: $|\mathbf{m}| = 1$. We have $\mathbf{m} = (1, 0)$ or $\mathbf{m} = (0, 1)$. If $\mathbf{m} = (1, 0)$,

$$\Gamma_j^{(\mathbf{m})}(\mathbf{y}) = -\text{cov} [X_{1t} - E_{\boldsymbol{\theta}}(X_{1t} | \mathcal{I}_{t-1}), \mathbf{1}(\mathbf{X}_{t-|j|} \leq \mathbf{y})].$$

If $\mathbf{m} = (0, 1)$, then

$$\Gamma_j^{(\mathbf{m})}(\mathbf{y}) = -\text{cov} [X_{2t} - E_{\boldsymbol{\theta}}(X_{2t} | \mathcal{I}_{t-1}), \mathbf{1}(\mathbf{X}_{t-|j|} \leq \mathbf{y})].$$

Thus, the choice of $|\mathbf{m}| = 1$ can be used to check misspecifications in the conditional mean dynamics of X_{1t} and X_{2t} respectively.

- Case 2: $|\mathbf{m}| = 2$. We have $\mathbf{m} = (2, 0)$, $(0, 2)$ or $(1, 1)$. If $\mathbf{m} = (2, 0)$,

$$\Gamma_j^{(\mathbf{m})}(\mathbf{y}) = -\text{cov} [X_{1t}^2 - E_{\boldsymbol{\theta}}(X_{1t}^2 | \mathcal{I}_{t-1}), \mathbf{1}(\mathbf{X}_{t-|j|} \leq \mathbf{y})].$$

If $\mathbf{m} = (0, 2)$,

$$\Gamma_j^{(\mathbf{m})}(\mathbf{y}) = -\text{cov} [X_{2t}^2 - E_{\boldsymbol{\theta}}(X_{2t}^2 | \mathcal{I}_{t-1}), \mathbf{1}(\mathbf{X}_{t-|j|} \leq \mathbf{y})].$$

Finally, if $\mathbf{m} = (1, 1)$,

$$\Gamma_j^{(\mathbf{m})}(\mathbf{y}) = -\text{cov}[\text{cov}(X_{1t}, X_{2t} | \mathcal{I}_{t-1}) - \text{cov}_{\boldsymbol{\theta}}(X_{1t}, X_{2t} | \mathcal{I}_{t-1}), \mathbf{1}(\mathbf{X}_{t-|j|} \leq \mathbf{y})].$$

Thus, the choice of $|\mathbf{m}| = 2$ can be used to check model misspecifications in the conditional variances of X_{1t} and X_{2t} , as well as their conditional correlation.

Then the CDF-based generalized spectral integral is defined as

$$F^{\mathbf{m}}(\omega, \mathbf{y}) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma_j^{\mathbf{m}}(\mathbf{y}) \exp(-ij\omega). \quad (5.1)$$

We now obtain the class of diagnostic test statistics as follows:

$$\hat{Q}_1^{\mathbf{m}} = \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int \hat{\Gamma}_j^{\mathbf{m}}(\mathbf{y})^2 dW(\mathbf{y}) - \hat{C}_1^{\mathbf{m}} \right] / \sqrt{\hat{D}_1^{\mathbf{m}}}, \quad (5.2)$$

where the centering and scaling factors

$$\begin{aligned} \hat{C}_1^{\mathbf{m}} &= \sum_{j=1}^{T-1} k^2(j/p)(T-j)^{-1} \sum_{t=j+1}^T Z_t^{\mathbf{m}}(\hat{\boldsymbol{\theta}})^2 \int \hat{\psi}_{t-j}^2(\mathbf{y}) dW(\mathbf{y}), \\ \hat{D}_1^{\mathbf{m}} &= 2 \sum_{j=1}^{T-2} \sum_{l=1}^{T-2} k^2(j/p)k^2(l/p) \int \int \\ &\times \left\{ [T - \max(j, l)]^{-1} \sum_{t=\max(j, l)+1}^T Z_t^{\mathbf{m}}(\hat{\boldsymbol{\theta}})^2 \hat{\psi}_{t-j}(\mathbf{y}_1) \hat{\psi}_{t-l}(\mathbf{y}_2) \right\}^2 dW(\mathbf{y}_1) dW(\mathbf{y}_2), \end{aligned}$$

with

$$Z_t^{\mathbf{m}}(\hat{\boldsymbol{\theta}}) = - \left\{ \prod_{m_c \neq 0} [X_{ct}^{m_c} - E_{\hat{\boldsymbol{\theta}}}(X_{ct}^{m_c} | \mathcal{I}_{t-1})] - E_{\hat{\boldsymbol{\theta}}} \left\{ \prod_{m_c \neq 0} [X_{ct}^{m_c} - E_{\hat{\boldsymbol{\theta}}}(X_{ct}^{m_c} | \mathcal{I}_{t-1})] | \mathcal{I}_{t-1} \right\} \right\}.$$

Alternatively, we can obtain the integral of the generalized covariance function, which is the auto-covariance of the generalized residuals

$$\tilde{\Gamma}_j^{\mathbf{m}}(\mathbf{y}) = -\text{cov} \left\{ \prod_{m_c \neq 0} [X_{ct}^{m_c} - E_{\boldsymbol{\theta}}(X_{ct}^{m_c} | \mathcal{I}_{t-1})] - E_{\boldsymbol{\theta}} \left\{ \prod_{m_c \neq 0} [X_{ct}^{m_c} - E_{\boldsymbol{\theta}}(X_{ct}^{m_c} | \mathcal{I}_{t-1})] | \mathcal{I}_{t-1} \right\}, Z_{t-|j|}(\mathbf{y}, \boldsymbol{\theta}) \right\}.$$

Following similar derivations, we obtain a new class of diagnostic test statistics:

$$\tilde{Q}_1^{\mathbf{m}} = \left\{ \sum_{j=1}^{T-1} k^2(j/p)(T-j) \int [\hat{\tilde{\Gamma}}_j^{\mathbf{m}}(\mathbf{y})]^2 dW(\mathbf{y}) - \tilde{C}_1^{\mathbf{m}} \right\} / \sqrt{\tilde{D}_1^{\mathbf{m}}}, \quad (5.3)$$

where the centering and scaling factors

$$\begin{aligned} \tilde{C}_1^{\mathbf{m}} &= \sum_{j=1}^{T-1} k^2(j/p)(T-j)^{-1} \sum_{t=j+1}^T \left\{ \prod_{m_c \neq 0} [X_{ct}^{m_c} - E_{\hat{\boldsymbol{\theta}}}(X_{ct}^{m_c} | \mathcal{I}_{t-1})]^2 \right. \\ &\quad \left. - \left\{ E_{\hat{\boldsymbol{\theta}}} \prod_{m_c \neq 0} [X_{ct}^{m_c} - E_{\hat{\boldsymbol{\theta}}}(X_{ct}^{m_c} | \mathcal{I}_{t-1})] | \mathcal{I}_{t-1} \right\}^2 \right\} \int Z_{t-j}^2(\mathbf{y}, \hat{\boldsymbol{\theta}}) dW(\mathbf{y}), \\ \tilde{D}_1^{\mathbf{m}} &= 2 \sum_{j=1}^{T-2} k^4(j/p) \left\{ T^{-1} \sum_{t=1}^T \left\{ \prod_{m_c \neq 0} [X_{ct}^{m_c} - E_{\hat{\boldsymbol{\theta}}}(X_{ct}^{m_c} | \mathcal{I}_{t-1})]^2 - \left\{ E_{\hat{\boldsymbol{\theta}}} \prod_{m_c \neq 0} [X_{ct}^{m_c} - E_{\hat{\boldsymbol{\theta}}}(X_{ct}^{m_c} | \mathcal{I}_{t-1})] | \mathcal{I}_{t-1} \right\}^2 \right\} \right\}^2 \\ &\quad \times \int \left\{ T^{-1} \sum_{t=1}^T [P(\mathbf{y}_1 \wedge \mathbf{y}_2 | \mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}}) - P(\mathbf{y}_1 | \mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}}) P(\mathbf{y}_2 | \mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}})] \right\}^2 dW(\mathbf{y}_1) dW(\mathbf{y}_2). \end{aligned}$$

To derive the asymptotic distribution of $\hat{Q}_1^{\mathbf{m}}$, we impose the following moment conditions:

Assumption A.8: (i) $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta} \prod_{m_c \neq 0} E_{\theta} (X_{ct}^{m_c} | \mathcal{I}_{t-1}) \right\|^2 \leq C;$

(ii) $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E \sup_{\theta \in \Theta} \left\| \frac{\partial^2}{\partial \theta \partial \theta'} \prod_{m_c \neq 0} E_{\theta} (X_{ct}^{m_c} | \mathcal{I}_{t-1}) \right\|^2 \leq C;$

(iii) $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E \sup_{\theta \in \Theta} \left\| \prod_{m_c \neq 0} E_{\theta} (X_{ct}^{m_c} | \mathcal{I}_{t-1}) \right\|^4 \leq C;$ (iv) $E(\prod_{m_c \neq 0} X_{ct}^{4m_c}) \leq C.$

Theorem 3: *Suppose Assumption A.1– A.8 hold for some pre-specified \mathbf{m} . Then $\hat{Q}_1^{\mathbf{m}} \xrightarrow{d} N(0, 1)$ under \mathbb{H}_0 as $T \rightarrow \infty$.*

Like the GCM test \hat{Q}_1 , the generalized spectral integral test $\hat{Q}_1^{\mathbf{m}}$ has a convenient asymptotic $N(0, 1)$ distribution and parameter estimation uncertainty in $\hat{\theta}$ has no impact on the asymptotic distribution of $\hat{Q}_1^{\mathbf{m}}$. Any \sqrt{T} -consistent estimator can be used. Moreover, different choices of \mathbf{m} can examine various aspects of the dynamics of random vector \mathbf{X}_t and thus may provide information on how well a time series conditional distribution model fits various aspects of the dynamics of \mathbf{X}_t .

For any given m , the diagnostic test statistic $\tilde{Q}_1^{\mathbf{m}}$ also follows the same asymptotic $N(0, 1)$ distribution under \mathbb{H}_0 .

6. FINITE SAMPLE PERFORMANCE

It is unclear how well the asymptotic theory can provide reliable reference and guidance in finite samples when applied to actual economic and financial time series data, which usually display conditional heteroskedasticity and serial dependence in higher moments. We now investigate the finite sample performance of the proposed tests for the adequacy of some conditional distribution models. For simplicity, we focus on two GCM tests \hat{Q}_1 and \tilde{Q}_1 in both univariate and bivariate contexts.

6.1 Univariate Models

6.1.1 Simulation Design

To examine the size of our tests under \mathbb{H}_0 , we consider the following DGP:

DGP0 [MA(1)-GARCH(1,1)- $N(0, 1)$]:

$$\begin{cases} X_t = u_t + 0.5u_{t-1}, \\ u_t = h_t^{\frac{1}{2}} \varepsilon_t, \\ h_t = 0.05 + 0.15u_{t-1}^2 + 0.8h_{t-1}, \\ \{\varepsilon_t\} \sim i.i.d.N(0, 1). \end{cases} \quad (6.1)$$

The MA(1)-GARCH(1,1) model is commonly used in empirical finance. We simulate 1,000 data sets of a random sample $\{X_t\}_{t=1}^T$ for $T = 100, 250, 500, 1,000, 2,500$ respectively. For each iteration, we first generate $T + 500$ observations and then discard the first 500 to reduce the impact of initial values. Under DGP0, the conditional distribution of X_t given \mathcal{I}_{t-1} is normal with mean $0.5u_{t-1}$ and variance h_t . For each data set, we estimate the model parameters via MLE and then compute our statistics.

To investigate the power of our test, we consider the following DGPs:

DGP1 [ARMA(1,1)-GARCH(1,1)- $N(0,1)$]:

$$\begin{cases} X_t = 0.3X_{t-1} + u_t + 0.5u_{t-1}, \\ u_t = h_t^{\frac{1}{2}}\varepsilon_t, \\ h_t = 0.05 + 0.15u_{t-1}^2 + 0.8h_{t-1}, \\ \text{where } \varepsilon_t \sim i.i.d.N(0, 1). \end{cases} \quad (6.2)$$

DGP2 [MA(1)-EGARCH(1,1)- $N(0,1)$]:

$$\begin{cases} X_t = u_t + 0.5u_{t-1} \\ u_t = h_t^{\frac{1}{2}}\varepsilon_t \\ \ln h_t = 0.05 + 0.8 \ln h_{t-1} + 0.15 \left(|\varepsilon_{t-1}| - \frac{2}{\sqrt{\pi}} \right) - 0.8\varepsilon_{t-1}, \\ \text{where } \varepsilon_t \sim i.i.d.N(0, 1). \end{cases} \quad (6.3)$$

In DGP3-6 below, the individual mean and variance are of the same forms as those in DGP0.

DGP3 [MA(1)-GARCHK- t]:

$$\begin{cases} \varepsilon_t \sim \sqrt{\frac{\nu_t - 2}{\nu_t}} t(\nu_t), \\ k_t = 5.041 + \frac{0.412u_{t-1}^4}{h_{t-1}^2} + 0.171k_{t-1}, \\ \nu_t = \frac{2(2k_t - 3)}{k_t - 3}. \end{cases} \quad (6.4)$$

DGP4 [MA(1)-GARCH(1,1)- $\chi^2(5)$]:

$$\varepsilon_t \sim i.i.d.[\chi^2(5) - 5]/\sqrt{10}. \quad (6.5)$$

DGP5 [MA(1)-GARCH(1,1)- $t(5)$]:

$$\varepsilon_t \sim i.i.d.\sqrt{3/5}t(5). \quad (6.6)$$

DGP6 [MA(1)-GARCH(1,1)-time varying skewed Student's t]:

$$\begin{cases} \varepsilon_t \sim p(\varepsilon|\nu_t, \lambda_t), \\ \nu_t = -1.2 - 0.4u_{t-1} - 0.5u_{t-1}^2, \\ \lambda_t = -0.5 - 0.5u_{t-1} - 0.6u_{t-1}^2. \end{cases} \quad (6.7)$$

DGP1 is an ARMA(1,1)-GARCH(1,1) process with $i.i.d.N(0, 1)$ innovations. Under DGP1, model (6.1) is misspecified for the conditional mean but is correctly specified for the conditional variance and higher moments. DGP2 is Nelson's (1991) EGARCH model with $i.i.d.N(0, 1)$ innovations. Under DGP 2, model (6.1) is correctly specified for the conditional mean but is misspecified for the conditional variance because it fails to capture the asymmetric effects in volatility. DGP3 is Brooks *et al*'s (2005) GARCHK model, which allows the conditional variance and kurtosis to vary over time separately via the time-varying degrees of freedom. If we use model (6.1) to fit the data generated from DGP3, the first three conditional moments are correctly specified, but there exists dynamic misspecifications in

the conditional kurtosis since it ignores the time-varying conditional fourth moment. Under DGPs 4-6, model (6.1) is correctly specified for both the conditional mean and the conditional variance, but the distribution of the innovation ε_t is misspecified. Among them, DGP4 and DGP5 assume that ε_t is generated from the time-invariant $\chi^2(5)$ and $t(5)$ respectively, while DGP6 assumes that ε_t is generated from Hansen’s (1994) time-varying skewed Student’s t distribution, whose degrees of freedom ν_t and skew parameter λ_t change over time.⁹ As suggested by Hansen (1994), we bound ν_t between 2.1 and 30, and λ_t between -0.9 and 0.9 by a logistic transformation.

For each of DGPs 1-6, we generate 500 data sets of the random sample $\{X_t\}_{t=1}^T$ for $T = 250, 500, 1,000$ and $2,500$ respectively. For each iteration, we generate $T + 500$ observations and then discard the first 500 to reduce the impact of the choice of some initial values. For each data set, we estimate model (6.1) via MLE. Because model (6.1) is misspecified under all six DGPs, our tests \hat{Q}_1 and \tilde{Q}_1 are expected to have nontrivial power under DGPs 1-6, provided the sample size T is sufficiently large.

6.1.2 Monte Carlo Evidence

We choose the $N(0,1)$ CDF for $W(\cdot)$ and the Bartlett kernel for $k(\cdot)$, which has bounded support and is computationally efficient. Our simulation experience suggests that the choices of $W(\cdot)$ and $k(\cdot)$ have little impact on both size and power of the tests.¹⁰ Like Hong (1999), we use a data-driven \hat{p} via a plug-in method that minimizes the asymptotic integrated mean squared error of the generalized spectral density estimator $\hat{F}(\omega, \mathbf{x}, \mathbf{y})$, with the Bartlett kernel $\bar{k}(\cdot)$ used in some preliminary generalized spectral density estimators. To examine the sensitivity of the choice of the preliminary bandwidth \bar{p} on the size and power of the tests, we consider \bar{p} in the range of 10 to 40. We consider the empirical rejection rates using the asymptotic critical values (1.28 and 1.65) at the 10% and 5% significance levels respectively.

Table 1 reports the rejection rates (in terms of percentage) of \hat{Q}_1 and \tilde{Q}_1 under DGP0 at the 10% and 5% levels. Both tests have reasonable sizes for sample sizes as small as $T = 100$, at both 10% and 5% levels. Both tests, especially \hat{Q}_1 , tend to underreject a little but the underrejection is not excessive. The sizes of \hat{Q}_1 and \tilde{Q}_1 are not sensitive to the choice of the preliminary lag order \bar{p} .

Table 2 reports the rejection rates of \tilde{Q}_1 under DGPs 1-6 at the 10% and 5% levels respectively. Under DGP1, model (6.1) ignores the autoregressive part in the conditional mean dynamics. The \tilde{Q}_1 test has good power in detecting such misspecification in the conditional mean. The rejection rate of \tilde{Q}_1 increases significantly with the sample size T and approaches unity when $T = 2,500$. Under DGP2, model (6.1) ignores the asymmetric effects in the conditional variance. The \tilde{Q}_1 test has excellent power when (6.1) is used to fit data generated from DGP2. The rejection rate is around 50% at the 5% level when $T = 250$ and approaches unity when $T = 1,000$. Under DGP3, model (6.1) is correctly specified for the conditional mean, conditional variance and conditional skewness, but is misspecified for the conditional kurtosis. The \tilde{Q}_1 test has no power when the sample size T is small but the rejection rate increases with the sample size. The reason why \tilde{Q}_1 has worse power under DGP3 than under DGPs1 and 2 is that \tilde{Q}_1 checks model misspecification in all directions while under DGP3, the conditional mean, conditional variance and conditional skewness are all correctly specified. On the other hand, when we examine the data generated from DGP3, we find that the degrees of freedom are all around 5, which enhances the difficulty in distinguishing the normal innovation with the Student’s t innovation.

⁹ $p(\varepsilon|\nu_t, \lambda_t)$ is a skewed student’s t distribution, whose PDF is given in (2.2).

¹⁰We have tried the Parzen kernel for $k(\cdot)$, obtaining similar results (not reported here).

Under DGPs 4-6, model (6.1) is correctly specified for the conditional mean and conditional variance but is misspecified for the entire distribution. It is well known that when the degrees of freedom ν are large, the standardized t_ν or χ_ν^2 random variable ε_t is approximately standard normal. Thus the power of the test decreases as ν increases. Here we only report the results for $\nu = 5$, which is close to the empirical findings for high-frequency asset returns in the literature. The \tilde{Q}_1 test has better power under DGP4 than under DGP5, with the rejection rates approaching unity when $T = 2500$. This is expected because χ^2 is a skewed distribution. We also conjecture that part of the heavy tail generated by the t or χ^2 distribution has been captured by the GARCH model, which complicates the detection of the misspecification in the distribution. Under DGP 6, ε_t is generated from a time-varying skewed Student's t distribution, and so \tilde{Q}_1 , which is sensitive to the shape of the distribution, is expected to have better power than under DGP5. This is indeed confirmed in our simulation, with the rejection rates approaching unity when $T = 2500$ under DGP6.

Table 3 reports the rejection rates of \hat{Q}_1 under DGPs 1-6 at the 10% and 5% levels respectively. The general patterns are similar to those of \tilde{Q}_1 , with the rejection rates increasing significantly with the sample size T . Although \hat{Q}_1 has higher rejection rates under DGPs 1, 3 and 5, the overall performances of \tilde{Q}_1 and \hat{Q}_1 are close to each other. But in terms of the computational cost, \tilde{Q}_1 is much less time-consuming than \hat{Q}_1 , because a $4d$ dimensional integration is reduced to a $2d$ dimensional integration in calculating \tilde{Q}_1 . We thus suggest using \tilde{Q}_1 in practice.

6.2 Bivariate Distribution models

6.2.1 Simulation Design

To examine the size of our tests for multivariate distributional models, we consider the following bivariate DGP:

DGP B0 [AR(1)-BGARCH(1,1)-BN($\mathbf{0}, \mathbf{I}$)]

$$\begin{cases} X_{1t} = 0.3X_{1t-1} + u_{1t}, \\ X_{2t} = 0.2X_{2t-1} + u_{2t}, \end{cases} \quad (6.8)$$

$$\mathbf{u}_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \mathbf{H}_t^{1/2} \boldsymbol{\varepsilon}_t,$$

where $\mathbf{H}_t = \begin{bmatrix} H_{11t} & H_{12t} \\ H_{21t} & H_{22t} \end{bmatrix}$, $\boldsymbol{\varepsilon}_t \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$, and

$$\begin{cases} H_{11t} = 0.05 + 0.09u_{1t-1}^2 + 0.8H_{11t-1}^2, \\ H_{22t} = 0.3 + 0.11u_{2t-1}^2 + 0.7H_{22t-1}^2, \\ \rho = \frac{H_{12t}}{\sqrt{H_{11t}H_{22t}}} = \frac{H_{21t}}{\sqrt{H_{11t}H_{22t}}} = 0.2. \end{cases}$$

DGP B0 is a bivariate Gaussian GARCH model with a constant conditional correlation. The volatilities of two components are not dynamically related but they are contemporaneously correlated. Similar to the univariate case, we simulate 1,000 data sets of $\{\mathbf{X}_t\}_{t=1}^T$ for $T = 100, 250, 500$ and 1,000 respectively. For each data set, we estimate the model parameters via MLE.

To investigate the power of our tests for multivariate models, we consider the following DGPs:

DGP B1 [DCC]:

The conditional mean and the dynamics of H_{11t} and H_{22t} are the same as DGP B0 but with time-varying conditional correlation:

$$\begin{aligned} \mathbf{H}_t &= \begin{bmatrix} \sqrt{H_{11t}} & 0 \\ 0 & \sqrt{H_{22t}} \end{bmatrix} \mathbf{R}_t \begin{bmatrix} \sqrt{H_{11t}} & 0 \\ 0 & \sqrt{H_{22t}} \end{bmatrix}, \\ \mathbf{Q}_t &= 0.1\mathbf{R}_0 + 0.7(\mathbf{R}_t^{1/2}\boldsymbol{\varepsilon}_{t-1})(\mathbf{R}_t^{1/2}\boldsymbol{\varepsilon}_{t-1})' + 0.2\mathbf{Q}_{t-1}, \\ \mathbf{R}_t &= \text{diag}(\mathbf{Q}_t)^{-1} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1}, \end{aligned} \quad (6.9)$$

where $\mathbf{R}_0 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$, $\text{diag}(\cdot)$ denotes the diagonals of a matrix.

DGP B2 [Granger causality in mean]:

$$\begin{cases} X_{1t} = 0.3X_{1t-1} + u_{1t} + 0.3X_{2t-1}, \\ X_{2t} = 0.2X_{2t-1} + u_{2t}, \end{cases} \quad (6.10)$$

where \mathbf{u}_t has the same dynamics as that of DGP B0.

DGP B3 [Granger causality in variance]: The conditional mean dynamics has the same forms as DGP B0.

$$\begin{cases} H_{11t} = 0.05 + 0.15u_{1t-1}^2 + 0.8H_{11t-1}^2 + 0.3u_{2t-1}^2, \\ H_{22t} = 0.5 + 0.2u_{2t-1}^2 + 0.5H_{22t-1}^2 + 0.3u_{1t-1}^2, \\ H_{12t} = H_{21t} = 0.3\sqrt{H_{11t}H_{22t}}. \end{cases} \quad (6.11)$$

DGP B4 [Granger causality in distribution]: The conditional mean and variance have the same forms as DGP B0, with

$$\varepsilon_{lt} \sim p(\varepsilon_l | \nu_{lt}, \lambda_{lt}), \quad (6.12)$$

where $l = 1, 2$ and $p(\cdot | \cdot, \cdot)$ is Hansen's (1994) time-varying skewed Student's t distribution, whose degrees of freedom ν_{lt} and skew parameter λ_{lt} change over time as

$$\begin{cases} \lambda_{lt} = \delta_{l1} + \delta_{l2}u_{1t-1} + \delta_{l3}u_{2t-1} + \delta_{l4}\lambda_{1t-1} + \delta_{l5}\lambda_{2t-1}, \\ \nu_{lt} = \tau_{l1} + \tau_{l2}u_{1t-1} + \tau_{l3}u_{2t-1} + \tau_{l4}\nu_{1t-1} + \tau_{l5}\nu_{2t-1}, \end{cases}$$

where

$$\begin{cases} (\delta_{11}, \delta_{12}, \delta_{13}, \delta_{14}, \delta_{15}, \tau_{11}, \tau_{12}, \tau_{13}, \tau_{14}, \tau_{15}) = (-0.2, 1, -5, 0, -0.9, -0.2, 1, -5, 0, -0.9), \\ (\delta_{21}, \delta_{22}, \delta_{23}, \delta_{24}, \delta_{25}, \tau_{21}, \tau_{22}, \tau_{23}, \tau_{24}, \tau_{25}) = (-0.2, 0, 1, 0, 0, -0.2, 0, 1, 0, 0). \end{cases}$$

DGP B1 is Engle's (2002) DCC model. A logistic function is used to bound conditional correlation ρ_t between -1 and 1. Under DGP B1, model (6.8) is correctly specified for the conditional mean and conditional variance but misspecified for the conditional correlation. Specifically, model (6.8) assumes a constant conditional correlation, while under DGP B1, there exists a time-varying conditional correlation. Under DGP B2, there exists Granger causality in mean from X_{2t} to X_{1t} as the conditional mean of X_{1t} is determined by both X_{1t-1} and X_{2t-1} . If we use model (6.8) to fit data generated from DGP B2, the conditional mean of X_{1t} is misspecified but the conditional variances and higher moments

are correctly specified. Under DGP B3, the conditional variances of X_{1t} and X_{2t} are misspecified because they fail to capture Granger causality in variance from both directions. This model can be used to characterize volatility spillover between different financial markets. Under DGP B4, model (6.8) is correctly specified for the conditional mean and variance but is misspecified for the distribution of ε_t as it ignores Granger causality in higher moments from X_{2t} to X_{1t} .

Similar to the univariate case, for each of DGPs B1-B4, we generate 500 data sets of the random sample $\{\mathbf{X}_t\}_{t=1}^T$ for $T = 250, 500$ and $1,000$ respectively. For each data set, we estimate model (6.8) via MLE and check power performances. For computational simplicity, we just focus on \tilde{Q}_1 in bivariate cases.

6.2.2 Monte Carlo Evidence

To reduce computational costs, we generate $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ from an $N(\mathbf{0}, \mathbf{I}_2)$ distribution, with each $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ having 15 grid points in \mathbb{R}^2 respectively, and let $\mathbf{x} = (\hat{\mathbf{x}}', -\hat{\mathbf{x}})'$ and $\mathbf{y} = (\hat{\mathbf{y}}', -\hat{\mathbf{y}})'$ to ensure their symmetry. The choices of $k(\cdot)$, \hat{p} and \bar{p} are the same as in the univariate case.

Table 4 reports the rejection rates of \tilde{Q}_1 under DGPs B0-B4 at the 10% and 5% significance levels. Model (6.8) is correctly specified under DGP B0. The \tilde{Q}_1 test tends to overreject a little when $T = 100$, but the overrejection is not excessive and becomes weaker when the sample size increases. We conjecture that the overrejection is due to the estimation uncertainty of small samples. Similar to the univariate case, the size of \tilde{Q}_1 is not sensitive to the choice of the preliminary lag order \bar{p} .

Under DGP B1, model (6.8) ignores the time-varying conditional correlation. The \tilde{Q}_1 test has good power in detecting such misspecification in the conditional correlation. The rejection rate is around 16% at the 5% level when the sample size T is as small as 100, and increases significantly with the sample size. Under DGP B2, model (6.8) ignores the Granger causality in mean from X_{2t} to X_{1t} . The \tilde{Q}_1 test has excellent power when model (6.8) is used to fit data generated from DGP B2. The rejection rate is around 25% at the 5% level when $T = 100$ and approaches unity when $T = 1,000$. Under DGP B3, model (6.8) ignores the Granger causality in variance from both directions. The \tilde{Q}_1 test has good power and the rejection rate approaches 85% at the 5% level when $T = 1000$. Under DGP B4, model (6.8) ignores the Granger causality in distribution. Since misspecification only exists in higher moments, we conjecture that it may be difficult to be captured. However, our \tilde{Q}_1 test has rather good power when model (6.8) is used to fit the data generated from DGP B4. The rejection rate increases significantly with the sample size and approaches 80% at the 5% level when $T = 1,000$.

To sum up, we observe:

- Both GCM tests \hat{Q}_1 and \tilde{Q}_1 have reasonable sizes for sample sizes as small as $T = 100$. The sizes of tests are robust to the choice of a preliminary lag order.
- Both \hat{Q}_1 and \tilde{Q}_1 have good omnibus powers in detecting various model misspecifications, which demonstrates the nice feature of the proposed indicator function approach embedded in a frequency domain framework. Although the powers may vary with the degree of discrepancy between the null and the alternative models, the power performances are satisfactory for sample sizes often encountered in finance and economics.

- The finite sample performances of \hat{Q}_1 and \tilde{Q}_1 are close to each other under various univariate and bivariate alternatives but the computational costs differ. The test statistic \tilde{Q}_1 is computationally more efficient.

7. CONCLUSION

Conditional distribution models in time series have becoming increasingly important in studying various applications in economics and finance, such as macroeconomic control, asset allocation, option pricing, risk management and hedging. We propose a new class of GCM tests for dynamic conditional distribution models in time series, where the conditional information set may depend on the entire history of the data. Thanks to the use of the empirical distribution function embedded in a frequency domain framework, both univariate and multivariate conditional distribution models are covered in a unified framework and our GCM tests can detect a variety of linear and nonlinear misspecifications. Our frequency domain approach can check a large number of lags without suffering from the curse of dimensionality, and naturally discount higher order lags. When applied to multivariate conditional distribution models, our tests can fully exploit the information in the joint dynamics of variables and thus can capture misspecification in modeling joint dynamics, which may be easily missed by existing procedures. Our tests are applicable to both discrete and continuous distributions. They are supplemented by a class of diagnostic procedures, which are obtained by integrating the CDF and focus on various specific aspects of the dynamics such as whether there exists neglected structures in conditional mean, conditional variance, conditional correlation, conditional skewness and conditional kurtosis respectively. Such information is useful for practitioners in reconstructing a misspecified model. Unlike the traditional CM and KS tests, which also use the empirical distribution function but have nonstandard distributions, our test statistics all follow a convenient asymptotic $N(0, 1)$ distribution, and they are applicable to various estimation methods, including suboptimal but consistent estimators. Moreover, parameter estimation uncertainty has no impact on the asymptotic distribution of the test statistics. Simulation studies show that the proposed tests perform reasonably well in finite samples.

REFERENCES

- Ait-Sahalia, Y., J. Fan and H. Peng (2009): "Nonparametric Transition-Based Tests for Diffusions," *Journal the of American Statistical Association*, forthcoming.
- Andersen, T. and J. Lund (1997): "Estimating Continuous-time Stochastic Volatility Models of the Short-term Interest Rate," *Journal of Econometrics*, 77, 343-377.
- Andrews D. (1997): "A Conditional Kolmogorov Test," *Econometrica*, 65, 1097-1128.
- Ang, A. and J. Chen (2002): "Asymmetric Correlations of Equity Portfolios," *Journal of Financial Economics*, 63, 443-494.
- Bai, J. (2003): "Testing Parametric Conditional Distributions of Dynamic Models," *Review of Economics and Statistics*, 85, 531-549.
- Bai, J. and Z. Chen (2008): "Testing Multivariate Distributions in GARCH Models," *Journal of Econometrics*, 143, 19-36.
- Barberis, N. and M. Huang (2001): "Mental Accounting, Loss Aversion, and Individual Stock Returns," *Journal of Finance*, 56, 1247-1292.

- Barberis, N., M. Huang and T. Santos (2001): "Prospect Theory And Asset Prices," *The Quarterly Journal of Economics*, 116, 1-53.
- Bauwens, L. and S. Laurent (2002): "A New Class of Multivariate Skew Densities, with Application to GARCH Models," working paper, CORE.
- Bhardwaj, G., V. Corradi, and N. R. Swanson (2008) "A Simulation Based Specification Test for Diffusion Processes", *Journal of Business and Economic Statistics*, 26, 176-193.
- Black, F. and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637-654.
- Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307-327.
- Bollerslev, T., R.F. Engle. and J.M. Wooldridge (1988): "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy* 96, 116-131.
- Bond, S.A. (2001): "A Review of Asymmetric Conditional Density Functions in Autoregressive Conditional Heteroscedasticity Models," *Return Distributions in Finance*, Oxford, UK.
- Bowsher, C. (2007): "Modelling Security Market Events in Continuous Time: Intensity Based, Multivariate Point Process Models," *Journal of Econometrics*, 141, 867-912.
- Brooks, C., S. Burke and G. Persaud (2005): "Autoregressive Conditional Kurtosis," *Journal of Financial Econometrics*, 3, 399-421.
- Christoffersen, P. F. and F. Diebold (1997): "Optimal Prediction Under Asymmetric Loss," *Econometric Theory*, 13, 808-817.
- Chunhachinda, P., K. Dandapani, S. Hamid, and A. Prakash. (1997): "Portfolio Selection and Skewness: Evidence from International Stock Markets," *Journal of Banking and Finance*, 21, 143-167.
- Corradi, V. and N. Swanson (2003): "Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification," Working paper, Rutgers University.
- (2006): "Predictive Density Evaluation," *Handbook of Economic Forecasting*, eds. Granger, C., G. Elliot and A. Timmerman, Elsevier, Amsterdam, pp. 197-284.
- Corrado C. and T. Su (1996): "S&P 500 Index Option Tests of Jarrow and Rudd's Approximate Option Valuation Formula," *Journal of Futures Markets*, 16, 611-629.
- (1997): "Implied Volatility Skews and Stock Index Skewness and Kurtosis Implied by S&P 500 Index Option Prices," *Journal of Derivatives*, 4, 8-19.
- Dai, Q., A. Le and J. Singleton (2005): "Discrete-time Dynamic Term Structure Models with Generalized Market Prices of Risk," working paper, University of North Carolina at Chapel Hill.
- Davies, R. B. (1977): "Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 64, 247-254.
- (1987): "Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 1A, 33-43.
- DeJong, D. and C. Dave (2007): *Structural Macroeconometrics*, Princeton University Press, New Jersey.
- Diebold, F. X., T. Gunther, and A. Tay (1998): "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863-883.
- Duffie, D. and J. Pan (1997): "An Overview of Value at Risk," *Journal of Derivatives*, 7-49.
- Engle, R.F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of

- United Kingdom Inflation," *Econometrica*, 50, 987-1007.
- (2002a): "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business and Economic Statistics*, 20, 339-350.
- (2002b): "New Frontiers for ARCH Models," *Journal of Applied Econometrics*, 17, 425-446.
- Engle, R.F. and K.F. Kroner (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122-150.
- Engle, R.F. and Gonzalez-Rivera, G. (1991): "Semiparametric ARCH Models," *Journal of Business and Economic Statistics*, 9, 345-359.
- Engle, R.F., Lilien, D.M. and Robins, R.P. (1987): "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391-407.
- Engle, R.F. and J.R. Russell (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127-1162.
- (2005): "A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The Autoregressive Conditional Multinomial-Autoregressive Conditional Duration Model," *Journal of Business & Economic Statistics*, 23, 166-180.
- Fan, Y., Q. Li and I. Min (2006): "A Nonparametric Bootstrap Test of Conditional Distributions," *Econometric Theory*, 22, 587-613.
- Gallant, A.R., D. Hsieh and G. Tauchen (1997): "Estimation of Stochastic Volatility Models with Diagnostics," *Journal of Econometrics*, 81, 159-192.
- Geweke, J. and G. Amisano (2007): "Hierarchical Markov Normal Mixture Models with Applications to Financial Asset Returns," Working Paper, University of Brescia.
- Granger, C. (1999): *Empirical Modeling in Economics*, Cambridge University Press, Cambridge.
- (2003): "Time Series Concepts for Conditional Distributions," *Oxford Bulletin of Economics and Statistics*, 65, 639-701.
- Granger, C. and M.H. Pesaran (2000): "Economic and Statistical measures of Forecast Accuracy," *Journal of Forecasting*, 19, 537-560.
- Granger, C., T. Teravirta and A. Patton (2006): "Common Factors in Conditional Distributions," *Journal of Econometrics*, 132, 43-57.
- Hamilton, J. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.
- (1990): "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, 39-70.
- (1994): *Time Series Analysis*, Princeton University Press, New Jersey.
- Hansen, B.E. (1994): "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705-730.
- (1996): "Inference When A Nuisance Parameter is Notidentified Under the Null Hypothesis," *Econometrica*, 413-430.
- Harvey, C.R. and Siddique, A. (1999): "Autoregressive Conditional Skewness," *Journal of Financial and Quantitative Analysis*, 34, 465-487.
- (2000): "Conditional Skewness in Asset Pricing," *Journal of Finance*, 55, 1263-1295.

- Hoeffding, W. (1948): "A Nonparametric Test of Independence," *Annals of Mathematical Statistics* 58, 546-557.
- Hong, Y. (1998): "Testing for Pairwise Independence via the Empirical Distribution Function," *Journal of the Royal Statistical Society, Ser. B*, 60, 429-453.
- (1999): "Hypothesis Testing in Time Series via the Empirical Characteristic Function: a Generalized Spectral Density Approach," *Journal of the American Statistical Association*, 94, 1201-1220.
- Hu, L. (2006): "Dependence Patterns across Financial Markets: A Mixed Copula Approach," *Applied Financial Economics*, 16, 717-729.
- Jarrow R. and A. Rudd (1982): "Approximate Option Valuation for Arbitrary Stochastic Processes," *Journal of Financial Economics*, 10, 347-369.
- Khmaladze, E.V. (1981): "Martingale Approach in the Theory of Goodness-of-fit Tests," *Theory of Probability and its Application*, 26, 240-257.
- Kraus, A. and R.H. Litzenberger (1976): "Skewness Preference and the Valuation of Risk Assets," *Journal of Finance*, 31, 1085-1100.
- Lai, T.Y., (1991): "Portfolio Selection with Skewness: A Multiple-objective Approach," *Review of Quantitative Finance and Accounting*, 1, 293-305.
- Lambert, P. and Laurent, S. (2001): "Modelling Financial Time Series Using GARCH-type Models and a Skewed Student Density," *Mimeo*, Banque de France and HEC School of Management.
- Lee, T.H. and X. Long (2009): "Copula-based Multivariate GARCH Model with Uncorrelated Dependent Errors," *Journal of Econometrics*, 150, 207-218.
- Li, F. and G. Tkacz (2006): "A Consistent Test for Conditional Density Functions with Time Dependent Data," *Journal of Econometrics*, 133, 863-886.
- Lien, D. and Y.K. Tse, (2002): "Some Recent Developments in Futures Hedging," *Journal of Economic Surveys*, 16, 357-396.
- Linton, O. and P. Gozalo (1997): "Conditional Independence Restrictions: Testing and Estimation," Discussion Paper, Cowles Foundation for Research in Economics, Yale University.
- Longin, F. and B. Solnik (2001): "Extreme Correlation of International Equity Markets," *Journal of Finance*, 56, 649-676.
- McNeil, A.J. and R. Frey (2000): "Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach," *Journal of Empirical Finance*, 271-300.
- Mills, F.C. (1927): "The Behavior of Prices," National Bureau of Economic Research, New York.
- Nelson, D.B. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.
- Patton, A. (2004): "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation," *Journal of Financial Econometrics*, 2, 130-168.
- Premaratne, G. and A. Bera (2001): "Modeling Asymmetry and Excess Kurtosis in Stock Return Data," working paper, University of Illinois.
- Rubinstein M. (1998): "Edgeworth Binomial Trees," *Journal of Derivatives*, 5, 20-27.
- Rothschild, M. and J. Stiglitz (1971): "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2, 225-243.
- (1972): "Increasing Risk II: Its Economic Consequences," *Journal of Economic Theory*, 3, 66-84.

- Shephard, N. (2005): *Stochastic Volatility: Selected Readings*. Oxford University Press, London.
- Stinchcombe, M. and H. White (1998): "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14, 295-324.
- Taylor, S. (1986): *Modelling Financial Times Series*, John Wiley & Sons Ltd, Chichester.
- Tsay, R. (2005): *Analysis of Financial Time Series*, John Wiley & Sons, New Jersey.
- Tse, Y.K. and A.K. Tsui (2002): "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-varying Correlations," *Journal of Business and Economic Statistics*, 20, 351-362.
- Zheng, X. (2000): "A Consistent Test of Conditional Parametric Distributions," *Econometric Theory*, 16, 667-691.

Table 1. Sizes of specification tests under DGP0

T	Lag order	100		250		500		1000	
α		.10	.05	.10	.05	.10	.05	.10	.05
Tests based on the covariance between the generalized residual and its lag term									
\tilde{Q}_1	10	.063	.038	.062	.040	.067	.039	.058	.037
	20	.090	.052	.093	.056	.089	.062	.074	.042
	30	.108	.060	.105	.063	.102	.069	.078	.043
	40	.105	.072	.107	.063	.110	.065	.076	.044
Tests based on the covariance between the generalized residual and the lag indicator function									
\hat{Q}_1	10	.076	.043	.065	.043	.078	.052	.062	.037
	20	.061	.038	.065	.037	.077	.048	.070	.037
	30	.056	.034	.068	.040	.074	.045	.070	.035
	40	.053	.029	.067	.034	.068	.044	.073	.034

Notes: (1) DGP0 is: $X_t = u_t + 0.5u_{t-1}$, $u_t = h_t^{1/2}\varepsilon_t$, $h_t = 0.05 + 0.15X_{t-1}^2 + 0.8h_{t-1}$, where $\varepsilon_t \sim \text{i.i.d.N}(0,1)$;

(2) \hat{Q}_1 and \tilde{Q}_1 are tests based on the covariance between the generalized residual and the lag indicator function and tests based on the covariance between the generalized residual and its lag term, given in equation (3.11) and (3.13) respectively;

(3) 1000 iterations.

Table 2. Powers of \tilde{Q}_1 under DGPs1-6

T	Lag order	250		500		1000		2500	
α		.10	.05	.10	.05	.10	.05	.10	.05
DGP1 [ARMA -GARCH -N(0,1)]	10	.426	.330	.838	.762	.994	.992	1.00	1.00
	20	.356	.274	.758	.690	.988	.968	1.00	1.00
	30	.306	.238	.686	.608	.966	.936	1.00	1.00
	40	.278	.218	.644	.548	.942	.914	1.00	1.0
DGP2 [EGARCH -N(0,1)]	10	.812	.730	.994	.986	1.00	1.00	1.00	1.00
	20	.676	.558	.978	.958	1.00	1.00	1.00	1.00
	30	.564	.452	.952	.924	1.00	1.00	1.00	1.00
	40	.480	.348	.918	.882	1.00	1.00	1.00	1.00
DGP3 [MA - GARCHK]	10	.040	.030	.070	.042	.142	.090	.400	.294
	20	.046	.030	.070	.038	.100	.054	.260	.168
	30	.050	.034	.066	.034	.082	.044	.182	.126
	40	.054	.036	.062	.030	.070	.036	.152	.092
DGP4 [MA -GARCH -Chi(5)]	10	.220	.154	.416	.318	.840	.780	1.00	1.00
	20	.180	.126	.278	.184	.670	.570	1.00	1.00
	30	.150	.110	.220	.140	.546	.434	.996	.994
	40	.144	.110	.186	.120	.464	.344	.986	.968
DGP5 [MA -GARCH -t(5)]	10	.066	.034	.096	.054	.138	.074	.464	.364
	20	.054	.032	.060	.030	.078	.048	.296	.210
	30	.054	.034	.042	.068	.044	.069	.204	.144
	40	.050	.030	.040	.054	.026	.065	.148	.102
DGP6 [MA -GARCH -time varying t]	10	.152	.110	.322	.236	.708	.616	1.00	1.00
	20	.124	.080	.224	.148	.548	.400	.994	.980
	30	.108	.070	.162	.106	.394	.278	.958	.920
	40	.098	.058	.134	.082	.314	.214	.916	.838

Notes: (1) \tilde{Q}_1 is based on the covariance between the generalized residual and its lag term, given in equation (3.13);

(2) 500 iterations.

Table 3. Powers of \hat{Q}_1 under DGPs 1-6

T	Lag order	250		500		1000		2500	
α		.10	.05	.10	.05	.10	.05	.10	.05
DGP1 [ARMA -GARCH -N(0,1)]	10	.636	.552	.882	.836	.994	.988	1.00	1.00
	20	.548	.438	.806	.752	.982	.980	1.00	1.00
	30	.486	.390	.762	.714	.978	.964	1.00	1.00
	40	.444	.362	.748	.672	.966	.948	1.00	1.00
DGP2 [EGARCH -N(0,1)]	10	.686	.580	.940	.906	.998	.998	1.00	1.00
	20	.584	.488	.900	.832	.996	.992	1.00	1.00
	30	.512	.412	.856	.784	.994	.986	1.00	1.00
	40	.452	.340	.820	.736	.990	.978	1.00	1.00
DGP3 [MA - GARCHK]	10	.108	.058	.132	.098	.228	.160	.444	.358
	20	.094	.050	.103	.069	.218	.136	.362	.272
	30	.074	.044	.104	.061	.188	.122	.328	.238
	40	.066	.040	.104	.059	.168	.114	.304	.224
DGP4 [MA -GARCH -Chi(5)]	10	.206	.136	.404	.292	.746	.666	.986	.984
	20	.160	.102	.302	.212	.616	.534	.974	.962
	30	.142	.082	.258	.174	.542	.438	.956	.926
	40	.126	.068	.224	.140	.482	.378	.936	.874
DGP5 [MA -GARCH -t(5)]	10	.116	.072	.146	.106	.250	.188	.504	.402
	20	.096	.050	.124	.074	.232	.150	.426	.326
	30	.084	.044	.108	.066	.200	.138	.392	.274
	40	.076	.038	.094	.064	.180	.112	.356	.268
DGP6 [MA -GARCH -time varying t]	10	.222	.150	.328	.244	.694	.558	.988	.972
	20	.182	.133	.258	.176	.552	.424	.964	.918
	30	.172	.112	.240	.150	.478	.336	.902	.830
	40	.157	.083	.206	.136	.438	.302	.844	.778

Notes: (1) \hat{Q}_1 is based on the covariance between the generalized residual and the lag indicator function, given in equation (3.11);

(2) 500 iterations.

Table 4. Size and powers of \tilde{Q}_1 under DGPs B0-B4

T	Lag order	100		250		500		1000	
α		.10	.05	.10	.05	.10	.05	.10	.05
Size									
DGPB0 AR(1)- BGARCH (1,1)- BN(0,1)	10	.101	.068	.085	.061	.088	.052	.074	.038
	20	.136	.087	.119	.077	.102	.064	.083	.048
	30	.156	.121	.127	.085	.103	.070	.087	.051
	40	.169	.128	.136	.090	.112	.079	.095	.055
Powers									
DGPB1 (DCC)	10	.220	.140	.442	.342	.644	.554	.894	.852
	20	.170	.118	.386	.260	.558	.466	.848	.810
	30	.146	.108	.316	.236	.504	.388	.812	.758
	40	.144	.090	.300	.212	.460	.356	.780	.714
DGPB2 (Granger causality in mean)	10	.506	.366	.930	.876	.994	.992	1.00	1.00
	20	.412	.250	.870	.810	.978	.954	1.00	1.00
	30	.314	.206	.828	.774	.958	.922	1.00	1.00
	40	.276	.158	.798	.742	.928	.876	1.00	1.00
DGPB3 (Granger causality in variance)	10	.202	.112	.652	.625	.792	.778	.834	.830
	20	.218	.126	.680	.657	.820	.810	.848	.844
	30	.220	.116	.696	.677	.838	.828	.860	.856
	40	.218	.118	.712	.690	.844	.836	.866	.866
DGPB4 (Granger causality in distribution)	10	.224	.130	.436	.296	.720	.614	.972	.930
	20	.174	.096	.288	.190	.570	.446	.902	.814
	30	.158	.094	.234	.162	.490	.344	.820	.738
	40	.144	.076	.206	.144	.424	.278	.744	.656

Notes: (1) \tilde{Q}_1 is based on the covariance between the generalized residual and its lag term, given in equation (3.13);

(2) Results of DGP B0 are based on 1000 iterations; results of DGPs B1-B4 are based on 500 iterations.