Generalized spectral testing for multivariate continuous-time models

Bin Chen a, *, Yongmiao Hong b, c

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

ARTICLE INFO

Article history:
Received 17 December 2008
Received in revised form 21 May 2011
Accepted 1 June 2011
Available online 15 June 2011

JEL classification:
C4
E4
G0

Keywords:
Affine jump–diffusion model
Conditional characteristic function
Discrete-time distribution model
Generalized cross-spectrum
Lévy processes
Model specification test
Multivariate continuous-time model

ABSTRACT

We develop an omnibus specification test for multivariate continuous-time models using the conditional characteristic function, which often has a convenient closed-form or can be accurately approximated for many multivariate continuous-time models in finance and economics. The proposed test fully exploits the information in the joint conditional distribution of underlying economic processes and hence is expected to have good power in a multivariate context. A class of easy-to-interpret diagnostic procedures is supplemented to gauge possible sources of model misspecification. Our tests are also applicable to discrete-time distribution models. Simulation studies show that the tests provide reliable inference in finite samples.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Multivariate continuous-time models have proved to be versatile and productive tools in finance and economics (e.g., Andersen et al. (2002), Chernov et al. (2003), Dai and Singleton (2003), Pan and Singleton (2008), Jarrow et al. (2010), and Piazzesi (2010)). Compared with that of discrete-time models, the econometric analysis of continuous-time models is often challenging. In the past decade or so, substantial progress has been made in developing estimation methods for continuous-time models. However, relatively little effort has been devoted to specification and evaluation of continuous-time models. A continuous-time model essentially specifies the transition density of underlying processes. Model misspecification generally renders inconsistent parameter estimators and their variance–covariance matrix estimators, yielding misleading conclusions in inference and hypothesis testing.

Correct model specification is also crucial for valid economic interpretations of model parameters. More importantly, a misspecified model can lead to large errors in pricing, hedging and managing risk. However, economic theories usually do not suggest any concrete functional form for continuous-time models; the choice of a model is somewhat arbitrary. A prime example of this practice is the pricing and hedging literature, where continuous-time models are generally assumed to have a functional form that results in a closed-form pricing formula. The models, though convenient, are often incorrect or suboptimal. To avoid this pitfall, the development of reliable specification tests for continuous-time models is necessary.

In a pioneer paper, Ait-Sahalia (1996a) develops a nonparametric test for univariate diffusion models. By observing that the drift and diffusion functions completely characterize the stationary (i.e., marginal) density of a diffusion model, Ait-Sahalia (1996a) compares the model-implied stationary density with a smoothed kernel density estimator based on discretely sampled data. Gao and King (2004) develop a simulation procedure to improve the finite sample performance of Ait-Sahalia’s (1996a) test. These tests are...
convenient to implement. Nevertheless, they may pass over a misspecified model that has a correct stationary density.

Hong and Li (2005) develop an omnibus nonparametric specification test for continuous-time models. The test uses the transition density, which depicts the full dynamics of a continuous-time process. When a univariate continuous-time model is correctly specified, the probability integral transform (PIT) of data via the model-implied transition density is i.i.d. \( U(0, 1) \). Hong and Li (2005) check the joint hypothesis of i.i.d. \( U(0, 1) \) by using a smoothed kernel estimator of the joint density of the PIT series.

However, Hong and Li’s (2005) approach cannot be extended to a multivariate model. The PIT series with respect to a model-implied multivariate transition density are no longer i.i.d \( U(0, 1) \), even if the model is correctly specified. In an application, Hong and Li (2005) evaluate multivariate affine term structure models (ATSMs) for interest rates by using the marginal PIT for each state variable. This is analogous to Singleton’s (2001) conditional characteristic function-based limited-information maximum likelihood (LML-CCF) estimator in a related context, which is based on the conditional density function of an individual state variable. As pointed out by Singleton (2001), “the LML-CCF estimator fully exploits the information in the conditional likelihood function of the individual \( y_{i,t} \) but not the information in the joint conditional distribution of \( y_{i,t} \).” This generally leads to an asymptotic efficiency loss in estimation. Similarly, Hong and Li’s (2005) test, when applied to each state variable of a multivariate process, may fail to detect misspecification in the joint dynamics of state variables. In particular, the test may easily overlook misspecification in the conditional correlations between state variables. Moreover, the use of the transition density may not be computationally convenient because the transition densities of most continuous-time models have no closed-form.

Gallant and Tauchen (1996) propose a class of Efficient Method of Moments (EMM) tests that can be used to test multivariate continuous-time models. They propose a \( y^2 \) test for model misspecification, and a class of appealing diagnostic \( t \)-tests that can be used to gauge possible sources for model failure. Since these tests are by-products of the EMM algorithm, they cannot be used when the model is estimated by other methods. This may limit the scope of these tests’ otherwise useful applications. Bharadwaj et al. (2008) consider a simulation-based test, which is an extension of Andrews’ (1997) conditional Kolmogorov test, for multivariate diffusion processes. The limiting distribution of the test statistic is not nuisance parameter free and hence asymptotically critical values must be obtained via the block bootstrap, which may be time-consuming.

There have been other tests for univariate diffusion models in the recent literature. Ait-Sahalia et al. (2009) propose some tests by comparing the model-implied transition density and distribution function with their nonparametric counterparts. Chen et al. (2008) also propose a transition density-based test using a nonparametric empirical likelihood approach. Li (2007) focuses on the parametric specification of the diffusion function by measuring the distance between the model-implied diffusion function and its kernel estimator. These approaches could be extended to multivariate continuous-time models. However, all these tests maintain the Markov assumption for the data generating process (DGP), and consider the finite order lag only. If the DGP is non-Markov, these tests may miss some dynamic misspecifications.

This paper proposes a new approach to testing the adequacy of a multivariate continuous-time model that uses the full information of the joint dynamics of state variables. In a multivariate context, modeling the joint dynamics of state variables is important in many applications (e.g., Jacobson (1988)). For example, as the conditional correlations between asset returns change over time, the specific weight allocated to each asset within a portfolio should be adjusted accordingly. Similarly, hedging requires knowledge of conditional correlations between the returns on different assets within the hedge. Conditional correlations are also important in pricing structured products such as rainbow options, which are based on multiple underlying assets whose prices are correlated. In the term structure literature, models of interest rate term structure impose dynamic cross-sectional restrictions, as implied by the no-arbitrage condition on bond yields of different maturities. This joint dynamics can be used to investigate the transmission mechanism that transfers the impact of government policy from spot rates to longer-term yields. There is a conflict between the flexibilities of modeling instantaneous conditional variances and instantaneous conditional correlations of bond yields. Dai and Singleton (2000) find that not only do swap rate data consistently call for negative conditional correlations between bond yields, but also factor correlations help explain the shape of the term structure of bond yields’ volatilities. Indeed, as Engle (2002) points out, “the quest for reliable estimates of correlations between financial variables has been the motivation for countless academic articles, practitioner conferences and Wall Street research”.

There has been a long history of using the CF in estimation and hypotheses testing in statistics and econometrics. To name a few, Koutriouvelis (1980) constructs a chi-squared goodness-of-fit test for simple null hypotheses with empirical characteristic function (ECF). Fan (1997) takes the ECF approach to testing multivariate distributions. But both tests maintain the i.i.d. assumption and hence are not suitable for the time series data. Recently, Su and White (2007) test conditional independence by comparing the unrestricted and restricted CCFs via a kernel regression. All above works deal with discrete-time models, in recent years, the CF approach has attracted an increasing attention in the continuous-time literature. For most continuous-time models, the transition density has no closed-form, which makes estimation of and testing for continuous-time models rather challenging. However, for a general class of affine jump–diffusion (AJD) models (e.g., Duffie et al. (2000)) and time-changed Lévy processes (e.g., Carr and Wu (2003, 2004)), the CCF has a closed-form as an exponential-affine function of state variables up to a system of ordinary differential equations. This fact has been exploited to develop new estimation methods for multifactor continuous-time models in the literature. Specifically, Chacko and Viceira (2003) suggest a spectral GMM estimator based on the average of the differences between the ECF and the model-implied CF. Jiang and Knight (2002) derive the unconditional joint CF of an AJD model and use it to develop some GMM and ECF estimation procedures. Singleton (2001) proposes both time-domain estimators based on the Fourier transform of the CCF, and frequency-domain estimators directly based on the CCF. By extending Carrasco and Florens (2000), Carrasco et al. (2007) propose GMM estimators with a continuum of moment conditions (C-GMM) via the CF. All these estimation methods differ in their uses of the conditional information set. As of yet, no attempt has been made in the literature to use the CF to test continuous-time models, although Carrasco et al. (2007) do mention that “the future work will have to refine our results on estimation of non-Markovian processes and latent states as well as develop tests in the framework of CF-based continuum of moment conditions”.

In light of the convenient closed-form of the CCF of AJD models, we provide a new test of the adequacy of multivariate continuous-time models. The multivariate continuous-time models can include jumps and the underlying DGPs of observable state variables need not be Markov. Naturally, as a special case, our test can be used to check univariate continuous-time models. Compared with the existing tests for continuous-time models in the literature, our approach has several main advantages.

First, because the CCF is the Fourier transform of the transition density, our omnibus test fully exploits the information in the
joint transition density of state variables rather than only the information in the transition densities of individual state variables. Thus, it can detect misspecifications in the joint transition density even if the transition density of each state variable is correctly specified. When the underlying multivariate continuous-time process is Markov, our omnibus test is consistent against any model misspecification. This is unattainable by some existing tests for multivariate continuous-time models. Moreover, we have used a novel generalized cross-spectral approach, which embeds the CCF in a spectral framework, thus enjoying the appealing features of spectral analysis. For example, it checks many lags. This is particularly useful when the DGP is non-Markov.

Second, besides the omnibus test, we propose a class of diagnostic tests by differentiating the generalized cross-spectrum of the state vector. These tests can evaluate how well a continuous-time model captures various specific aspects of the joint dynamics and they are easy to interpret. In particular, these tests can provide valuable information about neglected dynamics in conditional means, conditional variances, and conditional correlations of state variables, respectively. Therefore, they complement Gallant and Tauchen's (1996) popular EMM-based individual test. All our omnibus test and diagnostic tests are derived from a unified framework.

Third, our tests are applicable to a wide variety of continuous-time models and discrete-time multivariate distribution models, since we impose regularity conditions on the CCF of discretely observed samples with some fixed sample frequency, rather than on stochastic differential equations (SDEs). By using the CCF to characterize the adequacy of a model, our tests are most convenient whenever the model has a closed-form CCF and many popular continuous-time models in finance (e.g., the class of multivariate AJD models and the class of time-changed Lévy processes) have a closed-form CCF, although they have no closed-form transition density. Of course, our tests can also evaluate the multivariate continuous-time models with no closed-form CCF. In this case, we need to recover the model-implied CCF by using inverse Fourier transforms or simulation methods. Unlike tests based on CFS in the statistical literature, which often have nonstandard asymptotic distributions, our tests have a convenient null asymptotic distribution.

Fourth, we do not require a particular estimation method. Any \( \sqrt{T} \)-consistent parametric estimators can be used. Parameter estimation uncertainty does not affect the asymptotic distribution of our test statistics. One can proceed as if true model parameters were known and equal to parameter estimates. This makes our tests easy to implement, particularly in view of the notorious difficulty of estimating multivariate continuous-time models. The only inputs needed to calculate the test statistics are the discretely observed data and the model-implied CCF.

In Section 2, we introduce the framework, state the hypotheses, and characterize the correct specification of a multivariate continuous-time model. In Section 3, we propose a generalized cross-spectral omnibus test, and in Section 4 we derive the asymptotic null distribution of our omnibus test and discuss its asymptotic power property. In Section 5, we develop a class of generalized cross-spectral derivative tests that focuses on various specific aspects of the joint dynamics of a time series model. In Section 6, we assess the reliability of the asymptotic theory in finite samples by simulation. Section 7 concludes our work. All mathematical proofs are collected in Appendix. A GAUSS code to implement our tests is available from the authors upon request. Throughout the paper, we will use \( C \) to denote a generic bounded constant, \( \| \cdot \| \) for the Euclidean norm, and \( A^* \) for the complex conjugate of \( A \).

2. Hypotheses of interest

For concreteness, we focus on a multivariate continuous-time setup. For a given complete probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and an information filtration \(\{\mathcal{F}_t\}\), we assume that a \(N \times 1\) state vector \(X_t\) is a continuous-time DGP in some state space \( \mathbb{D} \subset \mathbb{R}^N \). We permit but do not require \( X_t \) to be Markov, which is often assumed in the continuous-time modeling in finance and macroeconomics.

In financial modeling, the following class \( \mathcal{M} \) of continuous-time models is often used to capture the dynamics of \( X_t^j \):

\[
dX_t^j = \mu(X_t^j, \theta) \, dt + \sigma(X_t^j, \theta) \, dW_t^j + df_t^j(\theta), \quad \theta \in \Theta, \tag{2.1}
\]

where \( W_t \) is an \( N \times 1 \) standard Brownian motion in \( \mathbb{R}^N \), \( \Theta \) is a finite-dimensional parameter space, \( \mu: \mathbb{D} \times \Theta \to \mathbb{R}^N \) is a drift function (i.e., instantaneous conditional mean), \( \sigma: \mathbb{D} \times \Theta \to \mathbb{R}^{N \times N} \) is a diffusion function (i.e., instantaneous conditional standard deviation), and \( f_t \) is a pure jump process whose jump size follows a probability distribution \( v: \mathbb{D} \times \Theta \to \mathbb{R}^+ \) and whose jump times arrive with intensity \( \lambda: \mathbb{D} \times \Theta \to \mathbb{R}^+ \). We allow some state variables to be unobservable. One example is the stochastic volatility (SV) model, where the volatility, which is a proxy for the information inflow, is a latent process.

The setup (2.1) is a general multivariate specification that nests most existing continuous-time models in finance and economics. For example, suppose we restrict the drift \( \mu(\cdot, \cdot) \), the instantaneous covariance matrix \( \sigma(\cdot, \cdot) \sigma(\cdot, \cdot)' \) and the jump intensity \( \lambda(\cdot, \cdot) \) to be affine functions of the state vector \( X_t \); namely,

\[
\begin{align*}
\mu(X_t^j, \theta) &= K_0 + K_j X_t^j, \\
\left[ \sigma(X_t^j, \theta) \sigma(X_t^j, \theta)' \right]_t &= \left[ H_0 \right]_t + \left[ H_1 \right]_t X_t, \\
\lambda(X_t^j, \theta) &= L_0 + L_j' X_t,
\end{align*}
\tag{2.2}
\]

where \( K_0 \in \mathbb{R}^N, K_j \in \mathbb{R}^{N \times N}, H_0 \in \mathbb{R}^{N \times N}, H_j \in \mathbb{R}^{N \times N \times N}, L_0 \in \mathbb{R}^N \), and \( L_j \in \mathbb{R}^{N \times N} \) are unknown parameters. Then we obtain the class of AJD models of Duffie et al. (2000).

It is well known that for a continuous-time model characterized by a SDE, the specification of the drift \( \mu(X_t, \theta) \), the diffusion \( \sigma(X_t, \theta) \) and the jump process \( f_t(\theta) \) completely determines the joint transition density of \( X_t \). We use \( p(x, t|\mathcal{F}_s, \theta_0) \) to denote the model-implied transition density of \( X_t = x \) given \( \mathcal{F}_s \), where \( s < t \). Suppose \( X_t \) has a true transition density, say \( p_0(x, t|\mathcal{F}_s) \). Then the continuous-time model is correctly specified for the full dynamics of \( X_t \) if there exists some parameter value \( \theta_0 \in \Theta \) such that

\[
\|p_0(\cdot, t|\mathcal{F}_s, \theta_0) - p(\cdot, t|\mathcal{F}_s, \theta)\|_\infty \to 0 \quad \text{almost surely (a.s.)}
\tag{2.3}
\]

\[\] Our test is applicable to both continuous-time and discrete-time models but we focus on a continuous-time setup due to the following reasons. Our approach is most convenient when the CCF has a closed-form and many popular continuous-time models in finance have no closed-form transition density, but do have a closed-form CCF. On the other hand, to our knowledge, no CF-based test is available to check the specification of continuous-time models, although the CF approach has been used in the estimation of continuous-time models. Hence, our test nicely fills the gap in the literature.

4 However, Easterly and O'Hara (1992) develop an economic structural model and show that financial time series, such as prices and volumes, are likely non-Markov. Despite the fact that most continuous-time models characterized by SDEs in the literature are Markov, it is still important to allow \( X_t \) to be non-Markov. Even the null model is Markov, to allow for non-Markov DGPs under the alternative will ensure the power of the test against a wider range of misspecification, particularly, dynamic misspecification. On the other hand, the model may be Markov but involves some latent variables, as is the case of SV models. As a result, observable state variables themselves will be non-Markov, even under the null.

6 We assume that the functions \( \mu, \sigma, \nu \) and \( \lambda \) are regular enough to have a unique strong solution to (2.1). See (e.g.) Art-Sahalia (1996a), Duffie et al. (2000) and Genon-Catalot et al. (2000) for more discussions.
Alternatively, if for all $\theta \in \Theta$, we have

\[ h \models \Phi \models \quad \text{for some } s < t \text{ with positive probability measure,} \]

then the continuous-time model is misspecified for the full dynamics of $X_t$.

The transition density characterization can be used to test correct specification of a continuous-time model. When $X_t$ is univariate, Hong and Li (2005) propose a kernel-based test for a continuous-time model by checking whether

\[ Z_t(\theta_0) = \int_{-\infty}^{X_1} \cdots \int_{-\infty}^{X_N} \int_{-\infty}^{X_t} \cdots \int_{-\infty}^{X_t} p(x, t | I_{t-\Delta}, \theta_0) \; dx \sim \text{i.i.d.} U[0,1]. \tag{2.5} \]

which holds under $H_0$, where $I_{t-\Delta} = \{X_{t-\Delta}, X_{t-2\Delta}, \ldots \}$ is the information set available at time $t-\Delta$, where $\Delta$ is a fixed sampling interval for an observed sample. The i.i.d. $U[0,1]$ property for the PIT has also been used in other contexts (e.g., Diebold et al. (1998)). However, there are some limitations to this approach. For example, for most continuous-time diffusion models (except such simple diffusion models as Vasicek’s (1977) model), the transition densities, have no closed-form. Most importantly, the PIT cannot be applied to the multivariate joint transition density $p(x, t | F_T, \theta)$, because when $N > 1$,

\[ Z_t(\theta_0) = \int_{-\infty}^{X_1} \cdots \int_{-\infty}^{X_N} p(x, t | I_{t-\Delta}, \theta_0) \; dx \tag{2.6} \]

is no longer i.i.d. $U[0,1]$ even if $H_0$ holds. Hong and Li (2005) suggest using the PIT for each state variable. This is valid, but it does not make full use of the information contained in the joint distribution of $X_t$. In particular, it may miss important model misspecification in the joint dynamics of $X_t$. For example, consider the DGP

\[ d \begin{pmatrix} X_{t+1,1} \\ X_{t+1,2} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix} \begin{pmatrix} \theta_1 - X_{t,1} \\ \theta_2 - X_{t,2} \end{pmatrix} dt \\
+ \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} \begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}, \]

where $W_{1,t}, W_{2,t}$ are independent standard Brownian motions and $\kappa_{21} \neq 0$. Suppose we fit the data using the model

\[ d \begin{pmatrix} X_{t+1,1} \\ X_{t+1,2} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ 0 & \kappa_{22} \end{pmatrix} \begin{pmatrix} \theta_1 - X_{t,1} \\ \theta_2 - X_{t,2} \end{pmatrix} dt \\
+ \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} \begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}, \]

Then this model is misspecified because it ignores correlations in drift. Now, following Hong and Li (2005), we calculate the generalized residuals $\{Z_{t,i}, Z_{t,2,i}, Z_{t,1-i}, Z_{t,2-i}, \ldots \}$, where $Z_{t,i}$ and $Z_{t,2-i}$ are the PITs of $X_{t,1}$ and $X_{t,2}$, with respect to the conditional density models $p(X_{t,1}, t | X_{t-\Delta}, \theta)$ and $p(X_{t,2}, t | X_{t-\Delta}, \theta)$ respectively. Then

\[ \text{Hong and Li’s (2005) test will have no power because each of these PITs is an i.i.d. } U[0,1] \text{ sequence respectively.} \]

As the Fourier transform of the transition density, the CCF can capture the full dynamics of $X_t$. Let $\psi(u, t | F_T, \theta)$ be the model-implied CCF of $X_t$ conditional on $F_T$ at time $s < t$; that is,

\[ \psi(u, t | F_T, \theta) \equiv E_{\theta} \left[ \exp \{iu(X_{t} - X_{s}) \} | F_T \right] \]

\[ = \int_{\mathbb{R}^N} \exp \{iu(x) \} p(x, t | F_T, \theta) \; dx, \quad u \in \mathbb{R}^N, \quad i = \sqrt{-1}, \tag{2.7} \]

where $E_{\theta} (| F_T)$ denotes the conditional expectation under the model-implied transition density $p(x, t | F_T, \theta)$. Note that for a Markov model, the filtration $F_T$ can be replaced by $X_t$.

Given the equivalence between the transition density and the CCF, we can write the hypotheses of interest $H_0$ in (2.3) versus $H_A$ in (2.4) as follows:

\[ H_0 : E \left[ \exp \{iuX_{t} \} | F_T \right] = \varphi(u, t | F_T, \theta_0) \quad \text{a.s. for all } u \in \mathbb{R}^N \]

versus

\[ H_A : E \left[ \exp \{iuX_{t} \} | F_T \right] \neq \varphi(u, t | F_T, \theta) \quad \text{with positive probability for all } \theta \in \Theta. \tag{2.8} \]

Suppose we have a discrete random sample $\{X_t\}_{t=1}^T$ of size $T$. For notational simplicity, we set $\Delta = 1$ below, where time is measured in units of the sampling interval of data.\(^7\) Also, we denote $I_{t-1} = \{X_{t-1}, X_{t-2}, \ldots \}$, the information set available at time $t-1$. Define the process

\[ Z_t(u, \theta) \equiv \exp \{iuX_{t} \} - \varphi(u, t | I_{t-1}, \theta_0), \quad u \in \mathbb{R}^N. \tag{2.9} \]

Then $H_0$ is equivalent to the following martingale difference sequence (MDS) characterization:

\[ E \left[ Z_t(u, \theta_0) | I_{t-1} \right] = 0 \quad \text{for all } u \in \mathbb{R}^N \]

and some $\theta_0 \in \Theta$, a.s.\(^8\)

We may call $Z_t(u, \theta)$ a “CCF-based generalized residual” of the continuous-time model $M$. This can be seen obviously from the auxiliary regression

\[ \exp \{iuX_{t} \} = \varphi(u, t | I_{t-1}, \theta_0) + Z_t(u, \theta_0), \tag{2.12} \]

where $\varphi(u, t | I_{t-1}, \theta_0)$ is the regression model for the dependent variable $\exp \{iuX_{t} \}$ conditional on information set $I_{t-1}$, and $Z_t(u, \theta_0)$ is a MDS regression disturbance (under $H_0$).

In principle, we can always obtain the CCF by the inverse Fourier transform, provided the transition density of $X_t$ is given. Even if the CCF or the transition density has no closed-form, we can accurately approximate the model transition density by using (e.g.) the Hermite expansion method of Ait-Sahalia (2002), the simulation methods of Brandt and Santa-Clara (2002) and Pedersen (1995), or the closed-form approximation method of Duffie et al. (2003), and then calculating the Fourier transform. Nevertheless, our test is most convenient when the CCF has a closed-form.

AJD models are a class of continuous-time models with a closed-form CCF, developed and popularized by Dai and Singleton (2000), Duffie and Kan (1996), and Duffie et al. (2000). These models have proven fruitful in capturing the dynamics of economic variables, such as interest rates, exchange rates and stock prices. It has been shown (e.g., Duffie et al. (2000)) that for AJDs, the CCF of the vector conditional on $I_{t-1}$ is a closed-form exponential-affine function of $X_{t-1}$:

\[ \psi(u, t | I_{t-1}, \theta) = \exp \left[ a_{t-1}(u) + \beta_{t-1}(u) X_{t-1} \right], \tag{2.13} \]

where $a_{t-1} : \mathbb{R}^N \rightarrow \mathbb{R}$ and $\beta_{t-1} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ satisfy the complex-valued Riccati equations:

\[ \dot{\beta}_t = K_t \beta_t + \frac{1}{2} \beta_t H_t H_t^* \beta_t + L_t (g(\beta_t) - 1), \]

\[ \dot{a}_t = K_t \beta_t + \frac{1}{2} \beta_t H_t H_t^* \beta_t + L_0 (g(\beta_t) - 1), \tag{2.14} \]

with boundary conditions $\beta_{T} = i u_0$ and $a_T(u) = 0$.\(^8\)

\(^7\) Our approach can be adapted to the case of $\Delta \neq 1$ easily. See Singleton (2001, p. 117) for related discussion.

\(^8\) Assuming that the spot rate is an affine function of the state vector $X$, and that $X$ follows an affine diffusion, Duffie and Kan (1996) show that the yield of the zero coupon bond has a closed-form CCF. Assuming that the spot rate is a quadratic function of the normally distributed state vector, Ahn et al. (2002) also derive the closed-form CCF for the yield of the zero coupon bond.
Affine SV models are another popular class of multivariate continuous-time models with a closed-form CCF (e.g., Heston (1993), Bates (1996, 2000), Bakshi et al. (1997), Das and Sundaram (1999)). SV models can capture salient properties of volatility such as randomness and persistence. Basic affine SV models, affine SV models with Poisson jumps and affine SV models with Lévy processes have been widely used in modeling asset return dynamics as they allow for closed-form solutions for European option prices (e.g., Heston (1993), Das and Sundaram (1999), Chacko and Viceira (2003), Carr and Wu (2003, 2004), and Huang and Wu (2003)).

To test SV models, where $V_t$ is a latent process, we need to modify the characterization (2.11) to make it operable. Generally, we put $X_t \equiv (X_{t,1}, X_{t,2})'$, where $X_{t,1} \in \mathbb{R}^{N_1}$ denotes the observable state variables, $X_{t,2} \in \mathbb{R}^{N_2}$ denotes the unobservable state variables, and $N_1 + N_2 = N$. Also, partition $u$ conformably as $u = (u_1', u_2')$. Then we can define

$$Z_{t,1}(u_1, \theta) = \exp \left( (u_1' X_{t,1}) - \phi(u_1, t, I_{t,1-1}, \theta) \right),$$

where

$$\phi(u_1, t, I_{t,1-1}, \theta) = E_0[\exp \left( (u_1' X_{t,1}) | I_{t,1-1} \right)],$$

$$I_{t,1-1} = \{X_{t,1-1} = (x_{t,1-1-1}, x_{t,2-1-1}, \ldots) \text{ is the information set on the observables that are available at time } t \text{ and the second equality follows from the law of iterated expectations. Then we have}$$

$$E \left[ Z_{t,1}(u_1, \theta_0) | I_{t,1-1} \right] = 0 \quad \text{for all } u_1 \in \mathbb{R}^{N_1} \text{ and some } \theta_0 \in \Theta \text{ a.s.} \quad (2.15)$$

This provides a basis for constructing operational tests for multivariate continuous-time models with partially observable state variables.

Although the model-implied CCF $\psi(u, t, I_{t-1}, \theta)$, where $I_{t-1} = (I_{t-2}, I_{t-1-1}, \ldots)$, may have a closed-form, the conditional expectation $\phi(u_1, t, I_{t,1-1}, \theta)$ generally has no closed-form. However, one can approximate it accurately by using some simulation techniques. For almost all continuous-time models characterized by SDEs in the literature, the CCF $\psi(u, t, I_{t-1}, \theta) = \psi(u_1, t, X_{t-1}, \theta)$ is a Markov process. In this case, we find

$$E_0[\psi(u_1' X_{t,1}) | I_{t,1-1}]$$

$$= \int \psi(u_1' y', t, | X_{t,1-1}, x_{t,2-1}, \ldots) \frac{p(x_{t,2-1})}{p(x_{t,2-1})} dx_{t,2-1},$$

where $p(x_{t,2-1})$ is the model-implied transition density of the unobservable $X_{t,2-1}$. Given the observable information $I_{t,1-1}$, noting that models with latent variables are the leading examples that may yield non-Markov observables in the continuous-time literature, we discuss several popular methods to estimate the model-implied CCF based on observables. We first consider particle filters, which have been developed by Gordon et al. (1993), Pitt and Shephard (1999) and Johannes et al. (2009). The term “particle” was first used by Kitagawa (1996) in this literature to mean the simulated discrete data with random support. Particle filters are the class of simulation filters that recursively approximate the filtering random variable $X_{t,2-1} | I_{t,1-1}, \theta$ by “particles” $\tilde{X}_{t,1}^1, \tilde{X}_{t,2}^1, \ldots, \tilde{X}_{t,1}^N$ with discrete probability mass of $\pi_{t,1}^1, \pi_{t,1}^2, \ldots, \pi_{t,1}^N$. Hence a continuous variable is approximately a discrete one with random support. These discrete points are viewed as samples from $p(x_{t,2-1} | I_{t,1-1}, \theta)$ and as $f \to \infty$, the particles can approximate the conditional density increasingly well.

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

$$p(x_{t,2-1} | I_{t,1}, \theta)$$

$$= p(x_{t,2-1} | x_{t,2-1}, I_{t,2-1}, \theta) \frac{p(x_{t,2-1} | I_{t,2-1}, \theta)}{p(x_{t,2-1} | I_{t,1-1}, \theta)}.$$

We can approximate $p(x_{t,2-1} | I_{t,1}, \theta)$ up to some proportionality; namely,

$$\hat{p}(x_{t,2-1} | I_{t,1-1}, \theta) \propto \hat{p}(x_{t,1-1} | \tilde{X}_{t,2-1}, \tilde{I}_{t,2-1}, \theta)$$

$$\times \sum_{j=1}^L \pi_{t,1}^j \hat{p}(x_{t,2-1} | \tilde{X}_{t,2-1}^j | \tilde{I}_{t,2-1}, \theta),$$

where $\hat{p}(x_{t,1-1} | \tilde{X}_{t,2-1}, \tilde{I}_{t,2-1}, \theta)$ and $\sum_{j=1}^L \pi_{t,1}^j \hat{p}(x_{t,2-1} | \tilde{X}_{t,2-1}^j | \tilde{I}_{t,2-1}, \theta)$ can be viewed as the likelihood and prior respectively. As pointed out by Gordon et al. (1993), the particle filters require that the likelihood function can be evaluated and that $X_{t,2-1}$ can be sampled from $\hat{p}(x_{t,1-1} | \tilde{X}_{t,2-1}, \tilde{I}_{t,2-1}, \theta)$, these can be achieved by using time-discretized solutions to the SDEs.

To implement particle filters, we can use Johannes et al. (2009) and Pitt and Shephard (1999) algorithm. First we generate a simulated sample $\{X_{t,1}^{t,0}\}_{j=0}^n$, where $X_{t,1}^{t,0} = (x_{t,1-1}^{t,0}, x_{t,2-1}^{t,0}, \ldots, x_{t,2-1}^{t,0})$ and $M$ is an integer. Then we simulate them one step forward, evaluate the likelihood function, and set

$$\pi_{t,1}^{j-1} = \hat{p}(x_{t,1-1} | X_{t,1}^{t,0} | \tilde{I}_{t,2-1,0}, \theta), \quad j = 1, \ldots, M.$$
With a suitable choice of \( L \) via some information criteria such as AIC or BIC, we can approximate \( p(x_{2t-1}, t - 1 | I_{1t-1}, \theta) \) 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.

For models whose \( CCF \) is exponentially affine in \( X_{t-1} \), we can also adopt Bates’ (2007) approach to compute \( \hat{\phi}(u, t | I_{1t-1}, \theta) \). First, at time \( t = 1 \), we initialize the \( CCF \) of the latent vector \( X_{2t-1} \) conditional on \( I_{1t-1} \), its unconditional \( CCF \). Then, by exploiting the Markov property and the affine structure of the \( CCF \), we can evaluate the model-implied \( CCF \) conditional on data observed through period \( t \), namely, \( E_0[\psi(u, t | X_{t-1}, \theta) | I_{1t-1}] \), and thus an estimator for \( \phi(u, t | I_{1t-1}, \theta) \) is obtained.

### 3. Omnibus testing

We now propose omnibus and spectral envelope tests for the adequacy of a multivariate continuous-time model by using the \( CCF \) characterization in (2.11) or (2.15). For notational convenience, below we focus on the case where \( X_t \) is fully observable. The proposed procedures are readily applicable to the case where only \( X_{2t} \) is observable, with \( X_{2t} \rightarrow X_{1t} \) and \( I_{1t-1} \) replacing \( Z_t \) and \( I_{1t-1} \) respectively. It is not a trivial task to check (2.11) because the \( MDS \) property must hold for each \( u \in R^N \), and because \( Z_t \) may display serial dependence in higher order conditional moments. Any test for (2.11) should be robust to time-varying conditional heteroskedasticity and higher order moments of unknown form in \( Z_t \). To check the \( MDS \) property of \( Z_t \), we substantively extend Hong’s (1999) univariate generalized spectrum to a multivariate generalized spectrum. The general spectral test is an analytic tool for nonlinear time series that embeds the \( CCF \) in a spectral framework. It can capture nonlinear dynamics while maintaining the nice spectral properties.

Because the dimension of \( Z_{t-1} \) can be infinite, we encounter the so-called “curse of dimensionality” problem in checking the \( MDS \) property in (2.11). Fortunately, the generalized spectral approach provides a solution to tackle this difficulty. It checks many lags in a pairwise manner, thus avoiding the “curse of dimensionality”.

Define the generalized covariance function

\[
I_j(u, v) = \text{cov}[Z_t(u, \theta), \psi(iW_t - j)], \quad u, v \in R^N.
\]

With the generalized cross-covariance \( I_j(u, v) \), we can define the generalized cross-spectrum

\[
F(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} I_j(u, v) \exp(-ij\omega),
\]

\[
\omega \in [-\pi, \pi], \quad u, v \in R^N,
\]

(3.2)

where \( \omega \) is the frequency. This is the Fourier transform of the generalized covariance function \( \{I_j(u, v)\} \) and thus contains the same information as \( \{I_j(u, v)\} \). An advantage of generalized cross-spectral analysis is that it can capture cyclical patterns caused by both linear and nonlinear cross dependence. Examples include volatility spillover, the comovements of tail distribution clustering between state variables, and asymmetric spillover of business cycles across different sectors or countries. Another attractive feature of \( F(\omega, u, v) \) is that it does not require the existence of any moment condition on \( X_t \).

Under \( H_0 \), we have \( I_j(u, v) = 0 \) for all \( u, v \in R^N \) and all \( j \neq 0 \). Consequently, the generalized cross-spectrum \( F(\omega, u, v) \) becomes a “flat” spectrum:

\[
F(\omega, u, v) = F_0(\omega, u, v) = \frac{1}{2\pi} \Gamma(\omega, u, v),
\]

\[
\omega \in [-\pi, \pi], \quad u, v \in R^N.
\]

(3.3)

We can test \( H_0 \) by checking whether a consistent estimator for \( F(\omega, u, v) \) is flat with respect to frequency \( \omega \). Any significant deviation from a flat generalized cross-spectrum is evidence of model misspecification.

Suppose we have a discretely observed sample \( \{X_t\}_{t=1}^{T} \) of size \( T \). Then we estimate the generalized covariance \( \hat{I}_j(u, v) \) by its sample analogue

\[
\hat{I}_j(u, v) = \frac{1}{T} \sum_{t=|j|+1}^{T} \hat{Z}_t(u, \hat{\theta}) \left[ \exp(iW_t - j) - \psi(t) \right],
\]

\[
u, v \in R^N,
\]

(3.4)

where

\[
\hat{Z}_t(u, \hat{\theta}) \equiv \psi(iW_t) - \psi(t | I_{1t-1}, \hat{\theta}).
\]

\( \hat{I}_j(u, v) \) is the observed information set available at time \( t \) that may involve certain initial values, \( \hat{\theta} \) is a \( \sqrt{T} \)-consistent estimator for \( \theta_0 \), and \( \psi(t) = (T - |j|)^{-1} \sum_{t=|j|+1}^{T} \exp(iW_t - j) \) is the empirical unconditional \( CF \) of \( X_t \). We note that the information set \( I_{1t-1} = \{X_{1t-1}, X_{2t-1}, \ldots\} \) dates back to past infinity and is not feasible. Thus, when \( X_t \) is non-Markov, we may need to assume some initial values in computing \( \psi(t) \) at \( I_{1t-1} \). Therefore, we have to replace \( I_{1t-1} \) with a truncated information set \( I_{1t-1}^{j} \), which contains some initial values. We provide a condition (see Assumption A.5 in Section 4) to ensure that the use of initial values has no impact on the asymptotic distribution of the proposed test statistics.

A consistent estimator for \( \hat{F}_0(\omega, u, v) \) is

\[
\hat{F}_0(\omega, u, v) = \frac{1}{2\pi} \hat{I}_0(u, v), \quad \omega \in [-\pi, \pi], \quad u, v \in R^N.
\]

(3.5)

Consistent estimation for \( F(\omega, u, v) \) is more challenging. We use a smooth kernel estimator

\[
\hat{F}(\omega, u, v) = \frac{1}{2\pi} \sum_{t=-\infty}^{T-1} \left( 1 - |j|/T \right)^{1/2} k(j/p) \hat{I}_j(u, v) e^{-ij\omega},
\]

\[
\omega \in [-\pi, \pi], \quad u, v \in R^N,
\]

(3.6)

where \( p \equiv p(T) \rightarrow \infty \) is a bandwidth or an effective lag order, and \( k : R \rightarrow [-1, 1] \) is a kernel function, assigning weights to various lags. Examples of \( k(\cdot) \) include the Bartlett kernel, the Parzen kernel and the Quadratic-Spectral (QS) kernel. In (3.6), the factor \((1 - |j|/T)^{1/2}\) is a finite sample correction. It could be replaced by unity. Under regularity conditions, \( \hat{F}(\omega, u, v) \) and \( \hat{F}_0(\omega, u, v) \) are consistent for \( F(\omega, u, v) \) and \( F_0(\omega, u, v) \) respectively. These estimators converge to the same limit under \( H_0 \) but they generally converge to different limits under \( H_1 \), giving the power of the test.

We can measure the distance between \( \hat{F}(\omega, u, v) \) and \( \hat{F}_0(\omega, u, v) \) by the quadratic form

\[
\frac{\pi T}{2} \int \int_{-\pi}^{\pi} \left| \hat{F}(\omega, u, v) - \hat{F}_0(\omega, u, v) \right|^2 d\omega dW(u) dW(v),
\]

(3.7)

where the equality follows by Parseval’s identity. \( W : R^N \rightarrow R^+ \) is a positive nondecreasing right-continuous function, and the unspecified integrals are all taken over the support of \( W(\cdot) \). An example of \( W(\cdot) \) is the \( N(0, I_N) \) CDF, where \( I_N \) is a \( N \times N \) identity.
matrix. The function $W(\cdot)$ can also be a step function, analogous to the CDF of a discrete random vector.

Our omnibus test statistic for $H_0$ against $H_A$ is a standardized version of (3.7):

$$
\tilde{Q}(\mathbf{0}, \mathbf{0}) = \left[ \sum_{j=1}^{T-1} k(j/p)(T-j) \right] \times \int \int \left| \tilde{F}_{j}(\mathbf{u}, \mathbf{v})^2 \right| dW(\mathbf{u}) dW(\mathbf{v}) - \tilde{C}(\mathbf{0}, \mathbf{0}) \right] \sqrt{\tilde{D}(\mathbf{0}, \mathbf{0})},
$$

where the centering and scaling factors

$$
\tilde{C}(\mathbf{0}, \mathbf{0}) = \sum_{j=1}^{T-1} k(j/p)(T-j)^{-1} \sum_{l=j+1}^{T} \int \int \tilde{Z}_l(\mathbf{u}, \mathbf{v})^2 \times \left| \tilde{\psi}_{l-j}(\mathbf{v})^2 \right| dW(\mathbf{u}) dW(\mathbf{v}),
$$

$$
\tilde{D}(\mathbf{0}, \mathbf{0}) = 2 \sum_{j=1}^{T-1} \sum_{l=j+1}^{T} k(j/p)^2(l/p)
$$

The factors $\tilde{C}(\mathbf{0}, \mathbf{0})$ and $\tilde{D}(\mathbf{0}, \mathbf{0})$ are the approximately mean and variance of the quadratic form in (3.7). They have taken into account the impact of higher order structure in the generalized residual $\{Z_t(\mathbf{u}, \mathbf{v})\}$. As a result, the $\tilde{Q}(\mathbf{0}, \mathbf{0})$ test is robust to conditional heteroskedasticity and time-varying higher order conditional moments of unknown form in $\{Z_t(\mathbf{u}, \mathbf{v})\}$.

In practice, when $W(\cdot)$ is continuous, $\tilde{Q}(\mathbf{0}, \mathbf{0})$ can be calculated by numerical integration or simulation. This may be computationally costly when the dimension $N$ of $\mathbf{X}_t$ is large. Alternatively, one can use a finite number of grid points for $\mathbf{u}$ and $\mathbf{v}$, which is equivalent to using a discrete CDF. For example, we can generate finitely many numbers of $\mathbf{u}$ and $\mathbf{v}$ from the $N(\mathbf{0}, I_N)$ distribution. This will reduce the computational cost, but at a cost of some power loss.

4. Asymptotic theory

To derive the null asymptotic distribution of the test statistic $\tilde{Q}(\mathbf{0}, \mathbf{0})$ and invest its asymptotic power property, we impose following regularity conditions.

**Assumption A.1.** A discrete-time sample $\{\mathbf{X}_t\}_{t=\Delta}^T$, where $\Delta \equiv 1$ is the sampling interval, is observed at equally spaced discrete times.

**Assumption A.2.** Let $\psi(\mathbf{u}, t, t_{j-1}, \Theta)$ be the CCF of $\mathbf{X}_t$ given $t_{j-1} \equiv \{\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots\}$ for a time series model for $\mathbf{X}_t$. (i) For each $\Theta \in \Theta$, each $\mathbf{u} \in \mathbb{R}^N$, and each $t$, $\psi(\mathbf{u}, t, t_{j-1}, \Theta)$ is measurable with respect to $\mathcal{F}_{j-1}$; (ii) for each $\Theta \in \Theta$, each $\mathbf{u} \in \mathbb{R}^N$, and each $t$, $\psi(\mathbf{u}, t, t_{j-1}, \Theta)$ is twice continuously differentiable with respect to $\Theta$ with probability one; (iii) $\sup_{\Theta \in \Theta} \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E \sup_{\Theta \in \Theta} \left\| \frac{\partial^2}{\partial \Theta^2} \psi(\mathbf{u}, t, t_{j-1}, \Theta) \right\| = C$ and $\sup_{\Theta \in \Theta} \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E \sup_{\Theta \in \Theta} \left\| \frac{\partial}{\partial \Theta} \psi(\mathbf{u}, t, t_{j-1}, \Theta) \right\| = C$.

**Assumption A.3.** $\hat{\Theta}$ is a parameter estimator such that $\sqrt{T}(\hat{\Theta} - \Theta^*) = O_p(1)$, where $\Theta^* \equiv \lim_{T \to \infty} \Theta$ and $\Theta^* = \Theta_0$ under $H_0$.

**Assumption A.4.** $\{\mathbf{X}_t, \psi(\mathbf{u}, t, t_{j-1}, \Theta_0)\}, \frac{\partial}{\partial \Theta} \psi(\mathbf{u}, t, t_{j-1}, \Theta_0)$ is a strictly stationary $\beta$-mixing process with the mixing coefficient $|\beta(\lambda)| \leq \beta^* \leq C^{-1}$ for some constant $\beta > 2$.

**Assumption A.5.** Let $\mathbf{J}_t$ be the observed feasible information set available at time $t$ that may involve certain initial values. Then $\sup_{\Theta \in \Theta^0} \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E \sup_{\Theta \in \Theta^0} \left\| \psi(\mathbf{u}, t, t_{j-1}, \Theta) - \psi(\mathbf{u}, t, t_{j-1}, \Theta^*) \right\|^2 \leq C$.

**Assumption A.6.** $k : \mathbb{R} \to [-1, 1]$ 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 $k(z) \leq c |z|^{\alpha b}$ for some $b > \frac{1}{2}$ as $z \to \infty$.

**Assumption A.7.** $W : \mathbb{R}^{K} \to \mathbb{R}^{+}$ is a nondecreasing right-continuous function with $\int_{\mathbb{R}^{K}} dW(\mathbf{u}) < \infty$ and $\int_{\mathbb{R}^{K}} \left\| \mathbf{u} \right\| dW(\mathbf{u}) < \infty$. Furthermore, $W = \int w dv$ for some nonnegative weighting function $w$ and some measure $v$, where $W$ is absolutely continuous with respect to $v$ and $w$ is symmetric about the origin.

**Assumption A.1** imposes some regularity conditions on the discretely observed random sample. Both univariate and multivariate continuous-time or discrete-time processes are covered, and we allow but do not require $\mathbf{X}_t$ to be Markov. This distinguishes our test from all other tests in the literature. We emphasize that it is important to allow the DGP to be non-Markov, even if one is interested in testing the adequacy of a Markov model. This is because the misspecification of a Markov model may come from not only the improper specification in functional forms, but also the violation of the Markov assumption. The non-Markov assumption for the DGPs of observable state variables is also necessary when some state variables are latent, as is the case of SV models.

There are two kinds of asymptotics in the literature on continuous-time models. The first is to let the sampling interval $\Delta \to 0$. This implies that the number of observations per unit of time tends to infinity. The second is to let the time horizon $T \to \infty$. As argued by Ait-Sahalia (1996b), the first approach hardly matches the way in which new data are added to the sample. Even if such ultra-high-frequency data are available, market micro-structural problems are likely to complicate the analysis considerably. Hence, like Ait-Sahalia (1996b) and Singleton (2001), we fix the sampling interval $\Delta$ and derive the asymptotic properties of our test for an expanding sampling period (i.e., $T \to \infty$). Unlike Ait-Sahalia (1996a,b), however, we do not impose additional conditions on the SDEs for $\mathbf{X}_t$. Instead, we impose conditions directly on the model-implied CCF, which is equivalent to imposing conditions on the model-implied transition density. For these reasons, our approach is more general and the proposed tests are applicable to both continuous-time and discrete-time models.

**Assumption A.2** imposes regularity conditions on the CCF of the multivariate time series model. As the CCF is the Fourier transform of the transition density, we can ensure **Assumption A.2** by imposing the following conditions on the model-implied transition density $p(\mathbf{x}, t, t_{j-1}, \Theta)$; (i) for each $\mathbf{x} \in \mathbf{D}$, and each $\Theta \in \Theta$, $p(\mathbf{x}, t, t_{j-1}, \Theta)$ is measurable with respect to $\mathcal{F}_{j-1}$; (ii) for each $\Theta \in \Theta$, each $\mathbf{u} \in \mathbb{R}^N$, and each $t$, $\psi(\mathbf{u}, t, t_{j-1}, \Theta)$ is twice continuously differentiable with respect to $\Theta$ with probability one; and (iii) $\sup_{\Theta \in \Theta} \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E \sup_{\Theta \in \Theta} \left\| \frac{\partial}{\partial \Theta} \psi(\mathbf{u}, t, t_{j-1}, \Theta) \right\| = C$.

**Assumption A.3** requires a $\sqrt{T}$-consistent estimator $\hat{\Theta}$ under $H_0$. Examples include asymptotically optimal and suboptimal estimators, such as Gallant and Tauchen’s (1996) EMM, Singleton’s (2001) ML-CCF and GMM-CCF, Ait-Sahalia’s (2002) and Ait-Sahalia and Kimmel’s (2010) approximated MLE, Carrasco et al.’s (2007) C-GMM, and Chib et al.’s (2010) MCMC method. We do not require any asymptotically most efficient estimator or a specified estimator. This is attractive for practitioners given the notorious
difficulty of efficient estimation of many multivariate nonlinear time series models.

**Assumption A.4** is a regularity condition on the temporal dependence of the process \( \{X_t, \psi(u, t|J_{t-1}, \theta_0), \frac{d}{dt}\psi(u, t|J_{t-1}, \theta_0)\} \). The \( \beta \)-mixing assumption is a standard condition for discrete time series analysis. In the continuous-time context, almost all continuous-time models in the literature are Markov. For Markov models, \( \psi(u, t|J_{t-1}, \theta_0) \) and \( \frac{d}{dt}\psi(u, t|J_{t-1}, \theta_0) \) are functions of \( X_{t-1} \). Thus, **Assumption A.4** holds if the discretely observed random sample \( \{X_{t|1}\} \) is a strictly stationary \( \beta \)-mixing process with \( |\beta (t)| \leq C^{-t} \) for some constant \( \nu > 2.15 \).

**Assumption A.5** is a start-up value condition, which ensures that the impact of initial values (if any) assumed in \( J_{t-1} \) is asymptotically negligible. This condition holds automatically for Markov models and easily for many non-Markov models. For simplicity, we illustrate such an initial value condition by a discrete-time GARCH model: \( X_t = \epsilon_t \sqrt{h_t} \), where \( h_t = \omega + \alpha \epsilon_{t-1} + \beta X_{t-1}^2 \) and \( \{\epsilon_t\} \) is i.i.d. N(0, 1). Here we have \( \psi(u, t|J_{t-1}, \theta) = \exp(-\frac{1}{2}u^2h_t) \), where \( J_{t-1} = \{X_{t-1}, X_{t-2}, \ldots\} \). Furthermore, we have \( J_{t-1} = \{X_{t-1}, X_{t-2}, \ldots, X_0, h_0\} \), where \( h_0 \) is an initial value assumed for \( h_t \). By recourse substitution, we obtain

\[
\begin{align*}
\text{var}(X_t|J_{t-1}, \theta) - \text{var}(X_t|J_{t-1}, \theta) &= \omega + \beta \sum_{j=0}^{t-2} \alpha^j X_{t-1-j}^2 \\
+ \beta \alpha^{t-1} h_0 - \omega - \beta \sum_{j=0}^{t-2} \alpha^j X_{t-1-j}^2 - \beta \alpha^{t-1} h_0.
\end{align*}
\]

It follows that

\[
\sup_{\theta \in \Theta} \left\{ \sum_{t=0}^{T} \mathbb{E} \left[ \frac{1}{2} u^2 \left[ \text{var}(X_t|J_{t-1}, \theta) - \text{var}(X_t|J_{t-1}, \theta) \right] \right] \right\} = \sup_{\theta \in \Theta} \left\{ \sum_{t=0}^{T} \mathbb{E} \left[ \frac{1}{2} \left[ \exp\left(\frac{1}{2}u^2h_t\right) - \exp\left(\frac{1}{2}u^2h_0\right) \right] \right] \right\} \leq C
\]

provided \( \omega > 0, 0 < \alpha < 1, \beta < 1, \alpha + \beta < 1 \) and \( E h_0 < \infty \).

**Assumption A.6** is the regularity condition on the kernel function. The continuity of \( k(\cdot) \) at 0 and \( k(0) = 1 \) ensure that the bias of the generalized cross-spectral estimator \( \hat{F}(u, \theta, v) \) 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 \( \hat{F}(u, \theta, v) \). **Assumption A.6** covers most commonly used kernels. For kernels with bounded support, such as the Bartlett and Parzen kernels, \( b = \infty \). For kernels with unbounded support, \( b \) is some finite positive real number. For example, \( b = 2 \) for the Quadratic-Spectral kernel.

**Assumption A.7** imposes mild conditions on the function \( W(\cdot) \). The CDF of any symmetric distribution with finite fourth moment satisfies **Assumption A.7**. Note that \( W(\cdot) \) can be the CDF of continuous or discrete distribution and \( w(\cdot) \) is its Radon–Nikodym derivative with respect to \( v \). If \( W(\cdot) \) is the CDF of some continuous distribution, \( v \) is Lebesgue measure; if \( W(\cdot) \) is the CDF of some discrete distribution, \( v \) is some counting measure. 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 \( u \) and \( v \). This is equivalent to using the CDF of a discrete random vector.

We now state the asymptotic distribution of the omnibus test \( \hat{Q}(0, 0) \) under \( H_0 \).

**Theorem 1.** Suppose **Assumptions A.1–A.7** hold, and \( p = ct^k \) for \( \frac{x}{12} < c < (\frac{3}{4} + \frac{1}{2\nu})^{-1} \), where \( 0 < c, \delta < \infty \) and \( v \) is defined in **Assumption A.4**. Then \( \hat{Q}(0, 0) \to^p N(0, 1) \) under \( H_0 \) as \( T \to \infty \).

An important feature of \( \hat{Q}(0, 0) \) is that the use of the estimated generalized residuals \( \hat{Z}_t(\theta, \theta) \) in place of the true unobservable generalized residuals \( Z_t(\theta, \theta) \) has no impact on the limiting distribution of \( \hat{Q}(0, 0) \). One can proceed as if the true parameter value \( \theta_0 \) were known and equal to \( \theta \). Intuitively, the parametric estimator \( \hat{\theta} \) converges to \( \theta_0 \) faster than the nonparametric estimator \( \hat{f}(\omega, u, v) \) converges to \( f(\omega, u, v) \) as \( T \to \infty \). Consequently, the limiting distribution of \( \hat{Q}(0, 0) \) is solely determined by \( \hat{F}(\omega, u, v) \), and replacing \( \theta_0 \) by \( \hat{\theta} \) 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), \( \lambda < \frac{1}{4} \) because \( b = \infty \). For the QSK kernel (\( b = 2 \)), \( \lambda < \frac{5}{18} \). These conditions are mild.

Next, we investigate the asymptotic power of \( \hat{Q}(0, 0) \) under \( H_A \).

**Theorem 2.** Suppose **Assumptions A.1–A.7** hold, and \( p = ct^k \) for \( 0 < \lambda < \frac{1}{2} \) and \( 0 < c < \infty \). Then as \( T \to \infty \),

\[
\frac{p^2}{T} \hat{Q}(0, 0) \to^p \frac{1}{\sqrt{D(0, 0)}} \sum_{j=1}^{\infty} \int |f_j(u, v)|^2 \, dW(u) \, dW(v)
\]

\[
= \frac{\pi}{2 \sqrt{D(0, 0)}} \int_0^\pi \int_0^\pi |F(\omega, u, v) - F_0(\omega, u, v)|^2 \, d\omega \, dW(u) \, dW(v),
\]

where

\[
D(0, 0) = 2 \int_0^\infty k^4(\omega) \, d\omega \int_0^\pi |\Sigma_0(u_1, u_2)|^2 \, dW(u_1) \, dW(u_2)
\]

\[
\times \left[ \int_0^\pi \sum_{j=-\infty}^{\infty} |\Omega_j(v_1, v_2)|^2 \, dW(v_1) \, dW(v_2),
\]

and \( \Sigma_0(u, v) = \text{cov}(Z_t(u, \theta^0), Z_t(v, \theta^0)) \) and \( \Omega_j(u, v) = \text{cov}(\text{cov}(u^jX_t, v^jX_t), i) \). The function \( G(\omega, u, v) \) is the generalized spectral density of the state vector \( X_t \). It captures temporal dependence in \( X_t \). The dependence of the constant \( D(0, 0) \) on \( G(\omega, u, v) \) is due to the fact that the conditioning variable \( \text{exp}(ivX_t) \) is a time series process.

Following Bierens (1982) and Stinchcombe and White (1998), we have that for \( j > 0 \), \( f_j(u, v) = 0 \) for all \( u, v \in \mathbb{R}^n \) if and only if \( E Z_t(u, \theta^0) |X_t| = 0 \) a.s. for all \( u, v \in \mathbb{R}^n \). Suppose \( E Z_t(u, \theta^0) |X_t| \neq 0 \) at some lag \( j > 0 \) under \( H_A \). Then we have

\[
\int |f_j(u, v)|^2 \, dW(u) \, dW(v) \geq C > 0 \text{ for any function } W(\cdot)
\]

that is positive, monotonically increasing and continuous, with unbounded support on \( \mathbb{R}^n \). As a result, \( P[\hat{Q}(0, 0) > C(T)] \to 1 \) for
any sequence of constants \(\{C(T) = o(T/p^{1/2})\}\). Thus \(\hat{Q}(0,0)\) has an asymptotic unit power at any given significance level \(\alpha \in (0,1)\), whenever \(E[Z(t, \theta^0)|X_{t-j}|]\) is nonzero at some lag \(j > 0\) under \(H_A\). Note that for a multivariate Markov process \(X_t\), we always have \(E[Z(t, \theta^0)|X_{t-j}|] \neq 0\) at least for some \(j > 0\) under \(H_A\). Hence, \(\hat{Q}(0,0)\) is consistent against \(H_A\) when \(X_t\) is Markov. Gallant and Tauchen’s (1996) EMM test does not have this property.

For a non-Markovian process \(X_t\), the hypothesis that \(E[Z(t, \theta_0)|X_{t-j}|] = 0\) a.s. for all \(u \in R^N\) and some \(\theta_0 \in \Theta\) and all \(j > 0\) is not equivalent to the hypothesis that \(E[Z(t, \theta_0)|X_{t-j-1}|] = 0\) a.s. for all \(u \in R^N\) and some \(\theta_0 \in \Theta\). The latter implies the former but not vice versa. This is the price we have to pay for dealing with the difficulty of the “curse of dimensionality”. Nevertheless, our 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 degrees of freedom. Fortunately, such power loss is substantially alleviated for \(Q(0,0)\), thanks to the downward weighting by \(k^2(v)\) for higher order lags. Generally speaking, the state vector \(X_t\) 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 discount 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 studies below.

5. Diagnostic testing

When a multivariate time series model \(\mathcal{M}\) is rejected by the omnibus test \(Q(0,0)\), it would be interesting to explore possible sources of the rejection. For example, one may like to know whether the misspecification comes from conditional mean/drift dynamics, conditional variance/diffusion dynamics, or conditional correlations between state variables. Such information, if any, will be valuable in reconstructing the model.

The CCF is a convenient and useful tool to gauge possible sources of model misspecification. As is well known, the CCF can be differentiated to obtain conditional moments. We now develop a class of diagnostic tests by differentiating the generalized cross-spectrum \(F^\omega(\omega, u, v)\). This class of diagnostic tests can provide useful information about how well a multivariate time series model can capture the dynamics of various conditional moments and conditional cross-moments of state variables.

Define an \(N \times 1\) index vector \(m = (m_1, m_2, \ldots, m_N)^T\), where \(m_c \geq 0\) for all \(1 \leq c \leq N\), and put \(|m| = \sum_{c=1}^{N} m_c\). Then we define the generalized cross-spectral derivative

\[
F^{0,m,0}(\omega,0,v) = \left[ \frac{\partial m_1}{\partial u_{c1}} \right] \cdots \left[ \frac{\partial m_N}{\partial u_{cN}} \right] F(\omega, u, v) \bigg|_{u=0}
\]

\[
= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} f_j^{(m,0)}(0, v) \exp(-ij\omega),
\]

where the derivative of the generalized cross-covariance function

\[
f_j^{(m,0)}(0, v) = \text{cov}\left\{ \sum_{c=1}^{N} (iX_{ct})^{m_c} \right\}
\]

\[
- \theta^0 \left[ \sum_{c=1}^{N} (iX_{ct})^{m_c} \bigg| J_{t-1} \right] \exp(i\psi X_{t-j}),
\]

Here, as before, \(E_\theta (|J_{t-1}|)\) is the conditional expectation under the model-implied transition density \(p(x, t|J_{t-1}, \theta)\).

To gain insight into the generalized cross-spectral derivative \(F^{0,m,0}(\omega,0,v)\), we consider a bivariate process \(X_t = (X_{1t}, X_{2t})\) and examine the cases of \(|m| = 1\) and \(|m| = 2\) respectively:

- **Case 1:** \(|m| = 1\). We have \(m = (1,0)\) or \(m = (0,1)\). If \(m = (1,0)\),

\[
\gamma_j^{(m,0)}(0,v) = \text{ico} v\left\{ X_{1t} - \theta_0 X_{1t} \big| J_{t-1} \right\} \exp(i\psi X_{t-j}).
\]

If \(m = (0,1)\), then

\[
\gamma_j^{(m,0)}(0,v) = \text{ico} v\left\{ X_{2t} - \theta_0 X_{2t} \big| J_{t-1} \right\} \exp(i\psi X_{t-j}).
\]

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

- **Case 2:** \(|m| = 2\). We have \(m = (2,0)\), \((0,2)\) or \((1,1)\). If \(m = (2,0)\),

\[
\gamma_j^{(m,0)}(0,v) = i^2 \text{cov}\left\{ X_{1t}^2 - \theta_0 X_{1t}^2 \big| J_{t-1} \right\} \exp(i\psi X_{t-j}).
\]

Finally, if \(m = (1,1)\),

\[
\gamma_j^{(m,0)}(0,v) = i^2 \text{cov}\left\{ X_{1t} X_{2t} - \theta_0 X_{1t} X_{2t} \big| J_{t-1} \right\} \exp(i\psi X_{t-j}).
\]

Thus, the choice of \(|m| = 2\) can be used to check model misspecifications in the conditional volatility of state variables as well as their conditional correlations.

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

\[
\hat{Q}(m,0) = \left[ \sum_{j=1}^{T-1} k_j^2 (j/p)(T-j) \right]^2 \int \left| \hat{f}_j^{(m,0)}(0,v) \right|^2 dW(v) - \hat{C}(m,0),
\]

\[
\hat{C}(m,0) = \sum_{j=1}^{T-1} k_j^2 (j/p) \frac{1}{T-j} \times \sum_{t=m(j)+1}^{T} \left| \hat{Z}_t^{(m)}(0, \theta) \right|^2 dW(v),
\]

\[
\hat{D}(m,0) = 2 \sum_{j=1}^{T-2} \sum_{t=m(j)+1}^{T-1} \left| \hat{Z}_t^{(m)}(0, \theta) \right|^2 dW(v_1)dW(v_2),
\]

and

\[
\hat{Z}_t^{(m)}(0, \theta) = \prod_{c=1}^{N} (iX_{ct})^{m_c} - \theta_0 \prod_{c=1}^{N} (iX_{ct})^{m_c} |J_{t-1}|.
\]

To derive the limit distribution of \(\hat{Q}(m,0)\) under \(H_0\), we impose some moment conditions.

**Assumption A8.** (i) \(\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E\sup_{\theta \in \Theta} \left| \frac{\partial m_1}{\partial u_{c1}} \right| \cdots \left| \frac{\partial m_N}{\partial u_{cN}} \psi(u, t|J_{t-1}, \theta) \big|_{u=0} \right|^2 \leq C\);
\( \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E \sup_{u, \theta} \left( \frac{\partial}{\partial \theta} \left( \frac{\partial \psi}{\partial \theta} \right) \right) \leq C; \)

\( \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E \sup_{u, \theta} \left( \frac{\partial}{\partial \theta} \left( \frac{\partial \psi}{\partial \theta} \right) \right) \leq C; \)

Assumption A.9. \( \{ \mathbf{X}_t, \frac{\partial \psi}{\partial \theta} \} \) is a strictly stationary \( \beta \)-mixing process with the mixing coefficient \( |\beta(I)| \leq C^{-s} \) for some constant \( s > 2 \).

Assumption A.10. Let \( J_1 + 1 \) be the observed feasible information set available at time \( t \) that may involve certain initial values. Then

\( \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E \sup_{u, \theta} \left( \frac{\partial}{\partial \theta} \left( \frac{\partial \psi}{\partial \theta} \right) \right) \leq C. \)

Theorem 3. Suppose Assumptions A.1, A.3 and A.6–A.10 hold for some pre-specified \( \mathbf{m} \) and \( p = \mathcal{C}^{T \lambda} \) for \( 0 < \lambda < (3 + \frac{1}{2m})^{-1} \), where \( 0 < \nu < \infty \) and \( \nu \) is defined in Assumption A.9. Then

\( \hat{Q}(\mathbf{m}, 0) \) converges in distribution to \( \mathcal{N}(0, 1) \) under \( \mathbb{H}_0 \).

It should be pointed out that the generalized cross-spectral derivative tests may fail to detect certain specific model misspecifications. In particular, they will fail to detect some time-invariant model misspecifications. For example, suppose a time series model assumes a zero conditional correlation between state variables, while there exists a nonzero but time-invariant conditional correlation. Then the derivative of the generalized covariance in (5.2) is identically zero. In this case, the correlation test will fail to detect such time-invariant conditional correlation. Of course, the tests \( \hat{Q}(\mathbf{m}, 0) \) will generally have power against time-varying misspecifications.

6. Finite sample performance

It is unclear how well the asymptotic theory can provide reliable reference and guidance when applied to financial time series data, which is well known to display highly persistent serial dependence. We now investigate the finite sample performance of the proposed tests for the adequacy of a multivariate affine diffusion model.

6.1. Simulation design

To examine the size of \( \hat{Q}(\mathbf{m}, 0) \) under \( \mathbb{H}_0 \), we consider the following DGP:

- **DGP0** (Uncorrelated Gaussian Diffusion):

\[
\begin{align*}
\mathbf{d} X_{1t} &= \left( \kappa_{11} 0 0 \right) \left( \theta_1 - X_{1t} \right) dt \\
\mathbf{d} X_{2t} &= \left( 0 \kappa_{22} 0 \right) \left( \theta_2 - X_{2t} \right) dt \\
\mathbf{d} X_{3t} &= \left( 0 0 \kappa_{33} \right) \left( \theta_3 - X_{3t} \right) dt
\end{align*}
\]

We set \( \left( \kappa_{11}, \kappa_{22}, \kappa_{33}, \theta_1, \theta_2, \theta_3, \sigma_{11}, \sigma_{22}, \sigma_{33} \right) = \left( 0.5, 1, 2, 0, 0, 1, 1, 1 \right) \). We use the Euler scheme to simulate 1000 data sets of the random sample \( \{ \mathbf{X}_t \}_{t=1}^{T} \) at the monthly frequency for \( T = 250, 500, 1000 \) respectively. These sample sizes correspond to about twenty to one hundred years of monthly data. Each simulated sample path is generated using 120 intervals per month. We then discard 119 out of every 120 observations, obtaining discrete observations at the monthly frequency. For each data set, we use MLE to estimate model (6.1), with no restrictions on the intercept coefficients. All data are generated using the Matlab 6.1 random number generator on a PC.

With a diagonal matrix \( \mathbf{\kappa} = \text{diag}(\kappa_{11}, \kappa_{22}, \kappa_{33}) \), DGP0 is an uncorrelated 3-factor Gaussian diffusion process. As shown in Diebold (2002), the Gaussian diffusion model has analytic expressions for the conditional mean and the conditional variance respectively:

\[
E(\mathbf{X}_t | \mathbf{X}_s) = \left[ 1 - e^{-\kappa(t-s)} \right] \mathbf{\theta} + e^{-\kappa(t-s)} \mathbf{X}_s,
\]

\[
\text{var}(\mathbf{X}_t | \mathbf{X}_s) = e^{-\kappa(t-s)} \sum \Sigma e^{-\kappa(s-m)} dm
\]

where \( \mathbf{\theta} = \left( \theta_1, \theta_2, \theta_3 \right)^T \) and \( \Sigma = \text{diag} \left( \sigma_{11}, \sigma_{22}, \sigma_{33} \right) \).

To investigate the power of \( \hat{Q}(\mathbf{m}, 0) \) in distinguishing model (6.1) from alternative processes, we also generate data from five affine diffusion processes respectively:

- **DGP1** (Correlated Gaussian Diffusion, with Constant Correlation in Drift):

\[
\begin{align*}
\mathbf{d} X_{1t} &= \left( 0.5 0 0 \right) \left( -X_{1t} \right) dt \\
\mathbf{d} X_{2t} &= \left( -0.5 1 0 \right) \left( -X_{2t} \right) dt \\
\mathbf{d} X_{3t} &= \left( 0.5 0.5 2 \right) \left( -X_{3t} \right) dt
\end{align*}
\]

- **DGP2** (Uncorrelated CIR (Cox et al., 1985) Diffusion):

\[
\begin{align*}
\mathbf{d} X_{1t} &= \left( 0.5 0 0 \right) \left( 2 - X_{1t} \right) dt \\
\mathbf{d} X_{2t} &= \left( 0 1 0 \right) \left( 1 - X_{2t} \right) dt \\
\mathbf{d} X_{3t} &= \left( 0 0 2 \right) \left( 1 - X_{3t} \right) dt
\end{align*}
\]
If we use model \( DGP3 \) to fit data generated from \( DGP1 \), we expect that the model tests, \( \hat{Q}(m, 0) \), are correctly specified, but there are dynamic misspecifications in conditional variances and conditional covariances of state variables.

DGP5 is a mixture of Gaussian and CIR diffusion processes, where \( \{X_{1t}, X_{2t}\} \) is a correlated 2-factor CIR diffusion process and \( \{X_{3t}\} \) is a univariate Gaussian process. Under DGP5, model \((6.1)\) is misspecified for the conditional variances of \( X_{1t} \) and \( X_{2t} \) and the conditional covariance between \( X_{1t} \) and \( X_{2t} \). The conditional means of \( X_{1t}, X_{2t} \), and \( X_{3t} \) and the conditional covariances between \( X_{1t} \) and \( X_{2t} \), \( X_{1t} \) and \( X_{2t} \), and \( X_{3t} \) are correctly specified.

For each of DGPs 1–5, we generate 500 data sets of the random sample \( \{X_{1t}\}_{t=1}^{T} \) by the Euler scheme, for \( T = 250 \), 500 and 1000 respectively at the monthly sample frequency. For each data set, we estimate model \((6.1)\) via MLE. Because model \((6.1)\) is misspecified under all five DGPs, our omnibus test \( \hat{Q}(0, 0) \) is expected to have nontrivial power under DGPs 1–5, provided the sample size \( T \) is sufficiently large. We will also examine how diagnostic tests \( \hat{Q}(m, 0) \) for \( |m| > 0 \) can reveal information about various model misspecifications.

### 6.2. Monte Carlo evidence

To reduce computational costs, we generate \( \mathbf{u} \) and \( \mathbf{v} \) from a \( N(0, I_{3}) \) distribution, with each \( \mathbf{u} \) and \( \mathbf{v} \) having 30 grid points in \( \mathbb{R}^{3} \) respectively. We use the Bartlett kernel, which has a 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 \( \hat{Q}(0, 0) \) tests.\(^{(17)}\) 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 cross-spectral \( \hat{F}(\omega, u, v) \), with the Bartlett kernel \( k(\cdot) \) used in some preliminary generalized cross-spectral estimators. To examine the sensitivity of the choice of the preliminary bandwidth \( \hat{p} \) on the size and power of the tests, we consider \( \hat{p} \) in the range of 10–40.\(^{(18)}\)

Table 1 reports the rejection rates (in terms of percentage) of \( \hat{Q}(m, 0) \) under DGP0 at the 10% and 5% significance levels, using the asymptotic theory. The omnibus \( \hat{Q}(0, 0) \) test tends to underreject when \( T = 250 \), but it improves as \( T \) increases. The size \( \hat{Q}(0, 0) \) is not very sensitive to the preliminary lag order \( \hat{p} \). For example, when \( T = 250 \), the rejection rate at the 5% level attains its maximum 3.8% at \( \hat{p} = 10 \), and attains its minimum 2.5% at \( \hat{p} = 31, 33, 34 \). For \( T = 1000 \), the rejection rate at the 5% level attains its maximum 6.2% at \( \hat{p} = 17, 18, 19, 35 \), and attains its minimum 5.6% at \( \hat{p} = 11 \). We also consider the diagnostic tests \( \hat{Q}(m, 0) \) for \( |m| = 1, 2 \), which check model misspecifications in conditional means, conditional variances and conditional covariances of state variables. The \( \hat{Q}(m, 0) \) tests have similar size patterns to \( \hat{Q}(0, 0) \), except that some of them tend to overreject a bit when \( T = 1000 \). Overall, both omnibus and diagnostic tests have reasonable sizes at both the 10% and 5% levels and the sizes are robust to the choice of the preliminary lag order \( \hat{p} \).

Next, we turn to examine the power of \( \hat{Q}(m, 0) \). Tables 2–6 report the rejection rates of \( \hat{Q}(m, 0) \) under DGPs 1–5 at the 10% and 5% levels respectively. Under DGP1, model \((6.1)\) ignores nonzero constant correlations in drifts. The omnibus test \( \hat{Q}(0, 0) \) is able to detect such model misspecification. The rejection rate of \( \hat{Q}(0, 0) \) is about 65% at the 5% level when \( T = 1000 \). Because only the misspecifications of the conditional means in \( X_{u} \) and \( X_{v} \) are time-varying when model \((6.1)\) is used to fit the data from DGP1, we expect that the mean tests, \( \hat{Q}(0, 0, 0, 0) \) and \( \hat{Q}(0, 0, 1, 0) \) and

\[\begin{aligned}
&\left( \begin{array}{ccc}
\sqrt{X_{1t}} & 0 & 0 \\
0 & \sqrt{X_{2t}} & 0 \\
0 & 0 & \sqrt{X_{3t}}
\end{array} \right) \cdot d \left( \begin{array}{c}
W_{1t} \\
W_{2t} \\
W_{3t}
\end{array} \right),
\end{aligned}\]

\[\begin{aligned}
\cdot DGP3 \text{ [Correlated Gaussian Diffusion, with Constant Correlation in Diffusion]}: & \\
\d X_{1t} &= \left( \begin{array}{ccc}
0.5 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 2
\end{array} \right) \cdot \left( \begin{array}{c}
-X_{1t} \\
-X_{2t} \\
-X_{3t}
\end{array} \right) dr \\
&+ \left( \begin{array}{ccc}
1 & 0 & 0 \\
0.5 & 1 & 0 \\
0.5 & 0.5 & 1
\end{array} \right) \cdot d \left( \begin{array}{c}
W_{1t} \\
W_{2t} \\
W_{3t}
\end{array} \right).
\end{aligned}\]

\[\begin{aligned}
\cdot DGP4 \text{ [Correlated CIR Diffusion]}: & \\
\d X_{1t} &= \left( \begin{array}{ccc}
0.5 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 2
\end{array} \right) \cdot \left( \begin{array}{c}
2 - X_{1t} \\
1 - X_{2t} \\
1 - X_{3t}
\end{array} \right) dr \\
&+ \left( \begin{array}{ccc}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 1
\end{array} \right) \cdot d \left( \begin{array}{c}
W_{1t} \\
W_{2t} \\
W_{3t}
\end{array} \right).
\end{aligned}\]

\[\begin{aligned}
\cdot DGP5 \text{ [Mixture of Gaussian and CIR Processes]}: & \\
\d X_{1t} &= \left( \begin{array}{ccc}
0.5 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 2
\end{array} \right) \cdot \left( \begin{array}{c}
2 - X_{1t} \\
1 - X_{2t} \\
1 - X_{3t}
\end{array} \right) dr \\
&+ \left( \begin{array}{ccc}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 1
\end{array} \right) \cdot d \left( \begin{array}{c}
W_{1t} \\
W_{2t} \\
W_{3t}
\end{array} \right).
\end{aligned}\]
Table 1
Sizes of specification tests under DGP0.

<table>
<thead>
<tr>
<th>Lag order</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ell )</td>
<td>250 500 1000 250 500 1000</td>
<td>250 500 1000 250 500 1000</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.10 0.05 0.05 0.10 0.05 0.05</td>
<td>0.10 0.05 0.05 0.10 0.05 0.05</td>
</tr>
<tr>
<td>( \hat{Q}_1 )</td>
<td>0.056</td>
<td>0.038</td>
</tr>
<tr>
<td>( \hat{Q}_2 )</td>
<td>0.053</td>
<td>0.041</td>
</tr>
<tr>
<td>( \hat{Q}_3 )</td>
<td>0.060</td>
<td>0.038</td>
</tr>
<tr>
<td>( \hat{Q}_4 )</td>
<td>0.060</td>
<td>0.038</td>
</tr>
<tr>
<td>( \hat{Q}_5 )</td>
<td>0.060</td>
<td>0.038</td>
</tr>
<tr>
<td>( \hat{Q}_6 )</td>
<td>0.060</td>
<td>0.038</td>
</tr>
<tr>
<td>( \hat{Q}_7 )</td>
<td>0.060</td>
<td>0.038</td>
</tr>
<tr>
<td>( \hat{Q}_8 )</td>
<td>0.060</td>
<td>0.038</td>
</tr>
<tr>
<td>( \hat{Q}_9 )</td>
<td>0.060</td>
<td>0.038</td>
</tr>
<tr>
<td>( \hat{Q}_{10} )</td>
<td>0.060</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Notes: (i) DGP0 is an uncorrelated Gaussian diffusion process, given in Eq. (6.1); (ii) \( \hat{Q}(0,0) \) is the omnibus test; \( \hat{Q}_1, \hat{Q}_2, \hat{Q}_3, \hat{Q}_4, \hat{Q}_5, \hat{Q}_6, \hat{Q}_7, \hat{Q}_8, \hat{Q}_9, \hat{Q}_{10} \) are conditional mean tests, conditional variance tests and conditional correlation tests respectively; (iii) The \( p \) values are based on the results of 1000 iterations.

Table 2
Powers of specification tests under DGP1.

<table>
<thead>
<tr>
<th>Lag order</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ell )</td>
<td>250 500 1000 250 500 1000</td>
<td>250 500 1000 250 500 1000</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.10 0.05 0.05 0.10 0.05 0.05</td>
<td>0.10 0.05 0.05 0.10 0.05 0.05</td>
</tr>
<tr>
<td>( \hat{Q}_1 )</td>
<td>0.120</td>
<td>0.078</td>
</tr>
<tr>
<td>( \hat{Q}_2 )</td>
<td>0.097</td>
<td>0.064</td>
</tr>
<tr>
<td>( \hat{Q}_3 )</td>
<td>0.076</td>
<td>0.053</td>
</tr>
<tr>
<td>( \hat{Q}_4 )</td>
<td>0.076</td>
<td>0.053</td>
</tr>
<tr>
<td>( \hat{Q}_5 )</td>
<td>0.076</td>
<td>0.053</td>
</tr>
<tr>
<td>( \hat{Q}_6 )</td>
<td>0.076</td>
<td>0.053</td>
</tr>
<tr>
<td>( \hat{Q}_7 )</td>
<td>0.076</td>
<td>0.053</td>
</tr>
<tr>
<td>( \hat{Q}_8 )</td>
<td>0.076</td>
<td>0.053</td>
</tr>
<tr>
<td>( \hat{Q}_9 )</td>
<td>0.076</td>
<td>0.053</td>
</tr>
<tr>
<td>( \hat{Q}_{10} )</td>
<td>0.076</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Notes: (i) DGP1 is a correlated Gaussian diffusion process with correlation in drift, given in Eq. (6.4); (ii) \( \hat{Q}(0,0) \) is the omnibus test; \( \hat{Q}_1, \hat{Q}_2, \hat{Q}_3, \hat{Q}_4, \hat{Q}_5, \hat{Q}_6, \hat{Q}_7, \hat{Q}_8, \hat{Q}_9, \hat{Q}_{10} \) are conditional mean tests, conditional variance tests and conditional correlation tests respectively; (iii) The \( p \) values are based on the results of 500 iterations.
Table 3
Powers of specification tests under DGP2.

<table>
<thead>
<tr>
<th>Lag order</th>
<th>10</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>10</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T )</td>
<td>250</td>
<td>500</td>
<td>1000</td>
<td>250</td>
<td>500</td>
<td>1000</td>
<td>250</td>
<td>500</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>.10</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
<td>.10</td>
<td>.05</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td>( Q(0, 0) )</td>
<td>.204</td>
<td>.150</td>
<td>.412</td>
<td>.360</td>
<td>.832</td>
<td>.770</td>
<td>.186</td>
<td>.116</td>
</tr>
</tbody>
</table>

\( \hat{Q}_1 \) \( \hat{Q}((1, 0, 0), \alpha) \) .050 .024 .056 .024 .044 .032 .046 .024 .048 .020 .060 .038
\( \hat{Q}((0, 1, 0), \alpha) \) .058 .034 .066 .040 .048 .034 .048 .028 .054 .026 .048 .030
\( \hat{Q}((0, 1, 0), \alpha) \) .070 .046 .066 .044 .058 .038 .074 .054 .066 .036 .068 .042

\( \hat{Q}_2 \) \( \hat{Q}((2, 0, 0), \alpha) \) .952 .922 .100 .996 .100 .100 .932 .900 .100 .996 .100 .100
\( \hat{Q}((2, 0, 0), \alpha) \) .958 .924 .100 .100 .100 .100 .922 .882 .100 .100 .100 .100
\( \hat{Q}((0, 2, 0), \alpha) \) .764 .678 .980 .970 .100 .100 .680 .576 .968 .942 .100 .100

\( \hat{Q}_3 \) \( \hat{Q}((1, 1, 0), \alpha) \) .080 .044 .082 .050 .080 .056 .062 .038 .062 .038 .086 .060
\( \hat{Q}((0, 1, 1), \alpha) \) .080 .050 .082 .062 .094 .066 .066 .040 .080 .048 .090 .058
\( \hat{Q}((0, 1, 1), \alpha) \) .072 .038 .094 .064 .106 .075 .062 .034 .078 .042 .102 .070

Notes: (i) DGP2 is an uncorrelated CIR diffusion process, given in Eq. (6.5); (ii) \( \hat{Q}((0, 0), \alpha) \) is the omnibus test; \( \hat{Q}_1 \), \( \hat{Q}_2 \), and \( \hat{Q}_3 \) are conditional mean tests, conditional variance tests and conditional correlation tests respectively; (iii) The \( p \) values are based on the results of 500 iterations.

Table 4
Powers of specification tests under DGP3.

<table>
<thead>
<tr>
<th>Lag order</th>
<th>10</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>10</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T )</td>
<td>250</td>
<td>500</td>
<td>1000</td>
<td>250</td>
<td>500</td>
<td>1000</td>
<td>250</td>
<td>500</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>.10</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
<td>.10</td>
<td>.05</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td>( Q(0, 0) )</td>
<td>.681</td>
<td>.366</td>
<td>.686</td>
<td>.792</td>
<td>1.00</td>
<td>1.00</td>
<td>.406</td>
<td>.310</td>
</tr>
</tbody>
</table>

\( \hat{Q}_1 \) \( \hat{Q}((1, 0, 0), \alpha) \) .032 .014 .062 .030 .054 .030 .036 .024 .044 .022 .048 .032
\( \hat{Q}((0, 1, 0), \alpha) \) .082 .056 .080 .052 .072 .038 .090 .054 .076 .050 .072 .044
\( \hat{Q}((0, 1, 0), \alpha) \) .092 .054 .102 .064 .084 .056 .086 .052 .098 .068 .092 .066

\( \hat{Q}_2 \) \( \hat{Q}((2, 0, 0), \alpha) \) .074 .060 .076 .040 .128 .078 .066 .038 .076 .036 .122 .084
\( \hat{Q}((2, 0, 0), \alpha) \) .080 .052 .122 .088 .116 .072 .068 .040 .110 .080 .094 .056
\( \hat{Q}((0, 2, 0), \alpha) \) .094 .060 .096 .060 .084 .052 .084 .040 .072 .042 .074 .046

\( \hat{Q}_3 \) \( \hat{Q}((1, 1, 0), \alpha) \) .080 .044 .124 .078 .094 .056 .056 .042 .126 .072 .086 .050
\( \hat{Q}((0, 1, 1), \alpha) \) .082 .044 .094 .068 .112 .074 .054 .032 .102 .062 .102 .070
\( \hat{Q}((0, 1, 1), \alpha) \) .094 .052 .118 .082 .088 .058 .078 .042 .100 .064 .092 .064

Notes: (i) DGP3 is a correlated Gaussian diffusion process with correlation in diffusion, given in Eq. (6.6); (ii) \( \hat{Q}((0, 0), \alpha) \) is the omnibus test; \( \hat{Q}_1 \), \( \hat{Q}_2 \), and \( \hat{Q}_3 \) are conditional mean tests, conditional variance tests and conditional correlation tests respectively; (iii) The \( p \) values are based on the results of 500 iterations.
Table 5
Powers of specification tests under DGP4.

<table>
<thead>
<tr>
<th>Lag order</th>
<th>( \hat{Q}, (1, 0, 0, 0) )</th>
<th>( \hat{Q}, (0, 1, 0, 0) )</th>
<th>( \hat{Q}, (0, 0, 1, 0) )</th>
<th>( \hat{Q}, (0, 0, 0, 1) )</th>
<th>( \hat{Q}, (1, 1, 0, 0) )</th>
<th>( \hat{Q}, (1, 0, 1, 0) )</th>
<th>( \hat{Q}, (0, 1, 1, 0) )</th>
<th>( \hat{Q}, (0, 1, 0, 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>20</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Notes: (i) DGP4 is a correlated CIR diffusion process, given in Eq. (6.7);
(ii) \( \hat{Q}, (0, 0) \) is the omnibus test; \( \hat{Q}, \hat{Q} \) and \( \hat{Q} \) are conditional mean tests, conditional variance tests and conditional correlation tests respectively;
(iii) The \( p \) values are based on the results of 500 iterations.

Table 6
Powers of specification tests under DGP5.

<table>
<thead>
<tr>
<th>Lag order</th>
<th>( \hat{Q}, (1, 0, 0, 0) )</th>
<th>( \hat{Q}, (0, 1, 0, 0) )</th>
<th>( \hat{Q}, (0, 0, 1, 0) )</th>
<th>( \hat{Q}, (0, 0, 0, 1) )</th>
<th>( \hat{Q}, (1, 1, 0, 0) )</th>
<th>( \hat{Q}, (1, 0, 1, 0) )</th>
<th>( \hat{Q}, (0, 1, 1, 0) )</th>
<th>( \hat{Q}, (0, 1, 0, 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>20</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Notes: (i) DGP5 is a mixture of Gaussian and CIR Processes, given in Eq. (6.8);
(ii) \( \hat{Q}, (0, 0) \) is the omnibus test; \( \hat{Q}, \hat{Q} \) and \( \hat{Q} \) are conditional mean tests, conditional variance tests and conditional correlation tests respectively;
(iii) The \( p \) values are based on the results of 500 iterations.
The diagnostic tests

\[ \hat{Q}(\mathbf{m}, \mathbf{0}) \]

will be powerful but the variance and covariance tests \( \hat{Q}(\mathbf{m}, \mathbf{0}) \), where \( |\mathbf{m}| = 2 \), will have no power. This is indeed confirmed in our simulation. Table 2 shows that the \( \hat{Q}((0, 1), \mathbf{0}) \) and \( \hat{Q}((0, 0, 1), \mathbf{0}) \) tests are able to capture time-varying mean misspecifications in \( X_{\tau_1} \) and \( X_{\tau_2} \), as their rejection rates are about 75% and 99% respectively at the 5% level when \( T = 1000 \).

However, the rejection rates of all variance and covariance tests \( \hat{Q}(\mathbf{m}, \mathbf{0}) \) for \( |\mathbf{m}| = 2 \) are close to significance levels. Under DGP2, model (6.1) ignores the so-called “level effect” in diffusions. The omnibus test \( \hat{Q}(\mathbf{0}, \mathbf{0}) \) has good power under DGP2, with a rejection rate of 75% at the 5% level when \( T = 1000 \). Note that the omnibus test \( \hat{Q}(\mathbf{0}, \mathbf{0}) \) is less powerful than the conditional variance tests, because \( \hat{Q}(\mathbf{0}, \mathbf{0}) \) has to check all possible directions whereas only the conditional variances are misspecified under DGP2. The variance tests \( \hat{Q}((0, 0, 0), \mathbf{0}) \), \( \hat{Q}((0, 0, 1), \mathbf{0}) \) and \( \hat{Q}((0, 0, 2), \mathbf{0}) \) have excellent power, which increases with \( T \) and approaches unity when \( T = 1000 \). Interestingly, the powers of the diagnostic tests for conditional means and conditional covariances are close to significance levels, indicating that these diagnostic tests do not overreject correctly specified conditional means and conditional covariances of state variables.

Under DGP3, model (6.1) is correctly specified for both conditional means and conditional variances of state variables but is misspecified for conditional correlations between state variables, because it ignores the nonzero but time-invariant correlations in diffusions. As expected, the omnibus test \( \hat{Q}(\mathbf{0}, \mathbf{0}) \) has excellent power when model (6.1) is used to fit the data generated from DGP3. The power of \( \hat{Q}(\mathbf{0}, \mathbf{0}) \) increases significantly with the sample size \( T \) and approaches unity when \( T = 1000 \). However, our correlation diagnostic tests fail to capture the misspecifications in correlations, because they are time-invariant.

Under DGP4, model (6.1) is correctly specified for conditional means of state variables but is misspecified for conditional correlations of state variables, because it ignores both the “level effect” and time-varying conditional correlations in diffusions. The omnibus test \( \hat{Q}(\mathbf{0}, \mathbf{0}) \) has excellent power when model (6.1) is used to fit data generated from DGP4. This is consistent with the fact that DGP4 deviates most from model (6.1). The \( \hat{Q}(\mathbf{m}, \mathbf{0}) \) tests with \( |\mathbf{m}| = 2 \) have good power in detecting misspecifications in conditional variances and correlations of state variables. The mean tests \( \hat{Q}(\mathbf{m}, \mathbf{0}) \) with \( |\mathbf{m}| = 1 \) have no power, because the conditional means of state variables are correctly specified.

Under DGP5, model (6.1) is correctly specified for conditional means of state variables but is misspecified for conditional variances of \( X_{\tau_1} \) and \( X_{\tau_2} \) and the conditional correlation between \( X_{\tau_1} \) and \( X_{\tau_2} \). Again, the omnibus test \( \hat{Q}(\mathbf{0}, \mathbf{0}) \) has good power under DGP5, with a rejection rate of 95% at the 5% level when \( T = 1000 \). The variance tests \( \hat{Q}((0, 0, 0), \mathbf{0}) \), \( \hat{Q}((0, 0, 1), \mathbf{0}) \) and the covariance test \( \hat{Q}((0, 0, 1), \mathbf{0}) \) have excellent power, which increases with \( T \) and attains unity when \( T = 1000 \). The powers of other diagnostic tests are close to significance levels, because those conditional moments are correctly specified.

To sum up, we observe:

- The \( \hat{Q}(\mathbf{m}, \mathbf{0}) \) tests have reasonable sizes in finite samples. Although the omnibus test \( \hat{Q}(\mathbf{0}, \mathbf{0}) \) tends to underreject when \( T = 250 \), it improves dramatically as the sample size \( T \) increases. The sizes of all tests \( \hat{Q}(\mathbf{m}, \mathbf{0}) \) are robust to the choice of a preliminary lag order used to estimate the generalized cross-spectrum.
- The omnibus test \( \hat{Q}(\mathbf{0}, \mathbf{0}) \) has good omnibus power in detecting various model misspecifications. It has reasonable power even when the sample size \( T \) is as small as 250. This demonstrates the nice feature of the proposed cross-spectral approach, which can capture various model misspecifications.
- The diagnostic tests \( \hat{Q}(\mathbf{m}, \mathbf{0}) \) can check various specific aspects of model misspecifications. Generally speaking, the mean tests, \( \hat{Q}(\mathbf{m}, \mathbf{0}) \), with \( |\mathbf{m}| = 1 \), can detect misspecification in drifts; the variance and covariance tests \( \hat{Q}(\mathbf{m}, \mathbf{0}) \), with \( |\mathbf{m}| = 2 \), can check misspecifications in variances and covariances respectively. However, the correlation tests fail to detect neglected nonzero but time-invariant conditional correlations in diffusions.

7. Conclusion

The CCF-based estimation of multivariate continuous-time models has attracted increasing attention in econometrics. We have complemented this literature by proposing a CCF-based omnibus specification test for the adequacy of a multivariate continuous-time model, which has not been attempted in the previous literature. The proposed test can be applied to a variety of univariate and multivariate continuous-time models, including those with jumps. It is also applicable to testing the adequacy of discrete-time dynamic multivariate distribution models. The most appealing feature of our omnibus test is that it fully exploits the information in the joint dynamics of state variables and thus can capture misspecification in modeling the joint dynamics, which may be easily missed by existing procedures. Indeed, when the underlying economic process is Markov, our omnibus test is consistent against any type of model misspecification. We assume that the DGP of state variables may not be Markov. This not only ensures the power of the proposed tests against a wider range of misspecification but also makes our approach applicable to testing models with latent variables, such as SV models. Our omnibus test is supplemented by a class of diagnostic procedures, which is obtained by differentiating the CCF and focuses on various specific aspects of the joint dynamics such as whether there exists neglected dynamics in conditional means, conditional variances, and conditional correlations of state variables respectively. Such information is useful for practitioners in reconstructing a misspecified model. Our procedures are most useful when the CCF of a multivariate time series model has a closed-form, as are the class of AJD models and the class of time-changed Lévy processes that have been commonly used in the literature. All test statistics follow a convenient asymptotic \( N(0, 1) \) distribution, and they are applicable to various estimation methods, including suboptimal 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 co-editor, A. Ronald Gallant, and two referees for careful and constructive comments. We also thank Yacine Aït-Sahalia, Zongwu Cai, Jianqiong Fan, Jerry Hausman, Roger Klein, Chung-ming Kuan, Arthur Lewbel, Norm Swanson, Yiu Kuen Tse, Chunchi Wu and Zhijie Xiao, and seminar participants at Boston College, Rutgers University and Singapore Management University for their comments and discussions. Any remaining errors are solely ours.

---

19 The power of the mean test \( \hat{Q}((1, 0, 0), \mathbf{0}) \) for the first state variable \( X_{\tau_1} \) is close to the significance level. This is expected because the drift function of \( X_{\tau_1} \) is correctly specified.
Appendix. Mathematical appendix

Throughout the appendix, we let $\hat{Q}(\mathbf{0}, \mathbf{0})$ be defined in the same way as $\hat{Q}(\mathbf{0}, \mathbf{0})$ in (3.8) with the unobservable generalized residual sample $\{\mathbf{Z}_t(\mathbf{u}, \mathbf{u}_0)\}_{t=1}^T$, replacing the estimated generalized residual sample $\{\hat{Z}_t(\mathbf{u}, \hat{\mathbf{u}})\}_{t=1}^T$. 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 below. □

Theorem A.1. Under the conditions of Theorem 1, $\hat{Q}(\mathbf{0}, \mathbf{0}) \rightarrow 0$.

Theorem A.2. Under the conditions of Theorem 1 and $q = p^{1+ \frac{1}{2}}(\ln^2 T)\rightarrow \infty$, $\hat{Q}(\mathbf{0}, \mathbf{0}) \rightarrow 0$.

Proof of Theorem A.1. Put $T_j = T - |j|$, and let $\hat{f}_j(\mathbf{u}, \mathbf{v})$ be defined in the same way as $\hat{f}_j(\mathbf{u}, \mathbf{v})$ in (3.4), with $\hat{Z}_t(\mathbf{u}, \mathbf{u}_0)$ replaced by $Z_t(\mathbf{u}, \mathbf{u}_0)$. To show $\hat{Q}(\mathbf{0}, \mathbf{0}) - \hat{Q}(\mathbf{0}, \mathbf{0}) \rightarrow 0$, it suffices to show

$$
\hat{D}\hat{f}_j(\mathbf{u}, \mathbf{v}) \int \sum_{t=1}^{T_j} k^2(j/p)T_j \left[ \hat{f}_j(\mathbf{u}, \mathbf{v})^2 - |\hat{f}_j(\mathbf{u}, \mathbf{v})|^2 \right] \times dW(\mathbf{u})dW(\mathbf{v}) \rightarrow 0.
$$

(A.1)

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

$$
\hat{f}_j(\mathbf{u}, \mathbf{v}) = -\hat{f}_j(\mathbf{u}, \mathbf{v})T_j \sum_{t=j+1}^{T} \frac{\partial}{\partial \mathbf{u}} \hat{f}_j(\mathbf{u}, \mathbf{v}) | \mathbf{u} = I_{t-1}, \mathbf{v} \rangle \psi_j - \hat{f}_j(\mathbf{u}, \mathbf{v})
$$

$$
\leq C \sqrt{T(j)} \sup_{\mathbf{u}} \sup_{\mathbf{u}_0} \left\| \frac{\partial^2}{\partial \theta \partial \theta} \right\| \times \psi(\mathbf{u}, t|I_{t-1}, \mathbf{v}) \times dW(\mathbf{u})dW(\mathbf{v}) = O_p (p/T).
$$

(A.4)

For some $\hat{\theta}$ between $\hat{\theta}$ and $\theta_0$.

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

$$
\sum_{j=1}^{T-1} k^2(j/p)T_j \int \left| \hat{B}_{12}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v})
$$

$$
\leq C \sqrt{T(\hat{\theta} - \theta_0)} \sup_{\mathbf{u}} \sup_{\mathbf{u}_0} \left\| \frac{\partial^2}{\partial \theta \partial \theta} \right\| \times \psi(\mathbf{u}, t|I_{t-1}, \hat{\theta}) \times dW(\mathbf{u})dW(\mathbf{v}) = O_p (p/T),
$$

where we made use of the fact that

$$
\sum_{j=1}^{T-1} a_j(j) \int \left| \hat{B}_{12}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) = O(p/T),
$$

(A.5)

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

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

$$
\sum_{j=1}^{T-1} k^2(j/p)T_j \int \left| \hat{B}_{13}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v})
$$

$$
\leq 4 \sum_{j=1}^{T-1} a_j(j) \int \left[ \sum_{t=j+1}^{T} \psi(\mathbf{u}, t|I_{t-1}, \hat{\theta}) - \psi(\mathbf{u}, t|I_{t-1}, \hat{\theta}) \right]^2 dW(\mathbf{u}) dW(\mathbf{v}) = O_p (p/T).
$$

(A.6)

where we have used Assumptions A.5–A.7.
For the first term in Eq. (A.4), we have
\[
\sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{b}_{1ij}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) \\
\leq \left\| \sqrt{T} \theta - \theta_0 \right\|^2 \sum_{j=1}^{T-1} k^2(j/p) \\
\times \int \left( \int T_j^{-1} \sum_{t=j+1}^{T} \frac{\partial \theta}{\partial \mathbf{v}} (\mathbf{u}, t|\mathbf{x}_{t-1}, \theta_0) \psi_{t-j}(\mathbf{v}) \right)^2 \\
\times dW(\mathbf{u}) dW(\mathbf{v}) \\
= O_p(1), \\
\text{as is shown below: Put } \eta_j(\mathbf{u}, \mathbf{v}) \equiv E\left[ \frac{\partial}{\partial \mathbf{v}} \psi(\mathbf{u}, t|\mathbf{x}_{t-1}, \theta_0) \psi_{t-j}(\mathbf{v}) \right] \\
= \text{cov} \left( \frac{\partial}{\partial \mathbf{v}} \psi(\mathbf{u}, t|\mathbf{x}_{t-1}, \theta_0), \psi_{t-j}(\mathbf{v}) \right). \text{Then we have } \sup_{\mathbf{u}, \mathbf{v} \in [2N]} \left\| \eta_j(\mathbf{u}, \mathbf{v}) \right\| \leq C \text{ by Assumption A.4. Next, expressing the moments by cumulants via well-known formulas (e.g., } \text{Hannan (1970, p. 23), for real-valued processes), we can obtain}
\]
\[
T_j E \left\| J^{-1} \sum_{t=j+1}^{T} \frac{\partial}{\partial \mathbf{v}} \psi(\mathbf{u}, t|\mathbf{x}_{t-1}, \theta_0) \psi_{t-j}(\mathbf{v}) - \eta_j(\mathbf{u}, \mathbf{v}) \right\|^2 \\
\leq \sum_{t=j}^{T} \text{cov} \left( \frac{\partial}{\partial \mathbf{v}} \psi(\mathbf{u}, \text{max}(0, \tau) + 2|J_{\text{max}(0, \tau) + 1}, \theta_0) \psi_{t-j}(\mathbf{v}) \right) \\
\times \left\| \frac{\partial}{\partial \mathbf{v}} \psi(\mathbf{u}, \text{max}(0, \tau) + 2|J_{\text{max}(0, \tau) + 1}, \theta_0) \psi_{t-j}(\mathbf{v}) \right\|^2 \\
\leq C, \\
\text{given Assumption A.4, where } k_{j,t}(\mathbf{v}) \text{ is the fourth order cumulant of the joint distribution of the process } \left\{ \frac{\partial}{\partial \mathbf{v}} \psi(\mathbf{u}, t|\mathbf{x}_{t-1}, \theta_0) \psi_{t-j}(\mathbf{v}) \right\} \text{. See also (A.7) of Hong (1999, p. 1212). Consequently, from (A.5), (A.8), } |k_j(\tau)| \leq 1, \text{ and } p/T \to 0, \text{ we have}
\]
\[
\sum_{j=1}^{T-1} k^2(j/p) E \int \left( \int T_j^{-1} \sum_{t=j+1}^{T} \frac{\partial}{\partial \mathbf{v}} \psi(\mathbf{u}, t|\mathbf{x}_{t-1}, \theta_0) \psi_{t-j}(\mathbf{v}) \right)^2 \\
\times dW(\mathbf{u}) dW(\mathbf{v}) \\
\leq C \sum_{j=1}^{T-1} \left\| \eta_j(\mathbf{u}, \mathbf{v}) \right\|^2 dW(\mathbf{u}) dW(\mathbf{v}) + C \sum_{j=1}^{T-1} a_T(j) \\
= O_p(1) + O_p(p/T) = O_p(1).
\]
Hence (A.7) is $O_p(1)$. The desired result of Lemma A.1 follows from (A.5)–(A.7). \(\square\)

**Proof of Lemma A.2.** We first decompose $\hat{b}_{2ij}(\mathbf{u}, \mathbf{v})$ as $\hat{b}_{2ij}(\mathbf{u}, \mathbf{v}) = \hat{b}_{21ij}(\mathbf{u}, \mathbf{v}) + \hat{b}_{22ij}(\mathbf{u}, \mathbf{v})$, say.

For the first term, we have
\[
\sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{b}_{21ij}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) \\
\leq \sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{b}_{22ij}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) \\
\times dW(\mathbf{u}) dW(\mathbf{v}) \\
= O_p(1), \\
\text{where we made use of the fact that } E \left| \hat{b}_{21ij}(\mathbf{u}, \mathbf{v}) \right|^2 \leq C T_j^{-1} \text{ given Assumption A.4 and}
\]
\[
\sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{b}_{22ij}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) \\
\leq \sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{b}_{22ij}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) \\
\times dW(\mathbf{u}) dW(\mathbf{v}) \\
= O_p(1). \\
\]
For the second term, we have
\[
\sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{b}_{21ij}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) \\
\leq \sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{b}_{22ij}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) \\
\times dW(\mathbf{u}) dW(\mathbf{v}) \\
= O_p(1), \\
\text{where we made use of the fact that } E \left| \hat{b}_{21ij}(\mathbf{u}, \mathbf{v}) \right|^2 \leq C T_j^{-1} \text{ given Assumptions A.4–A.7. The desired result of Lemma A.2 follows from (A.9) and (A.10).} \(\square\)

**Proof of Proposition A.2.** Given the decomposition in (A.3), we have
\[
\left| \hat{f}_j(\mathbf{u}, \mathbf{v}) - \hat{f}_j(\mathbf{u}, \mathbf{v}) \right|^2 \leq \sum_{a=1}^{2} \left\| \hat{b}_{2a} \right\|^2 \left| \hat{f}_j(\mathbf{u}, \mathbf{v}) \right| \\
\text{where the } \hat{b}_{2a} \text{ are defined in (A.3), } a = 1, 2.
\]
We first consider the term with $a = 2$. By the Cauchy–Schwarz inequality, we have
\[
\sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{b}_{2a}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) \\
\leq \left( \sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{b}_{2a}(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) \right)^{1/2} \\
\times \left( \sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{f}_j(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) \right)^{1/2} \\
= O_p(1). \\
\]
given Lemma A.2, and $p/T \to 0$, where $p^{-1} \sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| \hat{f}_j(\mathbf{u}, \mathbf{v}) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) = O_p(1)$ by Markov’s inequality, the m.d.s. hypothesis of $Z_t(\mathbf{u}, \theta_0)$, and (A.5).
For $a = 1$, by (A.4) and the triangular inequality, we have
\[
\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\tilde{B}_{ij}(u, v)\| \tilde{f}_j(u, v)\| dW(u) dW(v)
\leq \sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\tilde{B}_{ij}(u, v)\| dW(u) dW(v)
+ \sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\tilde{f}_j(u, v)\| dW(u) dW(v)
\times \iint |\tilde{f}_j(u, v)\| dW(u) dW(v).
\] (A.12)

For the first term in (A.12), we have
\[
\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\tilde{B}_{ij}(u, v)\| dW(u) dW(v)
\leq \|\hat{\theta} - \theta_0\| \sum_{j=1}^{T-1} k^2(j/p)T_j
\times \iint \bigg| \frac{\partial}{\partial \theta} \psi(u, t|I_{t-1}, \theta_0) \bigg| dW(u) dW(v)
\times \iint |\tilde{f}_j(u, v)\| dW(u) dW(v)
= O_p(1 + p/T^{1/2}) = o_p(p^{1/2}),
\]
given $p \to \infty$, $p/T \to 0$, Assumptions A.2, A.3, A.6 and A.7, and $T \bar{E} |\tilde{f}_j(u, v) - T_j \leq C$. Note that we have made use of the fact that
\[
E \left[ \left( \sum_{t_{j+1}}^{T-1} \frac{\partial}{\partial \theta} \psi(u, t|I_{t-1}, \theta_0) \right)^2 \right]^{1/2}
\leq \left[ \sum_{t_{j+1}}^{T-1} \frac{\partial}{\partial \theta} \psi(u, t|I_{t-1}, \theta_0) \right] \tilde{f}_j(u, v)
\leq C \left[ \sum_{t_{j+1}}^{T-1} \left( \frac{\partial}{\partial \theta} \psi(u, t|I_{t-1}, \theta_0) \right)^2 \right]^{1/2},
\]
by (A.7), and consequently,
\[
\|\hat{\theta} - \theta_0\| \sum_{j=1}^{T-1} k^2(j/p)T_j \iint \bigg| \frac{\partial}{\partial \theta} \psi(u, t|I_{t-1}, \theta_0) \bigg| dW(u) dW(v)
\times \iint |\tilde{f}_j(u, v)\| dW(u) dW(v)
\leq C \sum_{j=1}^{T-1} \left( \left< \frac{\partial}{\partial \theta} \psi(u, t|I_{t-1}, \theta_0) \right)^2 \right]^{1/2}
+ CT^{-1/2} \sum_{j=1}^{T-1} k^2(j/p) = O(1 + p/T^{1/2}),
\]
given $|k(|) \leq 1$ and Assumption A.7.

For the second term in (A.12), we have
\[
\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\tilde{B}_{ij}(u, v)\| dW(u) dW(v)
\leq CT \|\hat{\theta} - \theta_0\|^2 \left[ \sum_{t_{j+1}}^{T-1} \sup_{u \in \mathbb{R}} \sup_{t \in \mathbb{R}} \left( \frac{\partial^2}{\partial \theta^2} \psi(u, t|I_{t-1}, \theta) \right) \right]
\times \sum_{j=1}^{T-1} k^2(j/p) \iint |\tilde{f}_j(u, v)\| dW(u) dW(v)
= O_p\left( \frac{p}{T^{1/2}} \right),
\]
by the Cauchy–Schwarz inequality, Markov’s inequality, Assumptions A.2, A.3, A.6 and A.7, and $E |\tilde{f}_j(u, v) |^2 \leq CT^{-1}$. Hence, we have
\[
\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\tilde{B}_{ij}(u, v)\| dW(u) dW(v)
\leq C \sum_{t_{j+1}}^{T-1} \sup_{u \in \mathbb{R}} \sup_{t \in \mathbb{R}} \left( \frac{\partial}{\partial \theta} \psi(u, t|I_{t-1}, \theta) - \psi(u, t|I_{t-1}, \theta) \right) \iint |\tilde{f}_j(u, v)\| dW(u) dW(v)
\times \sum_{j=1}^{T-1} k^2(j/p) \iint |\tilde{f}_j(u, v)\| dW(u) dW(v)
= O_p\left( \frac{p}{T^{1/2}} \right),
\]
by the Cauchy–Schwarz inequality, Markov’s inequality, Assumptions A.2, A.3 and A.5–A.7, and $E |\tilde{f}_j(u, v) |^2 \leq CT^{-1}$. Hence, we have
\[
\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\tilde{B}_{ij}(u, v)\| dW(u) dW(v)
\leq O_p\left( \frac{1 + p/T^{1/2}}{T^{1/2}} \right) + O_p\left( \frac{p}{T^{1/2}} \right) + O_p\left( \frac{p}{T^{1/2}} \right)
= O_p\left( \frac{p}{T^{1/2}} \right).
\] (A.13)

Combining (A.11) and (A.13) then yields the result of this proposition. □

Proof of Theorem A.2. We first state a lemma, which will be used in the proof. □

Lemma A.3. Suppose that $I_n^m$ are the σ-fields generated by a stationary β-mixing process $X_t$ with mixing coefficient $\beta(i)$. For some positive integers $m$ let $y_i \in I_n^m$ where $s_1 < t_1 < s_2 < t_2 < \ldots < s_m$ and suppose $t_i - s_i > r$ for all $i$. Assume that
\[
\|y_i\|_{\ell_2} = E \|y_i\| < \infty,
\]
for some $p_i > 1$ for which $Q = \sum_{i=1}^{\infty} \frac{1}{p_i} < 1$. Then
\[
E \left( \prod_{i=1}^{\infty} E(y_i) \right) \leq 10 (1 - \frac{1}{Q}) \left( \frac{1}{p_i} \right) \prod_{i=1}^{\infty} \|y_i\|_{p_i}
\]


Let $q = p^{1+\frac{1}{p-1}} \left( \ln^2 T \right)^{\frac{p-1}{p}}$. We shall show Proposition A.3 and A.4 below.

Proposition A.3. Under the conditions of Theorem 1,
\[
p^{-\frac{1}{2}} \sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\tilde{f}_j(u, v) |^2 dW(u) dW(v)
= p^{-1/2} \hat{C}(0, 0) + p^{-1/2} \hat{V}_q + o_p(1),
\]
where $\hat{V}_q = \sum_{j=1}^{T-1+2q+1} \iint \mathbb{E} \{ \bar{Z}_j \} (u, \theta_0) \sum_{i=1}^{q} \bar{a}_j(i) \psi_{\tilde{y}-j}(v) |\sum_{r=1}^{2q+1} Z_{\tilde{y}}(u, \theta_0) \psi_{\tilde{y}-j}(v) dW(u) dW(v)$.

Proposition A.4. Under the conditions of Theorem 1,
Proposition A.4. Under the conditions of Theorem 1, \( \tilde{D}^{-1/2}(0,0) \)
\( \tilde{V}_{q} \overset{d}{\to} N(0,1) \).

Proof of Proposition A.3. We first decompose
\[
\sum_{j=1}^{T-1} k^2(j/p) T \int \int |\tilde{F}_j(u,v)|^2 \, dW(u) \, dW(v)
\]
\[
= \sum_{j=1}^{T-1} \alpha_T(j) \int \int \sum_{t=j+1}^{T} Z_t(u, \theta_0) \psi_{t-j}(v) \, dW(u) \, dW(v)
\]
\[
\times \alpha_T(j) \int \int \sum_{t=j+1}^{T} Z_t(u, \theta_0) \psi_{t-j}(v) \, dW(u) \, dW(v)
\]
\[
+ \sum_{j=1}^{T-1} \alpha_T(j) \int \int \sum_{t=j+1}^{T} Z_t(u, \theta_0) \psi_{t-j}(v) - \tilde{\psi}_t(v) \, dW(u) \, dW(v)
\]
\[
\times \psi'(v) - \tilde{\psi}_t(v) \, dW(u) \, dW(v)
\]
\[
+ 2 \operatorname{Re} \sum_{j=1}^{T-1} \alpha_T(j) \int \int \sum_{t=j+1}^{T} Z_t(u, \theta_0) \psi_{t-j}(v)
\]
\[
\times \left\{ \left[ \sum_{t=j+1}^{T} Z_t(u, \theta_0) \left( \psi(v) - \tilde{\psi}_t(v) \right) \right]^* \right\} \, dW(u) \, dW(v)
\]
\[
\equiv \tilde{M} + \tilde{R}_3 + 2 \operatorname{Re}(\tilde{R}_2).
\]

Next we write
\[
\tilde{M}_q = \sum_{j=1}^{T-1} \alpha_T(j) \int \int \sum_{t=j+1}^{T} |Z_t(u, \theta_0)|^2 |\psi_{t-j}(v)|^2 \, dW(u) \, dW(v)
\]
\[
\times \alpha_T(j) \int \int \sum_{t=j+1}^{T} Z_t(u, \theta_0) Z_t^*(u, \theta_0) \psi_{t-j}(v) \, dW(u) \, dW(v)
\]
\[
\times \alpha_T(j) \int \int \sum_{t=j+1}^{T} Z_t(u, \theta_0) \psi_{t-j}(v) \, dW(u) \, dW(v)
\]
\[
\equiv \tilde{U}_1 + \tilde{R}_3,
\]

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 \( \tilde{R}_3, \) we have \( 0 < t - s \leq 2q. \) Finally, we write
\[
\tilde{U}_1 = \sum_{t=2q+1}^{T} \int \int Z_t(u, \theta_0) \sum_{j=1}^{q} \alpha_T(j) \psi_{t-j}(v)
\]
\[
\times \sum_{s=j+1}^{T-1} Z_s^*(u, \theta_0) \psi_{s-j}(v) \, dW(u) \, dW(v)
\]
\[
+ \sum_{t=2q+1}^{T} \int \int Z_t(u, \theta_0) \sum_{j=q+1}^{T-1} \alpha_T(j) \psi_{t-j}(v)
\]
\[
\times \sum_{s=j+1}^{T-1} Z_s^*(u, \theta_0) \psi_{s-j}(v) \, dW(u) \, dW(v)
\]
\[
\equiv \tilde{U}_1 + \tilde{R}_3.
\]

where the first term \( \tilde{U}_1 \) is contributed by the lag orders \( j \) from 1 to \( q \); and the second term \( \tilde{R}_3 \) is contributed by the lag orders \( j > q. \) It follows from (A.14)-(A.17) that
\[
\sum_{j=1}^{T-1} k^2(j/p) T \int \int |\tilde{F}_j(u,v)|^2 \, dW(u) \, dW(v)
\]
\[
= \tilde{C}_1(0,0) + 2 \operatorname{Re}(\tilde{V}_q) + \tilde{R}_1 - 2 \operatorname{Re}(\tilde{R}_2 - \tilde{R}_3 - \tilde{R}_4).
\]

It suffices to show Lemmas A.4-A.8 below, which imply \( \tilde{C}_1(0,0) = \tilde{C}_1(0,0) = O(p) \) and \( p^{1/2} \tilde{R}_4 = o_p(1) \) given \( q = p^{1/2} \ln(T) \) and \( p = \ln(T) \) for \( 0 < \kappa < (3 + 1/4) \). □

Lemma A.4. Let \( \tilde{C}_1(0,0) \) be defined as in (A.15). Then \( \tilde{C}_1(0,0) \) = \( O(p/T^{1/2}) \).

Lemma A.5. Let \( \tilde{R}_1 \) be defined as in (A.14). Then \( \tilde{R}_1 = O(p/T) \).

Lemma A.6. Let \( \tilde{R}_2 \) be defined as in (A.14). Then \( \tilde{R}_2 = O(p/T^{1/2}) \).

Lemma A.7. Let \( \tilde{R}_3 \) be defined as in (A.16). Then \( \tilde{R}_3 = O(p/\alpha q T^{1/2}) \).

Lemma A.8. Let \( \tilde{R}_4 \) be defined as in (A.17). Then \( \tilde{R}_4 = O(p T^{2q}) \).
Further, it can be shown in a similar way that
\[
\sum_{j=1}^{t-1} \left| E \left[ \sum_{m=j+1}^{T} Z_{j-1} \psi_{m-j} \right] \right|^2 \leq C \left( T \right)^2,
\]
where we have used the fact \( E \sum_{m=j+1}^{T} Z_{j-1} \psi_{m-j} \leq C T^2 \) by Rosenthal’s inequality and \( E \left| \psi_{m-j} \right|^4 = O \left( T^{-2} \right) \). \( \square \)

**Proof of Lemma A.5.** By the m.d.s. property of \( Z_t \), the Cauchy–Schwarz inequality, we have
\[
E \left[ \tilde{R}_1 \right] = \sum_{j=1}^{T-1} \left| \int \left[ E \left[ \sum_{m=j+1}^{T} Z_{j-1} \psi_{m-j} \right] \right|^2 \right| \leq \sum_{j=1}^{T-1} \left| \int \left[ E \left[ \sum_{m=j+1}^{T} Z_{j-1} \psi_{m-j} \right] \right|^2 \right| \leq C T^2.
\]
where we have used the fact \( E \sum_{m=j+1}^{T} Z_{j-1} \psi_{m-j} \leq C T^2 \) by Rosenthal’s inequality and \( E \left| \psi_{m-j} \right|^4 = O \left( T^{-2} \right) \). \( \square \)

**Proof of Lemma A.6.** By the m.d.s. property of \( Z_t \), the Cauchy–Schwarz inequality, we have
\[
E \left[ \tilde{R}_2 \right] \leq \sum_{j=1}^{T-1} \left| \int \left[ E \left[ \sum_{m=j+1}^{T} Z_{j-1} \psi_{m-j} \right] \right|^2 \right| \leq \sum_{j=1}^{T-1} \left| \int \left[ E \left[ \sum_{m=j+1}^{T} Z_{j-1} \psi_{m-j} \right] \right|^2 \right| \leq C T^2.
\]
where we have used the fact \( E \sum_{m=j+1}^{T} Z_{j-1} \psi_{m-j} \leq C T^2 \) by Rosenthal’s inequality and \( E \left| \psi_{m-j} \right|^4 = O \left( T^{-2} \right) \). \( \square \)

**Proof of Proposition A.4.** We rewrite \( \tilde{V}_q = \sum_{t=2q+2}^{T} V_q(t) \), where
\[
V_q(t) = \int \left[ Z_{j} \psi_{j-1} \psi_{j} \right] \frac{a(t) \psi_{j-1} \psi_{j}}{\sqrt{p}} \int dV_q(t) dW(u) dW(v),
\]
and \( \psi_{j-1} \psi_{j} \sum_{t=2q+2}^{T} V_q(t) \) we apply Brown’s (1971) martingale limit theorem, which states \( \text{var}(2 \text{Re} \tilde{V}_q)^{-1} \text{Re} 2 \text{Re} \tilde{V}_q \to N(0, 1) \) if
\[
\text{var}(2 \text{Re} \tilde{V}_q)^{-1} \sum_{t=2q+2}^{T} \left[ \text{Re} V_q(t) \right]^2 \leq \frac{T}{2}.
\]
and
\[
\text{var}(2 \text{Re} \tilde{V}_q)^{-1} \sum_{t=2q+2}^{T} \left[ \text{Re} V_q(t) \right]^2 \leq \frac{T}{2}.
\]
and
\[
\text{var}(2 \text{Re} \tilde{V}_q)^{-1} \sum_{t=2q+2}^{T} \left[ \text{Re} V_q(t) \right]^2 \leq \frac{T}{2}.
\]
First, we compute \( \text{var}(2 \text{Re } \tilde{V}_q) \). By the m.d.s. property of \( Z_t(u, \theta_0) \) under \( H_{00} \), we have

\[
E(V_q^2) = \sum_{t=2q+2}^T E \left[ \int \int Z_t(u, \theta_0) \sum_{j=1}^q a_j(\psi_{t-j}(v)) \times \sum_{s=j+1}^{t-2q-1} Z^*_s(u, \theta_0) \psi_{s-j}(v) dW(u) dW(v) \right]^2 \\
= \sum_{j=1}^q \sum_{l=1}^q a_j(l) a_l(l) \\
\times \int \int \int \int_{t=2q+2}^T \sum_{s=j+1}^{t-2q-1} E[Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{t-j}(v) \psi_{t-s}(v)] \\
\times \psi_{t-s}(v_1) \psi_{t-s}(v_2) \\
= \sum_{j=1}^q \sum_{l=1}^q a_j(l) a_l(l) \\
\times \int \int \int \int_{t=2q+2}^T \sum_{s=j+1}^{t-2q-1} \psi_{t-s}(v_1) \psi_{t-s}(v_2) \\
\times E[Z^*_s(u_1, \theta_0) Z^*_s(u_2, \theta_0) \psi_{s-j}(v_1) \psi_{s-j}(v_2)] \\
\times dW(u_1) dW(u_2) dW(v_1) dW(v_2) + \sum_{j=1}^q \sum_{l=1}^q a_j(l) a_l(l) \\
\times \int \int \int \int_{t=2q+2}^T \sum_{s=j+1}^{t-2q-1} E[Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{t-s}(v_1) \psi_{t-s}(v_2)] \\
\times \psi_{t-s}(v_1) \psi_{t-s}(v_2) \\
\times Z^*_s(u_1, \theta_0) Z^*_s(u_2, \theta_0) \psi_{s-j}(v_1) \psi_{s-j}(v_2) \\
\times dW(u_1) dW(u_2) dW(v_1) dW(v_2) + 2 \sum_{j=1}^q \sum_{l=1}^q a_j(l) a_l(l) \\
\times \int \int \int \int_{t=2q+2}^T \sum_{s=j+1}^{t-2q-1} E[Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{t-s}(v_1) \psi_{t-s}(v_2)] \\
\times \psi_{t-s}(v_1) \psi_{t-s}(v_2) \\
\times Z^*_s(u_1, \theta_0) Z^*_s(u_2, \theta_0) \psi_{s-j}(v_1) \psi_{s-j}(v_2) \\
\times dW(u_1) dW(u_2) dW(v_1) dW(v_2) \\
= \frac{1}{2} \sum_{j=1}^q \sum_{l=1}^q k^2(j/p) k^2(l/p) \\
\times \int \int \int \int |E[Z_{q+j}(u_1, \theta_0) Z_{q+j}(u_2, \theta_0) \psi_{q+j}(v_1) \psi_{q+j}(v_2)]|^2 \\
\times dW(u_1) dW(u_2) dW(v_1) dW(v_2) [1 + o(1)] \\
+ \sum_{t=2q+2}^T V_1(t) + \sum_{t=2q+2}^T V_2(t) + \sum_{t=2q+2}^T V_3(t), \quad (A.20)
\]

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

\[
\sum_{t=2q+2}^T \sum_{j=1}^q \sum_{l=1}^q a_j(l) a_l(l) \\
\times \int \int \int \int_{t=2q+2}^T \sum_{s=j+1}^{t-2q-1} E[Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{t-s}(v_1) \psi_{t-s}(v_2)] \\
\times \psi_{t-s}(v_1) \psi_{t-s}(v_2) \\
\times Z^*_s(u_1, \theta_0) Z^*_s(u_2, \theta_0) \psi_{s-j}(v_1) \psi_{s-j}(v_2) \\
\times dW(u_1) dW(u_2) dW(v_1) dW(v_2) \\
\leq 2 \sum_{j=1}^q \sum_{l=1}^q a_j(l) a_l(l) \\
\times \int \int \int \int_{t=2q+2}^T \sum_{s=j+1}^{t-2q-1} \psi_{t-s}(v_1) \psi_{t-s}(v_2) \\
\times E[Z^*_s(u_1, \theta_0) Z^*_s(u_2, \theta_0) \psi_{s-j}(v_1) \psi_{s-j}(v_2)] \\
\times \psi_{s-j}(v_1) \psi_{s-j}(v_2) \\
\times Z^*_s(u_1, \theta_0) Z^*_s(u_2, \theta_0) \psi_{s-j}(v_1) \psi_{s-j}(v_2) \\
\times dW(u_1) dW(u_2) dW(v_1) dW(v_2) \\
= O \left( p^2 T^{-2q - 3(1 + 1)} \right), \quad (A.21)
\]

and

\[
\sum_{t=2q+2}^T \sum_{j=1}^q \sum_{l=1}^q a_j(l) a_l(l) \\
\times \int \int \int \int_{t=2q+2}^T \sum_{s=j+1}^{t-2q-1} E[Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{t-s}(v_1) \psi_{t-s}(v_2)] \\
\times \psi_{t-s}(v_1) \psi_{t-s}(v_2) \\
\times Z^*_s(u_1, \theta_0) Z^*_s(u_2, \theta_0) \psi_{s-j}(v_1) \psi_{s-j}(v_2) \\
\times dW(u_1) dW(u_2) dW(v_1) dW(v_2) \\
\leq 4 \sum_{j=1}^q \sum_{l=1}^q a_j(l) a_l(l) \\
\times \int \int \int \int_{t=2q+2}^T \sum_{s=j+1}^{t-2q-1} \psi_{t-s}(v_1) \psi_{t-s}(v_2) \\
\times E[Z^*_s(u_1, \theta_0) Z^*_s(u_2, \theta_0) \psi_{s-j}(v_1) \psi_{s-j}(v_2)] \\
\times \psi_{s-j}(v_1) \psi_{s-j}(v_2) \\
\times Z^*_s(u_1, \theta_0) Z^*_s(u_2, \theta_0) \psi_{s-j}(v_1) \psi_{s-j}(v_2) \\
\times dW(u_1) dW(u_2) dW(v_1) dW(v_2) \\
= O \left( p^2 T^{-2q - 3(1 + 1)} \right). \quad (A.23)
\]
Similarly, we can obtain
\[ E(\tilde{V}_q^2) = \frac{1}{2} \sum_{j=1}^{2q} \sum_{k=1}^{q} k^2 (j/p) k^2 (l/p) \]
\[ \times \int \int \int \left| E \left[ Z_{j+1} \left( \mathbf{u}, \mathbf{\theta}_0 \right) Z_{j+1} \left( \mathbf{u}, \mathbf{\theta}_0 \right) \psi_{j}(\mathbf{v}_1) \psi_{j}(\mathbf{v}_2) \right] \right|^2 \]  
\[ \times dW(\mathbf{u}_1) dW(\mathbf{u}_2) dW(\mathbf{v}_1) dW(\mathbf{v}_2) \left[ 1 + o(1) \right], \]
\[ \int \int \int \left| E \left[ Z_{j+1} \left( \mathbf{u}, \mathbf{\theta}_0 \right) Z_{j+1} \left( \mathbf{u}, \mathbf{\theta}_0 \right) \psi_{j}(\mathbf{v}_1) \psi_{j}(\mathbf{v}_2) \right] \right|^2 \]  
\[ \times dW(\mathbf{u}_1) dW(\mathbf{u}_2) dW(\mathbf{v}_1) dW(\mathbf{v}_2) \left[ 1 + o(1) \right]. \]

Because \( dW(\cdot) \) weighs sets symmetric about zero equally, we have \( E(\tilde{V}_q^2) = E(\tilde{V}^2) = E(\tilde{V}_q^2) \). Hence,
\[ \text{var}(2 \text{Re} \tilde{V}_q) = E(\tilde{V}_q^2) + E(\tilde{V}_q^2) + 2E \left| \tilde{V}_q \right|^2 \]
\[ = 2 \sum_{j=1}^{2q} \sum_{k=1}^{q} k^2 (j/p) k^2 (l/p) \]
\[ \times \int \int \int \left| E \left[ Z_{j+1} \left( \mathbf{u}, \mathbf{\theta}_0 \right) Z_{j+1} \left( \mathbf{u}, \mathbf{\theta}_0 \right) \psi_{j}(\mathbf{v}_1) \psi_{j}(\mathbf{v}_2) \right] \right|^2 \]  
\[ \times dW(\mathbf{u}_1) dW(\mathbf{u}_2) dW(\mathbf{v}_1) dW(\mathbf{v}_2) \left[ 1 + o(1) \right]. \]  

Put \( C(0, j, l) \equiv E \left[ Z_{j+1} \left( \mathbf{u}, \mathbf{\theta}_0 \right) Z_{j+1} \left( \mathbf{u}, \mathbf{\theta}_0 \right) \psi_{j}(\mathbf{v}_1) \psi_{j}(\mathbf{v}_2) \right] \), where \( \Sigma_j \left( \mathbf{u}_1, \mathbf{u}_2; \mathbf{\theta}_0 \right) \equiv E \left[ Z_j \left( \mathbf{u}, \mathbf{\theta}_0 \right) Z_{j} \left( \mathbf{u}, \mathbf{\theta}_0 \right) \right] \). Then
\[ E \left[ Z_{j+1} \left( \mathbf{u}, \mathbf{\theta}_0 \right) Z_{j+1} \left( \mathbf{u}, \mathbf{\theta}_0 \right) \psi_{j}(\mathbf{v}_1) \psi_{j}(\mathbf{v}_2) \right] \]
\[ = C(0, j, l) + \Sigma_j \left( \mathbf{u}_1, \mathbf{u}_2; \mathbf{\theta}_0 \right) \Omega_{j} \left( \mathbf{v}_1, \mathbf{v}_2 \right). \]

Given \( \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |C(0, j, l)| \leq C \) and \( |k(\cdot)| \leq 1 \), we have
\[ \text{var}(2 \text{Re} \tilde{V}_q) = 2 \sum_{j=1}^{2q} \sum_{k=1}^{q} k^2 (j/p) k^2 (l/p) \]
\[ \times dW(\mathbf{v}_1) dW(\mathbf{v}_2) \int \int \left| \Omega_{j} \left( \mathbf{v}_1, \mathbf{v}_2 \right) \right|^2 \]  
\[ \times dW(\mathbf{u}_1) dW(\mathbf{u}_2) \left[ 1 + o(1) \right] \]
\[ = 2p \sum_{m=1}^{2-q} \left\{ \frac{p^{-1} \sum_{j=1}^{q} k^2 (j/p) k^2 (j-m)/p} {1 + o(1)} \right\} \]
\[ \times \int \int \left| \Omega_{m} \left( \mathbf{v}_1, \mathbf{v}_2 \right) \right|^2 dW(\mathbf{v}_1) dW(\mathbf{v}_2) \]
\[ \times \int \int \left| \Sigma_0 \left( \mathbf{u}_1, \mathbf{u}_2; \mathbf{\theta}_0 \right) \right|^2 dW(\mathbf{u}_1) dW(\mathbf{u}_2) \left[ 1 + o(1) \right] \]
\[ = 2p \int_{0}^{\infty} k^2(z)iz \int_{-\infty}^{\infty} \int \left| G(0, \mathbf{v}_1, \mathbf{v}_2) \right|^2 dW(\mathbf{v}_1) dW(\mathbf{v}_2) \]
\[ \times \int \int \left| \Sigma_0 \left( \mathbf{u}_1, \mathbf{u}_2; \mathbf{\theta}_0 \right) \right|^2 dW(\mathbf{u}_1) dW(\mathbf{u}_2) \left[ 1 + o(1) \right]. \]
\[ V_{\theta}(t) = E \left[ V_q^2(t) \right] - V_1(t) - V_2(t) - V_3(t), \]

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

\[ L_{l,j}^j(u_1, u_2, v_1, v_2) \equiv Z_1(u_1, \theta_0) Z_2(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \]

\[ - E \left[ Z_1(u_1, \theta_0) Z_2(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right]. \]  

(A.28)

Then we write

\[ S_2(t) = \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \int \int \int E \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1)dW(v_2) + \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \int \int \int E \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1)dW(v_2) \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \int \int \int E \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1) dW(v_2) \]

\[ \equiv V_3(t) + S_3(t), \quad \text{say,} \]  

(A.29)

where

\[ S_3(t) = \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \int \int \int E \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1) dW(v_2) \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \int \int \int E \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1)dW(v_2) \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \int \int \int E \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1) dW(v_2) \]

\[ \equiv V_3(t) + V_3(t), \quad \text{say.} \]  

(A.30)

It follows from (A.26) and (A.27) and (A.29) and (A.30) that

\[ \sum_{t=2q+2}^{T} E \left[ V_q^2(t) \right] - E \left[ V_q^2(t) \right] = \sum_{t=2q+2}^{T} \sum_{l=0}^{q} a_l(j) a_l(l) \int \int \int E \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1)dW(v_2) \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \int \int \int E \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1)dW(v_2) \]

\[ = 0 \left( T^3 \right) + O \left( T^4 q^{-\frac{1}{4}+}\right) + O \left( T^3 q \right) + O \left( T^2 q^{-\frac{1}{4}+} \right), \]

where we have made use of the fact that \( E \left[ H_{3q-2q-1}(u_1, v_1) \right] \] \( \leq C^q \) for \( 1 \leq j \leq q \). It follows by Minkowski’s inequality and (A.5) that

\[ E \left[ \sum_{t=2q+2}^{T} V_q(t) \right]^2 \leq \left\{ \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \right\}^2 \]

\[ \times \left[ \sum_{t=2q+2}^{T} \sum_{l=0}^{q} a_l(j) a_l(l) \int \int \int \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \right] \]

\[ \times dW(v_1)dW(v_2) \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \int \int \int \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1)dW(v_2) \]

\[ = O \left( q^4 T \right). \]  

\textbf{Proof of Lemma A.10.} Recalling the definition of \( L_{l,j}^j(u_1, v_1, u_2, v_2) \) in (A.28), we have

\[ E \left[ \sum_{t=2q+2}^{T} V_q(t) \right]^2 \leq \left\{ \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \right\}^2 \]

\[ \times \left[ \sum_{t=2q+2}^{T} \sum_{l=0}^{q} a_l(j) a_l(l) \int \int \int \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \right] \]

\[ \times dW(v_1)dW(v_2) \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \int \int \int \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1)dW(v_2) \]

\[ = O \left( q^4 T \right). \]  

\textbf{Lemma A.9.} Let \( V_4(t) \) be defined as in (A.26). Then \( E \left[ \sum_{t=2q+2}^{T} V_4(t) \right]^2 = O(p^q q^{-\frac{1}{4}+} T) \).

\textbf{Lemma A.10.} Let \( V_3(t) \) be defined as in (A.29). Then \( E \left[ \sum_{t=2q+2}^{T} V_3(t) \right]^2 = O(p^q q^{-\frac{1}{4}+} T) \).

\textbf{Lemma A.11.} Let \( V_2(t) \) be defined as in (A.30). Then \( E \left[ \sum_{t=2q+2}^{T} V_2(t) \right]^2 = O(p^q q^{-\frac{1}{4}+} T) \).

\textbf{Lemma A.12.} Let \( V_1(t) \) be defined as in (A.30). Then \( E \left[ \sum_{t=2q+2}^{T} V_1(t) \right]^2 = O(p^q q^{-\frac{1}{4}+} T) \).

\textbf{Proof of Lemma A.9.} Recalling the definition of \( \tilde{L}_{l,j}^j(u_1, v_1, u_2, v_2) \) in (A.26), we can obtain

\[ E \left[ \sum_{t=2q+2}^{T} \tilde{L}_{l,j}^j(u_1, v_1, u_2, v_2) \right]^2 \leq \left\{ \sum_{t=2q+2}^{T} \sum_{l=0}^{q} a_l(j) a_l(l) \int \int \int \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \right\}^2 \]

\[ \times \left[ \sum_{t=2q+2}^{T} \sum_{l=0}^{q} a_l(j) a_l(l) \int \int \int \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \right] \]

\[ \times dW(v_1)dW(v_2) \sum_{j=1}^{q} \sum_{l=1}^{q} a_l(j) a_l(l) \int \int \int \left[ Z_t(u_1, \theta_0) Z_t(u_2, \theta_0) \psi_{s,j}(v_1) \psi_{t,j}(v_2) \right] \]

\[ \times dW(v_1)dW(v_2) \]

\[ = O \left( T^4 q^{-\frac{1}{4}+} \right) + O \left( T^3 q \right) + O \left( T^2 q^{-\frac{1}{4}+} \right), \]
\[ E \left( \sum_{s=\text{max}(j,l)}^{\tau-2q-1} L^j_{l,s}(u_1, v_1, u_2, v_2) \right)^2 = \sum_{s=\text{max}(j,l)}^{\tau-2q-1} E \left( \sum_{s=\text{max}(j,l)}^{\tau-2q-1} \beta^{s+t} (2q) \left[ E \left| L^j_{l,s}(u_1, v_1, u_2, v_2) \right|^{2(1+\delta)} \right]^{\frac{1}{2(1+\delta)}} \right) \times \left[ E \left| L^j_{l,s}(u_1, v_1, u_2, v_2) \right|^{2(1+\delta)} \right]^{\frac{1}{2(1+\delta)}} = O(tq^t) . \]

It follows that
\[ E \left( \sum_{t=2q+2}^{\tau} V_s(t) \right)^2 \leq \left\{ \sum_{t=2q+2}^{\tau} \sum_{j=1}^{q} \sum_{l=1}^{q} a_r(j)a_l(l) \times \int \int \int \int \left| E[Z_s(u_1, \theta_0) Z_{t-s}(u_2, \theta_0) \psi_{t-s}(v_1) \psi_{t-s}(v_2)] \right| \times \left[ \sum_{s=\text{max}(j,l)}^{\tau-2q-1} Z_{s}(u_2, \theta_0) \psi_{s-t}(v_2) \right]^{\frac{1}{2}} \times dW(u_1)dW(u_2) \times dW(v_1)dW(v_2) \right\}^2 \leq C t q \left[ \sum_{j=1}^{q} a_r(j) \right]^4 = O \left( t q^t / T^4 \right) . \]

**Proof of Lemma A.11.** The result that \( E[\sum_{t=2q+2}^{\tau} V_s(t)]^2 = O(\sigma^2 / T) \) by Minkowski's inequality and
\[ E \left[ V_s(t) \right]^2 \leq \left\{ \sum_{t=2q+2}^{\tau} \sum_{j=1}^{q} \sum_{l=1}^{q} a_r(j)a_l(l) \times \int \int \int \int \left| E[Z_s(u_1, \theta_0) Z_{t-s}(u_2, \theta_0) \psi_{t-s}(v_1) \psi_{t-s}(v_2)] \right| \times \left[ \sum_{s=\text{max}(j,l)}^{\tau-2q-1} Z_{s}(u_2, \theta_0) \psi_{s-t}(v_2) \right]^{\frac{1}{2}} \times dW(u_1)dW(u_2) \times dW(v_1)dW(v_2) \right\}^2 \leq C t q \left[ \sum_{j=1}^{q} a_r(j) \right]^4 = O \left( t q^t / T^4 \right) . \]

**Proof of Lemma A.12.** The result that \( E[\sum_{t=2q+2}^{\tau} V_s(t)]^2 = O(\sigma^2 / T^4) \) follows from Minkowski's inequality, \( p \to \infty \), and the fact that
\[ E[|V_s(t)|^2] = E \left[ \sum_{s=\text{max}(j,l)}^{\tau-2q-1} L^j_{l,s}(u_1, v_1, u_2, v_2) \right]^2 = \sum_{s=\text{max}(j,l)}^{\tau-2q-1} \beta^{s+t} (2q) \left[ E \left| L^j_{l,s}(u_1, v_1, u_2, v_2) \right|^{2(1+\delta)} \right]^{\frac{1}{2(1+\delta)}} \times \left[ E \left| L^j_{l,s}(u_1, v_1, u_2, v_2) \right|^{2(1+\delta)} \right]^{\frac{1}{2(1+\delta)}} = O(tq^t) . \]

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

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

**Theorem A.4.** Under the conditions of Theorem 2,
\[ \left( \int_{-\pi}^{\pi} \left| F(\omega, \textbf{u}, \textbf{v}) \right|^2 d\theta d\textbf{u} d\textbf{v} \right)^{\frac{1}{2}} \to 0. \]

**Proof of Theorem A.3.** It suffices to show that
\[ T^{-1} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{j=1}^{p^3} k^2(j/p) T_j \left[ \left| \hat{F}_j(u, v) \right|^2 - \left| \hat{F}_j(u, v) \right|^2 \right] \times dW(u)dW(v) \to 0, \]

(A.31)
Proof of Theorem 3. 

From (A.31), the Cauchy–Schwarz inequality, and the fact that $T^{-1} \left\| \sum_{j=1}^T \kappa_j (p) \right\| \to 0$ are straightforward, we focus on the proof of (A.31). 

From (A.5), the Cauchy–Schwarz inequality, and the fact that $T^{-1} \left\| \sum_{j=1}^T \kappa_j (p) \right\| \to 0$ are straightforward, we focus on the proof of (A.31). 

Proof of Theorem 3. The proof is very similar to Hong (1999). Proof of Thm.5, for the case $(m, l) = (0, 0)$. The consistency result follows from (a) $p^{-1} \sum_{j=1}^T \kappa_j (p) \to \int_0^\infty \kappa (\omega) d\omega$; (b) $\int \int \int \phi (\omega, u, v) F (\omega, u, v) |d\omega dW (u) dW (v)| \to 0$; (c) $\hat{C} (0, 0) = O_p (p)$; (d) $\delta (0, 0) \to p D (0, 0)$. 

Proof of Theorem 3. The proof is very similar to Theorem 1, with $\hat{Z}_t (0, \theta), \hat{Z}_t (0, \theta), \hat{Z}_t (0, \theta), \hat{Z}_t (0, \theta)$, $\hat{Z}_t (0, \theta)$, $D (0, 0)$ respectively. 

References 


