ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A unified approach to validating univariate and multivariate conditional distribution models in time series

Bin Chen[a,*], Yongmiao Hong[b,c,d,e,**]

[a] Department of Economics, University of Rochester, Rochester, NY, 14627, United States
[b] Department of Economics, Cornell University, Ithaca, NY 14850, United States
[c] Department of Statistical Science, Cornell University, Ithaca, NY 14850, United States
[d] Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Xiamen 361005, China
[e] MOE Key Laboratory in Econometrics, Xiamen University, Xiamen 361005, China

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C4
G0

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

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.

1. Introduction

The modeling of conditional distributions in time series has been advancing rapidly, with a wide range of applications in economics and finance (e.g. Granger, 1999; Corradi and Swanson, 2006b). 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 economics and econometrics, effort has been devoted to using higher moments and the entire distribution. Rothschild and Stiglitz's (1971, 1972) seminal works have demonstrated that the risk or uncertainty should be characterized by the distribution function, rather than the first two moments. In particular, the

* Corresponding author.
** Corresponding author at: Department of Economics, Cornell University, Ithaca, NY 14850, United States.
E-mail addresses: bchen8@mail.rochester.edu (B. Chen), yh20@cornell.edu (Y. Hong).

[1] Empirical 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).

[2] A 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.

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.08.004

ranking of the cumulative distribution function (CDF) by certain rules always coincides with that of the risk-averter'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]

In time series analysis, the most popular models are ARMA models for conditional mean and 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''. 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 (GARCHS) in a conditional non-central t-distribution framework by explicitly modeling the conditional second and third moments jointly. Brooks et al. (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 (1998, 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 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.[5] For example, CAPM studies the relationship between individual asset returns and the market return, which has motivated the development of multivariate GARCH models (e.g., Bollerslev et al., 1988, Engle, 2002b). 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., Clements and Krolzig, 2003, Chauvet and Hamilton, 2006). 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 dates back to the Kolmogorov–Smirnov (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 et al. (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 (2006a) 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. Bierens and Wang (2012) propose a weighted integrated conditional moment (ICM) test of the validity of parametric specifications of conditional distribution models for stationary time series, extending Bierens' (1984) test. Their ICM test is consistent against all stationary alternatives, but its asymptotic distribution is case dependent and a bootstrap method has to be applied to obtain critical values, which is computationally intensive.

In a continuous-time diffusion framework, Ait-Sahalia et al. (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 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 et al. (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

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

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

[5] Geweke and Amisano (2007) argue that "while univariate models are a first step, there is an urgent need to move on to multivariate modeling of the time-varying distribution of asset returns''.

Please cite this article in press as: Chen, B., Hong, Y., A unified approach to validating univariate and multivariate conditional distribution models in time series. Journal of Econometrics (2013), http://dx.doi.org/10.1016/j.jeconom.2013.08.004

of the empirical distribution function, our approach can detect a variety of linear and nonlinear functional form misspecifications. Unlike the time domain approach (e.g., Corradi and Swanson, 2006a), our frequency domain approach can check a growing number of lags as the sample size increases without suffering severely 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 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. 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.

Third, 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. We derive the asymptotic distribution of the proposed test statistic in Section 4, and discuss its asymptotic power property in Section 5. 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

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

where $\boldsymbol{\mu}(\mathcal{I}_{t-1}, \boldsymbol{\theta})$ is a parametric model for $E(\mathbf{X}_t|\mathcal{I}_{t-1})$, $\mathbf{h}\left(\mathcal{I}_{t-1}, \boldsymbol{\theta}\right)$ is a parametric model for $\mathrm{var}(\mathbf{X}_t|\mathcal{I}_{t-1})$, $\boldsymbol{\varepsilon}_t$ has the conditional CDF $P_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathcal{I}_{t-1}, \boldsymbol{\theta})$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ 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 $P_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathcal{I}_{t-1}, \boldsymbol{\theta})$, it is straightforward to calculate the conditional CDF of $\mathbf{X}_t$

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

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 ε_t has a continuous distribution with the conditional PDF

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

where $\boldsymbol{\alpha}(\mathcal{I}_{t-1}, \boldsymbol{\theta}) = [\boldsymbol{\mu}(\mathcal{I}_{t-1}, \boldsymbol{\theta}), \mathbf{h}\left(\mathcal{I}_{t-1}, \boldsymbol{\theta}\right), \boldsymbol{\lambda}(\mathcal{I}_{t-1}, \boldsymbol{\theta}), \boldsymbol{\nu}(\mathcal{I}_{t-1}, \boldsymbol{\theta})]'$ 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

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

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

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

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 $\boldsymbol{\mu}(\mathcal{I}_{t-1}, \boldsymbol{\theta}) = 0$, $\mathbf{h}\left(\mathcal{I}_{t-1}, \boldsymbol{\theta}\right)$ 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

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

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 Engle and Russell'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,

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

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

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

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

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

where s is the number of states, $\tilde{y}_t$ takes the jth column of the $s \times s$ identity matrix if the jth 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 $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ such that

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

Alternatively, if for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, 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 $U_{t1}(\boldsymbol{\theta}_0) \equiv P_t(X_{t1}|\mathcal{I}_{t-1}, \boldsymbol{\theta}_0), U_{t2}(\boldsymbol{\theta}_0) \equiv P_t(X_{t2}|X_{t1}, \mathcal{I}_{t-1}, \boldsymbol{\theta}_0), \ldots, U_{td}(\boldsymbol{\theta}_0) \equiv P_t(X_{td}|X_{t1}, \ldots, X_{td-1}, \mathcal{I}_{t-1}, \boldsymbol{\theta}_0)$. 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

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

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

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

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 $t \to \infty$, 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:

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

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, Andersen and Lund, 1997 and Gallant et al., 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:

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

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 $\mathbf{x} = \left(\mathbf{x}'_1, \mathbf{x}'_2\right)'$. Let

$$\begin{aligned} P(\mathbf{x}_1|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}) &\equiv E_{\boldsymbol{\theta}}[1(\mathbf{X}_{1,t} \leq \mathbf{x}_1)|\mathcal{I}_{1,t-1}] \\ &= E_{\boldsymbol{\theta}}\{P[(\mathbf{x}'_1, \mathbf{0}')'|\mathcal{I}_{t-1}, \boldsymbol{\theta}]|\mathcal{I}_{1,t-1}\}, \end{aligned}$$

where $\mathcal{I}_{1,t-1} = \left\{\mathbf{X}_{1,t-1}, \mathbf{X}_{1,t-2}, \ldots, \mathbf{X}_{1,1}\right\}$ 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

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

Under $\mathbb{H}_0$, we have

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

This provides a basis for constructing operational tests for time series models with partially observable variables. Although the model-implied conditional distribution $P\left(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta}\right)$, where $\mathcal{I}_{t-1} = (\mathcal{I}_{1,t-1}, \mathcal{I}_{2,t-1})$, may have a closed-form, the conditional distribution $P(\mathbf{x}_1|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$ generally has no closed-form. However, one can approximate it accurately by using some simulation techniques. For both state-space models and SV models, the conditional distribution $P\left(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta}\right) = P\left(\mathbf{x}|\mathbf{X}_{t-1}, \boldsymbol{\theta}\right)$ is a Markov process. In this case,

$$\begin{aligned} &E_{\boldsymbol{\theta}}\{P[(\mathbf{x}'_1, \mathbf{0}')'|\mathcal{I}_{t-1}, \boldsymbol{\theta}]|\mathcal{I}_{1,t-1}\} \\ &\quad = \int P[(\mathbf{x}'_1, \mathbf{0}')'|\mathbf{X}_{1,t-1}, \mathbf{x}_{2,t-1}, \boldsymbol{\theta}]p(\mathbf{x}_{2,t-1}|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})d\mathbf{x}_{2,t-1}, \end{aligned}$$

where $p(\mathbf{x}_{2,t-1}|\mathcal{I}_{1,t-1},\boldsymbol{\theta})$ is the model-implied transition density of the unobservable $\mathbf{X}_{2,t-1}$ given the observable information $\mathcal{I}_{1,t-1}$. We could use particle filters to estimate the model-implied conditional distribution based on observables. The term "particle" was first used by Kitagawa (1996) in this literature to denote the simulated discrete data with random support. Particle filters are the class of simulation filters that recursively approximate the filtering random variable $\mathbf{x}_{2,t-1}|\mathcal{I}_{1,t-1},\boldsymbol{\theta}$ by "particles" $\hat{\mathbf{X}}^1_{2,t-1},\hat{\mathbf{X}}^2_{2,t-1},\ldots,\hat{\mathbf{X}}^J_{2,t-1}$ with discrete probability mass of $\pi^1_{t-1},\pi^2_{t-1},\ldots,\pi^J_{t-1}$ (see, e.g. Gordon et al., 1993; Pitt and Shephard, 1999). Hence a continuous variable is approximately a discrete one with random support. These discrete points are viewed as samples from $p\left(\mathbf{x}_{2,t-1}|\mathcal{I}_{1,t-1},\boldsymbol{\theta}\right)$ and as $J\to\infty$, the particles can approximate the conditional density increasingly well.

The key of this method is to propagate particles $\{\hat{\mathbf{X}}^j_{2,t-2}\}^J_{j=1}$ one step forward to get the new particles $\{\hat{\mathbf{X}}^j_{2,t-1}\}^J_{j=1}$. By the Bayes rule, we have

$$p\left(\mathbf{x}_{2,t-1}|\mathcal{I}_{1,t-1},\boldsymbol{\theta}\right)=\frac{p\left(\mathbf{x}_{1,t-1}|\mathbf{x}_{2,t-1},\mathcal{I}_{1,t-2},\boldsymbol{\theta}\right)p\left(\mathbf{x}_{2,t-1}|\mathcal{I}_{1,t-2},\boldsymbol{\theta}\right)}{p\left(\mathbf{x}_{1,t-1}|\mathcal{I}_{1,t-2},\boldsymbol{\theta}\right)},$$

where

$$p\left(\mathbf{x}_{2,t-1}|\mathcal{I}_{1,t-2},\boldsymbol{\theta}\right)=\int p\left(\mathbf{x}_{2,t-1}|\mathbf{x}_{2,t-2},\mathcal{I}_{1,t-2},\boldsymbol{\theta}\right)\times p\left(\mathbf{x}_{2,t-2}|\mathcal{I}_{1,t-2},\boldsymbol{\theta}\right)d\mathbf{x}_{2,t-2}.$$

We can approximate $p\left(\mathbf{x}_{2,t-1}|\mathcal{I}_{1,t-1},\boldsymbol{\theta}\right)$ up to some proportionality; namely,

$$\hat{p}\left(\mathbf{x}_{2,t-1}|\hat{\mathcal{I}}_{1,t-1},\boldsymbol{\theta}\right)\propto\hat{p}\left(\mathbf{x}_{1,t-1}|\hat{\mathbf{X}}_{2,t-1},\hat{\mathcal{I}}_{1,t-2},\boldsymbol{\theta}\right)\times\sum_{j=1}^{J}\pi^j_{t-1}\hat{p}\left(\mathbf{x}_{2,t-1}|\hat{\mathbf{X}}_{2,t-2},\hat{\mathcal{I}}_{1,t-2},\boldsymbol{\theta}\right),$$

where $\hat{p}(\mathbf{x}_{1,t-1}|\hat{\mathbf{X}}_{2,t-1},\hat{\mathcal{I}}_{1,t-2},\boldsymbol{\theta})$ and $\sum_{j=1}^{J}\pi^j_{t-1}\hat{p}(\mathbf{x}_{2,t-1}|\hat{\mathbf{X}}_{2,t-2},\hat{\mathcal{I}}_{1,t-2},\boldsymbol{\theta})$ can be viewed as the likelihood and prior respectively.

The second method to approximate $p(\mathbf{x}_{2,t-1}|\mathcal{I}_{1,t-1},\boldsymbol{\theta})$ is Gallant and Tauchen's (1998) SNP-based reprojection technique, which can characterize the dynamic response of a partially observed nonlinear system to its past observable history. First, we can generate simulated samples $\{\hat{\mathbf{X}}_{1,t-1}\}^J_{t=2}$ and $\{\hat{\mathbf{X}}_{2,t-1}\}^J_{t=2}$ from the conditional distribution model, where J is a large integer. Then, we project the simulated data $\{\hat{\mathbf{X}}_{2,t-1}\}^J_{t=2}$ onto a Hermite series representation of the transition density $p(\mathbf{x}_{2,t-1}|\hat{\mathbf{X}}_{1,t-1},\hat{\mathbf{X}}_{1,t-2},\ldots,\hat{\mathbf{X}}_{1,t-L})$, where L denotes a truncation lag order. With a suitable choice of L via some information criteria such as AIC or BIC, we can approximate $p(\mathbf{x}_{2,t-1}|\hat{\mathcal{I}}_{1,t-1},\boldsymbol{\theta})$ arbitrarily well. The final step is to evaluate the estimated density function at the observed data in the conditional information set. See Gallant and Tauchen (1998) for more discussion. 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. 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.

In a related context, Hong (1999) considers a generalized spectral test based on the characteristic function and Hong and Lee (2005) extend the test to check the specification of conditional-mean models. Our test is proposed in a unified framework, where Hong and Lee's (2005) test corresponds to our $\hat{Q}^{\mathbf{m}}_1$ test defined in (5.1) with $|\mathbf{m}|=1$. Moreover, our test is based on the conditional distribution function, which admits closed form for many conditional distribution models as shown in Section 2 and hence is convenient to use here. Last, we replace the q-dependence assumption (Assumption A.2, Hong and Lee, 2005) with the mixing condition, which is commonly assumed in the testing literature of time series models (e.g., Ait-Sahalia et al., 2009). Escanciano and Velasco (2006) propose a test for the MDS property based on the generalized spectral distribution function. The asymptotic null distribution of their test depends on the DGP and is nonstandard. Moreover, it is difficult to account for the estimation uncertainty of the conditional distribution model with their method.

To introduce our test, we first define a generalized covariance function

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

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

$$F(\omega,\mathbf{x},\mathbf{y})=\frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\Gamma_j(\mathbf{x},\mathbf{y})\exp(-ij\omega),\quad\omega\in[-\pi,\pi],\ \mathbf{x},\mathbf{y}\in\mathbb{R}^d,\tag{3.1}$$

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 $j\neq0$. Consequently, the generalized spectral density $F(\omega,\mathbf{x},\mathbf{y})$ becomes a "flat" spectrum (i.e., a constant function of frequency ω):

$$F(\omega,\mathbf{x},\mathbf{y})=F_0(\omega,\mathbf{x},\mathbf{y})\equiv\frac{1}{2\pi}\Gamma_0(\mathbf{x},\mathbf{y}),\quad\omega\in[-\pi,\pi],\mathbf{x},\mathbf{y}\in\mathbb{R}^d.\tag{3.2}$$

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

$$\hat{\Gamma}_j(\mathbf{x},\mathbf{y})=\frac{1}{T-|j|}\sum_{t=|j|+1}^{T}Z_t(\mathbf{x},\hat{\boldsymbol{\theta}})\left[\mathbf{1}\left(\mathbf{X}_{t-|j|}\le\mathbf{y}\right)-\hat{\varphi}_j(\mathbf{y})\right],\quad\mathbf{x},\mathbf{y}\in\mathbb{R}^d,\tag{3.3}$$

where $\hat{\boldsymbol\theta}$ is a $\sqrt{T}$-consistent estimator for $\boldsymbol\theta_0$ and $\hat\varphi_j(\mathbf{y}) = (T-|j|)^{-1}\sum_{t=|j|+1}^{T}\mathbf{1}\left(\mathbf{X}_{t-|j|}\le\mathbf{y}\right)$ is the empirical distribution function of $\mathbf{X}_t$.

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

$$\hat F_0(\omega,\mathbf{x},\mathbf{y}) = \frac{1}{2\pi}\hat\Gamma_0(\mathbf{x},\mathbf{y}),\quad \omega\in[-\pi,\pi],\ \mathbf{x},\mathbf{y}\in\mathbb{R}^d. \tag{3.4}$$

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

$$\hat F(\omega,\mathbf{x},\mathbf{y}) = \frac{1}{2\pi}\sum_{j=1-T}^{T-1}(1-|j|/T)^{1/2}k(j/p)\hat\Gamma_j(\mathbf{x},\mathbf{y})e^{-\mathrm{i}j\omega},$$
$$\omega\in[-\pi,\pi],\mathbf{x},\mathbf{y}\in\mathbb{R}^d, \tag{3.5}$$

where $p\equiv p(T)\to\infty$ is a bandwidth or an effective lag order, and $k:\mathbb{R}\to[-1,1]$ is a kernel function, assigning weights to various lags. Examples of $k(\cdot)$ include the Bartlett kernel, the Parzen kernel and the Quadratic-Spectral kernel. In (3.5), 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

$$\begin{aligned}\hat L^2 &= \frac{\pi T}{2}\int_{-\pi}^{\pi}\iint\left|\hat F(\omega,\mathbf{x},\mathbf{y})-\hat F_0(\omega,\mathbf{x},\mathbf{y})\right|^2 d\omega dW(\mathbf{x})\,dW(\mathbf{y})\\ &= \sum_{j=1}^{T-1}k^2(j/p)(T-j)\iint\hat\Gamma_j^2(\mathbf{x},\mathbf{y})\,dW(\mathbf{x})\,dW(\mathbf{y}),\end{aligned} \tag{3.6}$$

where the equality follows by Parseval's identity, $W:\mathbb{R}^d\to\mathbb{R}^+$ 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 (3.6):

$$\hat Q_1 = \left[\sum_{j=1}^{T-1}k^2(j/p)(T-j)\times\iint\hat\Gamma_j^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})-\hat C_1\right]\Big/\sqrt{\hat D_1}, \tag{3.7}$$

where the centering and scaling factors

$$\hat C_1 = \sum_{j=1}^{T-1}k^2(j/p)(T-j)^{-1}\sum_{t=j+1}^{T}\times\int Z_t^2(\mathbf{x},\hat{\boldsymbol\theta})dW(\mathbf{x})\int\hat\psi_{t-j}^2(\mathbf{y})dW(\mathbf{y}),$$

$$\begin{aligned}\hat D_1 &= 2\sum_{j=1}^{T-2}\sum_{l=1}^{T-2}k^2(j/p)k^2(l/p)\\ &\times\iiiint dW(\mathbf{x}_1)dW(\mathbf{y}_1)\,dW(\mathbf{x}_2)dW(\mathbf{y}_2)\\ &\times\left\{[T-\max(j,l)]^{-1}\sum_{t=\max(j,l)+1}^{T}Z_t(\mathbf{x}_1,\hat{\boldsymbol\theta})\right.\\ &\left.\times Z_t(\mathbf{x}_2,\hat{\boldsymbol\theta})\hat\psi_{t-j}(\mathbf{y}_1)\hat\psi_{t-l}(\mathbf{y}_2)\right\}^2,\end{aligned}$$

where $\hat\psi_t(\mathbf{y}) = \mathbf{1}(\mathbf{X}_t\le\mathbf{y})-\hat\varphi(\mathbf{y})$, and $\hat\varphi(\mathbf{y}) = T^{-1}\sum_{t=1}^{T}\mathbf{1}(\mathbf{X}_t\le\mathbf{y})$. The factors $\hat C_1$ and $\hat D_1$ are the approximately mean and variance of the quadratic form in (3.6). When $W(\cdot)$ is continuous, $\hat Q_1$ can be calculated by numerical integration or simulation.[6]

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

$$\bar\Gamma_j(\mathbf{x},\mathbf{y}) = \operatorname{cov}\left[Z_t(\mathbf{x},\boldsymbol\theta),Z_{t-|j|}(\mathbf{y},\boldsymbol\theta)\right],\quad \mathbf{x},\mathbf{y}\in\mathbb{R}^d,$$

and estimate it by its sample analogue

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

Following similar derivations, we can obtain a new test statistics:

$$\bar Q_1 = \left[\sum_{j=1}^{T-1}k^2(j/p)(T-j)\times\iint\hat{\bar\Gamma}_j^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})dW(\mathbf{y})-\bar C_1\right]\Big/\sqrt{\bar D_1}, \tag{3.8}$$

where the centering and scaling factors

$$\bar C_1 = \sum_{j=1}^{T-1}k^2(j/p)(T-j)^{-1}\sum_{t=j+1}^{T}\int P(\mathbf{x}|\mathcal{I}_{t-1},\hat{\boldsymbol\theta})\times\left[1-P(\mathbf{x}|\mathcal{I}_{t-1},\hat{\boldsymbol\theta})\right]dW(\mathbf{x})\int Z_{t-j}^2(\mathbf{y},\hat{\boldsymbol\theta})dW(\mathbf{y}),$$

$$\bar D_1 = 2\sum_{j=1}^{T-2}k^4(j/p)\left\{\iint\left\{T^{-1}\sum_{t=1}^{T}\left[P(\mathbf{x}\wedge\mathbf{y}|\mathcal{I}_{t-1},\hat{\boldsymbol\theta})-P(\mathbf{x}|\mathcal{I}_{t-1},\hat{\boldsymbol\theta})P(\mathbf{y}|\mathcal{I}_{t-1},\hat{\boldsymbol\theta})\right]\right\}^2dW(\mathbf{x})dW(\mathbf{y})\right\}^2,$$

where $\mathbf{x}\wedge\mathbf{y}\equiv\min(\mathbf{x},\mathbf{y})$. We note that $\bar Q_1$ is computationally simpler than $\hat Q_1$. In particular, the integration for $\bar D_1$ is reduced from 4d dimensions to 2d dimensions. The key difference between $\bar Q_1$ and $\hat Q_1$ is the use of different conditioning variables. We will further examine the finite sample performance of $\bar Q_1$ and $\hat Q_1$ in Section 6.

One could also consider a test based on the supremum norm

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

This delivers a generalized KS test for dynamic conditional distribution models. In this paper, we focus on the test based on (3.6). 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 $\hat Q_1$ and $\bar Q_1$ and investigate their asymptotic power property, we impose following regularity conditions.

[6] Alternatively, the empirical distribution function also provides a 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 $\hat Q_2 = [\sum_{j=1}^{T-1}k^2(j/p)(T-j)T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat\Gamma_j^2(\mathbf{X}_t,\mathbf{X}_s)-\hat C_2]/\sqrt{\hat D_2}$, where the centering and scaling factors are $\hat C_2 = \sum_{j=1}^{T-1}k^2(j/p)(T-j)^{-1}T^{-2}\sum_{m=j+1}^{T}\sum_{t=1}^{T}Z_m^2(\mathbf{X}_t,\hat{\boldsymbol\theta})\sum_{s=1}^{T}\hat\psi_{m-j}^2(\mathbf{X}_s)$ and $\hat D_2 = 2\sum_{j=1}^{T-2}\sum_{l=1}^{T-2}k^2(j/p)k^2(l/p)T^{-4}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\{[T-\max(j,l)]^{-1}\sum_{m=\max(j,l)+1}^{T}Z_m(\mathbf{X}_{t_1},\hat{\boldsymbol\theta})Z_m(\mathbf{X}_{t_2},\hat{\boldsymbol\theta})\hat\psi_{m-j}(\mathbf{X}_{s_1})\hat\psi_{m-l}(\mathbf{X}_{s_2})\}^2$. 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 $\hat Q_1$.

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 $\mathcal{I}_{t-1} \equiv \{\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots, \mathbf{X}_1\}$ 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 $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, 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 $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, 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 $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with probability one; (iii) $\sup_{\mathbf{x}\in\mathbb{R}^d} E[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\frac{\partial}{\partial\boldsymbol{\theta}} P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta})\|^2] \leq C$ and $\sup_{\mathbf{x}\in\mathbb{R}^d} E[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta})\|] \leq C$.

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

Assumption A.4. For each $\mathbf{x} \in \mathbb{R}^d$, $\{\mathbf{X}_t, P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta}_0), \frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)\}$ is a strictly stationary β-mixing process with the mixing coefficient $|\beta(l)| \leq Cl^{-\nu}$ for some constant $\nu > 2$.

Assumption A.5. $k : \mathbb{R} \to [-1, 1]$ 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 $z \to \infty$.

Assumption A.6. $W : \mathbb{R}^d \to \mathbb{R}^+$ 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.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 $P(\mathbf{x}|\mathcal{I}_{t-1}, \theta)$ of $\mathbf{X}_t$. The assumption that the conditional CDF is twice continuously differentiable with respect to θ is weaker than the requirement that the conditional parametric density be twice continuously differentiable in θ, since the integration is a smoothing operation. Bai (2003) imposes similar regularity conditions. We allow $P(\mathbf{x}|\mathcal{I}_{t-1}, \theta)$ 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 $\{\mathbf{X}_t, P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta}_0), \frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)\}$. The β-mixing assumption is a standard condition for discrete time series analysis. 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 $T \to \infty$. 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, 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. 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.

We now state the asymptotic distribution of the GCM test $\hat{Q}_1$ under $\mathbb{H}_0$. The test $\bar{Q}_1$ follows the same asymptotic $N(0, 1)$ distribution under $\mathbb{H}_0$.

Theorem 1. *Suppose Assumptions A.1–A.6 hold, and $p = cT^{\lambda}$ for $\frac{1+\delta}{\nu\delta} < \lambda < (3 + \frac{1}{4b-2})^{-1}$, where $0 < c, \delta < \infty$. Then $\hat{Q}_1 \xrightarrow{d} N(0, 1)$ under $\mathbb{H}_0$ as $T \to \infty$.*

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 $\left\{\iint \hat{\Gamma}_j^2(\mathbf{x}, \mathbf{y}) dW(\mathbf{x})\, dW(\mathbf{y})\right\}_{j=1}^{p}$, which are approximately independent under $\mathbb{H}_0$ when $p \to \infty$. 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 $T \to \infty$. 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. We allow for weakly dependent data and data dependence has some impact on the feasible range of the bandwidth p. The condition on the tail behavior of the kernel function $k(\cdot)$ also has some impact. For kernels with bounded support (e.g., the Bartlett and Parzen), $\lambda < \frac{1}{3}$ because $b = \infty$. For the QS kernel ($b = 2$), $\lambda < \frac{6}{19}$. These conditions are mild.

In practice, one may like to choose p via some data-driven methods, which can let data determine an appropriate lag order. One plausible choice of the data-driven bandwidth is the nonparametric plug-in method proposed by Hong (1999, Theorem 2.2). It minimizes an asymptotic integrated mean squared error (IMSE) criterion for the estimator $\hat{F}(\omega, \mathbf{x}, \mathbf{y})$. Consider some "pilot" generalized spectrum estimators based on a preliminary bandwidth $\bar{p}$:

$$\bar{F}(\omega, \mathbf{x}, \mathbf{y}) = \frac{1}{2\pi} \sum_{j=1-T}^{T-1} (1 - |j|/T)^{1/2}\, \bar{k}(j/\bar{p})\, \hat{\Gamma}_j(\mathbf{x}, \mathbf{y})\, e^{-ij\omega}$$

and

$$\bar{F}^{(q,0,0)}(\omega, \mathbf{x}, \mathbf{y}) = \frac{1}{2\pi} \sum_{j=1-T}^{T-1} (1 - |j|/T)^{1/2}\, \bar{k}(j/\bar{p})\, |j|^q\, \hat{\Gamma}_j(\mathbf{x}, \mathbf{y})\, e^{-ij\omega},$$

where $\bar{k}(\cdot)$ is a kernel not necessarily the same as that used in (3.5). For the kernel $k(\cdot)$, suppose there exists some $q \in (0, \infty)$ such that $0 < k^{(q)} \equiv \lim_{z\to 0} \frac{1-k^{(q)}}{|z|^q} < \infty$. Then we define the plug-in bandwidth $\hat{p}_0$ as shown in Box I.

The data-driven $\hat{p}_0$ involves the choice of a preliminary bandwidth $\bar{p}$, which can be fixed or grow with the sample size T. If it is fixed, $\hat{p}_0$ still generally grows at rate $T^{1/(2q+1)}$ under $\mathbb{H}_A$, but $\hat{c}_0$ does not converge to the optimal tuning constant that minimizes the IMSE of $\hat{F}(\omega, \mathbf{x}, \mathbf{y})$. This is a parametric plug-in method. Alternatively, following Hong (1999), we can show that when $\bar{p}$ grows with T properly, the data-driven bandwidth $\hat{p}_0$ will minimize an asymptotic IMSE of $\hat{F}(\omega, \mathbf{x}, \mathbf{y})$. The choice of

$$\hat{p}_0 = \hat{c}_0 T^{1/(2q+1)},$$

where

$$\hat{c}_0 = \left\{ \frac{2q\left(k^{(q)}\right)}{\int_{-\infty}^{\infty} k^2(z)dz} \frac{\iint_{-\pi}^{\pi} \left|\bar{F}^{(q,0,0)}(\omega, \mathbf{x}, \mathbf{y})\right|^2 d\omega dW(\mathbf{x}, \mathbf{y})}{\mathrm{Re} \iint_{-\pi}^{\pi} \bar{F}(\omega, \mathbf{x}, -\mathbf{x}) \bar{F}(\omega, \mathbf{y}, -\mathbf{y}) d\omega dW(\mathbf{x}, \mathbf{y})} \right\}^{1/(2q+1)}$$
$$= \left\{ \frac{2q\left(k^{(q)}\right)}{\int_{-\infty}^{\infty} k^2(z)dz} \frac{\sum_{j=1-T}^{T-1} (T-|j|)\bar{k}^2(j/\bar{p})|j|^{2q} \int \left|\hat{\Gamma}_j(\mathbf{x}, \mathbf{y})\right|^2 dW(\mathbf{x}, \mathbf{y})}{\sum_{j=1-T}^{T-1} (T-|j|)\bar{k}^2(j/\bar{p}) \mathrm{Re} \int \hat{\Gamma}_j(\mathbf{x}, -\mathbf{x}) \hat{\Gamma}_j(\mathbf{y}, -\mathbf{y}) dW(\mathbf{x}, \mathbf{y})} \right\}^{1/(2q+1)}$$

Box I.

$\bar{p}$ is somewhat arbitrary, but we expect that it is of secondary importance. This is confirmed in our simulation below.

5. Asymptotic power

Our test is derived without assuming a specific alternative to $\mathbb{H}_0$. To get insights into the nature of the alternatives that our test is able to detect, we now examine the asymptotic power of $\hat{Q}_1$ under $\mathbb{H}_A$.

Theorem 2. *Suppose Assumptions A.1–A.6 hold, and $p = cT^{\lambda}$ for $0 < \lambda < \frac{1}{2}$ and $0 < c < \infty$. Then as $T \to \infty$,*

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

where

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

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

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 $1\left(\mathbf{X}_{t-|j|} \le \mathbf{y}\right)$ 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 $E\left[Z_t(\mathbf{x}, \boldsymbol{\theta}^*)|\mathbf{X}_{t-j}\right] = 0$ *a.s.* for all $\mathbf{x} \in \mathbb{R}^d$. Suppose $E\left[Z_t(\mathbf{x}, \boldsymbol{\theta}^*)|\mathbf{X}_{t-j}\right] \neq 0$ at some lag $j > 0$ under $\mathbb{H}_A$. Then we have $\iint \left|\Gamma_j(\mathbf{x}, \mathbf{y})\right|^2 dW(\mathbf{x}) dW(\mathbf{y}) \ge C > 0$ for any weighting function $W(\cdot)$ that is positive, monotonically increasing and continuous, with unbounded support on $\mathbb{R}$. As a result, $P[\hat{Q}_1 > C(T)] \to 1$ 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 $E\left[Z_t(\mathbf{x}, \boldsymbol{\theta}^*)|\mathbf{X}_{t-j}\right]$ is nonzero at some lag $j > 0$ under $\mathbb{H}_A$. Note that for a Markov process $\mathbf{X}_t$, we always have $E\left[Z_t\left(\mathbf{x}, \boldsymbol{\theta}^*\right)|\mathbf{X}_{t-j}\right] \neq 0$ 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 $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ 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 $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. The latter implies the former but not vice versa. This is the price we have to pay for dealing with the difficulty of the increasing information set.[7] 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 property of $\bar{Q}_1$ can be derived in a similar manner.

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. To gauge possible sources of model misspecification, we can construct a sequence of tests by integrating the generalized model residual $Z_t(x, \boldsymbol{\theta})$:

[7] To reduce the gap between $E[Z_t(\mathbf{x}, \boldsymbol{\theta}_0)|\mathcal{I}_{t-1}] = 0$ for all $\mathbf{x} \in \mathbb{R}^d$ and $E[Z_t(\mathbf{x}, \boldsymbol{\theta}_0)|\mathbf{X}_{t-j}] = 0$ for all $\mathbf{x} \in \mathbb{R}^d$ and all $j \neq 0$, we can extend $F(\omega, \mathbf{x}, \mathbf{y})$ to a generalized bispectrum

$$B(\omega_1, \omega_2, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \frac{1}{(2\pi)^2} \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} C_{j,l}(\mathbf{x}, \mathbf{y}, \mathbf{z}) e^{-ij\omega_1 - il\omega_2},$$

where

$$C_{j,l}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = Z_t(\mathbf{x}, \boldsymbol{\theta}) \left[1\left(\mathbf{X}_{t-|j|} \le \mathbf{y}\right) - \hat{P}(\mathbf{y})\right] \times \left[1\left(\mathbf{X}_{t-|l|} \le \mathbf{z}\right) - \hat{P}(\mathbf{z})\right], \quad \omega_1, \omega_2 \in [-\pi, \pi], \ \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^d,$$

is a generalized third order central cumulant function. This is equivalent to the use of $E[Z_t(\mathbf{x}, \boldsymbol{\theta}_0)|\mathbf{X}_{t-j}, \mathbf{X}_{t-l}]$. With $C_{j,l}(\mathbf{x}, \mathbf{y}, \mathbf{z})$, we can detect a larger class of alternatives to $E[Z_t(\mathbf{x}, \boldsymbol{\theta}_0)|\mathcal{I}_{t-1}] = 0$. Note that the nonparametric generalized bispectrum approach can check many pairs of lags (j, l), while still avoiding the "curse of dimensionality". Nevertheless, in this paper, we focus on $\Gamma_j(\mathbf{x}, \mathbf{y})$ for simplicity.

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

where $\mathbf{m} = (m_1, m_2, \ldots, m_d)'$, $m_c \geq 0$ for all $1 \leq c \leq d$ and $|\mathbf{m}| = \sum_{c=1}^{d} m_c$. For the univariate time series (i.e., $d = 1$), the choices of $\mathbf{m} = 1, 2, 3, 4$ corresponds to tests for misspecifications in the first four conditional moments respectively. For each $\mathbf{m}$, the resulting test statistic is given by:

$$\hat{Q}_1^{\mathbf{m}} = \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j)\int \hat{\Gamma}_j^{\mathbf{m}}(\mathbf{y})^2 dW(\mathbf{y}) - \hat{C}_1^{\mathbf{m}}\right] \Big/ \sqrt{\hat{D}_1^{\mathbf{m}}}, \quad (5.1)$$

where the centering and scaling factors

$$\hat{C}_1^{\mathbf{m}} = \sum_{j=1}^{T-1} k^2(j/p)\,(T-j)^{-1} \sum_{t=j+1}^{T} Z_t^{\mathbf{m}}(\hat{\boldsymbol{\theta}})^2 \int \hat{\psi}_{t-j}^2(\mathbf{y})dW(\mathbf{y}),$$

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

with

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

This set of diagnostic tests is similar to the moment-based tests used in Brooks et al. (2005) and Harvey and Siddique (1999). Compared with those conditional moment tests, our tests have several advantages: first, our GCM test $\hat{Q}_1$ essentially checks every moment, which is not obtainable by their 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.

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 $\bar{Q}_1$ 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)-$N(0, 1)$]:

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

The MA(1)-GARCH(1, 1) model is commonly used in empirical finance. We simulate 1000 data sets of a random sample $\{X_t\}_{t=1}^T$ for $T = 100, 250, 500, 1000$ 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)-$N(0, 1)$]:

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

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

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

In DGP3-6 below, the individual mean and variance are of the same forms as those in DGP0.DGP3 [MA(1)-GARCHK-t]:

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

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

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

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

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

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

Table 1
Sizes of specification tests under DGP0.

T	Lag order	100		250		500		1000	
α		0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
Tests based on the covariance between the generalized residual and its lag term									
$\bar{Q}_1$	10	0.063	0.038	0.062	0.040	0.067	0.039	0.058	0.037
	20	0.090	0.052	0.093	0.056	0.089	0.062	0.074	0.042
	30	0.108	0.060	0.105	0.063	0.102	0.069	0.078	0.043
	40	0.105	0.072	0.107	0.063	0.110	0.065	0.076	0.044
Tests based on the covariance between the generalized residual and the lag indicator function									
$\hat{Q}_1$	10	0.076	0.043	0.065	0.043	0.078	0.052	0.062	0.037
	20	0.061	0.038	0.065	0.037	0.077	0.048	0.070	0.037
	30	0.056	0.034	0.068	0.040	0.074	0.045	0.070	0.035
	40	0.053	0.029	0.067	0.034	0.068	0.044	0.073	0.034

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 $\varepsilon_t \sim$ i.i.d. $N(0, 1)$.
(2) $\hat{Q}_1$ and $\bar{Q}_1$ 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 Eqs. (3.7) and (3.8) respectively.
(3) 1000 iterations.

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.[8] 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 $T = 250, 500, 1000$ and 2500 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 $\bar{Q}_1$ 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.[9] 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 10–40.

Table 1 reports the rejection rates (in terms of percentage) of $\hat{Q}_1$ and $\bar{Q}_1$ 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 $\bar{Q}_1$ are not sensitive to the choice of the preliminary lag order $\bar{p}$.

Table 2 reports the rejection rates of $\bar{Q}_1$ 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 $\bar{Q}_1$ test has good power in detecting such a misspecification in the conditional mean. The rejection rate of $\bar{Q}_1$ increases significantly with the sample size T and approaches unity when $T = 2500$. Under DGP2, model (6.1) ignores the asymmetric effects in the conditional variance. The $\bar{Q}_1$ 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 $T = 1000$. 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 $\bar{Q}_1$ test has no power when the sample size T is small but the rejection rate increases with the sample size. The reason why $\bar{Q}_1$ has worse power under DGP3 than under DGPs 1 and 2 is that $\bar{Q}_1$ 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.

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^2_ν 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 $\bar{Q}_1$ 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 DGP6, ε_t is generated from a time-varying skewed Student's t distribution, and so $\bar{Q}_1$, 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 $\bar{Q}_1$, 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 $\bar{Q}_1$ and $\hat{Q}_1$ are close to each other. But in terms of the computational cost, $\bar{Q}_1$ is much less time-consuming than $\hat{Q}_1$, because a $4d$ dimensional integration is reduced to a $2d$ dimensional integration in calculating $\bar{Q}_1$. We thus suggest using $\bar{Q}_1$ in practice.

8 $p(\varepsilon|\nu_t, \lambda_t)$ is a skewed Student's t distribution, whose PDF is given in (2.2).
9 We have tried the Parzen kernel for $k(\cdot)$, obtaining similar results (not reported here).

Table 2
Powers of $\bar{Q}_1$ under DGPs 1–6.

T	Lag order	250		500		1000		2500	
α		0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
DGP1 [ARMA-GARCH-$N(0, 1)$]	10	0.426	0.330	0.838	0.762	0.994	0.992	1.00	1.00
	20	0.356	0.274	0.758	0.690	0.988	0.968	1.00	1.00
	30	0.306	0.238	0.686	0.608	0.966	0.936	1.00	1.00
	40	0.278	0.218	0.644	0.548	0.942	0.914	1.00	1.0
DGP2 [EGARCH-$N(0, 1)$]	10	0.812	0.730	0.994	0.986	1.00	1.00	1.00	1.00
	20	0.676	0.558	0.978	0.958	1.00	1.00	1.00	1.00
	30	0.564	0.452	0.952	0.924	1.00	1.00	1.00	1.00
	40	0.480	0.348	0.918	0.882	1.00	1.00	1.00	1.00
DGP3 [MA-GARCHK]	10	0.040	0.030	0.070	0.042	0.142	0.090	0.400	0.294
	20	0.046	0.030	0.070	0.038	0.100	0.054	0.260	0.168
	30	0.050	0.034	0.066	0.034	0.082	0.044	0.182	0.126
	40	0.054	0.036	0.062	0.030	0.070	0.036	0.152	0.092
DGP4 [MA-GARCH-Chi(5)]	10	0.220	0.154	0.416	0.318	0.840	0.780	1.00	1.00
	20	0.180	0.126	0.278	0.184	0.670	0.570	1.00	1.00
	30	0.150	0.110	0.220	0.140	0.546	0.434	0.996	0.994
	40	0.144	0.110	0.186	0.120	0.464	0.344	0.986	0.968
DGP5 [MA-GARCH-$t(5)$]	10	0.066	0.034	0.096	0.054	0.138	0.074	0.464	0.364
	20	0.054	0.032	0.060	0.030	0.078	0.048	0.296	0.210
	30	0.054	0.034	0.042	0.068	0.044	0.069	0.204	0.144
	40	0.050	0.030	0.040	0.054	0.026	0.065	0.148	0.102
DGP6 [MA-GARCH-time varying t]	10	0.152	0.110	0.322	0.236	0.708	0.616	1.00	1.00
	20	0.124	0.080	0.224	0.148	0.548	0.400	0.994	0.980
	30	0.108	0.070	0.162	0.106	0.394	0.278	0.958	0.920
	40	0.098	0.058	0.134	0.082	0.314	0.214	0.916	0.838

Notes: (1) $\bar{Q}_1$ is based on the covariance between the generalized residual and its lag term, given in Eq. (3.8).
(2) 500 iterations.

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

T	Lag order	250		500		1000		2500	
α		0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
DGP1 [ARMA-GARCH-$N(0, 1)$]	10	0.636	0.552	0.882	0.836	0.994	0.988	1.00	1.00
	20	0.548	0.438	0.806	0.752	0.982	0.980	1.00	1.00
	30	0.486	0.390	0.762	0.714	0.978	0.964	1.00	1.00
	40	0.444	0.362	0.748	0.672	0.966	0.948	1.00	1.00
DGP2 [EGARCH-$N(0, 1)$]	10	0.686	0.580	0.940	0.906	0.998	0.998	1.00	1.00
	20	0.584	0.488	0.900	0.832	0.996	0.992	1.00	1.00
	30	0.512	0.412	0.856	0.784	0.994	0.986	1.00	1.00
	40	0.452	0.340	0.820	0.736	0.990	0.978	1.00	1.00
DGP3 [MA-GARCHK]	10	0.108	0.058	0.132	0.098	0.228	0.160	0.444	0.358
	20	0.094	0.050	0.103	0.069	0.218	0.136	0.362	0.272
	30	0.074	0.044	0.104	0.061	0.188	0.122	0.328	0.238
	40	0.066	0.040	0.104	0.059	0.168	0.114	0.304	0.224
DGP4 [MA-GARCH-Chi(5)]	10	0.206	0.136	0.404	0.292	0.746	0.666	0.986	0.984
	20	0.160	0.102	0.302	0.212	0.616	0.534	0.974	0.962
	30	0.142	0.082	0.258	0.174	0.542	0.438	0.956	0.926
	40	0.126	0.068	0.224	0.140	0.482	0.378	0.936	0.874
DGP5 [MA-GARCH-$t(5)$]	10	0.116	0.072	0.146	0.106	0.250	0.188	0.504	0.402
	20	0.096	0.050	0.124	0.074	0.232	0.150	0.426	0.326
	30	0.084	0.044	0.108	0.066	0.200	0.138	0.392	0.274
	40	0.076	0.038	0.094	0.064	0.180	0.112	0.356	0.268
DGP6 [MA-GARCH-time varying t]	10	0.222	0.150	0.328	0.244	0.694	0.558	0.988	0.972
	20	0.182	0.133	0.258	0.176	0.552	0.424	0.964	0.918
	30	0.172	0.112	0.240	0.150	0.478	0.336	0.902	0.830
	40	0.157	0.083	0.206	0.136	0.438	0.302	0.844	0.778

Notes: (1) $\hat{Q}_1$ is based on the covariance between the generalized residual and the lag indicator function, given in Eq. (3.7).
(2) 500 iterations.

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($\mathbf{0}$, $\mathbf{I}$)]

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

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

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

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

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 1000 data sets of $\{\mathbf{X}_t\}_{t=1}^T$ for $T = 100, 250, 500$ and 1000 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:

$$\mathbf{H}_t = \begin{bmatrix} \sqrt{H_{11t}} & 0 \\ 0 & \sqrt{H_{22t}} \end{bmatrix} \mathbf{R}_t \begin{bmatrix} \sqrt{H_{11t}} & 0 \\ 0 & \sqrt{H_{22t}} \end{bmatrix}, \tag{6.9}$$

$$\mathbf{Q}_t = 0.1\mathbf{R}_0 + 0.7(\mathbf{R}_t^{1/2}\boldsymbol{\varepsilon}_{t-1})(\mathbf{R}_t^{1/2}\boldsymbol{\varepsilon}_{t-1})' + 0.2\mathbf{Q}_{t-1},$$

$$\mathbf{R}_t = \operatorname{diag}(\mathbf{Q}_t)^{-1}\,\mathbf{Q}_t \operatorname{diag}(\mathbf{Q}_t)^{-1},$$

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

DGP B2 (Granger causality in mean):

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

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.

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

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

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

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

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

DGP B1 is Engle's (2002a, 2002b) 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 $\boldsymbol{\varepsilon}_t$ 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 1000 respectively. For each data set, we estimate model (6.8) via MLE and check power performances. For computational simplicity, we just focus on $\bar{Q}_1$ 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 $\bar{Q}_1$ under DGPs B0–B4 at the 10% and 5% significance levels. Model (6.8) is correctly specified under DGP B0. The $\bar{Q}_1$ 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 $\bar{Q}_1$ 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 $\bar{Q}_1$ 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 $\bar{Q}_1$ 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 $T = 1000$. Under DGP B3, model (6.8) ignores the Granger causality in variance from both directions. The $\bar{Q}_1$ 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 $\bar{Q}_1$ 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 $T = 1000$.

To sum up, we observe:

- Both GCM tests $\hat{Q}_1$ and $\bar{Q}_1$ 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 $\bar{Q}_1$ 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 $\bar{Q}_1$ are close to each other under various univariate and bivariate alternatives but the computational costs differ. The test statistic $\bar{Q}_1$ is computationally more efficient.

7. Conclusion

Conditional distribution models in time series have become 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,

Table 4
Size and powers of $\bar{Q}_1$ under DGPs B0–B4.

T	Lag order	100		250		500		1000	
α		0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
Size									
DGPB0 AR(1)-BGARCH (1, 1)- BN(0,I)	10	0.101	0.068	0.085	0.061	0.088	0.052	0.074	0.038
	20	0.136	0.087	0.119	0.077	0.102	0.064	0.083	0.048
	30	0.156	0.121	0.127	0.085	0.103	0.070	0.087	0.051
	40	0.169	0.128	0.136	0.090	0.112	0.079	0.095	0.055
Powers									
DGPB1 (DCC)	10	0.220	0.140	0.442	0.342	0.644	0.554	0.894	0.852
	20	0.170	0.118	0.386	0.260	0.558	0.466	0.848	0.810
	30	0.146	0.108	0.316	0.236	0.504	0.388	0.812	0.758
	40	0.144	0.090	0.300	0.212	0.460	0.356	0.780	0.714
DGPB2 (Granger causality in mean)	10	0.506	0.366	0.930	0.876	0.994	0.992	1.00	1.00
	20	0.412	0.250	0.870	0.810	0.978	0.954	1.00	1.00
	30	0.314	0.206	0.828	0.774	0.958	0.922	1.00	1.00
	40	0.276	0.158	0.798	0.742	0.928	0.876	1.00	1.00
DGPB3 (Granger causality in variance)	10	0.202	0.112	0.652	0.625	0.792	0.778	0.834	0.830
	20	0.218	0.126	0.680	0.657	0.820	0.810	0.848	0.844
	30	0.220	0.116	0.696	0.677	0.838	0.828	0.860	0.856
	40	0.218	0.118	0.712	0.690	0.844	0.836	0.866	0.866
DGPB4 (Granger causality in distribution)	10	0.224	0.130	0.436	0.296	0.720	0.614	0.972	0.930
	20	0.174	0.096	0.288	0.190	0.570	0.446	0.902	0.814
	30	0.158	0.094	0.234	0.162	0.490	0.344	0.820	0.738
	40	0.144	0.076	0.206	0.144	0.424	0.278	0.744	0.656

Notes: (1) $\bar{Q}_1$ is based on the covariance between the generalized residual and its lag term, given in Eq. (3.8).
(2) Results of DGP B0 are based on 1000 iterations; results of DGPs B1–B4 are based on 500 iterations.

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 severely 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 exist neglected structures in conditional mean, conditional variance, conditional correlation, conditional skewness and conditional kurtosis respectively. 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.

Acknowledgments

We thank the Guest Editor, Zongwu Cai, and two referees for careful and constructive comments. We also thank Anil Bera, Bruce Hansen, Oliver Linton and Peter Robinson for their comments and discussions. Chen's research was supported, in part, by the National Nature Science Foundation of China grant #71131008. Any remaining errors are solely ours.

Appendix. Mathematical appendix

Throughout the appendix, we let $\tilde{Q}$ be defined in the same way as $\hat{Q}_1$ in (3.7) with the unobservable generalized residual sample $\{Z_t(\mathbf{x}, \boldsymbol{\theta}_0)\}_{t=1}^T$ replacing the estimated generalized residual sample $\{Z_t(\mathbf{x}, \hat{\boldsymbol{\theta}})\}_{t=1}^T$. Also, $C \in (1, \infty)$ denotes a generic bounded constant.

Proof of Theorem 1. The proof of Theorem 1 consists of the proofs of Theorems A.1 and A.2.

Theorem A.1. *Under the conditions of Theorem 1,* $\hat{Q}_1 - \tilde{Q} \xrightarrow{p} 0$.

Theorem A.2. *Under the conditions of Theorem 1,* $\tilde{Q} \xrightarrow{d} N(0, 1)$.

Proof of Theorem A.1. Put $T_j \equiv T - |j|$, and let $\tilde{\Gamma}_j(\mathbf{x}, \mathbf{y})$ be defined in the same way as $\hat{\Gamma}_j(\mathbf{x}, \mathbf{y})$ in (3.3), with $Z_t(\mathbf{x}, \hat{\boldsymbol{\theta}})$ replaced by $Z_t(\mathbf{x}, \boldsymbol{\theta}_0)$. To show $\hat{Q}_1 - \tilde{Q} \xrightarrow{p} 0$, it suffices to show

$$\hat{D}_1^{-\frac{1}{2}} \iint \sum_{j=1}^{T-1} k^2(j/p) T_j \left[\hat{\Gamma}_j^2(\mathbf{x}, \mathbf{y}) - \tilde{\Gamma}_j^2(\mathbf{x}, \mathbf{y})\right] \times dW(\mathbf{x})\, dW(\mathbf{y}) \xrightarrow{p} 0, \tag{A.1}$$

$p^{-1}(\hat{C}_1 - \tilde{C}) = O_P(T^{-\frac{1}{2}})$, and $p^{-1}(\hat{D}_1 - \tilde{D}) = o_p(1)$, where $\tilde{C}$ and $\tilde{D}$ are defined in the same way as $\hat{C}_1$ and $\hat{D}_1$ in (3.7), with $Z_t(\mathbf{x}, \hat{\boldsymbol{\theta}})$ replaced by $Z_t(\mathbf{x}, \boldsymbol{\theta}_0)$. For space, we focus on the proof of (A.1); the proofs for $p^{-1}(\hat{C}_1 - \tilde{C}) = O_P(T^{-\frac{1}{2}})$ and $p^{-1}(\hat{D}_1 - \tilde{D}) = o_p(1)$ are straightforward. We note that it is necessary to obtain the convergence rate $O_P(pT^{-\frac{1}{2}})$ for $\hat{C}_1 - \tilde{C}$ to ensure that replacing $\hat{C}_1$ with $\tilde{C}$ has asymptotically negligible impact given $p/T \to 0$.

To show (A.1), we first decompose

$$\iint \sum_{j=1}^{T-1} k^2(j/p) T_j \left[\hat{\Gamma}_j^2(\mathbf{x}, \mathbf{y}) - \tilde{\Gamma}_j^2(\mathbf{x}, \mathbf{y})\right] \times dW(\mathbf{x})\, dW(\mathbf{y}) = \hat{A}_1 + 2\hat{A}_2, \tag{A.2}$$

where

$$\hat{A}_1 = \iint \sum_{j=1}^{T-1} k^2(j/p) T_j [\hat{\Gamma}_j(\mathbf{x}, \mathbf{y}) - \tilde{\Gamma}_j(\mathbf{x}, \mathbf{y})]^2 dW(\mathbf{x})\, dW(\mathbf{y}),$$

$$\hat{A}_2 = \iint \sum_{j=1}^{T-1} k^2(j/p)T_j \left[\hat{\Gamma}_j(\mathbf{x},\mathbf{y}) - \tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right] \times \tilde{\Gamma}_j(\mathbf{x},\mathbf{y}) dW(\mathbf{x})\, dW(\mathbf{y}).$$

Then, (A.1) follows from Propositions A.1 and A.2, and $p \to \infty$ as $T \to \infty$.

Proposition A.1. *Under the conditions of Theorem 1, $\hat{A}_1 = O_P(1)$.*

Proposition A.2. *Under the conditions of Theorem 1, $p^{-\frac{1}{2}}\hat{A}_2 \xrightarrow{p} 0$.*

Proof of Proposition A.1. Put $\psi_t(\mathbf{y}) \equiv 1(\mathbf{X}_t \le \mathbf{y}) - \varphi(\mathbf{y})$ and $\varphi(\mathbf{y}) \equiv E[1(\mathbf{X}_t \le \mathbf{y})]$. Then straightforward algebra yields that for $j > 0$,

$$\begin{aligned}\hat{\Gamma}_j(\mathbf{x},\mathbf{y}) - \tilde{\Gamma}_j(\mathbf{x},\mathbf{y}) &= T_j^{-1}\sum_{t=j+1}^{T}\left[Z_t(\mathbf{x},\hat{\theta}) - Z_t(\mathbf{x},\boldsymbol{\theta}_0)\right]\psi_{t-j}(\mathbf{y}) \\ &\quad + \left[\varphi(\mathbf{y}) - \hat{\varphi}_j(\mathbf{y})\right]T_j^{-1} \\ &\quad \times \sum_{t=j+1}^{T}\left[Z_t(\mathbf{x},\hat{\theta}) - Z_t(\mathbf{x},\boldsymbol{\theta}_0)\right] \\ &= \hat{B}_{1j}(\mathbf{x},\mathbf{y}) + \hat{B}_{2j}(\mathbf{x},\mathbf{y}), \quad \text{say.} \end{aligned} \tag{A.3}$$

It follows that $\hat{A}_1 \le 2\sum_{a=1}^{2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint \hat{B}_{aj}^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})\, dW(\mathbf{y})$. Proposition A.1 follows from Lemmas A.1 and A.2, and $p/T \to 0$.

Lemma A.1. $\sum_{j=1}^{T-1} k^2(j/p)T_j \iint \hat{B}_{1j}^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})\, dW(\mathbf{y}) = O_P(1)$.

Lemma A.2. $\sum_{j=1}^{T-1} k^2(j/p)T_j \iint \hat{B}_{2j}^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})\, dW(\mathbf{y}) = O_P(p/T)$.

We now show these lemmas. Throughout, we put $a_T(j) \equiv k^2(j/p)T_j^{-1}$.

Proof of Lemma A.1. A second order Taylor series expansion yields

$$\begin{aligned}\hat{B}_{1j}(\mathbf{x},\mathbf{y}) &= -(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'T_j^{-1}\sum_{t=j+1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y}) \\ &\quad - \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'T_j^{-1}\sum_{t=j+1}^{T}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}P\left(\mathbf{x}|\mathcal{I}_{t-1},\bar{\boldsymbol{\theta}}\right) \\ &\quad \times \psi_{t-j}(\mathbf{y})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= -\hat{B}_{11j}(\mathbf{x},\mathbf{y}) - \hat{B}_{12j}(\mathbf{x},\mathbf{y}), \quad \text{say,}\end{aligned} \tag{A.4}$$

for some $\bar{\boldsymbol{\theta}}$ between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$.

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

$$\begin{aligned}&\sum_{j=1}^{T-1}k^2(j/p)T_j\iint \hat{B}_{12j}^2(\mathbf{x},\mathbf{y})\,dW(\mathbf{x})\,dW(\mathbf{y}) \\ &\le C\left\|\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right\|^4 \iint\left[T^{-1}\sum_{t=1}^{T}\sup_{\mathbf{x}\in\mathbb{R}^d}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta})\right\|\right]^2\left[\sum_{j=1}^{T-1}\alpha_T(j)\right] \\ &\quad \times dW(\mathbf{x})dW(\mathbf{y}) = O_p(p/T),\end{aligned}$$

where we made use of the fact that

$$\sum_{j=1}^{T-1}a_T(j) = \sum_{j=1}^{T-1}k^2(j/p)T_j^{-1} = O(p/T), \tag{A.5}$$

given $p = cT^{\lambda}$ for $\lambda \in (0,1)$, as shown in Hong (1999, (A.15), p. 1213).

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

$$\begin{aligned}&\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\hat{B}_{11j}^2(\mathbf{x},\mathbf{y})\,dW(\mathbf{x})\,dW(\mathbf{y}) \\ &\le \left\|\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right\|^2\sum_{j=1}^{T-1}k^2(j/p) \\ &\quad \times \iint\left\|T_j^{-1}\sum_{t=j+1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y})\right\|^2 dW(\mathbf{x})\,dW(\mathbf{y}) \\ &= O_p(1),\end{aligned} \tag{A.6}$$

as is shown below: Put $\eta_j(\mathbf{x},\mathbf{y}) \equiv E[\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y})] = \operatorname{cov}[\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0),\psi_{t-j}(\mathbf{y})]$. Then we have $\sup_{\mathbf{x},\mathbf{y}\in\mathbb{R}^{2d}}\Sigma_{j=1}^{\infty}\|\eta_j(\mathbf{x},\mathbf{y})\| \le C$ by Assumption A.4. Next, expressing the moments by cumulants via well-known formulas (e.g., Hannan, 1970, (5.1), p. 23), we can obtain

$$\begin{aligned}&T_jE\left\|T_j^{-1}\sum_{t=j+1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y}) - \eta_j(\mathbf{x},\mathbf{y})\right\|^2 \\ &\leqslant \sum_{\tau=-T_j}^{T_j}\left\|\operatorname{cov}\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\varphi\left[\mathbf{x},\max(0,\tau)+2|\mathcal{I}_{\max(0,\tau)+1},\boldsymbol{\theta}_0\right],\right.\right. \\ &\quad \left.\left.\frac{\partial}{\partial\boldsymbol{\theta}}\varphi\left[\mathbf{x},\max(0,\tau)+2-\tau|\mathcal{I}_{\max(0,\tau)+1-\tau},\boldsymbol{\theta}_0\right]'\right\}\right\| \\ &\quad \times|\Omega_\tau(\mathbf{y},-\mathbf{y})| + \sum_{\tau=-T_j}^{T_j}\left\|\eta_{j+|\tau|}(\mathbf{x},-\mathbf{y})\right\|\left\|\eta_{j-|\tau|}(\mathbf{x},\mathbf{y})\right\| \\ &\quad + \sum_{\tau=-T_j}^{T_j}\left\|\kappa_{j,|\tau|,j+|\tau|}(\mathbf{x},\mathbf{y})\right\| \\ &\leqslant C,\end{aligned} \tag{A.7}$$

given Assumption A.4, where $\kappa_{j,l,\tau}(v)$ is the fourth order cumulant of the joint distribution of the process $\{\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0),\psi_{t-j}(\mathbf{y}),\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0),\psi_{t-j}(\mathbf{y})\}$. See also (A.7) of Hong (1999, p. 1212). Consequently, from (A.5), (A.6), $|k(\cdot)| \le 1$, and $p/T \to 0$, we have

$$\begin{aligned}&\sum_{j=1}^{T-1}k^2(j/p)E\iint\left\|T_j^{-1}\sum_{t=j+1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y})\right\|^2 \\ &\quad \times dW(\mathbf{x})dW(\mathbf{y}) \\ &\le C\sum_{j=1}^{T-1}\iint\left\|\eta_j(\mathbf{x},\mathbf{y})\right\|^2 dW(\mathbf{x})\,dW(\mathbf{y}) + C\sum_{j=1}^{T-1}a_T(j) \\ &= O_p(1) + O_p(p/T) = O_p(1).\end{aligned}$$

Hence (A.6) is $O_P(1)$. The desired result of Lemma A.1 follows from (A.5) and (A.6). ■

Proof of Lemma A.2. We have

$$\begin{aligned}&\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\hat{B}_{2j}^2(\mathbf{x},\mathbf{y})\,dW(\mathbf{x})\,dW(\mathbf{y}) \\ &\le \sum_{j=1}^{T-1}k^2(j/p)T_j\iint[\varphi(\mathbf{y}) - \hat{\varphi}_j(\mathbf{y})]^2T_j^{-1} \\ &\quad \times\sum_{t=j+1}^{T}\left[P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0) - P\left(\mathbf{x}|\mathcal{I}_{t-1},\hat{\boldsymbol{\theta}}\right)\right]^2 dW(\mathbf{x})\,dW(\mathbf{y}) \\ &= O_p(p/T),\end{aligned} \tag{A.8}$$

where we made use of the fact that $E\left|\varphi(\mathbf{y})-\hat{\varphi}_j(\mathbf{y})\right|^2 \leq CT_j^{-1}$ given Assumption A.4 and

$$\sum_{t=j+1}^{T}\left[P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)-P\left(\mathbf{x}|\mathcal{I}_{t-1},\hat{\boldsymbol{\theta}}\right)\right]^2$$

$$\leq T_j\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right\|^2 T_j^{-1}\sum_{t=1}^{T}\sup_{\mathbf{x}\in\mathbb{R}^d}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}P\left(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}\right)\right\|^2$$

$$=O_P(1). \quad \blacksquare$$

Proof of Proposition A.2. Given the decomposition in (A.3), we have

$$\left|[\hat{\Gamma}_j(\mathbf{x},\mathbf{y})-\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})]\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right| \leq \sum_{a=1}^{2}\left|\hat{B}_{aj}(\mathbf{x},\mathbf{y})\right|\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|,$$

where the $\hat{B}_{aj}(\mathbf{x},\mathbf{y})$ are defined in (A.3).

We first consider the term with $a=2$. By the Cauchy–Schwarz inequality, we have

$$\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{2j}(\mathbf{x},\mathbf{y})\right|\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|dW(\mathbf{x})\,dW(\mathbf{y})$$

$$\leq\left[\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\hat{B}_{2j}^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})\right]^{\frac{1}{2}}$$

$$\times\left[\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\tilde{\Gamma}_j^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})\right]^{\frac{1}{2}}$$

$$=O_P(p^{\frac{1}{2}}/T^{\frac{1}{2}})O_P(p^{\frac{1}{2}})=O_P(p^{\frac{1}{2}}), \tag{A.9}$$

given Lemma A.2, and $p/T\to 0$, where $p^{-1}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\tilde{\Gamma}_j^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})=O_P(1)$ by Markov's inequality, the MDS hypothesis of $Z_t(\mathbf{x},\boldsymbol{\theta}_0)$, and (A.5).

For $a=1$, by (A.4) and the triangular inequality, we have

$$\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{1j}(\mathbf{x},\mathbf{y})\right|\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|dW(\mathbf{x})\,dW(\mathbf{y})$$

$$\leq\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{11j}(\mathbf{x},\mathbf{y})\right|\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|dW(\mathbf{x})\,dW(\mathbf{y})$$

$$+\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{12j}(\mathbf{x},\mathbf{y})\right|$$

$$\times\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|dW(\mathbf{x})\,dW(\mathbf{y}). \tag{A.10}$$

For the first term in (A.10), we have

$$\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{11j}(\mathbf{x},\mathbf{y})\right|\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|dW(\mathbf{x})\,dW(\mathbf{y})$$

$$\leq\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right\|\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left\|T_j^{-1}\sum_{t=j+1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}\right.$$

$$\left.\times P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y})\right\|\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|dW(\mathbf{x})\,dW(\mathbf{y})$$

$$=O_P(1+p/T^{\frac{1}{2}})=O_P(p^{\frac{1}{2}}), \tag{A.11}$$

given $p\to\infty$, $p/T\to 0$, Assumptions A.2, A.3, A.5, A.6, and $T_jE\tilde{\Gamma}_j^2(\mathbf{x},\mathbf{y})\leq C$. Note that we have made use of the fact that

$$E\left[\left\|T_j^{-1}\sum_{t=j+1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y})\right\|\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|\right]$$

$$\leq\left[E\left\|T_j^{-1}\sum_{t=j+1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y})\right\|^2\right]^{\frac{1}{2}}\left[E\tilde{\Gamma}_j^2(\mathbf{x},\mathbf{y})\right]^{\frac{1}{2}}$$

$$\leq C\left[\left\|\eta_j(\mathbf{x},\mathbf{y})\right\|+CT_j^{-\frac{1}{2}}\right]T_j^{-\frac{1}{2}},$$

by (A.6), and consequently,

$$\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right\|\sum_{j=1}^{T-1}k^2(j/p)T_jE$$

$$\times\iint\left\|T_j^{-1}\sum_{t=j+1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}P(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y})\right\|$$

$$\times\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|dW(\mathbf{x})\,dW(\mathbf{y})$$

$$\leq C\sum_{j=1}^{T-1}\iint\left\|\eta_j(\mathbf{x},\mathbf{y})\right\|dW(\mathbf{x})\,dW(\mathbf{y})$$

$$+CT^{-\frac{1}{2}}\sum_{j=1}^{T-1}k^2(j/p)=O(1+p/T^{\frac{1}{2}}),$$

given $|k(\cdot)|\leq 1$ and Assumption A.6.

For the second term in (A.10), we have

$$\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{12j}(\mathbf{x},\mathbf{y})\right|\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|dW(\mathbf{x})\,dW(\mathbf{y})$$

$$\leq CT\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right\|^2\left[T^{-1}\sum_{t=j+1}^{T}\sup_{\mathbf{x}\in\mathbb{R}^d}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}P\left(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}\right)\right\|\right]$$

$$\times\sum_{j=1}^{T-1}k^2(j/p)\iint\left|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})\right|dW(\mathbf{x})\,dW(\mathbf{y})$$

$$=O_P(p/T^{\frac{1}{2}}), \tag{A.12}$$

by Cauchy–Schwarz inequality, Markov inequality, Assumptions A.2, A.3, A.5, A.6, and $E\tilde{\Gamma}_j^2(\mathbf{x},\mathbf{y})\leq CT_j^{-1}$.

Hence, we have

$$\sum_{j=1}^{T-1}k^2(j/p)T_j\iint|\hat{B}_{1j}(\mathbf{x},\mathbf{y})|\,|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})|dW(\mathbf{x})dW(\mathbf{y})$$

$$=O_P\left(1+p/T^{\frac{1}{2}}\right)+O_P\left(p/T^{\frac{1}{2}}\right)=O_P\left(p^{\frac{1}{2}}\right). \tag{A.13}$$

Combining (A.9) and (A.13) then yields the result of this proposition. $\blacksquare$

Proof of Theorem A.2. Let $q=p^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$. We shall show Propositions A.3 and A.4.

Proposition A.3. *Under the conditions of Theorem 1,*

$$p^{-\frac{1}{2}}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\tilde{\Gamma}_j^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})$$

$$=p^{-1/2}\tilde{C}+p^{-1/2}\tilde{V}+O_P(1),$$

where $\tilde{V}_q=\sum_{t=2q+2}^{T}\iint Z_t(\mathbf{x},\boldsymbol{\theta}_0)\sum_{j=1}^{q}a_T(j)\psi_{t-j}(\mathbf{y})[\sum_{s=1}^{t-2q-1}Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})]dW(\mathbf{x})\,dW(\mathbf{y})$.

Proposition A.4. *Under the conditions of Theorem 1,* $\tilde{D}^{-1/2}\tilde{V}_q\xrightarrow{d}N(0,1)$.

Proof of Proposition A.3. We first decompose

$$\sum_{j=1}^{T-1} k^2(j/p)T_j \iint \tilde{\Gamma}_j^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})$$
$$= \sum_{j=1}^{T-1} a_T(j) \iint \left[\sum_{t=j+1}^{T} Z_t(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y})\right]^2 dW(\mathbf{x})\,dW(\mathbf{y})$$
$$+ \sum_{j=1}^{T-1} a_T(j) \iint \left[\sum_{t=j+1}^{T} Z_t(\mathbf{x},\boldsymbol{\theta}_0)\right]^2 \left[\varphi(\mathbf{y}) - \hat{\varphi}_j(\mathbf{y})\right]^2 \times dW(\mathbf{x})\,dW(\mathbf{y})$$
$$+2\sum_{j=1}^{T-1} \alpha_T(j) \iint \left[\sum_{t=j+1}^{T} Z_t(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y})\right] \times \left\{\sum_{t=j+1}^{T} Z_t(\mathbf{x},\boldsymbol{\theta}_0)\left[\varphi(\mathbf{y}) - \hat{\varphi}_j(\mathbf{y})\right]\right\} dW(\mathbf{x})\,dW(\mathbf{y})$$
$$\equiv \tilde{M} + \tilde{R}_1 + 2\tilde{R}_2. \tag{A.14}$$

Next we write

$$\tilde{M}_q = \sum_{j=1}^{T-1} a_T(j) \iint \sum_{t=j+1}^{T} Z_t^2(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{t-j}^2(\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})$$
$$+2\sum_{j=1}^{T-1} a_T(j) \iint \sum_{t=j+2}^{T}\sum_{s=j+1}^{t-1} Z_t(\mathbf{x},\boldsymbol{\theta}_0) \times Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y})\psi_{s-j}(\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})$$
$$\equiv \tilde{C}_1 + 2\tilde{U}, \tag{A.15}$$

where we further decompose

$$\tilde{U} = \sum_{t=2q+2}^{T} \iint Z_t(\mathbf{x},\boldsymbol{\theta}_0) \sum_{j=1}^{t-1} a_T(j)\psi_{t-j}(\mathbf{y}) \times \sum_{s=j+1}^{t-2q-1} Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})$$
$$+ \sum_{t=2}^{T} \iint Z_t(\mathbf{x},\boldsymbol{\theta}_0) \sum_{j=1}^{t-1} a_T(j)\psi_{t-j}(\mathbf{y}) \times \sum_{s=\max(j+1,t-2q)}^{t-1} Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})$$
$$\equiv \tilde{U}_1 + \tilde{R}_3, \tag{A.16}$$

where in the first term $\tilde{U}_1$, we have $t - s > 2q$ so that we can bound it with the mixing inequality. In the second term $\tilde{R}_{3q}$, we have $0 < t - s \le 2q$. Finally, we write

$$\tilde{U}_1 = \sum_{t=2q+2}^{T} \iint Z_t(\mathbf{x},\boldsymbol{\theta}_0) \sum_{j=1}^{q} a_T(j)\psi_{t-j}(\mathbf{y}) \times \sum_{s=j+1}^{t-2q-1} Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})$$
$$+ \sum_{t=2q+2}^{T} \iint Z_t(\mathbf{x},\boldsymbol{\theta}_0) \sum_{j=q+1}^{t-1} a_T(j)\psi_{t-j}(\mathbf{y}) \times \sum_{s=j+1}^{t-2q-1} Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})$$
$$\equiv \tilde{V}_q + \tilde{R}_4, \tag{A.17}$$

where the first term $\tilde{V}_q$ is contributed by the lag orders j from 1 to q; and the second term $\tilde{R}_4$ is contributed by the lag orders $j > q$. It follows from (A.14)–(A.17) that

$$\sum_{j=1}^{T-1} k^2(j/p)T_j \iint \tilde{\Gamma}_j^2(\mathbf{x},\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})$$
$$= \tilde{C}_1 + 2\tilde{V}_q + \tilde{R}_{1q} - 2(\tilde{R}_{2q} - \tilde{R}_{3q} - \tilde{R}_{4q}).$$

It suffices to show Lemmas A.3–A.7, which imply $p^{-\frac{1}{2}}(\tilde{C}_1 - \tilde{C}) = O_P(1)$ and $p^{-\frac{1}{2}}\tilde{R}_a = O_P(1)$ given $q = p^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$ and $p = cT^{\lambda}$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$. ■

Lemma A.3. *Let $\tilde{C}_1$ be defined as in (A.15). Then $\tilde{C}_1 - \tilde{C} = O_P(p/T^{\frac{1}{2}})$.*

Lemma A.4. *Let $\tilde{R}_1$ be defined as in (A.14). Then $\tilde{R}_1 = O_P(p/T)$.*

Lemma A.5. *Let $\tilde{R}_2$ be defined as in (A.14). Then $\tilde{R}_2 = O_P(p/T^{\frac{1}{2}})$.*

Lemma A.6. *Let $\tilde{R}_3$ be defined as in (A.16). Then $\tilde{R}_3 = O_P(pq/T^{\frac{1}{2}})$.*

Lemma A.7. *Let $\tilde{R}_4$ be defined as in (A.17). Then $\tilde{R}_4 = O_P(p^{2b}\ln T/q^{2b-1})$.*

Proof of Lemma A.3. By Markov's inequality and $E\left|\tilde{C}_1 - \tilde{C}\right| \le Cp/T^{1/2}$ given $\sum_{j=1}^{T-1}(j/p)a_T(j) = O(p/T)$ and $E[\varphi(\mathbf{y}) - \hat{\varphi}_j(\mathbf{y})]^4 = O\left(T^{-2}\right)$, which are shown in Hong (1999) and below respectively.

Let $t_1,\ldots,t_4$ be distinct integers and $|j|+1 \le t_i \le T$, let $|j|+1 \le r_1 < \cdots < r_4 \le T$ be the permutation of $t_1,\ldots,t_4$ in ascending order and let d_c be the c-th largest difference among $r_{l+1} - r_l,\ l = 1, 2, 3$.

By Lemma 1 of Yoshihara (1976), we have

$$\sum_{\substack{|j|+1\le r_1<\cdots<r_4\le T\\ r_2-r_1=d_1}} \Big|E\left\{\left[1(\mathbf{X}_{r_1-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\left[1(\mathbf{X}_{r_2-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\right.$$
$$\left.\times\left[1(\mathbf{X}_{r_3-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\left[1(\mathbf{X}_{r_4-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\right\}\Big|$$
$$\le \sum_{r_1=|j|+1}^{T-3}\ \sum_{r_2=r_1+\max_{j\ge3}(r_j-r_{j-1})}^{T-2}\ \sum_{r_3=r_2+1}^{T-1}\ \sum_{r_4=r_3+1}^{T} 4C^{\frac{1}{1+\delta}}\beta^{\frac{\delta}{1+\delta}}(r_2-r_1)$$
$$\le 4C^{\frac{1}{1+\delta}} \sum_{r_1=|j|+1}^{T-3}\ \sum_{r_2=r_1+1}^{T-2} (r_2-r_1)^2\beta^{\frac{\delta}{1+\delta}}(r_2-r_1)$$
$$\le 4TC^{\frac{1}{1+\delta}} \sum_{r=|j|+1}^{T} r^2\beta^{\frac{\delta}{1+\delta}}(r) = O(T).$$

Similarly,

$$\sum_{\substack{|j|+1\le r_1<\cdots<r_4\le T\\ r_4-r_3=d_1}} \Big|E\left\{\left[1(\mathbf{X}_{r_1-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\left[1(\mathbf{X}_{r_2-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\right.$$
$$\left.\times\left[1(\mathbf{X}_{r_3-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\left[1(\mathbf{X}_{r_4-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\right\}\Big| = O(T).$$

Further, it can be shown in a similar way that

$$\sum_{\substack{|j|+1\le r_1<\cdots<r_4\le T\\ r_3-r_2=d_1}} \Big|E\left\{\left[1(\mathbf{X}_{r_1-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\left[1(\mathbf{X}_{r_2-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\right.$$
$$\left.\times\left[1(\mathbf{X}_{r_3-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\left[1(\mathbf{X}_{r_4-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\right\}\Big| = O\left(T^2\right).$$

Similar to the above, we can show that if r_s are not distinct to each other, we have

$$\sum_{\substack{|j|+1\le r_1,r_2,r_3\le T\\ r_1,r_2,r_3\ \text{different}}} \Big|E\left\{\left[1(\mathbf{X}_{r_1-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]^2\right.$$
$$\left.\times\left[1(\mathbf{X}_{r_2-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\left[1(\mathbf{X}_{r_3-j}\le\mathbf{y}) - \varphi(\mathbf{y})\right]\right\}\Big| = O(T^2),$$

and

$$\sum_{\substack{|j|+1\le r_1,r_2\le T\\ r_1,r_2\ \text{different}}} \left| E\left\{\left[1(\mathbf{X}_{r_1-j}\le \mathbf{y})-\varphi(\mathbf{y})\right]^2 \times \left[1(\mathbf{X}_{r_2-j}\le \mathbf{y})-\varphi(\mathbf{y})\right]^2\right\}\right| = O\left(T^2\right).$$

Therefore, $E\left[\varphi(\mathbf{y})-\hat{\varphi}_j(\mathbf{y})\right]^4 = O\left(T^{-2}\right)$. ■

Proof of Lemma A.4.

$$E|\tilde{R}_1| \le \sum_{j=1}^{T-1} a_T(j) \iint \left\{E\left[\sum_{t=j+1}^{T} Z_t(\mathbf{x},\boldsymbol{\theta}_0)\right]^4\right\}^{\frac{1}{2}} \times \left\{E\left[\varphi(\mathbf{y})-\hat{\varphi}_j(\mathbf{y})\right]^4\right\}^{\frac{1}{2}} dW(\mathbf{x})\,dW(\mathbf{y}) = O(p/T),$$

where we have used the fact $E\left[\sum_{t=j+1}^{T} Z_t(\mathbf{x},\boldsymbol{\theta}_0)\right]^4 \le CT_j^2$ by Rosenthal's inequality and $E\left[\varphi(\mathbf{y})-\hat{\varphi}_j(\mathbf{y})\right]^4 = O\left(T^{-2}\right)$. ■

Proof of Lemma A.5. By the MDS property of $Z_t(\mathbf{x},\boldsymbol{\theta}_0)$, the Cauchy–Schwarz inequality, we have

$$\begin{aligned} E\left|\tilde{R}_2\right| &\le 2\sum_{j=1}^{T-1} a_T(j) \iint \left\{E\left[\sum_{t=j+1}^{T} Z_t(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y})\right]^2\right\}^{\frac{1}{2}} \\ &\quad \times \left\{E\left\{\sum_{t=j+1}^{T} Z_t(\mathbf{x},\boldsymbol{\theta}_0)\left[\varphi(\mathbf{y})-\hat{\varphi}_j(\mathbf{y})\right]\right\}^2\right\}^{\frac{1}{2}} \\ &\quad \times dW(\mathbf{x})\,dW(\mathbf{y}) \\ &\le 2\sum_{j=1}^{T-1} a_T(j) \iint \left[E\sum_{t=j+1}^{T} Z_t^2(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{t-j}^2(\mathbf{y})\right]^{\frac{1}{2}} \\ &\quad \times \left\{E\left[\sum_{t=j+1}^{T} Z_t(\mathbf{x},\boldsymbol{\theta}_0)\right]^4\right\}^{\frac{1}{4}} \\ &\quad \times \left\{E\left[\varphi(\mathbf{y})-\hat{\varphi}_j(\mathbf{y})\right]^4\right\}^{\frac{1}{4}} dW(\mathbf{x})\,dW(\mathbf{y}) \\ &= O\left(p/T^{\frac{1}{2}}\right), \end{aligned}$$

given $E[\varphi(\mathbf{y})-\hat{\varphi}_j(\mathbf{y})]^4 = O\left(T^{-2}\right)$. ■

Proof of Lemma A.6. By the MDS property of $Z_t(\mathbf{x},\boldsymbol{\theta}_0)$, Minkowski's inequality and (A.5), we have

$$\begin{aligned} E\left|\tilde{R}_3\right|^2 &= \sum_{t=2}^{T} E\left|\sum_{j=1}^{t-1} a_T(j) \iint Z_t(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y}) \times \sum_{s=\max(j+1,t-2q)}^{t-1} Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})\right|^2 \\ &\leqslant \sum_{t=2}^{T}\left\{\sum_{j=1}^{t-1} a_T(j) \iint \left[E\left|Z_t(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y}) \times \sum_{s=\max(j+1,t-2q)}^{t-1} Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})\right|^2\right]^{1/2} dW(\mathbf{x})\,dW(\mathbf{y})\right\}^2 \\ &\le CTq^2\left[\sum_{j=1}^{T-1} a_T(j)\right]^2 = O(p^2q^2/T). \quad ■ \end{aligned}$$

Proof of Lemma A.7. By the MDS property of $Z_t(\mathbf{x},\boldsymbol{\theta}_0)$ and Minkowski's inequality, we have

$$\begin{aligned} E\tilde{R}_4^2 &= \sum_{t=2q+2}^{T} E\left[\sum_{j=q+1}^{t-1} a_T(j) \iint Z_t(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y}) \times \sum_{s=j+1}^{t-2q-1} Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y})\right]^2 \\ &\leqslant \sum_{t=2q+2}^{T}\left\{\sum_{j=q+1}^{t-1} a_T(j) \iint \left[E\left|Z_t(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y})\right|^4\right]^{1/4} \times \left[E\left|\sum_{s=j+1}^{t-2q-1} Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})\right|^4\right]^{1/4} dW(\mathbf{x})\,dW(\mathbf{y})\right\}^2 \\ &\leqslant CT^2\left[\sum_{j=q+1}^{T-1} a_T(j)\right]^2 \leqslant CT^2\left[\sum_{j=q+1}^{T-1} (j/p)^{-2b}T_j^{-1}\right]^2 \\ &= O(p^{4b}\ln^2 T/q^{4b-2}), \end{aligned}$$

given Assumption A.5 (i.e., $k(z) \leqslant C|z|^{-b}$ as $z\to\infty$). ■

Proof of Proposition A.4. We rewrite $\tilde{V}_q = \sum_{t=2q+2}^{T} V_q(t)$, where

$$V_q(t) = \iint Z_t(\mathbf{x},\boldsymbol{\theta}_0)\sum_{j=1}^{q} a_T(j)\psi_{t-j}(\mathbf{y})H_{j,t-2q-1} \times (\mathbf{x},\mathbf{y})dW(\mathbf{x})\,dW(\mathbf{y}),$$

and $H_{j,t-2q-1}(\mathbf{x},\mathbf{y}) = \sum_{s=j+1}^{t-2q-1} Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})$. We apply Brown's (1971) martingale limit theorem, which states $\mathrm{var}(2\tilde{V}_q)^{-\frac{1}{2}}\, 2\tilde{V}_q \xrightarrow{d} N(0,1)$ if

$$\mathrm{var}(2\tilde{V}_q)^{-1}\sum_{t=2q+2}^{T}\left[2V_q(t)\right]^2 \mathbf{1} \times \left[\left|2V_q(t)\right| > \eta\cdot \mathrm{var}(2\tilde{V}_q)^{\frac{1}{2}}\right] \to 0 \quad \forall\,\eta>0, \tag{A.18}$$

$$\mathrm{var}(2\tilde{V}_q)^{-1}\sum_{t=2q+2}^{T} E\left\{\left[2V_q(t)\right]^2|\mathcal{F}_{t-1}\right\} \xrightarrow{p} 1. \tag{A.19}$$

First, we compute $\mathrm{var}(2\tilde{V}_q)$. By the MDS property of $Z_t(\mathbf{x},\boldsymbol{\theta}_0)$ under $\mathbb{H}_0$, we have

$$\begin{aligned} E(\tilde{V}_q^2) &= \sum_{t=2q+2}^{T} E\left[\iint Z_t(\mathbf{x},\boldsymbol{\theta}_0)\sum_{j=1}^{q} a_T(j)\psi_{t-j}(\mathbf{y}) \times \sum_{s=j+1}^{t-2q-1} Z_s(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y})dW(\mathbf{x})dW(\mathbf{y})\right]^2 \\ &= \sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l) \\ &\quad \times \iiiint \sum_{t=2q+2}^{T}\sum_{s=j+1}^{t-2q-1} E[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)\,Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0) \end{aligned}$$

$$
\begin{aligned}
&\times \psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)]\\
&= \sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l)\\
&\quad\times \iiiint \sum_{t=2q+2}^{T}\sum_{s=j+1}^{t-2q-1} E[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\\
&\quad\times \psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)]E[Z_s(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_s(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{s-j}(\mathbf{y}_1)\\
&\quad\times \psi_{s-l}(\mathbf{y}_2)]dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&\quad+ \sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l)\\
&\quad\times \iiiint \sum_{t=2q+2}^{T}\sum_{s=j+1}^{t-2q-1} \operatorname{cov}[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\\
&\quad\times \psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)Z_s(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_s(\mathbf{x}_2,\boldsymbol{\theta}_0)\\
&\quad\times \psi_{s-j}(\mathbf{y}_1)\psi_{s-l}(\mathbf{y}_2)]dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&\quad+2\sum_{j=1}^{q} a_T^2(j)\iiiint \sum_{t=2q+2}^{T}\sum_{s_1=j+1}^{t-2q-1}\sum_{s_2=j+1}^{s_1-1}\\
&\quad\times E[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y}_1)\psi_{t-j}(\mathbf{y}_2)\\
&\quad\times Z_{s_1}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{s_2}(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{s_1-j}(\mathbf{y}_1)\psi_{s_2-j}(\mathbf{y}_2)]\\
&\quad\times dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&\quad+4\sum_{j=2}^{q}\sum_{l=1}^{j-1} a_T(j)a_T(l)\iiiint \sum_{t=2q+2}^{T}\sum_{s_1=j+1}^{t-2q-1}\sum_{s_2=l+1}^{s_1-1}\\
&\quad\times E[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\\
&\quad\times Z_{s_1}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{s_2}(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{s_1-j}(\mathbf{y}_1)\psi_{s_2-l}(\mathbf{y}_2)]\\
&\quad\times dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&= \frac{1}{2}\sum_{j=1}^{q}\sum_{l=1}^{q} k^2(j/p)k^2(l/p)\\
&\quad\times \iiiint \left|E\left[Z_{q,j+l}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{q,j+l}(\mathbf{x}_2,\boldsymbol{\theta}_0)\right.\right.\\
&\quad\times \left.\left.\psi_{q,l}(\mathbf{y}_1)\psi_{q,j}(\mathbf{y}_2)\right]\right|^2 dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\\
&\quad\times dW(\mathbf{y}_1)dW(\mathbf{y}_2)[1+o(1)] + \sum_{t=2q+2}^{T} V_1(t)\\
&\quad+\sum_{t=2q+2}^{T} V_2(t) + \sum_{t=2q+2}^{T} V_3(t),
\end{aligned}\tag{A.20}
$$

where the first term is $O(p)$ as shown in (A.25) and the remaining terms are $o(p)$ following the arguments below

$$
\begin{aligned}
\sum_{t=2q+2}^{T}|V_1(t)| &\le \sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l)\\
&\quad\times \iiiint \sum_{t=2q+2}^{T}\sum_{s=j+1}^{t-2q-1}\beta^{\delta/(1+\delta)}(t-s)\\
&\quad\times \left[E\left|Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\right.\right.\\
&\quad\times \left.\left.\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\right|^{2(1+\delta)}\right]^{\frac{1}{2(1+\delta)}}\\
&\quad\times \left[E\left|Z_s(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_s(\mathbf{x}_2,\boldsymbol{\theta}_0)\right.\right.\\
&\quad\times \left.\left.\psi_{s-j}(\mathbf{y}_1)\psi_{s-l}(\mathbf{y}_2)\right|^{2(1+\delta)}\right]^{\frac{1}{2(1+\delta)}}\\
&\quad\times dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&= O\left(p^2T^{-1}q^{-\nu\delta/(1+\delta)+1}\right),
\end{aligned}\tag{A.21}
$$

$$
\begin{aligned}
\sum_{t=2q+2}^{T}|V_2(t)| &\le 2\sum_{j=1}^{q} a_T^2(j)\iiiint \sum_{t=2q+2}^{T}\sum_{s_1=j+1}^{t-2q-1}\sum_{s_2=j+1}^{s_1-1}\\
&\quad\times E\left\{\left\{Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y}_1)\psi_{t-j}(\mathbf{y}_2)\right.\right.\\
&\quad\left.-E\left[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y}_1)\psi_{t-j}(\mathbf{y}_2)\right]\right\}\\
&\quad\times \left.Z_{s_1}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{s_2}(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{s_1-j}(\mathbf{y}_1)\psi_{s_2-j}(\mathbf{y}_2)\right\}\\
&\quad\times dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&\le 2\sum_{j=1}^{q} a_T^2(j)\iiiint \sum_{t=2q+2}^{T}\sum_{s_1=j+1}^{t-2q-1}\sum_{s_2=j+1}^{s_1-1}\\
&\quad\times \beta^{\delta/(1+\delta)}(t-s_1)\left[E\left|Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\right.\right.\\
&\quad\times \left.\left.\psi_{t-j}(\mathbf{y}_1)\psi_{t-j}(\mathbf{y}_2)\right|^{2(1+\delta)}\right]^{\frac{1}{2(1+\delta)}}\\
&\quad\times \left[E\left|Z_{s_1}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{s_2}(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{s_1-j}(\mathbf{y}_1)\right.\right.\\
&\quad\times \left.\left.\psi_{s_2-j}(\mathbf{y}_2)\right|^{2(1+\delta)}\right]^{\frac{1}{2(1+\delta)}} dW(\mathbf{x}_1)\\
&\quad\times dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&= O\left(pq^{-\nu\delta/(1+\delta)+1}\right),
\end{aligned}\tag{A.22}
$$

and

$$
\begin{aligned}
\sum_{t=2q+2}^{T}|V_3(t)| &\le 4\sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l)\\
&\quad\times \iiiint \sum_{t=2q+2}^{T}\sum_{s_1=j+1}^{t-2q-1}\sum_{s_2=j+1}^{s_1-1}\\
&\quad\times E\left\{\left\{Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\right.\right.\\
&\quad\left.-E\left[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\right]\right\}\\
&\quad\times \left.Z_{s_1}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{s_2}(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{s_1-j}(\mathbf{y}_1)\psi_{s_2-l}(\mathbf{y}_2)\right\}\\
&\quad\times dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&\le 4\sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l)\\
&\quad\times \iiiint \sum_{t=2q+2}^{T}\sum_{s_1=j+1}^{t-2q-1}\sum_{s_2=j+1}^{s_1-1}\beta^{\delta/(1+\delta)}\\
&\quad\times (t-s_1)\left[E\left|Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\right.\right.\\
&\quad\times \left.\left.\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\right|^{2(1+\delta)}\right]^{\frac{1}{2(1+\delta)}}\\
&\quad\times \left[E\left|Z_{s_1}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{s_2}(\mathbf{x}_2,\boldsymbol{\theta}_0)\right.\right.\\
&\quad\times \left.\left.\psi_{s_1-j}(\mathbf{y}_1)\psi_{s_2-l}(\mathbf{y}_2)\right|^{2(1+\delta)}\right]^{\frac{1}{2(1+\delta)}}\\
&\quad\times dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&= O\left(p^2q^{-\nu\delta/(1+\delta)+1}\right).
\end{aligned}\tag{A.23}
$$

Hence,

$$
\begin{aligned}
\operatorname{var}(2\tilde{V}_q) &= 2\sum_{j=1}^{q}\sum_{l=1}^{q} k^2(j/p)k^2(l/p)\\
&\quad\times \iiiint \left\{E\left[Z_{j+l}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{j+l}(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_l(\mathbf{y}_1)\psi_j(\mathbf{y}_2)\right]\right\}^2\\
&\quad\times dW(\mathbf{x}_1,\mathbf{y}_1)\,dW(\mathbf{x}_2,\mathbf{y}_2)[1+o(1)].
\end{aligned}\tag{A.24}
$$

Put $C(0,j,l) \equiv E\{[Z_{j+l}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{j+l}(\mathbf{x}_2,\boldsymbol{\theta}_0) - \Sigma_0(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)]\psi_l(\mathbf{y}_1)\psi_j(\mathbf{y}_2)\}$, where $\Sigma_j(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0) \equiv E[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{t-j}(\mathbf{x}_2,\boldsymbol{\theta}_0)]$. Then

$$\begin{aligned}
&E\left[Z_{j+l}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{j+l}(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_l(\mathbf{y}_1)\psi_j(\mathbf{y}_2)\right]\\
&\quad = C(0,j,l) + \Sigma_0(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)\,\Omega_{l-j}(\mathbf{y}_1,\mathbf{y}_2),\\
&\left\{E\left[Z_{j+l}(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_{j+l}(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_l(\mathbf{y}_1)\psi_j(\mathbf{y}_2)\right]\right\}^2\\
&\quad = [C(0,j,l)]^2 + \left[\Sigma_0(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)\,\Omega_{l-j}(\mathbf{y}_1,\mathbf{y}_2)\right]^2\\
&\qquad + 2C(0,j,l)\Sigma_0(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)\,\Omega_{l-j}(\mathbf{y}_1,\mathbf{y}_2).
\end{aligned}$$

Given $\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|C(0,j,l)| \le C$ and $|k(\cdot)| \le 1$, we have

$$\begin{aligned}
\operatorname{var}(2\tilde{V}_q) &= 2\sum_{j=1}^{q}\sum_{l=1}^{q}k^2(j/p)k^2(l/p)\\
&\quad\times\iint\left[\Omega_{l-j}(\mathbf{y}_1,\mathbf{y}_2)\right]^2 dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&\quad\times\iint[\Sigma_0(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)]^2\\
&\quad\times dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,[1+o(1)]\\
&= 2p\sum_{m=1-q}^{q-1}\left\{p^{-1}\sum_{j=m+1}^{q}k^2(j/p)k^2[(j-m)/p]\right\}\\
&\quad\times\iint[\Omega_m(\mathbf{y}_1,\mathbf{y}_2)]^2\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&\quad\times\iint[\Sigma_0(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)]^2\\
&\quad\times dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,[1+o(1)]\\
&= 2p\int_0^\infty k^4(z)dz\sum_{m=-\infty}^{\infty}\iint[\Omega_m(\mathbf{y}_1,\mathbf{y}_2)]^2\\
&\quad\times dW(\mathbf{y}_1)dW(\mathbf{y}_2)\iint[\Sigma_0(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)]^2\\
&\quad\times dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,[1+o(1)]\\
&= 4\pi p\int_0^\infty k^4(z)dz\iint\int_{-\pi}^{\pi}[F(\omega,\mathbf{y}_1,\mathbf{y}_2)]^2\\
&\quad\times d\omega dW(\mathbf{y}_1)dW(\mathbf{y}_2)\iint|\Sigma_0(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)|^2\\
&\quad\times dW(\mathbf{x}_1)\,dW(\mathbf{x}_2)\,[1+o(1)],
\end{aligned}\tag{A.25}$$

where we used the fact that for any given m, $p^{-1}\sum_{j=m+1}^{q}k^2(j/p)k^2(\frac{j-m}{p}) \to \int_0^\infty k^4(z)dz$ as $p\to\infty$.

We now verify condition (A.18). Noting that $E\left[H_{j,t-2q-1}(\mathbf{x},\mathbf{y})\right]^8 \le Ct^4$ for $1\le j\le q$ given the MDS property of $Z_t(\mathbf{x},\boldsymbol{\theta}_0)$ and Rosenthal's inequality (cf. Hall and Heyde, 1980, p. 23), we have

$$\begin{aligned}
E\left[V_q(t)\right]^4 &\leqslant \left\{\sum_{j=1}^{q}a_T(j)\iint\left[E\left(Z_t(\mathbf{x},\boldsymbol{\theta}_0)\,\psi_{q,t-j}(\mathbf{y})\right.\right.\right.\\
&\qquad\left.\left.\left.\times H_{j,t-2q-1}(\mathbf{x},\mathbf{y})\right)^4\right]^{1/4}dW(\mathbf{x})dW(\mathbf{y})\right\}^4\\
&\leqslant Ct^2\left[\sum_{j=1}^{q}a_T(j)\right]^4 = O(p^4t^2/T^4).
\end{aligned}$$

It follows that $\sum_{t=2q+2}^{T}E\left[V_q(t)\right]^4 = O\left(p^4/T\right) = o(p^2)$ given $p^2/T\to 0$. Thus, (A.18) holds.

Next, we verify condition (A.19). Put $\Sigma_t(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0) \equiv E[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\,|\,\mathcal{I}_{t-1}]$. Then

$$\begin{aligned}
E[V_q^2(t)|\mathcal{I}_{t-1}] &= \sum_{j=1}^{q}\sum_{l=1}^{q}a_T(j)a_T(l)\\
&\quad\times\iiiint\Sigma_t(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\\
&\quad\times H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\\
&\quad\times dW(\mathbf{x}_1)dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&= \sum_{j=1}^{q}\sum_{l=1}^{q}a_T(j)a_T(l)\\
&\quad\times\iiiint E\left[\Sigma_t(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\right]\\
&\quad\times H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\\
&\quad\times dW(\mathbf{x}_1)dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&\quad+\sum_{j=1}^{q}\sum_{l=1}^{q}a_T(j)a_T(l)\\
&\quad\times\iiiint\tilde{L}_t^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)\\
&\quad\times H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\\
&\quad\times dW(\mathbf{x}_1)dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&\equiv S_1(t)+V_4(t),\quad\text{say,}
\end{aligned}\tag{A.26}$$

where $\tilde{L}_t^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2) \equiv \Sigma_t(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2) - E[\Sigma_t(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)]$. We further decompose

$$\begin{aligned}
S_1(t) &= \sum_{j=1}^{q}\sum_{l=1}^{q}a_T(j)a_T(l)\iiiint E\left[\Sigma_t(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)\right.\\
&\quad\left.\times\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\right]E\left[H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)\right.\\
&\quad\left.\times H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\right]dW(\mathbf{x}_1)dW(\mathbf{x}_2)\\
&\quad\times dW(\mathbf{y}_1)dW(\mathbf{y}_2)+\sum_{j=1}^{q}\sum_{l=1}^{q}a_T(j)a_T(l)\\
&\quad\times\iiiint E\left[\Sigma_t(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)\,\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\right]\\
&\quad\times\left\{H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\right.\\
&\quad\left.-E\left[H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\right]\right\}\\
&\quad\times dW(\mathbf{x}_1)dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&\equiv V_0(t)+S_2(t),\quad\text{say,}
\end{aligned}\tag{A.27}$$

where

$$\begin{aligned}
V_0(t) &= \sum_{j=1}^{q}\sum_{l=1}^{q}\min(t-2q-1-j,\,t-2q-1-l)a_T\\
&\quad\times(j)a_T(l)\iiiint|E\left[\Sigma_t(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta}_0)\right.\\
&\quad\left.\times\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\right]|^2\\
&\quad\times dW(\mathbf{x}_1)dW(\mathbf{x}_2)\,dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
&= E\left[V_q^2(t)\right]-V_1(t)-V_2(t)-V_3(t),
\end{aligned}$$

where $V_j(t)$, $j=1,2,3$, are defined in (A.20). Put

$$\begin{aligned}
L_s^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2) &\equiv Z_s(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_s(\mathbf{x}_2,\boldsymbol{\theta}_0)\,\psi_{s-j}(\mathbf{y}_1)\psi_{s-l}(\mathbf{y}_2)\\
&\quad-E\left[Z_s(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_s(\mathbf{x}_2,\boldsymbol{\theta}_0)\right.\\
&\quad\left.\times\psi_{s-j}(\mathbf{y}_1)\psi_{s-l}(\mathbf{y}_2)\right].
\end{aligned}\tag{A.28}$$

Then we write

$$
\begin{aligned}
S_2(t) = & \sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l) \iiiint E\left[Z_t\left(\mathbf{x}_1,\boldsymbol{\theta}_0\right)\right.\\
& \times Z_t\left(\mathbf{x}_2,\boldsymbol{\theta}_0\right)\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\big]\\
& \times \sum_{s=\max(j,l)}^{t-2q-1} L_s^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)\\
& \times dW(\mathbf{x}_1)dW(\mathbf{x}_2)dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
& + \sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l) \iiiint E\left[Z_t\left(\mathbf{x}_1,\boldsymbol{\theta}_0\right)Z_t\left(\mathbf{x}_2,\boldsymbol{\theta}_0\right)\right.\\
& \times \psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)\big] \sum_{s=\max(j,l)+1}^{t-2q-1}\sum_{\tau=l+1}^{s-1} Z_s\left(\mathbf{x}_1,\boldsymbol{\theta}_0\right)\\
& \times \psi_{s-j}(\mathbf{y}_1)Z_\tau\left(\mathbf{x}_2,\boldsymbol{\theta}_0\right)\psi_{\tau-l}(\mathbf{y}_2)dW(\mathbf{x}_1)\\
& \times dW(\mathbf{x}_2)dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
\equiv & V_5(t)+S_3(t), \quad \text{say,}
\end{aligned}
\tag{A.29}
$$

where

$$
\begin{aligned}
S_3(t) = & \sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l) \iiiint E[Z_t\left(\mathbf{x}_1,\boldsymbol{\theta}_0\right)\\
& \times Z_t\left(\mathbf{x}_2,\boldsymbol{\theta}_0\right)\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}](\mathbf{y}_2)\\
& \times \sum\sum_{0<s-\tau\leqslant 2q} Z_s\left(\mathbf{x}_1,\boldsymbol{\theta}_0\right)\psi_{s-j}(\mathbf{y}_1)Z_\tau\left(\mathbf{x}_2,\boldsymbol{\theta}_0\right)\\
& \times \psi_{\tau-l}(\mathbf{y}_2)dW(\mathbf{x}_1)dW(\mathbf{x}_2)dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
& + \sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l) \iiiint E[Z_t\left(\mathbf{x}_1,\boldsymbol{\theta}_0\right)\\
& \times Z_t\left(\mathbf{x}_2,\boldsymbol{\theta}_0\right)\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)]\\
& \times \sum\sum_{s-\tau>2q} Z_s\left(\mathbf{x}_1,\boldsymbol{\theta}_0\right)\psi_{s-j}(\mathbf{y}_1)Z_\tau\left(\mathbf{x}_2,\boldsymbol{\theta}_0\right)\\
& \times \psi_{\tau-l}(\mathbf{y}_2)dW(\mathbf{x}_1)dW(\mathbf{x}_2)dW(\mathbf{y}_1)dW(\mathbf{y}_2)\\
\equiv & V_6(t)+V_7(t), \quad \text{say.}
\end{aligned}
\tag{A.30}
$$

It follows from (A.26)–(A.27) and (A.29)–(A.30) that $\sum_{t=2q+2}^{T}\{E[V_q^2(t)|\mathcal{I}_{t-1}]-E[V_q^2(t)]\} = \sum_{t=2q+2}^{T}[\sum_{a=4}^{7}V_a(t)-\sum_{a=1}^{3}V_a(t)]$. It suffices to show Lemmas A.8–A.11, which imply $E|\{\sum_{t=2q+2}^{T}E[V_q^2(t)|\mathcal{I}_{t-1}]-E[V_q^2(t)]\}^2 = o(p^2)$ given $q=p^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$ and $p=cT^{\lambda}$ for $0<\lambda<(3+\frac{1}{4b-2})^{-1}$. Thus, condition (A.19) holds, and so $\tilde{Q}_q(\mathbf{0},\mathbf{0})\xrightarrow{d}N(0,1)$ by Brown's (1971) theorem. ■

Lemma A.8. *Let $V_4(t)$ be defined as in (A.26). Then $E[\sum_{t=2q+2}^{T}V_4(t)]^2=O(p/q^{\frac{\nu\delta}{1+\delta}})$.*

Lemma A.9. *Let $V_5(t)$ be defined as in (A.29). Then $E|[\sum_{t=2q+2}^{T}V_5(t)]^2=O(qp^4/T)$.*

Lemma A.10. *Let $V_6(t)$ be defined as in (A.30). Then $E[\sum_{t=2q+2}^{T}V_6(t)]^2=O(qp^4/T)$.*

Lemma A.11. *Let $V_7(t)$ be defined as in (A.30). Then $E[\sum_{t=2q+2}^{T}V_7(t)]^2=O(p)$.*

Proof of Lemma A.8. Recalling the definition of $\tilde{L}_t^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)$ as in (A.26), we can obtain

$$
\begin{aligned}
& E\left[\sum_{t=2q+2}^{T}\tilde{L}_{q,t}^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\right]^2\\
\le & \sum_{t=2q+2}^{T}\left\{E\left[\tilde{L}_{q,t}^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)\right]^4\right\}^{\frac{1}{2}}\\
& \times\left\{E\left[H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)\right]^8\right\}^{\frac{1}{4}}\left\{E\left[H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\right]^8\right\}^{\frac{1}{4}}\\
& +2\sum\sum_{t-\tau>2q}\beta^{\frac{\delta}{1+\delta}}(2q)\left\{E\left[\tilde{L}_{q,t}^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)\right]^{2(1+\delta)}\right\}^{\frac{1}{2(1+\delta)}}\\
& \times\left\{E\left[\tilde{L}_{q,\tau}^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,t-2q-1}\right.\right.\\
& \left.\left.\times(\mathbf{x}_2,\mathbf{y}_2)H_{j,\tau-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,\tau-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\right]^{2(1+\delta)}\right\}^{\frac{1}{2(1+\delta)}}\\
& +2\sum\sum_{0<t-\tau<2q}E\left[\tilde{L}_{q,t}^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)\tilde{L}_{q,\tau}^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)\right]\\
& \times\left\{E\left[H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)\right]^4\right\}^{\frac{1}{4}}\left\{E\left[H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\right]^4\right\}^{\frac{1}{4}}\\
& \times\left\{E\left[H_{j,\tau-2q-1}(\mathbf{x}_1,\mathbf{y}_1)\right]^4\right\}^{\frac{1}{4}}\left\{E\left[H_{l,\tau-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\right]^4\right\}^{\frac{1}{4}}\\
& +2\sum\sum_{0<t-\tau<2q}\beta^{\frac{\delta}{1+\delta}}(2q)\left\{E\left[\tilde{L}_{q,t}^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)\right.\right.\\
& \left.\left.\times\tilde{L}_{q,\tau}^{j,l}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)\right]^{2(1+\delta)}\right\}^{\frac{1}{2(1+\delta)}}\\
& \times\left\{E\left[H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\right.\right.\\
& \left.\left.\times H_{j,\tau-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,\tau-2q-1}(\mathbf{x}_2,\mathbf{y}_2)\right]^{2(1+\delta)}\right\}^{\frac{1}{2(1+\delta)}}\\
= & O\left(T^3\right)+O\left(T^4q^{-\frac{\nu\delta}{1+\delta}+1}\right)+O\left(T^3q\right)+O\left(T^3q^{-\frac{\nu\delta}{1+\delta}+1}\right),
\end{aligned}
$$

where we have made use of the fact that $E\left[H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)\right]^8\le Ct^4$ for $1\le j\le q$. It follows by Minkowski's inequality and (A.5) that

$$
\begin{aligned}
E\left[\sum_{t=2q+2}^{T}V_4(t)\right]^2 \leqslant & \left\{\sum_{j=1}^{q}\sum_{l=1}^{q}a_T(j)a_T(l)\right.\\
& \times\left\{E\left[\sum_{t=2q+2}^{T}\iiiint\tilde{L}_{q,t}^{j,l}(\mathbf{x}_1,\mathbf{y}_1,\mathbf{x}_2,\mathbf{y}_2)\right.\right.\\
& \times H_{j,t-2q-1}(\mathbf{x}_1,\mathbf{y}_1)H_{l,t-2q-1}(\mathbf{x}_2,\mathbf{y}_2)dW(\mathbf{x}_1)\\
& \left.\left.\left.\times dW(\mathbf{x}_2)dW(\mathbf{y}_1)dW(\mathbf{y}_2)\right]^2\right\}^{\frac{1}{2}}\right\}^2\\
= & O\left(qp^4/T\right). \quad ■
\end{aligned}
$$

Proof of Lemma A.9. Recalling the definition of $L_{q,s}^{j,l}(\mathbf{x}_1,\mathbf{y}_1,\mathbf{x}_2,\mathbf{y}_2)$ in (A.28), we have

$$
\begin{aligned}
& E\left[\sum_{s=\max(j,l)}^{t-2q-1}L_s^{j,l}(\mathbf{x}_1,\mathbf{y}_1,\mathbf{x}_2,\mathbf{y}_2)\right]^2\\
= & \sum_{|s-\tau|\le 2q}E\left[L_s^{j,l}(\mathbf{x}_1,\mathbf{y}_1,\mathbf{x}_2,\mathbf{y}_2)L_\tau^{j,l}(\mathbf{x}_1,\mathbf{y}_1,\mathbf{x}_2,\mathbf{y}_2)\right]\\
& +\sum_{|s-\tau|>2q}\beta^{\frac{\delta}{1+\delta}}(2q)\left\{E\left[L_s^{j,l}(\mathbf{x}_1,\mathbf{y}_1,\mathbf{x}_2,\mathbf{y}_2)\right]^{2(1+\delta)}\right\}^{\frac{1}{2(1+\delta)}}
\end{aligned}
$$

$$\times \left\{E\left[L_{\tau}^{j,l}(\mathbf{x}_1,\mathbf{y}_1,\mathbf{x}_2,\mathbf{y}_2)\right]^{2(1+\delta)}\right\}^{\frac{1}{2(1+\delta)}} = O(tq).$$

It follows that

$$E\left[\sum_{t=2q+2}^{T} V_5(t)\right]^2 \leqslant \left\{\sum_{t=2q+2}^{T}\sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l) \times \iiiint \{E[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0)Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0) \times \psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)]\} \times \left\{E\left[\sum_{s=\max(j,l)}^{t-2q-1} L_s^{j,l}(\mathbf{x}_1,\mathbf{y}_1,\mathbf{x}_2,\mathbf{y}_2)\right]^2\right\}^{\frac{1}{2}} \times dW(\mathbf{x}_1)dW(\mathbf{x}_2)dW(\mathbf{y}_1)dW(\mathbf{y}_2)\right\}^2 = O\left(qp^4/T\right). \quad \blacksquare$$

Proof of Lemma A.10. The result that $E\left[\sum_{t=2q+2}^{T} V_6(t)\right]^2 = O(qp^2/T)$ by Minkowski's inequality and

$$E[V_6(t)]^2 \leqslant \left\{\sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l)\iiiint \{E[Z_t(\mathbf{x}_1,\boldsymbol{\theta}_0) \times Z_t(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{t-j}(\mathbf{y}_1)\psi_{t-l}(\mathbf{y}_2)]\} \times \left\{\sum_{s=\max(j,l)}^{t-2q-1} E\left[Z_s(\mathbf{x}_1,\boldsymbol{\theta}_0)\psi_{s-j}(\mathbf{y}_1) \times \sum_{s-\tau\le 2q} Z_\tau(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{\tau-l}(\mathbf{y}_2)\right]^2\right\}^{\frac{1}{2}} \times dW(\mathbf{x}_1)dW(\mathbf{x}_2)dW(\mathbf{y}_1)dW(\mathbf{y}_2)\right\}^2 \leqslant Ctq\left[\sum_{j=1}^{q} a_T(j)\right]^4 = O\left(tqp^4/T^4\right). \quad \blacksquare$$

Proof of Lemma A.11. The result that $E|\left[\sum_{t=2q+2}^{T} V_7(t)\right]^2 = O(Tpq^{-\frac{\nu\delta}{1+\delta}+1})$ follows from Minkowski's inequality, $p\to\infty$, and the fact that

$$\begin{aligned}
E|[V_7(t)]^2 = {} & E\left\{\sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l)\iiiint E\left[Z_{q,j+l}(\mathbf{x}_1,\boldsymbol{\theta}_0)\right.\right. \\
& \times Z_{q,j+l}(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{q,l}(\mathbf{y}_1)\psi_{q,j}(\mathbf{y}_2)\big] \\
& \times \sum_{s=2q+2}^{t-2q-1} Z_{q,s}(\mathbf{x}_1,\boldsymbol{\theta}_0)\psi_{q,s-j}(\mathbf{y}_1) \\
& \times \sum_{\tau=l+1}^{s-2q-1} Z_{q,\tau}(\mathbf{x}_2,\boldsymbol{\theta}_0)\psi_{q,\tau-l}(\mathbf{y}_2)dW(\mathbf{x}_1) \\
& \left.\times dW(\mathbf{x}_2)dW(\mathbf{y}_1)dW(\mathbf{y}_2)\right\}^2 \\
= {} & \sum_{j_1=1}^{q}\sum_{j_2=1}^{q}\sum_{l_1=1}^{q}\sum_{l_2=1}^{q} a_T(j_1)a_T(j_2)a_T(l_1)a_T(l_2) \\
& \times \int_{\mathbb{R}^{8d}} E[Z_0(\mathbf{x}_{11},\boldsymbol{\theta}_0)Z_0(\mathbf{x}_{21},\boldsymbol{\theta}_0)\psi_{-j_1} \\
& \times (\mathbf{y}_{11})\psi_{-l_1}(\mathbf{y}_{21})] \\
& \times E[Z_0(\mathbf{x}_{12},\boldsymbol{\theta}_0)Z_0(\mathbf{x}_{22},\boldsymbol{\theta}_0)\psi_{-j_2}(\mathbf{y}_{12})\psi_{-l_2}(\mathbf{y}_{22})] \\
& \times \sum_{s=2q+2}^{t-2q-1} E\left[Z_s(\mathbf{x}_{11},\boldsymbol{\theta}_0)Z_s(\mathbf{x}_{12},\boldsymbol{\theta}_0)\right. \\
& \left.\times \psi_{s-j_1}(\mathbf{y}_{11})\psi_{s-j_2}(\mathbf{y}_{12})\right] \\
& \times \sum_{\tau=\max(l_1,l_2)+1}^{s-2q-1} E\left[Z_\tau(\mathbf{x}_{21},\boldsymbol{\theta}_0)Z_\tau(\mathbf{x}_{22},\boldsymbol{\theta}_0)\right. \\
& \left.\times \psi_{\tau-l_1}(\mathbf{y}_{21})\psi_{\tau-l_2}(\mathbf{y}_{22})\right] \\
& \times dW(\mathbf{x}_{11})dW(\mathbf{x}_{12})dW(\mathbf{x}_{21})dW(\mathbf{x}_{22}) \\
& \times dW(\mathbf{y}_{11})dW(\mathbf{y}_{12})dW(\mathbf{y}_{21})dW(\mathbf{y}_{22}) \\
& + \sum_{j_1=1}^{q}\sum_{j_2=1}^{q}\sum_{l_1=1}^{q}\sum_{l_2=1}^{q} a_T(j_1)a_T(j_2)a_T(l_1)a_T(l_2) \\
& \times \int_{\mathbb{R}^{8d}} E[Z_0(\mathbf{x}_{11},\boldsymbol{\theta}_0)Z_0(\mathbf{x}_{21},\boldsymbol{\theta}_0) \\
& \times \psi_{-j_1}(\mathbf{y}_{11})\psi_{-l_1}(\mathbf{y}_{21})]E[Z_0(\mathbf{x}_{12},\boldsymbol{\theta}_0)Z_0(\mathbf{x}_{22},\boldsymbol{\theta}_0) \\
& \times \psi_{-j_2}(\mathbf{y}_{12})\psi_{-l_2}(\mathbf{y}_{22})]\beta^{\frac{\delta}{1+\delta}}(2q) \\
& \times E\left\{\left[\sum_{s_1=2q+2}^{t-2q-1}\sum_{s_2=2q+2}^{t-2q-1} Z_{s_1}(\mathbf{x}_{11},\boldsymbol{\theta}_0)Z_{s_2}(\mathbf{x}_{12},\boldsymbol{\theta}_0)\right.\right. \\
& \left.\left.\times \psi_{s_1-j_1}(\mathbf{y}_{11})\psi_{s_2-j_2}(\mathbf{y}_{12})\right]^{2(1+\delta)}\right\}^{\frac{1}{2(1+\delta)}} \\
& \times E\left\{\left[\sum_{\tau_1=2q+2}^{s_1-2q-1}\sum_{\tau_2=2q+2}^{s_2-2q-1} Z_{\tau_1}(\mathbf{x}_{21},\boldsymbol{\theta}_0)Z_{\tau_2}(\mathbf{x}_{22},\boldsymbol{\theta}_0)\right.\right. \\
& \left.\left.\times \psi_{\tau_1-j_1}(\mathbf{y}_{21})\psi_{\tau_2-j_2}(\mathbf{y}_{22})\right]^{2(1+\delta)}\right\}^{\frac{1}{2(1+\delta)}} \\
& \times dW(\mathbf{x}_{11})dW(\mathbf{x}_{12})dW(\mathbf{x}_{21})dW(\mathbf{x}_{22}) \\
& \times dW(\mathbf{y}_{11})dW(\mathbf{y}_{12})dW(\mathbf{y}_{21})dW(\mathbf{y}_{22}) \\
= {} & O(t^2pT^{-4}) + O\left(t^3pq^{-\frac{\nu\delta}{1+\delta}+1}T^{-4}\right),
\end{aligned}$$

given Assumption A.6. $\blacksquare$

Proof of Theorem 2. The proof of Theorem 2 consists of the proofs of Theorems A.3 and A.4.

Theorem A.3. *Under the conditions of Theorem 2,* $(p^{\frac{1}{2}}/T)(\hat{Q}-\tilde{Q})\xrightarrow{p} 0$.

Theorem A.4. *Under the conditions of Theorem 2,*

$$(p^{\frac{1}{2}}/T)\tilde{Q}\xrightarrow{p} D^{-\frac{1}{2}}\iint\int_{-\pi}^{\pi}|F(\omega,\mathbf{x},\mathbf{y})-F_0(\omega,\mathbf{x},\mathbf{y})|^2 d\omega dW(\mathbf{x},\mathbf{y}).$$

Proof of Theorem A.3. It suffices to show that

$$T^{-1}\iint\sum_{j=1}^{T-1}k^2(j/p)T_j\left[|\hat{\Gamma}_j(\mathbf{x},\mathbf{y})|^2-|\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})|^2\right]dW(\mathbf{x},\mathbf{y})\xrightarrow{p}0, \tag{A.31}$$

$p^{-1}(\hat{C}-\tilde{C}) = O_P(1)$, and $p^{-1}(\hat{D}-\tilde{D})\xrightarrow{p}0$, where $\tilde{C}$ and $\tilde{D}$ are defined in the same way as $\hat{C}$ and $\hat{D}$ in (3.7), with $\boldsymbol{\theta}_0$ replaced by $\hat{\boldsymbol{\theta}}$. Since the proofs for $p^{-1}(\hat{C}-\tilde{C}) = O_P(1)$ and $p^{-1}(\hat{D}-\tilde{D})\xrightarrow{p}0$ are straightforward, we focus on the proof of (A.31). From (A.5), the Cauchy–Schwarz inequality, and the fact that $T^{-1}\iint\sum_{j=1}^{T-1}k^2(j/p)T_j[\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})]^2dW(\mathbf{x},\mathbf{y}) = O_P(1)$ as is implied by Theorem A.4 (the proof of Theorem A.4 does not depend on Theorem A.3), it suffices to show that $T^{-1}\hat{A}_1\xrightarrow{p}0$, where $\hat{A}_1$ is defined as in (A.2). Given (A.3), we shall show that $T^{-1}\iint\sum_{j=1}^{T-1}k^2(j/p)T_j[\hat{B}_{aj}(\mathbf{x},\mathbf{y})]^2\ dW(\mathbf{x},\mathbf{y})\xrightarrow{p}0$, $a = 1, 2$. We first consider $a = 1$. By the Cauchy–Schwarz inequality and $|\psi_{t-j}(\mathbf{y})|\le 2$, we have

$$\begin{aligned}\hat{B}_{1j}(\mathbf{x},\mathbf{y})^2 &\le CT_j^{-1}\sum_{t=j+1}^{T}\left[G(\mathbf{x}|\mathcal{I}_{t-1},\boldsymbol{\theta}_0)-G\left(\mathbf{x}|\mathcal{I}_{t-1},\hat{\boldsymbol{\theta}}\right)\right]^2\\ &\le CT_j^{-1}T\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right\|^2T^{-1}\\ &\quad\times\sum_{t=1}^{T}\sup_{\mathbf{x}\in\mathbb{R}^N}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}G\left(\mathbf{x}|\mathcal{I}_{t-1},\bar{\boldsymbol{\theta}}\right)\right\|^2\\ &= O_p\left(T^{-1}\right).\end{aligned}$$

It follows from (A.4) and Assumption A.6 that

$$T^{-1}\iint\sum_{j=1}^{T-1}k^2(j/p)T_j\hat{B}_{1j}^2(\mathbf{x},\mathbf{y})dW(\mathbf{x},\mathbf{y}) = O_P(p/T).$$

The proof for $a = 2$ is similar. This completes the proof for Theorem A.3. ■

Proof of Theorem A.4. The proof is similar to Hong (1999, Proof of Theorem 5), for the case $(m,l) = (0,0)$. The consistency result follows from (a) $p^{-1}\sum_{j=1}^{T}k^4(j/p)\to\int_0^\infty k^4(j/p)$; (b) $\iint\int_{-\pi}^{\pi}|\hat{F}(\omega,\mathbf{x},\mathbf{y})-F(\omega,\mathbf{x},\mathbf{y})|^2d\omega dW(\mathbf{x},\mathbf{y})\to 0$; (c) $\hat{C} = O_p(p)$; (d) $\hat{D}\to^p D$. Part (a) follows from Assumption A.5. The proof of (c) and (d) is straightforward by Markov's inequality. We will focus on the proof of (b). Define the pseudoestimator

$$\tilde{F}(\omega,\mathbf{x},\mathbf{y}) = \frac{1}{2\pi}\sum_{j=1-T}^{T-1}(1-|j|/T)^{1/2}k(j/p)\tilde{\Gamma}_j(\mathbf{x},\mathbf{y})e^{-ij\omega}.$$

By the C_r inequality,

$$\begin{aligned}&\frac{1}{4}E\iint\int_{-\pi}^{\pi}\left|\hat{F}(\omega,\mathbf{x},\mathbf{y})-F(\omega,\mathbf{x},\mathbf{y})\right|^2d\omega dW(\mathbf{x},\mathbf{y})\\ &\le E\iint\int_{-\pi}^{\pi}|\tilde{F}(\omega,\mathbf{x},\mathbf{y})-E\tilde{F}(\omega,\mathbf{x},\mathbf{y})|^2d\omega dW(\mathbf{x},\mathbf{y})\\ &\quad+\iint\int_{-\pi}^{\pi}|E\tilde{F}(\omega,\mathbf{x},\mathbf{y})-F(\omega,\mathbf{x},\mathbf{y})|^2d\omega dW(\mathbf{x},\mathbf{y})\\ &\quad+E\iint\int_{-\pi}^{\pi}|\hat{F}(\omega,\mathbf{x},\mathbf{y})-\tilde{F}(\omega,\mathbf{x},\mathbf{y})|^2d\omega dW(\mathbf{x},\mathbf{y}).\end{aligned}\tag{A.32}$$

For the first term,

$$\begin{aligned}&E\iint\int_{-\pi}^{\pi}|\tilde{F}(\omega,\mathbf{x},\mathbf{y})-E\tilde{F}(\omega,\mathbf{x},\mathbf{y})|^2d\omega dW(\mathbf{x},\mathbf{y})\\ &\le C(p/T)p^{-1}\sum_{|j|<T}k^2(j/p) = O(p/T).\end{aligned}\tag{A.33}$$

For the second term,

$$\begin{aligned}&\iint\int_{-\pi}^{\pi}|E\tilde{F}(\omega,\mathbf{x},\mathbf{y})-F(\omega,\mathbf{x},\mathbf{y})|^2d\omega dW(\mathbf{x},\mathbf{y})\\ &=\sum_{|j|<T}\left[(1-|j|/T)^{1/2}k(j/p)-1\right]^2\\ &\quad\times\int\left|\Gamma_j(\mathbf{x},\mathbf{y})\right|^2dW(\mathbf{x},\mathbf{y})\\ &\quad+\sum_{|j|>T}\int\left|\Gamma_j(\mathbf{x},\mathbf{y})\right|^2dW(\mathbf{x},\mathbf{y})\\ &= o(1).\end{aligned}\tag{A.34}$$

For the last term,

$$\begin{aligned}&E\iint\int_{-\pi}^{\pi}|\hat{F}(\omega,\mathbf{x},\mathbf{y})-\tilde{F}(\omega,\mathbf{x},\mathbf{y})|^2d\omega dW(\mathbf{x},\mathbf{y})\\ &\le Cp/T^2\left[p^{-1}\sum_{|j|<T}(1-|j|/T)^{-1}k^2(j/p)\right]\\ &= O\left(p/T^2\right).\end{aligned}\tag{A.35}$$

Part (b) follows from combining (A.32)–(A.35). ■

References

Ait-Sahalia, Y., Fan, J., Peng, H., 2009. Nonparametric transition-based tests for diffusions. Journal of the American Statistical Association 104, 1102–1116.
Andersen, T., Lund, J., 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., Chen, J., 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., Chen, Z., 2008. Testing multivariate distributions in GARCH models. Journal of Econometrics 143, 19–36.
Bhardwaj, G., Corradi, V., Swanson, N.R., 2008. A simulation based specification test for diffusion processe. Journal of Business and Economic Statistics 26, 176–193.
Bierens, H.J., 1984. Model specification testing of time series regressions. Journal of Econometrics 26, 323–353.
Bierens, H.J., Wang, L., 2012. Integrated conditional moment tests for parametric conditional distributions. Econometric Theory 28, 328–362.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.
Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. A capital asset pricing model with time-varying covariances. Journal of Political Economy 96, 116–131.
Bowsher, C., 2007. Modelling security market events in continuous time: intensity based, multivariate point process models. Journal of Econometrics 141, 867–912.
Brooks, C., Burke, S., Persand, G., 2005. Autoregressive conditional Kurtosis. Journal of Financial Econometrics 3, 399–421.
Brown, B.M., 1971. Martingale limit theorems. Annals of Mathematical Statistics 42, 59–66.
Chauvet, M., Hamilton, J., 2006. Dating business cycle turning points. In: Milas, C., Rothman, P., Dijk, D.V. (Eds.), Nonlinear Time Series Analysis of Business Cycles. North Holland.
Christoffersen, P.F., Diebold, F., 1997. Optimal prediction under asymmetric loss. Econometric Theory 13, 808–817.
Clements, M.P., Krolzig, H.M., 2003. Business cycle asymmetries: characterization and testing based on Markov-switching autoregressions. Journal of Business and Economic Statistics 21, 196–211.
Corradi, V., Swanson, N., 2006a. Bootstrap conditional distribution tests in the presence of dynamic misspecification. Journal of Econometrics 133, 779–806.
Corradi, V., Swanson, N., 2006b. Predictive density evaluation, in: Granger, C., G. Elliot and A. Timmerman (Eds.) Handbook of Economic Forecasting, Amsterdam, pp. 197–284.
Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64, 247–254.
Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 1A, 33–43.
DeJong, D., Dave, C., 2007. Structural Macroeconometrics. Princeton University Press, New Jersey.
Engle, R.F., 2002a. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business and Economic Statistics 20, 339–350.
Engle, R.F., 2002b. New Frontiers for ARCH models. Journal of Applied Econometrics 17, 425–446.

Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150.
Engle, R.F., Russell, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.
Engle, R.F., Russell, J.R., 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.
Escanciano, J., Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. Journal of Econometrics 134, 151–185.
Fan, Y., Li, Q., Min, I., 2006. A nonparametric bootstrap test of conditional distributions. Econometric Theory 22, 587–613.
Gallant, A.R., Hsieh, D., Tauchen, G., 1997. Estimation of stochastic volatility models with diagnostics. Journal of Econometrics 81, 159–192.
Gallant, A.R., Tauchen, G., 1998. Reprojecting partially observed systems with application to interest rate diffusions. Journal of the American Statistical Association 93, 10–24.
Geweke, J., Amisano, G., 2007. Hierarchical Markov normal mixture models with applications to financial asset returns. Working Paper, University of Brescia.
Gordon, N., Salmond, D., Smith, A., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings F140, 107–113.
Granger, C., 1999. Empirical Modeling in Economics. Cambridge University Press, Cambridge.
Granger, C., 2003. Time series concepts for conditional distributions. Oxford Bulletin of Economics and Statistics 65, 639–701.
Hall, P., Heyde, C., 1980. Martingale Limit Theory and Its Application. Academic Press.
Hamilton, J., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.
Hamilton, J., 1990. Analysis of time series subject to changes in regime. Journal of Econometrics 45, 39–70.
Hamilton, J., 1994. Time Series Analysis. Princeton University Press, New Jersey.
Hannan, E., 1970. Multiple Time Series. Wiley, New York.
Hansen, B.E., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.
Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.
Harvey, C.R., Siddique, A., 1999. Autoregressive conditional Skewness. Journal of Financial and Quantitative Analysis 34, 465–487.
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 Royal Statistical Society. Series B 60, 429–453.
Hong, Y., 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.
Hong, Y., Lee, Y.J., 2005. Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. Review of Economic Studies 72, 499–541.
Hu, L., 2006. Dependence patterns across financial markets: a mixed copula approach. Applied Financial Economics 16, 717–729.
Khmaladze, E.V., 1981. Martingale approach in the theory of goodness-of-fit tests. Theory of Probability and its Applications 26, 240–257.
Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. Journal of Computational and Graphical Statistics 5, 1–25.
Lee, T.H., Long, X., 2009. Copula-based multivariate GARCH model with uncorrelated dependent errors. Journal of Econometrics 150, 207–218.
Li, F., Tkacz, G., 2006. A consistent test for conditional density functions with time dependent data. Journal of Econometrics 133, 863–886.
Linton, O., Gozalo, P., 1997. Conditional independence restrictions: testing and estimation. Discussion Paper, Cowles Foundation for Research in Economics, Yale University.
Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets. Journal of Finance 56, 649–676.
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.
Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filters. Journal of the American Statistical Association 94, 590–599.
Rothschild, M., Stiglitz, J., 1971. Increasing risk: I. A definition. Journal of Economic Theory 2, 225–243.
Rothschild, M., Stiglitz, J., 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., White, H., 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.
Tse, Y.K., Tsui, A.K., 2002. A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. Journal of Business and Economic Statistics 20, 351–362.
Yoshihara, K., 1976. Limiting behavior of U-statistics for stationary, absolutely regular processes. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 35, 237–252.
Zheng, X., 2000. A consistent test of conditional parametric distributions. Econometric Theory 16, 667–691.

Please cite this article in press as: Chen, B., Hong, Y., A unified approach to validating univariate and multivariate conditional distribution models in time series. Journal of Econometrics (2013), http://dx.doi.org/10.1016/j.jeconom.2013.08.004