Characteristic Function-Based Testing for Multifactor Continuous-Time Markov Models via Nonparametric Regression

Bin Chen Yongmiao Hong

University of Rochester Cornell University & Xiamen University

June, 2008

We thank the co-editor (Oliver Linton) and two referees for careful and constructive comments. We also thank Jianqing Fan, Robert Jarrow, Joon Park, Peter Phillips, George Tauchen, Dag Tjøstheim, Qiwei Yao and seminar participants at the 2006 Symposium on Financial Engineering and Risk Management in Beijing, the 2006 International Symposium on Econometric Theory and Applications (SETA) in Xiamen, the North American Econometric Society Summer Meeting in Minneapolis, the 2006 Far Eastern Econometric Society Meeting in Beijing and the 2008 International Symposium on Recent Developments of Time Series Econometrics in Xiamen for their comments. All remaining errors are solely ours. Hong acknowledges support from the National Science Foundation of China and the Cheung Kong Visiting Scholarship by the Chinese Ministry of Education and Xiamen University, China. Correspondences: Bin Chen, Department of Economics, University of Rochester, Rochester, NY 14627; email: binchen.ur@gmail.com; Yongmiao Hong, Department of Economics & Department of Statistical Science, Cornell University, Ithaca, NY 14850, USA, and Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Xiamen 361005, China; email: yh20@cornell.edu.
Characteristic Function-Based Testing for Multifactor Continuous-Time Markov Models via Nonparametric Regression

We develop a nonparametric regression-based goodness-of-fit test for multifactor continuous-time Markov models using the conditional characteristic function, which often has a convenient closed-form or can be approximated accurately for many popular continuous-time Markov models in economics and finance. An omnibus test procedure fully utilizes the information in the joint conditional distribution of the underlying processes and hence has power against a vast class of continuous-time alternatives in the multifactor framework. A class of easy-to-interpret diagnostic procedures is also proposed to gauge possible sources of model misspecifications. All our test statistics have a convenient asymptotic $N(0,1)$ distribution under correct model specification. Simulations show that our tests have reasonable size, thanks to the dimension reduction in nonparametric regression, and good power against a variety of alternatives, including misspecifications in the joint dynamics even if the dynamics of each individual component is correctly specified. This feature is not attainable by some existing tests. A parametric bootstrap improves the finite sample performance of proposed tests, but with higher computational costs.

Key words: Conditional characteristic function, Goodness-of-fit, Multifactor continuous-time Markov model, Nonparametric regression

JEL Classifications: C4, E4, G0.
1. INTRODUCTION

Continuous-time Markov models are powerful analytic tools in modern finance and economics. Itô processes have been popularly adapted, and the more general Lévy processes have been the object of the recent research for derivatives pricing in the literature (e.g., Barndorff-Nielsen and Shephard 2001, Carr and Wu 2003, 2004, Chernov, Gallant, Ghysels and Tauchen 1999). Several reasons have contributed to the popularity of continuous-time Markov models in finance and economics. First, continuous information flows into financial markets provide a justification for using continuous-time models, and the development of stochastic calculus provides a powerful tool for elegant mathematical treatment of continuous-time models. Second, the Markov assumption, which is a maintained condition for almost all continuous-time models in finance and economics, simplifies greatly the involved mathematical derivation. Under the Markov assumption, the conditional probability distribution of future values of the underlying process, conditional on the currently available information, depends only on the current value of the process and the inclusion of any additional information available at the current time will not alter this conditional probability distribution. From an economic point of view, economic agents’ rationality provides a solid justification for the Markov assumption. Economic agents update beliefs and make decisions sequentially. Their subjective beliefs about future uncertainty and optimal decision rules depend on the past information only via the current state. In fact, Markov models are not as restrictive as they might first appear. As highlighted by Merton (1990), many non-Markovian processes can be transformed into Markov processes by the method of "expansion of the states".

Econometric analysis of continuous-time models is generally more challenging than that of discrete-time dynamic models. Much progress has been made in the literature in estimating continuous-time models. For example, Ait-Sahalia’s (2002) approximated MLE, Bates’ (2007) filtration-based MLE, Chib, Pitt and Shephard’s (2004) Markov Chain Monte Carlo (MCMC) method, Gallant and Tauchen’s (1996) efficient method of moments (EMM) method, and Singleton’s (2001) conditional characteristic function-based maximum likelihood estimation (MLE-CCF) and conditional characteristic function-based general method of moments (GMM-CCF) methods have been proposed. Compared with the vast literature on estimation, however, there has been relatively little effort devoted to specification analysis and evaluation of continuous-time models. In a continuous-time framework, model misspecification generally renders inconsistent parameter estimators and their conventional variance-covariance matrix estimators, which could lead to misleading conclusions on statistical inference. The validity of economic interpretations for model parameters also crucially depends on correct model specification.

1 This is the basic idea behind Markov decision processes (MDPs), which provide a broad framework for modeling sequential decision making under uncertainty. MDPs have been used extensively in both microeconomics and macroeconomics as well as in finance and marketing (see, e.g., Ljungqvist and Sargent 2000, Rust 1994, for excellent surveys). Applications include investment under uncertainty (Lucas and Prescott 1971, Sargent 1987), asset pricing models (Hall 1978, Hansen and Singleton 1983, Lucas 1978, Mehra and Prescott 1985), economic growth (Lucas 1988, Romer 1986, 1990), optimal taxation (Lucas and Stokey 1983, Zhu 1992), and equilibrium business cycles (Kydland and Prescott 1982).

2 In an important review, Sundaresan (2001) states that “perhaps the most significant development in the continuous-time field during the last decade has been the innovations in econometric theory and in the estimation techniques for models in continuous time.” For other reviews of related literature, see (e.g.) Tauchen (1997).
More importantly, a misspecified model can yield large errors in pricing, hedging and risk management. For many applications, it is important that a continuous-time Markov model fits data adequately.

Nevertheless, economic theories usually do not suggest a concrete functional form for continuous-time Markov models. The choice of a model is somewhat arbitrary, often based on convenience and empirical experience of the practitioner. For example, in the pricing and hedging literature, a continuous-time model is often assumed to have a functional form that yields a closed-form pricing formula, as is the case of multivariate affine term structure models (ATSMs) for interest rates (Dai and Singleton 2000, Duffie and Kan 1996). It is important to develop a reliable omnibus specification test for popular continuous-time Markov models. In addition, diagnostic procedures that focus on misspecification in certain directions (e.g., conditional mean, conditional variance and conditional correlation) will be also useful for guiding further improvement of the model.

There have been some works on testing continuous-time models. Ait-Sahalia (1996a) develops a nonparametric test for univariate diffusion models. Observing that the drift and diffusion functions completely characterize the stationary density of a diffusion model, Ait-Sahalia (1996a) checks the adequacy of the diffusion model by comparing the model-implied stationary density with a smoothed kernel density estimator based on discretely sampled data.\(^3\) 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, but they may overlook a misspecified model with a correct stationary density.

Hong and Li (2005) develop a specification test for continuous-time models using the transition density, which can capture the full dynamics of a continuous-time process. Observing the fact that when a continuous-time model is correctly specified, the probability integral transform (PIT) of the observed sample with respect to the model-implied transition density is i.i.d. \(U[0,1]\), they check the joint hypothesis of i.i.d. \(U[0,1]\) using a nonparametric density estimator. The most appealing feature of this test is its robustness to persistent dependence in data because the PIT series is always i.i.d. \(U[0,1]\) under correct model specification. This approach, however, cannot be extended to a multivariate joint transition density, because it is well-known that the PIT series with respect to a multivariate joint transition density is no longer i.i.d \(U[0,1]\) even if the model is correctly specified. In an empirical study, Hong and Li (2005) apply their test to evaluate multivariate continuous-time models by considering the PIT for each individual state variable, with a suitable partitioning. This practice is valid, but it may fail to detect model misspecification in the joint dynamics of state variables. In particular, it may miss misspecification of dynamic conditional correlations between individual state variables.

Alternative tests for univariate diffusion models have recently been suggested in the literature. Ait-Sahalia, Fan and Peng (2006) propose new tests by comparing the model-implied transition density and distribution functions with their nonparametric counterparts respectively. Chen, Gao and Tang (2007) develop a transition density-based test using a nonparametric empirical likelihood approach. Li (2007) tests on the parametric specification of the diffusion function by measuring the distance between the

\(^3\)Ait-Sahalia (1996a) also proposes a transition density-based test that exploits the "transition discrepancy" characterized by the forward and backward Kolmogorov equations, although the marginal density-based test is more emphasized there.
model-implied diffusion function and its kernel estimator. All these tests are constructed in a univariate framework although some of them may be extended to multivariate continuous-time models.

Gallant and Tauchen (1996) propose a class of Efficient Method of Moments (EMM) tests that can be used to test multivariate continuous-time models. Their idea is to match the model-implied moments to the moments implied by a seminonparametric (SNP) transition density of the observed sample. They propose a minimum $\chi^2$ test for model misspecification, and a class of diagnostic $t$-tests to gauge possible sources for model failure. Bhardwaj, Corradi and Swanson (2007) consider a simulation-based test, which is an extension of Andrews’ (1997) conditional Kolmogorov test, for multivariate diffusion models. The limit distribution of their test is not nuisance parameter free and asymptotic critical values must be obtained via a block bootstrap. Moreover, since these tests are by-products of the EMM and the simulated GMM algorithms respectively, they cannot be used when the model is estimated by other methods. This may limit the scope of these tests to otherwise very useful applications.

In a generalized cross-spectral non-Markov framework, Chen and Hong (2005) propose a new test for multivariate continuous-time models based on the CCF, which often has a closed form or can be approximated accurately for many popular multifactor continuous-time models. As the Fourier transform of the transition density, the CCF contains the full information of the joint dynamics of underlying processes. This provides a basis for constructing an omnibus test for multifactor continuous-time models. Unlike Hong and Li (2005), Chen and Hong (2005) fully exploit the information in the joint transition density of underlying processes and hence can capture model misspecifications in their joint dynamics. Chen and Hong (2005) do not assume that the data generating process (DGP) is Markov. They take a generalized cross-spectral density approach, which employs many lags simultaneously. For a Markov DGP (under both the null and alternative hypotheses), this test will not be most efficient, because it includes the past information of many lags which is redundant under the Markov assumption. In this case, it is more efficient to focus on the first lag order only. This is pursued in the present paper, which ignores the lag structure and focuses on functional form misspecification.

There has been a long history of using the characteristic function in estimation and hypotheses testing in econometrics and statistics. For example, Feuerverger and McDunnough (1981) and Feuerverger (1990) discuss parameter estimation using the joint empirical characteristic function (ECF) for stationary Markov time series models. Epps and Pulley (1983) propose an omnibus test of normality via a weighted integral of the squared modulus of the difference between the characteristic functions of the observed sample and of the normal distribution. Su and White (2007) test conditional independence by comparing the unrestricted and restricted CCFs via a kernel regression. We note that all above works deal with discrete-time models, but the characteristic function 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, Pan and Singleton 2000) and time-changed Lévy processes (e.g., Chernov et al. 1999), 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 characteristic function. Jiang
and Knight (2002) derive the unconditional joint characteristic function of an AJD model and use it
to develop some GMM and ECF estimators. Singleton (2001) proposes both time-domain estimators
based on the Fourier transform of the CCF, and frequency-domain estimators directly based on the
CCF. Carrasco, Chernov, Florens and Ghysels (2007) propose GMM estimators with a continuum of
moment conditions via the characteristic function. These estimation methods diﬀer in their ways of
using the conditional information set. Besides its convenient closed-form for many popular continuous-
time models, the CCF can be diﬀerentiated to generate moments, which provides powerful and intuitive
tools to check various speciﬁc aspects of a joint conditional distribution.

Motivated by these appealing features, we provide a CCF-characterization for the adequacy of a
continuous-time Markov model and use it to construct a speciﬁcation test for continuous-time Markov
models. The basic idea is that if a Markov model is correctly speciﬁed, prediction errors associated with
the model-implied CCF should be a martingale diﬀerence sequence (MDS). This characterization has
never been used in any goodness-of-ﬁt test for continuous-time models, although it has been used in
estimating them (e.g., Singleton 2001). To ensure the power of our test, we use nonparametric regression
to check whether these prediction errors are explainable by the current values of the underlying processes.
Our approach has several attractive properties.

First, our nonparametric omnibus test fully exploits the information in the joint transition density of
the state vector rather than only the information in the transition density of each individual component.
Hence, it can capture a vast range of model misspeciﬁcations in the joint dynamics of state variables.4
In particular, it can detect model misspeciﬁcations in the joint transition density even if the transition
density of each individual component is correctly speciﬁed.

Second, our test is applicable to a wide variety of continuous-time Markov models, such as diﬀusions,
jump diﬀusions and continuous-time Markov chains. Because we use the CCF to characterize the
adequacy of a continuous-time Markov model, our test is most convenient when the model has a closed-
form CCF, as is the case for a class of AJD models (e.g., Duffie et al. 2000) and a class of time-changed
Lévy processes (e.g., Chernov et al. 1999). However, we emphasize that our test is also applicable
to continuous-time Markov models with no closed-form CCF. In this case, we can use inverse Fourier
transforms or simulation techniques to calculate the CCF. See, e.g., Bates (2007) and Carrasco et al.
(2007) for discussion on simulation methods when the CCF has no closed-form.

Third, our test is applicable to partially observed multifactor continuous-time Markov models. An
example is the stochastic volatility (SV) models for interest rates and equity returns.

Fourth, we do not require any particular estimation method for model parameters. Any \( \sqrt{T} \)-
consistent estimators may be used. Parameter estimation uncertainty does not affect the asymptotic
distribution of our test. This makes our test easily implementable in light of the notorious difficulty
of obtaining asymptotically efficient estimators for multifactor continuous-time models. The inputs
needed to calculate the test statistics are the observed data and the model-implied CCF or its approx-
imation. Since we impose our regularity conditions on the CCF of a discretely observed sample
of a continuous-time Markov model, our test is readily applicable to discrete-time Markov probability
distribution models, in addition to continuous-time Markov models with discrete observations.

Fifth, in addition to the omnibus test, we also propose a class of diagnostic tests by differentiating
the CCF. These derivative tests provide useful information on how well a continuous-time Markov
model captures various specific aspects of the dynamics. In particular, they can reveal information on
neglected dynamics in conditional means, conditional variances and conditional correlations respectively.
These procedures complement Gallant and Tauchen’s (1996) EMM-based individual \( t \)-tests. All of our
omnibus and diagnostic tests are derived from a unified framework. They have a convenient null
asymptotic \( N(0,1) \) distribution.

In Section 2, we introduce the framework, state the hypotheses of interest, and provide a charac-
terization for correct specification of a continuous-time Markov model. In Section 3, we propose an
omnibus goodness-of-fit test using smoothed regression, and in Section 4 we derive the asymptotic null
distribution of our omnibus test and discuss its asymptotic power property. We then construct a class
of diagnostic procedures that focus on various specific aspects of the joint dynamics of a multifactor
continuous-time model in Section 5. In Section 6, we consider the tests for multifactor continuous-time
models with partially unobservable components. In Section 7, we apply our tests to both univariate and
bivariate continuous-time models in a simulation study. A conclusion follows in Section 8. All mathe-
matical proofs are collected in an appendix. Throughout, 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

Given a complete probability space \( (\Omega, \mathcal{F}, P) \) and an information filtration \( \mathcal{F}_t \), we assume that a
d \( \times 1 \) state vector \( X_t \) is a continuous-time Markov process in some state space \( D \subset \mathbb{R}^d \), where \( d \geq 1 \)
is an integer. In financial modelling, the following class \( \mathcal{M} \) of continuous-time models is often used to
capture the dynamics of \( X_t \):

\[
dX_t = \mu(X_t, \theta) \, dt + \sigma(X_t, \theta) \, dW_t + dJ_t(\theta), \quad \theta \in \Theta,
\]

where \( W_t \) is a \( d \times 1 \) standard Brownian motion in \( \mathbb{R}^d \), \( \Theta \) is a finite-dimensional parameter space,
\( \mu : D \times \Theta \to \mathbb{R}^d \) is a drift function (i.e., instantaneous conditional mean), \( \sigma : D \times \Theta \to \mathbb{R}^{d \times d} \)
is a diffusion function (i.e., instantaneous conditional standard deviation), and \( J_t \) is a pure jump process
whose jump size follows a probability distribution \( \nu : D \times \Theta \to \mathbb{R}^+ \) and whose jump times arrive with
intensity \( \lambda : D \times \Theta \to \mathbb{R}^+ \).

\(^5\)It is assumed that \( \mu, \sigma, \nu \) and \( \lambda \) are regular enough to have a unique strong solution to (2.1). See (e.g.) Ait-Sahalia
The above setup is a general multifactor framework that nests most existing continuous-time Markov models in finance. For example, suppose the drift $\mu (\cdot , \cdot )$, the instantaneous covariance matrix $\sigma (\cdot , \cdot ) \sigma (\cdot , \cdot )'$ and the jump intensity $\lambda (\cdot , \cdot )$ are all affine functions of the state vector $X_t$; namely,

$$
\begin{align*}
\mu (X_t, \theta) &= K_0 + K_1 X_t, \\
\sigma (X_t, \theta) \sigma (X_t, \theta)'[j,l] &= [H_0][j,l] + [H_1][j,l]X_t, & 1 \leq j, l \leq d, \\
\lambda (X_t, \theta) &= L_0 + L_1 X_t,
\end{align*}
$$

(2.2)

where $K_0 \in \mathbb{R}^d$, $K_1 \in \mathbb{R}^{d \times d}$, $H_0 \in \mathbb{R}^{d \times d}$, $H_1 \in \mathbb{R}^{d \times d \times d}$, $L_0 \in \mathbb{R}$, and $L_1 \in \mathbb{R}^d$ are unknown parameters.

Then we obtain the class of popular AJD models of Duffie et al. (2000).

It is well-known that for a continuous-time Markov model described by a stochastic differential equation (SDE), the specification of the drift $\mu (X_t, \theta)$, the diffusion $\sigma (X_t, \theta)$ and the jump process $J_t(\theta)$ together completely determines the joint transition density of the state vector $X_t$. We use $p(x, t|X_s, \theta)$ to denote the model-implied transition density function of $X_t = x$ given $X_s$, where $s < t$. Suppose $X_t$ has an unknown true transition density $p_0(x, t|X_s)$. Then the continuous-time Markov model is correctly specified for the full dynamics of $X_t$ if there exists some unknown parameter value $\theta_0 \in \Theta$ such that

$$
H_0 : p(x, t|X_s, \theta_0) = p_0(x, t|X_s) \quad \text{almost surely (a.s.) and for all } t, s, s < t.
$$

(2.3)

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

$$
H_A : p(x, t|X_s, \theta) \neq p_0(x, t|X_s) \text{ for some } t > s \text{ with positive probability measure,}
$$

(2.4)

then the continuous-time model is misspecified for the full dynamics of $X_t$. We maintain the Markov assumption for $X_t$ under both $H_0$ and $H_A$.

The transition density-based characterization can be used to test correct specification of the continuous-time model. When $X_t$ is univariate (i.e., $d = 1$), Hong and Li (2005) propose a kernel-based test for a continuous-time model by checking whether the PIT

$$
Z_t(\theta_0) = \int_{-\infty}^{X_t} p(x, t|X_{t-\Delta}, \theta_0) dx \sim \text{i.i.d.} U[0, 1] \text{ under } H_0,
$$

(2.5)

where $\Delta$ is the sampling interval for a discretely observed sample. The most appealing merit of this test is its robustness to persistent dependence in $\{X_t\}$. 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 multifactor joint transition density $p(x, t|X_{t-\Delta}, \theta)$, because when $d > 1$,

$$
Z_t(\theta_0) = \int_{-\infty}^{X_{1,t}} \cdots \int_{-\infty}^{X_{d,t}} p(x, t|X_{t-\Delta}, \theta_0) dx
$$

(2.6)

is no longer i.i.d. $U[0,1]$ even if $H_0$ holds, where $X_t = (X_{1,t}, ..., X_{d,t})'$. Hong and Li (2005) suggest using the PIT for each state variable with a suitable partitioning. 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 misspecifications in the joint dynamics of \( X_t \). For example, consider the DGP

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

where \( \{W_{1,t}, W_{2,t}\} \) are independent standard Brownian motions. Suppose we fit the data using the model

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

Then this model is misspecified because it ignores correlations in drift. Now, following Hong and Li (2005), we calculate the generalized residuals \( \{Z_t\} = \{Z_{1,t}, Z_{2,t}, Z_{1,t-\Delta}, Z_{2,t-\Delta}, \ldots\} \), where \( Z_{1,t} \) and \( Z_{2,t} \) are the PITs of \( X_{1,t} \) and \( X_{2,t} \) with respect to the conditional density models \( p(X_{1,t}, t|X_{-\Delta}, X_{2,t}, \theta) \) and \( p(X_{2,t}, t|X_{t-\Delta}, \theta) \) respectively, and \( \theta = (\kappa_{11}, \kappa_{22}, \theta_1, \theta_2, \sigma_{11}, \sigma_{22})' \). Then Hong and Li’s (2005) test will have no power because each of these PITs is an i.i.d. \( U[0,1] \) sequence. We will further investigate this issue via simulation in Section 7.

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

\[
    \varphi(u, t|X_s, \theta) \equiv E_\theta \left[ \exp \left( iu'X_t \right) | X_s \right] = \int_{\mathbb{R}^d} \exp \left( iu'x \right) p(x, t|X_s, \theta) dx, \ u \in \mathbb{R}^d, \ i = \sqrt{-1}, \ (2.7)
\]

where \( E_\theta (\cdot|X_s) \) denotes the expectation under the model-implied transition density \( p(x, t|X_s, \theta) \).

Given the equivalence between the transition density and the CCF, the hypotheses of interest \( \mathbb{H}_0 \) in (2.3) versus \( \mathbb{H}_A \) in (2.4) can be written as follows:

\[
    \mathbb{H}_0 : E \left[ \exp \left( iu'X_t \right) | X_s \right] = \varphi(u, t|X_s, \theta_0) \ a.s. \ \text{for all} \ u \in \mathbb{R}^d \ \text{and for some} \ \theta_0 \in \Theta \ (2.8)
\]

versus

\[
    \mathbb{H}_A : E \left[ \exp \left( iu'X_t \right) | X_s \right] \neq \varphi(u, t|X_s, \theta) \ \text{with positive probability measure for all} \ \theta \in \Theta. \ (2.9)
\]

Suppose we have a discretely observed sample \( \{X_t\}_{t=\Delta}^{T} \) of size \( T \), where for simplicity we set the sampling interval \( \Delta = 1 \). Define a complex-valued process

\[
    Z_t(u, \theta) \equiv \exp \left( iu'X_t \right) - \varphi(u, t|X_{t-1}, \theta), \ u \in \mathbb{R}^d \ \text{and} \ \theta \in \Theta. \ (2.10)
\]

Then \( \mathbb{H}_0 \) is equivalent to the following MDS characterization

\[
    E[Z_t(u, \theta_0)|X_{t-1}] = 0 \ a.s. \ \text{for all} \ u \in \mathbb{R}^d \ \text{and some} \ \theta_0 \in \Theta. \ (2.11)
\]

Thus, we can check (2.11) to test \( \mathbb{H}_0 \) versus \( \mathbb{H}_A \).

It is important to emphasize that (2.11) is not a simple MDS characterization. It is essentially a MDS process \( Z_t(\cdot, \theta_0) \) indexed by nuisance parameter \( u \in \mathbb{R}^d \) and we need to check all possible values for \( u \) in \( \mathbb{R}^d \). This is challenging, but it offers the omnibus property of the resulting test. Moreover, by
taking derivatives with respect to \(u\) at the origin, we are able to direct the test toward certain specific aspects of the joint dynamics of \(X_t\) (see Section 6).

To compute \(Z_t(u, \theta)\), we need to know the CCF. In principle, we can always recover the CCF by simulation when it has no closed form. For a given \(\theta\) and conditional on \(X_{t-1}\), we can generate a sequence \(\{\tilde{X}_{\theta,j}^{t-1}, j = 1, 2, \ldots, J\}\) via (e.g.) the Euler scheme or the generalized Milstein scheme (see, e.g., Kloeden, Platen and Schurz 1994 for more discussion) and then estimate the CCF by

\[
\hat{\varphi}(u, t|X_{t-1}, \theta) = \frac{1}{J} \sum_{j=1}^{J} \exp(iu\tilde{X}^{\theta,j}_{t-1}).
\]

It can be shown that for each \(t\), \(\hat{\varphi}(u, t|X_{t-1}, \theta) \to P \varphi(u, t|X_{t-1}, \theta)\) if \(J \to \infty\). Therefore, our CCF approach is generally applicable. Alternatively, we can accurately approximate the model transition density by using (e.g.) the Hermite expansion method of Ait-Sahalia (2002, 2007), the simulation methods of Brandt and Santa-Clara (2001) and Pedersen (1995), or the closed-form approximation method of Duffie, Pedersen and Singleton (2003), and then calculate the Fourier transform of the estimated transition density. Nevertheless, our test is most useful when the CCF has a closed-form, as is illustrated by the examples below:

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 AJD models, the CCF of \(X_t\) conditional on \(X_{t-1}\) is a closed-form exponential-affine function of \(X_{t-1}\):

\[
\varphi(u, t|X_{t-1}, \theta) = \exp \left[ \alpha_{t-1}(u) + \beta_{t-1}(u)X_{t-1} \right],
\]

where \(\alpha_{t-1}: \mathbb{R}^d \to \mathbb{R}\) and \(\beta_{t-1}: \mathbb{R}^d \to \mathbb{R}^d\) satisfy the complex-valued Riccati equations:

\[
\begin{aligned}
\dot{\alpha}_t &= K_0^* \alpha_t + i u B^* \beta_t + L_0 (g (\beta_t) - 1), \\
\dot{\beta}_t &= K_0^* \beta_t + i u B^* \beta_t + L_0 (g (\beta_t) - 1),
\end{aligned}
\]

with boundary conditions \(\beta_T(u) = iu\) and \(\alpha_T(u) = 0\).

AJD models have been widely used in finance. For example, in the interest rate term structure literature, Dai and Singleton (2000), Duffie and Kan (1996) and Duffie et al. (2000) have developed a class of ATSMs. Assuming that the spot rate \(r_t\) is an affine function of the state vector \(X_t\) and that \(X_t\) follows affine diffusions under an equivalent martingale measure, Duffie and Kan (1996) show that the yield of the zero coupon bond can be expressed as an affine function of \(X_t\):

\[
Y(X_t, \tau) \equiv -\frac{1}{\tau} \log P(X_t, \tau) = \frac{1}{\tau} \left[ -A(\tau) + B(\tau)X_t \right],
\]

where \(\tau\) is the remaining time to maturity, and the functions \(A: \mathbb{R}^+ \to \mathbb{R}\) and \(B: \mathbb{R}^+ \to \mathbb{R}^d\) either have a closed-form or can be easily solved via numerical methods. Since \(Y(X_t, \tau)\) is a linear transformation of \(X_t\), its CCF also has the closed-form solution; namely,

\[
\varphi_Y(u, t|X_{t-1}, \theta, \tau) = E_\theta \{ \exp[iuY(X_t, \tau)] | Y(X_{t-1}, \tau) \} = \exp \left\{ \frac{-iuA(\tau)}{\tau} + \alpha_{t-1} \left[ \frac{uB(\tau)}{\tau} \right] + \beta_{t-1} \left[ \frac{uB(\tau)}{\tau} \right]'X_{t-1} \right\},
\]

\[\tag{2.15}\]
where \( \alpha_{t-1} \) and \( \beta_{t-1} \) satisfy (2.13) and \( \mathbf{X}_t = [\mathbf{B} \tau \mathbf{X}_t]^{-1} [\tau Y (\mathbf{X}_t, \tau) + A (\tau)] \). In particular, for a multifactor Vasicek model, the CCF of the yield of the zero coupon bond has an analytical expression.

Ahn, Dittmar and Gallant (2002) break the tension between the specification of the instantaneous conditional volatility and that of the instantaneous conditional correlations in ATSMs by assuming that the spot rate is a quadratic function of the normally distributed state vector. They derive the yield of conditional volatility and that of the instantaneous conditional correlations in ATSMs by assuming that the Vasicek model, the CCF of the yield of the zero coupon bond has an analytical expression.

Models (QTSMs), for which the CCF of \( \mathbf{X}_t \) can be easily solved via numerical methods. This class of models is called the Quadratic Term Structure Models (QTSMs), for which the CCF of \( \mathbf{X}_t \) has an analytical expression.

where functions \( A : \mathbb{R}^+ \rightarrow \mathbb{R} \), \( \mathbf{B} : \mathbb{R}^+ \rightarrow \mathbb{R}^d \), and \( \mathbf{M} : \mathbb{R}^+ \rightarrow \mathbb{R}^{d \times d} \) either have a closed-form or can be easily solved via numerical methods. This class of models is called the Quadratic Term Structure Models (QTSMs), for which the CCF of \( \mathbf{X}_t \) is:

\[
\varphi_Y (u, t|\mathbf{X}_{t-1}, \theta, \tau) = \exp \left\{ \sum_{i=1}^{d} \frac{iu \lambda_j \omega_j^2}{1 - 2iu \lambda_j} \right\} (1 - 2i u \lambda_j)^{-\frac{1}{2}},
\]

where \( \lambda_j \) and \( \omega_j \) (\( j = 1, 2, \ldots, d \)) are some constants defined in Ahn et al. (2002).

### 3. NONPARAMETRIC REGRESSION-BASED CCF TESTING

We now propose a new test for the adequacy of a multivariate continuous-time Markov model using nonparametric regression. Recall that the CCF-based generalized residual \( Z_t (u, \theta) \) has the MDS property that

\[
m (u, X_{t-1}, \theta_0) = E \{ Z_t (u, \theta_0)|X_{t-1} \} = 0 \text{ a.s. for all } u \in \mathbb{R}^d \text{ and some } \theta_0 \in \Theta.
\]

To gain insight into the MDS characterization for \( Z_t (u, \theta_0) \), we take a Taylor series expansion of \( m(u, X_{t-1}, \theta) \) with respect to \( u \) around zero. This yields

\[
m (u, x, \theta) = \sum_{|\nu| = 0}^{\infty} \frac{m^{(\nu)} (0, x, \theta)}{\prod_{c=1}^{d} \nu_c !} u_1^{\nu_1} \cdots u_d^{\nu_d},
\]

where \( u = (u_1, \ldots, u_d)^T \),

\[
m^{(\nu)} (0, x, \theta) = \left. \frac{\partial^{\nu_1}}{\partial u_1^{\nu_1}} \cdots \frac{\partial^{\nu_d}}{\partial u_d^{\nu_d}} m(u, x, \theta) \right|_{u=0} = E \left[ \prod_{c=1}^{d} (iX_{t,c})^{\nu_c} \big| X_{t-1} \right] - E_{\theta} \left[ \prod_{c=1}^{d} (iX_{t,c})^{\nu_c} \big| X_{t-1} \right].
\]

Here, as before, \( E_{\theta} (\cdot|X_{t-1}) \) is the expectation under the model-implied transition density \( p (x, t|X_{t-1}, \theta) \), \( \nu = (\nu_1, \nu_2, \ldots, \nu_d)^T \), and \( |\nu| = \sum_{c=1}^{d} \nu_c \). Thus, checking the MDS condition for \( Z_t (u, \theta_0) \) is equivalent to checking whether the dynamics of various conditional moments and cross-moments of \( X_t \) has been adequately captured by the null continuous-time model. The MDS characterization thus provides a novel approach to constructing an omnibus test which does not have to use various conditional moments and cross-moments of \( X_t \).

Given a discretely observed sample \( \{X_t\}_{t=1}^{T} \), we can estimate the complex-valued regression function \( m(u, X_{t-1}, \theta_0) \) nonparametrically and check whether \( m(u, X_{t-1}, \theta_0) \) is identically zero for all \( u \in \mathbb{R}^d \).
and some \( \theta_0 \in \Theta \). Nonparametric estimation of \( m(u, X_{t-1}, \theta_0) \) is suitable in the present context because \( m(u, x, \theta_0) \) is potentially highly nonlinear under \( \mathbb{H}_A \). Various nonparametric regression methods could be used here. For concreteness, we use local linear regression. Local linear regression, and more generally local linear fitting, are introduced originally by Stone (1977) and studied by Cleveland (1979), Fan (1992, 1993), Masry (1996), and Ruppert and Wand (1994), among many others. Local linear fitting has significant advantages over the conventional Nadaraya–Watson estimator. It reduces the bias (Fan, 1992), and it adapts automatically to the boundary of design points (see Fan and Gijbels, 1996). Using a minimax argument, Fan (1993) shows that within the class of linear estimators that includes kernel and spline estimators, the local linear estimators achieve the best possible rates of convergence.

We consider the following local least squares problem:

\[
\min_{\beta \in \mathbb{R}^{d+1}} \sum_{t=2}^{T} \left| Z_t(u, \theta_0) - \beta_0 - \beta'_1 (X_t - x) \right|^2 K_h (x - X_t), \ x \in \mathbb{R}^d,
\]

where \( \beta = (\beta_0, \beta'_1)' \) is a \((d + 1) \times 1\) parameter vector, \( K_h (x) = h^{-1} K(x/h), \ K: \mathbb{R}^d \to \mathbb{R} \) is a kernel function, and \( h \) is a bandwidth. An example of \( K(\cdot) \) is a prespecified symmetric probability density function. We obtain the following solution:

\[
\hat{\beta} = \hat{\beta}(x, u) = \begin{bmatrix} \hat{\beta}_0 (x, u) \\ \hat{\beta}_1 (x, u) \end{bmatrix} = [X'WX]^{-1} X'WZ, \ x \in \mathbb{R}^d,
\]

where \( X \) is a \( dT \times 2 \) matrix with the \((t + 1)\) to \((t + d)\)th rows given by \([1, X_t - x], \ W = diag[K_h(X_1 - x), K_h(X_2 - x), ..., K_h(X_T - x)], \ Z = [Z_1 (u, \theta_0), Z_2 (u, \theta_0), ..., Z_T (u, \theta_0)]' \). Note that \( \hat{\beta} \) depends on the location \( x \) and parameter \( u \), but for notional simplicity, we have suppressed its dependence on \( x \) and \( u \).

Under suitable regularity conditions, \( m(u, x, \theta_0) \) can be consistently estimated by the local intercept estimator \( \hat{\beta}_0 (x, u) \). Specifically, we have

\[
\hat{m} (u, x, \hat{\theta}) = \sum_{t=2}^{T} \hat{W} \left( \frac{X_t - x}{h} \right) Z_t (u, \hat{\theta}),
\]

where \( \hat{W} (\cdot) \) is an effective kernel, defined as

\[
\hat{W} (t) \equiv e_1' S_T^{-1} [1, th, ..., th]' K(t)/h,
\]

\( e_1 = (1, 0, ..., 0)' \) is a \( d \times 1 \) unit vector, \( S_T = X'WX \) is a \((d + 1) \times (d + 1)\) matrix. As established by Hansen (2007) and Hjellvik, Yao and Tjøstheim (1998), for any compact set \( G \subset \mathbb{R}^d \), one has

\[
S_T^{-1} = g(x)^{-1} \begin{bmatrix} 1 & 0' \\ 0 & S_0 \end{bmatrix}^{-1} + o_P(1) \text{ uniformly for } x \in G,
\]

where \( g(x) \) is the true stationary density function of \( X_t \), \( 0 \) is a \( d \times 1 \) vector of zeros, and \( S_0 \) is the \( d \times d \) diagonal matrix whose diagonal element is \( \int_{\mathbb{R}^d} uu'K(u)du \). It follows that the effective kernel

\[
\hat{W} (t) = \frac{1}{Th^d g(x)} K(t) [1 + o_P(1)].
\]
Eq. (3.5) shows that the local linear estimator works like a kernel regression estimator based on the kernel $K(\cdot)$ with a known design density. Under certain conditions, $\hat{m}(u, x, \hat{\theta})$ is consistent for $m(u, x, \theta_0)$. It converges to a zero function under $H_0$ and a nonzero function under $H_A$. Thus, any significant difference of $\hat{m}(u, x, \hat{\theta})$ from zero is evidence of model misspecification.

We measure the distance between $\hat{m}(u, x, \hat{\theta})$ and a zero function by the quadratic form

$$L^2(\hat{m}) = \sum_{t=2}^{T} \int |\hat{m}(u, X_{t-1}, \hat{\theta})|^2 a(X_{t-1}) dW(u), \quad (3.6)$$

where $a : \mathbb{R}^d \to \mathbb{R}^+$ is a weighting function for the conditioning state vector $X_{t-1}$ and $W : \mathbb{R}^d \to \mathbb{R}^+$ is a nondecreasing weighting function for $u$ that weighs sets symmetric about the origin equally. The use of weighting function $a(X_{t-1})$ is not uncommon in the literature, see (e.g.) Ait-Sahalia, Bickel and Stoker (2001), Ait-Sahalia et al. (2006), Hjellvik et al. (1998), and Su and White (2007). This is often used to remove outliers or extreme observations. As noted by Ait-Sahalia et al. (2001), by choosing an appropriate $a(\cdot)$, one can focus on a particular empirical question of interest and reduce the influences of unreliable estimates. On the other hand, to ensure omnibus power, we have to consider many points for $u$. An example of $W(\cdot)$ is the $N(0, I_d)$ cumulative distribution function (CDF), where $I_d$ is a $d \times d$ identity matrix. Note that $W(\cdot)$ need not be continuous. They can be nondecreasing step functions such as discrete multivariate CDFs. This is equivalent to using finitely many or countable grid points for $u$. It will lead to a convenient implementation of our test, but possibly at a cost of power loss.

The omnibus test statistic for $H_0$ against $H_A$ is an appropriately standardized version of (3.6)

$$\hat{M} = \left[ h^2 \sum_{t=2}^{T} \int |\hat{m}(u, X_{t-1}, \hat{\theta})|^2 a(X_{t-1}) dW(u) - \hat{C} \right] / \sqrt{2\hat{D}}, \quad (3.7)$$

where the centering and scaling factors

$$\hat{C} = h^{-\frac{3}{2}} \iint \left[ 1 - \varphi(u, t|x, \hat{\theta}) \right]^2 a(x) dx dW(u) \int K^2(\tau) d\tau, \quad \varphi(u, t|x, \theta) \equiv \varphi(u, t|X_{t-1} = x, \theta),$$

$$\hat{D} = \iint \varphi(u + v, t|x, \hat{\theta}) - \varphi(u, t|x, \hat{\theta}) \varphi(v, t|x, \hat{\theta}) a^2(x) dx dW(u) dW(v)$$

$$\times \int \left[ \int K(\tau) K(\tau + \eta) d\tau \right]^2 d\eta.$$

The factors $\hat{C}$ and $\hat{D}$ are the approximate mean and variance of the quadratic form in (3.6).

In practice, $\hat{M}$ has to be calculated using numerical integration or approximated by simulation techniques. This may be computationally costly when the dimension $d$ of state vector $X_t$ is large. Alternatively, one can only use a finite number of grid points for $u$. For example, we can symmetrically generate finitely many numbers of $u$ from a $N(0, I_d)$ distribution. This will significantly reduce the computational cost but may lead to some power loss.

The centering and scaling factors $\hat{C}$ and $\hat{D}$ are derived under $H_0$ using an asymptotic argument. They may not approximate well the mean and variance of the quadratic form in (3.6). This may lead to

---

6We use the symmetry of $W(\cdot)$ to simplify the expression of the asymptotic variance of $L^2(\hat{m})$. 

---

11
poor size in finite samples, although not necessarily poor power. Motivated by obtaining a test statistic with better size in finite samples, we also consider the following modified test statistic

$$
\hat{M}_{FS} = \left[ h^2 \sum_{t=2}^{T} \int \left| \hat{m}(u, X_{t-1}, \hat{\theta}) \right|^2 \, a(X_{t-1}) \, dW(u) - \hat{C}_{FS} \right] / \sqrt{2\hat{D}}, \tag{3.8}
$$

where the centering factor

$$
\hat{C}_{FS} = h^2 \sum_{s=2}^{T} \int \left| Z_s(u, \hat{\theta}) \right|^2 \, dW(u) \sum_{t=2}^{T} \hat{W}^2 \left( \frac{X_{s-1} - X_{t-1}}{h} \right) a(X_{t-1})
$$

is a finite sample version of \( \hat{C} \). It is expected to give better approximation for the mean of \( L^2(\hat{m}) \) in finite samples. Similarly, we could also replace the scaling factor \( \hat{D} \) by its finite sample version

$$
h^2 \sum_{s=2}^{T} \sum_{s=2}^{T} \left\{ \sum_{t=2}^{T} \Re \left[ Z_s(u, \hat{\theta}) Z_s^{*}(u, \hat{\theta}) \right] dW(u) \hat{W}(\frac{X_{s-1} - X_{t-1}}{h}) \hat{W}(\frac{X_{s-1} - X_{t-1}}{h}) a(X_{t-1}) \right\}^2,
$$

but its computational cost is rather substantial when the sample size \( T \) is large. We are thus content with the test statistic \( \hat{M}_{FS} \) in (3.8).

We emphasize that although the CCF and the transition density are Fourier transforms of each other, our nonparametric regression-based CCF approach has an advantage over the nonparametric transition density-based approach that compares a nonparametric transition density with the model-implied transition density \( p(x, t|X_s, \hat{\theta}) \) via a quadratic form (e.g., Ait-Sahalia et al. 2006). This follows because our nonparametric regression estimator in (3.3) is only \( d \)-dimensional but the nonparametric transition density estimator is \( 2d \)-dimensional. We expect that such dimension reduction will give better size and power performance in finite samples.

4. ASYMPTOTIC THEORY

To derive the null asymptotic distribution of \( \hat{M} \), we impose the following regularity conditions.

**Assumption A.1:** Let \((\Omega, \mathcal{F}, P)\) be a complete probability space. (i) The stochastic time series vector process \( X_t \equiv X_t(\omega), \) where \( \omega \in \Omega \) and \( t \in [0, T] \subset \mathbb{R}^+ \), is a \( d \times 1 \) strictly stationary continuous-time Markov process with the marginal density \( g(x) \), which is positive and continuous for all \( x \in \mathcal{G} \), where \( \mathcal{G} \) is a compact set of \( \mathbb{R}^d \). Also, the joint density of \( (X_1, X_t) \) is continuous and bounded by some constant independent of \( l > 1 \). (ii) A discrete sample \( \{X_t\}_{t=1}^{T}, \) where \( \Delta \equiv 1 \) is the sampling interval, is observed at equally spaced discrete times and \( \{X_t\}_{t=1}^{T} \) is a \( \beta \)-mixing process with mixing coefficients satisfying \( \sum_{j=1}^{\infty} j^2 \beta(j) \frac{j}{1+\beta} < C \) for some \( 0 < \delta < 1 \).

**Assumption A.2:** Let \( \varphi(u, t|X_{t-1}, \theta) \) be the CCF of \( X_t \) given \( X_{t-1} \) of a continuous-time Markov model \( M = M(\theta) \) indexed by \( \theta \in \Theta \). (i) For each \( \theta \in \Theta, \) each \( u \in \mathbb{R}^d \) and each \( t, \varphi(u, t|X_{t-1}, \theta) \) is measurable with respect to \( X_{t-1} \); (ii) For each \( \theta \in \Theta, \) each \( u \in \mathbb{R}^d \), and each \( t, \varphi(u, t|X_{t-1}, \theta) \) is twice continuously differentiable with respect to \( \theta \) with probability one; and (iii) \( \sup_{u \in \mathbb{R}^d} E \sup_{\theta \in \Theta} || \frac{T}{\partial T} \varphi(u, t|X_{t-1}, \theta) ||^2 \leq C \) and \( \sup_{u \in \mathbb{R}^d} E \sup_{\theta \in \Theta} || \frac{T}{\partial T} \partial \varphi(u, t|X_{t-1}, \theta) || \leq C. \)
Assumption A.3: $\hat{\theta}$ is a parameter estimator such that $\sqrt{T}(\hat{\theta} - \theta^*) = O_P(1)$, where $\theta^* = p \lim_{T \to \infty} \hat{\theta}$ and $\theta^* = \theta_0$ under $\mathbb{H}_0$.

Assumption A.4: The function $K : \mathbb{R}^d \to \mathbb{R}^+$ is a product kernel of some univariate kernel $k$, i.e., $K(u) = \prod_{j=1}^d k(u_j)$, where $k : \mathbb{R} \to \mathbb{R}^+$ satisfies the Lipschitz condition and is a symmetric, bounded, and twice continuously differentiable function with $\int_{-\infty}^\infty k(u) du = 1$, $\int_{-\infty}^\infty uk(u) du = 0$, and $\int_{-\infty}^\infty u^2k(u) du < \infty$.

Assumption A.5: (i) $W : \mathbb{R}^d \to \mathbb{R}^+$ is a nondecreasing right-continuous weighting function that weighs sets symmetric about the origin equally, with $\int_{\mathbb{R}^d} \|u\|^4 dW(u) < \infty$; (ii) $a : \mathbb{G} \to \mathbb{R}^+$ is a bounded weighting function that is continuous over $\mathbb{G}$, where $\mathbb{G} \in \mathbb{R}^d$ is the compact support given in Assumption A.1.

Assumption A.1 imposes regularity conditions on the DGP. Both univariate and multivariate continuous-time processes are covered. Following Ait-Sahalia (1996a, 1996b), Gallant and Long (1997), Gallant and Tauchen (1996), we impose regularity conditions on a discretely observed random sample. There are two kinds of asymptotic results in the literature. 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. Moreover, even if such ultra high-frequency data are available, market microstructural problems are likely to complicate the analysis considerably. Hence, like Ait-Sahalia (1996a) and Singleton (2001), we fix the sampling interval $\Delta$ and derive the asymptotic properties of our test for an expanding sampling period. Unlike Ait-Sahalia (1996a, 1996b), however, we avoid imposing additional assumptions on the SDE, because we consider a more general framework. We allow but do not assume $X_t$ to be a diffusion process.

We assume that the DGP is Markov under both $\mathbb{H}_0$ and $\mathbb{H}_A$ and focus on testing functional form misspecification. Given the fact that the Markov is a maintained condition for almost all continuous-time models (e.g., diffusion, jump diffusion and levy processes), if these continuous-time models are correctly specified, the Markov assumption of the DGP is satisfied under $\mathbb{H}_0$. Hence our approach is applicable to these models.

The $\beta$-mixing condition in Assumption A.1 restricts the degree of temporal dependence in $\{X_t\}$. We say that $X_t$ is $\beta$-mixing (absolutely regular) if $\beta(j) = \sup_{s \geq 1} E \left[ \sup_{A \in \mathcal{F}_{s+j}} \left| P(A|\mathcal{F}_s) - P(A) \right| \right] \to 0$, as $j \to \infty$, where $\mathcal{F}_j^s$ is the $\sigma$-field generated by $\{X_{\tau} : \tau = j, ..., s\}$, $j \leq s$. Ait-Sahalia et al. (2006), Hjellvik et al. (1998) and Su and White (2007) also impose $\beta$-mixing conditions in related contexts. It is worth noting that our mixing condition is weaker than Ait-Sahalia et al. (2006), who assume a $\beta$-mixing with an exponential decay rate. Suggested by Hansen and Scheinkman (1995) and Ait-Sahalia (1996a), one set of sufficient conditions for the $\beta$-mixing when $d = 1$ is (i) $\lim_{x \to -l \text{ or } x \to u} \sigma(x, \theta) \pi(x, \theta) = 0$; and (ii) $\lim_{x \to -l \text{ or } x \to u} | \sigma(x, \theta) / \{2\mu(x, \theta) - \sigma(x, \theta) [\partial \sigma(x, \theta) / \partial x] \} | \to \infty$, where $l$ and $u$ are left and right boundaries of $X_t$ with possibly $l = -\infty$ and/or $u = +\infty$, and $\pi(x, \theta)$ is the model-implied marginal density.

Assumption A.2 provides regularity conditions on multifactor continuous-time Markov models. We
impose these conditions directly on the model-implied CCF, which cover other continuous-time processes not characterized by a SDE. As the CCF is the Fourier transform of the transition density, we can easily translate the conditions on the model-implied CCF into the conditions on the model-implied transition density \( p(x, t|X_{t-1}, \theta) \). In particular, Assumption A.2 holds if (i) for each \( t \), each \( x \in G \), and each \( \theta \in \Theta \), \( p(x, t|X_{t-1}, \theta) \) is measurable with respect to \( X_{t-1} \); (ii) \( p(x, t|X_{t-1}, \theta) \) is twice continuously differentiable with respect to \( \theta \in \Theta \) with probability one; and (iii) \( \sup_{x \in G} \frac{\partial^2}{\partial \theta^2} \ln p(x, t|X_{t-1}, \theta) \) \( \leq C \) and \( \sup_{x \in G} \frac{\partial^{2}}{\partial \theta^2} \ln p(x, t|X_{t-1}, \theta) \) \( \leq C \). An advantage of imposing regularity conditions on the model-implied CCF or transition density is that the asymptotic theory of our tests is readily applicable to test the validity of a discrete-time conditional distribution model.

Assumption A.3 requires a \( \sqrt{T} \)-consistent estimator \( \hat{\theta} \) under \( \mathbb{H}_0 \). We allow using both asymptotically optimal and suboptimal estimators, such as Ait-Sahalia’s (2002) approximated MLE, Chib et al.’s (2004) MCMC method, Gallant and Tauchen’s (1996) EMM, Singleton’s (2001) ML-CCF and GMM-CCF, and Quasi-MLE. We do not require any asymptotically most efficient estimator or a specified estimator. This is attractive for practitioners given the notorious difficulty of asymptotically efficient estimation of multifactor continuous-time models and may be viewed as an advantage over some existing tests which require a specific estimation method.

Assumption A.4 imposes regularity conditions on the kernel function used in local linear regression estimation. The condition on the boundedness of \( k(\cdot) \) is imposed for the brevity of proofs and could be removed at the cost of a more tedious proof.

Assumption A.5 imposes some mild conditions on the weighting functions \( W(u) \) and \( a(x) \) respectively. Any CDF with a finite fourth moment satisfies the condition for \( W(u) \). \( W(u) \) need not be continuous. This provides a convenient way to implement our tests, because we can avoid high dimensional numerical integrations by using finitely many or countable grid points for \( u \). For the simplicity of the proof, we assume that the weighting function \( a(x) \) has a compact support on \( \mathbb{R}^d \).

We now state the asymptotic distribution of \( \hat{M} \) under \( \mathbb{H}_0 \).

**Theorem 1:** Suppose Assumptions A.1–A.5 hold, and \( h = cT^{-\lambda} \) for \( 0 < \lambda < \frac{1}{2d} \) and \( 0 < c < \infty \). Then \( \hat{M} \overset{d}{\rightarrow} N(0, 1) \) under \( \mathbb{H}_0 \) as \( T \rightarrow \infty \).

The main idea in the proof of Theorem 1 is to perform the Hoeffding’s decomposition on the test statistic and apply asymptotic results for degenerate \( U \)-statistics (e.g., Hjellvik et al. 1998).

We require \( h \rightarrow 0 \) and \( Th^{2d} \rightarrow \infty \). This could be weakened by applying Collomb’s inequality, which is more involved. We do not do so because it still covers the optimal bandwidth \( h \propto T^{-\frac{1}{2d+3}} \) for \( d < 4 \). In practice, \( h \) is often chosen in an ad hoc manner. Alternatively, an automatic method such as cross-validation may be used. To describe this, define a “leave-one-out” estimator

\[
\hat{m}^-(u, X_t, \hat{\theta}) = \sum_{s=1, s \neq t}^{T} \hat{W} \left( \frac{X_s - X_t}{h} \right) Z_s(u, \hat{\theta}) \nonumber.
\]

\(^7\) The "optimality" is in the sense of minimizing the MSE of estimation, not necessarily maximizing the power of the proposed test.
Then a possible choice of $h$ is

$$
\hat{h}_{CV} = \arg \min_h CV(h) = \sum_{t=2}^T \left| Z_t \left( u, \hat{\theta} \right) - \hat{m} - \left( u, X_{t-1}, \hat{\theta} \right) \right|^2 a \left( X_{t-1} \right).
$$

As an important feature of $\hat{M}$, the use of the estimated generalized residuals $\{Z_t(u, \hat{\theta})\}$ in place of the true residuals $\{Z_t(u, \theta_0)\}$ has no impact on the limit distribution of $\hat{M}$. One can proceed as if the true parameter value $\theta_0$ were known and equal to $\hat{\theta}$. Intuitively, the parametric estimator $\hat{\theta}$ converges to $\theta_0$ faster than the nonparametric estimator $\hat{m}(u, x, \theta_0)$ to $m(u, x, \theta_0)$. Consequently, the limit distribution of $\hat{M}$ is solely determined by $\hat{m}(u, x, \theta_0)$, and replacing $\theta_0$ by $\hat{\theta}$ has no impact asymptotically.\(^8\) This delivers a convenient procedure, because any $\sqrt{T}$-consistent estimator can be used.

Next, we investigate the asymptotic power property of $\hat{M}$ under $\mathbb{H}_A$.

**Theorem 2:** Suppose Assumptions A.1–A.5 hold, and $h = cT^{-\lambda}$ for $0 < \lambda < \frac{2}{3d}$ and $0 < c < \infty$. Then as $T \to \infty$,

$$
T^{-1}h^{-\frac{d}{2}} \hat{M} \xrightarrow{p} (2D)^{-\frac{1}{2}} \iint |m(u, x, \theta^*)|^2 a(x) g(x) dxdW(u),
$$

where

$$
D = \iint \left| \varphi(u + v, t|x, \theta^*) - \varphi(u, t|x, \theta^*) \varphi(v, t|x, \theta^*) \right|^2 a^2(x) dxdW(u) dW(v)
\times \iint K(\tau) K(\tau + \eta) d\eta, \quad \varphi(u, t|x, \theta^*) \equiv \varphi(u, t|X_{t-1} = x, \theta).
$$

Under $\mathbb{H}_A$, we have $E[Z_t(u, \theta^*)|X_{t-1}] \neq 0$. Suppose $E[Z_t(u, \theta^*)|X_{t-1} = x] \neq 0$ with a positive Lebesgue measure on the support $\mathbf{G}$ of the weighting function $a(x)$. Then it follows that $\iint |m(u, x, \theta^*)|^2 a(x) g(x) dxdW(u) > 0$ for any weighting function $W(\cdot)$ that is positive, monotonically increasing and continuous, with unbounded support on $\mathbb{R}$, and for any continuous weighting function $a(\cdot)$. As a result, $P[\hat{M} > C(T)] \to 1$ for any sequence of constants $\{C(T) = o(Th^{d/2})\}$. Thus $\hat{M}$ has asymptotic unit power at any given significance level, whenever $E[Z_t(u, \theta^*)|X_{t-1}]$ is a nonzero function of $X_{t-1}$ on support $\mathbf{G}$. We note that under $\mathbb{H}_A$, $\hat{M}$ diverges to infinity at the rate of $Th^{d/2}$, which is faster than the rate $Th^d$ of a nonparametric transition density-based test (e.g., Ait-Sahalia et al. 2006, Hong and Li 2005, for $d = 1$). It could be shown that the $\hat{M}$ test is asymptotically more powerful than a nonparametric transition density-based test in terms of Bahadur’s (1960) asymptotic slope criterion, which is pertinent for power comparison under fixed alternatives.\(^9\) Similarly, although we do not examine the asymptotic local power, we expect that $\hat{M}$ can detect a class of local alternatives converging to $\mathbb{H}_0$ at the rate of $T^{-1/2}h^{-d/4}$, while the transition density-based test can only detect a class of local alternatives

---

\(^8\)Parameter estimation uncertainty may have impact on finite sample performance and a parametric bootstrap may be used to capture this impact. See simulation studies in Section 7.

\(^9\)The Bahadur relative efficiency is defined as the limiting ratio of the sample sizes required by the two tests under comparison to achieve the same asymptotic significance level ($p$-value) under the same fixed alternative.
with a slower rate of $T^{-1/2} h^{-d/2}$. This is an advantage of the nonparametric regression-based CCF
testing over the nonparametric transition density approach, due to the dimension reduction.

We note that unlike Chen and Hong (2005), we maintain the Markov property of $X_t$ under $\mathbb{H}_A$. In this case, the $\tilde{M}$ test is expected to have better power than Chen and Hong’s (2005) test in detecting functional misspecification of the drift, diffusion, jump and conditional correlation functions. In contrast, Chen and Hong’s (2005) test is expected to have better power when $X_t$ is not Markov under $\mathbb{H}_A$.

The finite sample omnibus test $M_{FS}$ in (3.8) has the same asymptotic $N(0,1)$ distribution under $\mathbb{H}_0$ and the same asymptotic power property under $\mathbb{H}_A$ as the $\tilde{M}$ test.

5. DIRECTIONAL DIAGNOSTIC PROCEDURES

When a multifactor continuous-time model $\mathcal{M}$ is rejected by the omnibus test, 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 dynamics, or conditional variance dynamics, or conditional correlations between state variables. Such information will be valuable in reconstructing the model.

The CCF is a convenient and useful tool to check 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 in a unified framework by differentiating $m(u,x,\theta)$ with respect to $u$ at the origin. This class of diagnostic tests can provide useful information about how well a continuous-time model captures the dynamics of various conditional moments and conditional cross-moments of state variables.

Recall the partial derivative of function $m(u,X_{t-1},\theta)$ at $u = 0$ :

$$m^{(\nu)}(0, X_{t-1}, \theta) = E \left[ \prod_{c=1}^{d} (iX_{c,t})^{\nu_c} | X_{t-1} \right] - E_{\theta} \left[ \prod_{c=1}^{d} (iX_{c,t})^{\nu_c} | X_{t-1} \right].$$

(5.1)

To get insight into $m^{(\nu)}(0, X_{t-1}, \theta)$, we consider a bivariate process $X_t = (X_{1,t}, X_{2,t})'$ for the cases of $|\nu| = 1$ and $|\nu| = 2$.

Case 1: $|\nu| = 1$. We have $\nu = (1,0)$ or $\nu = (0,1)$. If $\nu = (1,0)$,

$$m^{(\nu)}(0, X_{t-1}, \theta) = iE(X_{1,t} | X_{t-1}) - iE_{\theta}(X_{1,t} | X_{t-1}).$$

If $\nu = (0,1)$, then

$$m^{(\nu)}(0, X_{t-1}, \theta) = iE(X_{2,t} | X_{t-1}) - iE_{\theta}(X_{2,t} | X_{t-1}).$$

Thus, the choice of $|\nu| = 1$ can be used to check misspecifications in the conditional means for $X_{1,t}$ and $X_{2,t}$ respectively.

Case 2: $|\nu| = 2$. We have $\nu = (2,0), (0,2)$ or $(1,1)$. If $\nu = (2,0)$,

$$m^{(\nu)}(0, X_{t-1}, \theta) = -E(X_{1,t}^2 | X_{t-1}) + E_{\theta}(X_{1,t}^2 | X_{t-1}).$$

If $\nu = (0,2)$,

$$m^{(\nu)}(0, X_{t-1}, \theta) = -E(X_{2,t}^2 | X_{t-1}) + E_{\theta}(X_{2,t}^2 | X_{t-1}).$$
Finally, if \( \nu = (1, 1) \),
\[
m^{(\nu)}(0, X_{t-1}, \theta) = -E(X_{1,t}X_{2,t}\mid X_{t-1}) + E_\theta(X_{1,t}X_{2,t}\mid X_{t-1}).
\]
Thus, the choice of \(|\nu| = 2\) can be used to check model misspecifications in the conditional volatility of state variables and their conditional correlation.

We now define the class of diagnostic test statistics as follows:
\[
\hat{M}^{(\nu)} = \left[ h^{-2} \sum_{t=2}^{T} \hat{m}^{(\nu)}(0, X_{t-1}, \hat{\theta})^2 a(X_{t-1}) - \hat{C}^{(\nu)} \right] / \sqrt{2\hat{D}^{(\nu)}},
\]
where the centering and scaling factors
\[
\hat{C}^{(\nu)} = \int \left\{ E_\hat{\theta} \left[ \prod_{c=1}^{d} X_{c,t}^{\nu_c} \mid X_{t-1} = x \right] \right\} a(x) dx \int K^2(\tau) d\tau,
\]
\[
\hat{D}^{(\nu)} = \int \left\{ E_\hat{\theta} \left[ \prod_{c=1}^{d} X_{c,t}^{\nu_c} \mid X_{t-1} = x \right] \right\} a^2(x) dx \times \int \left[ \int K(\tau) K(\tau + \eta) d\tau \right]^2 d\eta.
\]
Here, \( E_\hat{\theta}(\cdot \mid X_{t-1}) \) is the expectation under the estimated model-implied transition density \( p(y, t \mid X_{t-1}, \hat{\theta}) \).

In general, we can differentiate the estimated CCF \( \varphi \left( u, t \mid X_{t-1}, \hat{\theta} \right) \) to compute \( E_\hat{\theta}(\cdot \mid X_{t-1}) \), namely,
\[
E_\hat{\theta} \left( \prod_{c=1}^{d} X_{c,t}^{\nu_c} \mid X_{t-1} \right) = (-1)^{\sum_{c=1}^{d} \nu_c} \frac{\partial^{2\nu_1}}{\partial u_1^{2\nu_1}} \cdots \frac{\partial^{2\nu_d}}{\partial u_d^{2\nu_d}} \varphi \left( u, t \mid X_{t-1}, \hat{\theta} \right) \bigg|_{u=0},
\]
\[
E_\hat{\theta} \left( \prod_{c=1}^{d} X_{c,t}^{\nu_c} \mid X_{t-1} \right) = \prod_{c=1}^{d} \frac{\partial^{\nu_c}}{\partial u_1^{\nu_c}} \cdots \frac{\partial^{\nu_c}}{\partial u_d^{\nu_c}} \varphi \left( u, t \mid X_{t-1}, \hat{\theta} \right) \bigg|_{u=0}.
\]
For the previous bivariate example, if we further assume that the DGP is the bivariate uncorrelated Gaussian model in (7.9), then for \( \nu = (1, 0) \)
\[
E_\hat{\theta} \left( X_{1,t}^{2\nu_1}X_{2,t}^{2\nu_2} \mid X_{t-1} \right) = \left\{ \left[ 1 - \exp(-\hat{\kappa}_{11}) \right] \hat{\theta}_1 + \exp(-\hat{\kappa}_{11}) X_{1,t-1} \right\}^2 + \frac{\hat{\sigma}_{11}^2}{2\hat{\kappa}_{11}} \left[ 1 - \exp(-2\hat{\kappa}_{22}) \right],
\]
\[
E_\hat{\theta} \left( X_{1,t}^{\nu_1}X_{2,t}^{\nu_2} \mid X_{t-1} \right) = \left\{ \left[ 1 - \exp(-\hat{\kappa}_{11}) \right] \hat{\theta}_1 + \exp(-\hat{\kappa}_{11}) X_{1,t-1} \right\}.
\]
To derive the limit distribution of \( \hat{M}^{(\nu)} \) under \( \mathbb{H}_0 \), we impose some moment conditions.

**Assumption A.6:** (i) \( E \sup_{\theta \in \Theta} \left\| \frac{\partial^{\nu_1}}{\partial u_1^{\nu_1}} \cdots \frac{\partial^{\nu_d}}{\partial u_d^{\nu_d}} \varphi \left( u, t \mid X_{t-1}, \theta \right) \bigg|_{u=0} \right\|^2 \leq C; \)
(ii) \( E \sup_{\theta \in \Theta} \left\| \frac{\partial^{\nu_1}}{\partial u_1^{\nu_1}} \cdots \frac{\partial^{\nu_d}}{\partial u_d^{\nu_d}} \varphi \left( u, t \mid X_{t-1}, \theta \right) \bigg|_{u=0} \right\|^2 \leq C; \)
(iii) \( E \sup_{\theta \in \Theta} \left\| \frac{\partial^{\nu_1}}{\partial u_1^{\nu_1}} \cdots \frac{\partial^{\nu_d}}{\partial u_d^{\nu_d}} \varphi \left( u, t \mid X_{t-1}, \theta \right) \bigg|_{u=0} \right\|^{4(1+\delta)} \leq C; \) and (iv) \( E \left\| \prod_{c=1}^{d} X_{c,t}^{\nu_c} \right\|^{4(1+\delta)} \leq C. \)

**Theorem 3:** Suppose Assumption A.1-A.6 hold for some prespecified derivative order vector \( \nu \), \( h = cT^{-\lambda} \) for \( 0 < \lambda < \frac{1}{2d} \) and \( 0 < c < \infty \). Then \( \hat{M}^{(\nu)} \overset{d}{\rightarrow} N(0, 1) \) under \( \mathbb{H}_0 \) as \( T \to \infty. \)
Like \( \hat{M} \), parameter estimation uncertainty in \( \hat{\theta} \) has no impact on the asymptotic distribution of \( \hat{M}(\nu) \). Any \( \sqrt{T} \)-consistent estimator can be used. Moreover, different choices of \( \nu \) allow one to examine various specific dynamic aspects of the underlying process and thus provide information on how well a multivariate continuous-time Markov model fits various aspects of the conditional distribution of \( X_t \).

These diagnostic tests are designed to test specification of various conditional moments, i.e., whether the conditional moments of \( X_t \) are correctly specified given the discrete sample information \( X_{t-1} \). The first two conditional moments differ from the instantaneous conditional mean (drift) and instantaneous conditional variance (squared diffusion). In general, the conditional moments tested here are functions of drift, diffusion and jump (see Section 7.1.2 for an example). Only when the sampling interval \( \Delta \to 0 \), the conditional mean and variance will coincide with drift and squared diffusion.\(^\text{10}\)

6. TESTS FOR MODELS WITH UNOBSERVABLE VARIABLES

So far we have assumed that all state variables are observable. However, there are continuous-time models with unobservable components in the literature. For example, within the family of asset pricing models, the problem of unobserved state variables typically arises when the dimension \( d \) of the state vector \( X_t \) exceeds the dimension \( p \) of the vector of observed prices or yields. In the context of ATSMs, if \( r_t \) is an affine function of \( d \) state variables and one estimates the model with only \( p \) \((<d)\) bond yields, then \( d-p \) remaining state variables are unobservable. Andersen and Lund (1996) estimate a three-factor model \((d = 3)\) of a single short-term interest rate \((p = 1)\) using Gallant and Tauchen’s (1996) EMM method. Singleton (2001) also proposes CCF-based simulated method of moments estimators as alternatives to exploit the special structure of ATSMs.

Another example is the class of SV models for equity returns and interest rates, see (e.g.) Bakshi, Cao and Chen (1997), Bates (1996, 2000), Das and Sundaram (1999) and Heston (1993). SV models can capture salient properties of volatility such as randomness and persistence. Affine SV models have been widely used in modelling asset return dynamics as they allow for closed-form solutions for European option prices. A basic version of SV models assumes:

\[
\begin{align*}
    dr_t &= \kappa_r (\bar{r} - r_t) \, dt + \sqrt{V_t} \, dW_{r,t}, \\
    dV_t &= \kappa_v (\bar{v} - V_t) \, dt + \sigma_v \sqrt{V_t} \, dW_{v,t},
\end{align*}
\]

where \( V_t \) is the latent volatility process, and \( \kappa_r, \kappa_v, \sigma_v, \bar{r}, \) and \( \bar{v} \) are all scalar parameters. It can be shown that the CCF of \( r_t \) is:

\[
    \varphi_r(u, t|\tau_{t-1}, V_{t-1}, \theta) = \exp \left[ A_{t-1}(u, 0) + B_{t-1}(u, 0) \tau_{t-1} + C_{t-1}(u, 0) V_{t-1} \right], \quad u \in \mathbb{R},
\]

\(^\text{10}\)Assuming that \( J_t \) in (2.1) is a Poisson process, Yu (2007) shows that \( \Pr(\{A_{t, \Delta}|X_{t-\Delta}, \theta\} = O(\Delta) \) and \( \Pr(\{A_{t, \Delta}^c|X_{t-\Delta}, \theta\} = O(1) \), where \( A_{t, \Delta} \) denotes the event of jump that occurs between time \( t \) and \( t - \Delta \) and \( A_{t, \Delta}^c \) denotes its compliment.
where $A_{t-1}, B_{t-1}, C_{t-1} : \mathbb{R}^2 \to \mathbb{R}$ satisfy the complex-valued Riccati equations:

$$\begin{cases} 
A_t = \kappa_r \bar{B}_t + \kappa_v \bar{C}_t, \\
B_t = -\kappa_r B_t, \\
C_t = -\kappa_v C_t + \frac{1}{2} (B_t^2 + 2B_tC_t + C_t^2). 
\end{cases}$$

To test SV models, where $V_t$ is a latent process, we need to modify the MDS characterization (2.11) to make it operational. Generally, we partition $X_t = (X_{1,t}', X_{2,t}')'$, where $X_{1,t} \subset \mathbb{R}^{d_1}$ denotes the observable state variables, $X_{2,t} \subset \mathbb{R}^{d_2}$ denotes the unobservable state variables, and $d_1 + d_2 = d$. Also, partition $u$ conformably as $u = (u_1', u_2')'$. Let $\phi(u_1, t|I_{1,t-1}, \theta) = E_{\theta} [\exp(iu_1'X_{1,t})|I_{1,t-1}]$, where $I_{1,t-1} = \{X_{1,t-1}, X_{1,t-2}, ..., X_{1,1}\}$ is the information set on the observables that is available at time $t-1$. Then we define

$$Z_{1,t} (u_1, \theta) \equiv \exp (iu_1'X_{1,t}) - \phi(u_1, t|I_{1,t-1}, \theta) = \exp(iu_1'X_{1,t}) - E_{\theta} \{ \varphi [(u_1', \theta)', t|X_{t-1}, \theta] |I_{1,t-1} \},$$

where the second equality follows from the law of iterated expectations and the Markov property of $X_t$. Then under $\mathbb{H}_0$, we have

$$E [Z_{1,t}(u_1, \theta_0)|I_{1,t-1}] = 0 \text{ a.s. for all } u_1 \in \mathbb{R}^{d_1} \text{ and some } \theta_0 \in \Theta. \tag{6.3}$$

This provides a basis for constructing operational tests for continuous-time Markov models with partially observable state variables. It has been used in Singleton (2001) to estimate continuous-time models with unobservable components. Note that although $Z_t(u, \theta_0)$ is a Markov process, the process $Z_{1,t}(u_1, \theta_0)$ is generally not Markov.\(^{11}\)

Although the model-implied CCF $\varphi[(u_1', \theta)', t|X_{t-1}, \theta]$ may have a closed-form, its conditional expectation $\phi(u_1, t|I_{1,t-1}, \theta) = E_{\theta} [\varphi [(u_1', \theta)', t|X_{t-1}, \theta] |I_{1,t-1}]$ generally has no closed-form. However, one can approximate it accurately by using simulation techniques.\(^{12}\) For continuous-time Markov models, the CCF $\varphi(u, t|X_{t-1}, \theta)$ is a function of $X_{t-1}$. It follows that

$$\phi(u_1, t|I_{1,t-1}, \theta) = \int \varphi[(u_1', \theta)', t|X_{t-1}, \theta] p[x_{2,t-1}, t - 1|I_{1,t-1}, \theta] dx_{2,t-1}, \tag{6.4}$$

where $p[x_{2,t-1}, t - 1|I_{1,t-1}, \theta]$ is the model-implied transition density of the unobservable $X_{2,t-1}$ given the past observable information $I_{1,t-1}$. Gordon, Salmond and Smith (1993), Kim, Shephard and Chib (1998), and Pitt and Shephard (1999) have developed a general method called particle filters that can sequentially approximate the conditional density $p(x_{2,t-1}, t - 1|I_{1,t-1}, \theta)$ by a set of particles $\{\tilde{X}_{2,t-1}^j\}^J_{j=1}$ with discrete probability masses $\{\pi_{t-1}^j\}^J_{j=1}$ for a large integer $J$. The key is to propagate particles $\{\tilde{X}_{2,t-2}^j\}^J_{j=1}$ one step forward to get the new particles $\{\tilde{X}_{2,t-1}^j\}^J_{j=1}$. By the Bayes rule, we have

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

\(^{11}\)Chen and Hong (2005) can be applied here since $Z_{1,t}(u_1, \theta_0)$ is generally not Markov. But we shall propose an alternative test, using a nonparametric regression approach.

\(^{12}\)As long as the sample size of simulated data is much bigger than that of the real data, the sampling variation of the simulated data is asymptotically negligible.
where \( p(x_{2,t-1}, t-1|\mathcal{I}_{1,t-2}, \theta) = \int p(x_{2,t-1}, t-1|x_{2,t-2}, \mathcal{I}_{1,t-2}, \theta) p(x_{2,t-2}, t-2|\mathcal{I}_{1,t-2}, \theta) \, dx_{2,t-2} \). We can then approximate \( p(x_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \theta) \) up to some proportionality; namely,

\[
\hat{p}(x_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \theta) \propto \hat{p}(x_{1,t-1}, t-1|\mathcal{X}_{2,t-1}, \mathcal{I}_{1,t-2}, \theta) \sum_{j=1}^{J} \pi_{t-1}^j \hat{p}(x_{2,t-1}, t-1|\mathcal{X}_{2,t-2}, \mathcal{I}_{1,t-2}, \theta),
\]

where \( \mathcal{I}_{1,t-1} = \{ \mathcal{X}_{1,t-1}, \mathcal{X}_{1,t-2}, ..., \mathcal{X}_{1}, \} \), and \( \hat{p}(x_{1,t-1}, t-1|\mathcal{X}_{2,t-1}, \mathcal{I}_{1,t-2}, \theta) \) and \( \sum_{j=1}^{J} \pi_{t-1}^j \hat{p}(x_{2,t-1}, t-1|\mathcal{X}_{2,t-2}, \mathcal{I}_{1,t-2}, \theta) \) 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 \( \mathcal{X}_{2,t-1} \) can be sampled from \( \hat{p}(x_{2,t-1}, t-1|\mathcal{X}_{2,t-2}, \mathcal{I}_{1,t-2}, \theta) \). These can be achieved by using time-discretized solutions to the SDEs.\(^{13}\)

To implement particle filters, we can use the algorithm developed by Johannes, Polson and Stroud (2006) and Pitt and Shephard (1999). First we generate a simulated sample \( \{ \mathcal{X}_{2,t-2}^M \}^J_{j=1} \), where \( \mathcal{X}_{2,t-2}^M = \{ \mathcal{X}_{2,t-2}^1, \mathcal{X}_{2,t-2}^2, ..., \mathcal{X}_{2,t-2}^M \} \) and \( M \) is an integer. Then we simulate them one step forward, evaluate the likelihood function, and set

\[
\pi_{t-1}^j = \frac{\hat{p}(x_{1,t-1}, t-1|\mathcal{X}_{2,t-1}^M, \mathcal{I}_{1,t-2}, \theta)}{\sum_{j=1}^{J} \hat{p}(x_{1,t-1}, t-1|\mathcal{X}_{2,t-1}^M, \mathcal{I}_{1,t-2}, \theta)}, \quad j = 1, ..., J.
\]

Finally, we resample \( J \) particles with weights \( \{ \pi_{t-1}^j \}^J_{j=1} \) to obtain a new random sample of size \( J \).\(^{14}\) It has been shown (e.g., Bally and Talay 1996 and Del Moral, Jacod and Protter 2001) that with both large \( J \) and \( M \), this algorithm sequentially generates valid simulated samples from \( p(x_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \theta) \).\(^{15}\)

Hence \( \phi(u_1, t|\mathcal{I}_{1,t-1}, \theta) \) can be calculated by Monte Carlo averages.\(^{16}\)

For models whose CCF is exponentially affine in \( \mathcal{X}_{t-1} \),\(^{17}\) we can also adopt Bates’ (2007) approach to compute \( \phi(u_1, t|\mathcal{I}_{1,t-1}, \theta) \). First, at time \( t = 1 \), initialize the CCF of the latent vector \( \mathcal{X}_{2,t-1} \) conditional on \( \mathcal{I}_{1,t-1} \) at its unconditional characteristic function. Then, by exploiting the Markov property and the affine structure of the CCF, we can evaluate the model-implied CCF conditional on data observed

\(^{13}\)There are many discretization methods in practice, such as the Euler scheme, the Milstein scheme, and the explicit strong scheme. See (e.g.) Kloeden et al. (1994) for more discussion.

\(^{14}\)This is called sampling/importance resampling (SIR) in the literature. Alternative methods include rejection sampling and the MCMC algorithm. See Doucet, Freitas and Gordon (2001) and Pitt and Shephard (1999) for more discussion.

\(^{15}\)An alternative method to obtain \( p(x_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \theta) \) is Gallant and Tauchen’s (1998) SNP-based reprojection technique, which is a general purpose technique for characterizing the dynamic response of a partially observed nonlinear system to its past observable history. First, we can generate simulated samples \( \{ \mathcal{X}_{1,t-1} \}^J_{t=2} \) and \( \{ \mathcal{X}_{2,t-1} \}^J_{t=2} \) from the continuous-time model, where \( J \) is a large integer. Then, we project the simulated data \( \{ \mathcal{X}_{2,t-1} \}^J_{t=2} \) onto a Hermite series representation of the transition density \( p(x_{2,t-1}, t-1|\mathcal{X}_{1,t-1}, \mathcal{X}_{1,t-2}, ..., \mathcal{X}_{1,L}) \), where \( L \) denotes a lag order. With a suitable choice of \( L \) via some information criteria such as AIC or BIC, we can approximate \( p(x_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \theta) \) arbitrarily well. The final step is to evaluate the estimated density function at the observed data. See Gallant and Tauchen (1998) for more discussion.

\(^{16}\)In a related estimation context, Chacko and Viceira (2003), Jiang and Knight (2002) and Singleton (2001) derive analytical expressions for \( \phi(u_1, t|\mathcal{I}_{1,t-1}, \theta) \) for some suitable subset \( \mathcal{I}_{1,t-1} \) of \( \mathcal{I}_{1,t-1} \). For example, Chacko and Viceira (2003) obtain a closed form expression for \( E \left[ \varphi_{\log \delta}(u, t|\mathcal{I}_{1,t-1}, \theta) \log S_t \right] \), by integrating out \( V_{t-1} \). This is computationally more convenient, but it will deliver a less powerful test for our purpose.

\(^{17}\)Examples include AJD models (Duffie et al. 2000) and time-changed Lévy processes (Carr and Wu 2003, 2004).
through period \( t \), namely, \( E_\theta[\varphi(u, t | X_{t-1}, \theta) | I_{1,t-1}] \) and back out the density function \( p(x_{1,t}, t | I_{1,t-1}, \theta) \) by Fourier inversion. Last, using Bayes’ rule, the CCF of the latent vector \( X_{2,t} \) conditional on \( I_{1,t} \) can be obtained. Repeating the previous two steps, we can estimate \( \phi(u_1, t | I_{1,t-1}, \theta) \) for all \( t \).

For notational simplicity, we define a new vector \( Y_t = (X_{1,t}, X_{1,t-1}, ..., X_{1,t-l+1})' \in \mathbb{R}^{ld_1} \), where \( l \) is a lag truncation order. Based on the MDS characterization in (6.3), we can use a nonparametric estimator for \( m(u_1, Y_{t-1}, \theta_0) \equiv E[Z_{1,t}(u_1, \theta_0) | Y_{t-1}] \). Similar to (3.1), we consider the following local least squares problem:

\[
\min_\beta \sum_{t=l+1}^{T} |Z_{1,t}(u_1, \theta_0) - \beta_0 - \beta_1(Y_t - y)|^2 K_h(y - Y_t), \quad y \in \mathbb{R}^{ld_1},
\]

where \( \beta = (\beta_0, \beta_1)' \). We obtain the following solution:

\[
\hat{\beta} \equiv \hat{\beta}(y, u_1) = \left[ \begin{array}{c} \hat{\beta}_0(y, u_1) \\ \hat{\beta}_1(y, u_1) \end{array} \right] = [Y' W Y]^{-1} Y' W Z_1,
\]

where \( Y \) is a \( Tld_1 \times 2 \) matrix with the \((t+1)\) to \((t+d)\)th row \([1, Y_t - y], W_y = \text{diag}[K_h(Y_1 - y), K_h(Y_2 - y), ..., K_h(Y_T - y)] \), and \( Z_1 = [Z_{1,1}(u_1, \theta_0), ..., Z_{1,T}(u_1, \theta_0)]' \).

Under regularity conditions, \( m(u, y, \theta_0) \) can be consistently estimated by \( \hat{\beta}_0(y, u_1) \); namely,

\[
\hat{m}(u_1, y, \hat{\theta}) = \sum_{t=l+1}^{T} \hat{W} \left( \frac{Y_t - y}{h} \right) Z_{1,t}(u_1, \hat{\theta}), \quad y \in \mathbb{R}^{ld_1},
\]

where \( \hat{W} \) is defined in the same way as in (3.4).

The omnibus test statistic for \( H_0 \) against \( H_A \) is a modified version of (3.7), namely,

\[
\hat{M}_u = \left[ h^{ld_1} \sum_{t=l+1}^{T} \int \left| \hat{m}(u_1, Y_{t-1}, \hat{\theta}) \right|^2 a(Y_{t-1}) dW(u_1) - \hat{C}_u \right] / \sqrt{2\hat{D}_u},
\]

where the centering and scaling factors

\[
\begin{align*}
\hat{C}_u &= h^{ld_1} \int \int \left\{ 1 - E_\theta \left[ \left| \phi(u_1, t | I_{1,t-1}, \hat{\theta}) \right|^2 \right| Y_{t-1} = y \right\} a(y) dy dW(u_1) \int K^2(\tau) d\tau, \\
\hat{D}_u &= \int \int \left[ E_\theta \left[ \phi(u_1 + v_1, t | I_{1,t-1}, \hat{\theta}) - \phi(u_1, t | I_{1,t-1}, \hat{\theta}) \phi(v_1, t | I_{1,t-1}, \hat{\theta}) \right| Y_{t-1} = y \right]^2 a^2(y) dy dW(u_1) dW(v_1) \int \left[ \int K(\tau) K(\tau + \eta) d\tau \right]^2 d\eta.
\end{align*}
\]

where \( \phi(u_1, t | I_{1,t-1}, \hat{\theta}) \equiv E_\theta \left\{ \varphi \left( u_1, 0' \right)' | X_{t-1}, \hat{\theta} | I_{1,t-1} \right\} \), \( W : \mathbb{R}^{ld_1} \to \mathbb{R}^+ \) is a nondecreasing weighting function that weighs sets symmetric about the origin equally, \( a : F \to \mathbb{R}^+ \) is a bounded weighting function, and \( F \in \mathbb{R}^{ld_1} \) is a compact support. The conditional expectations \( E_\theta(\cdot | Y_{t-1}) \) in \( \hat{C}_u \) and \( \hat{D}_u \) can be estimated via a nonparametric regression, but its implementation may be tedious.

Alternatively, we also consider the following finite sample version of the test statistic

\[
\hat{M}^{FS}_u = \left[ h^{ld_1} \sum_{t=l+1}^{T} \int \left| \hat{m}(u_1, Y_{t-1}, \hat{\theta}) \right|^2 a(Y_{t-1}) dW(u_1) - \hat{C}^{FS}_u \right] / \sqrt{2\hat{D}^{FS}_u},
\]

21
the use of the estimated generalized residuals property under $\mathbb{H}_0$ and $\mathbb{H}_1$. Similar to (3.8), $\hat{C}_u^{FS}$ and $\hat{D}_u^{FS}$ are expected to give better approximation for the mean and variance of $h^{\frac{1}{2}} \sum_{t=1}^{T} \frac{1}{T} \sum_{s=t+1}^{T} |Z_{1,s}(u_1, \hat{\theta})|^2 dW(u_1) \sum_{t=1+1}^{T} \hat{W}^2 \left( \frac{Y_{s-1} - Y_{t-1}}{h} \right) a(Y_{t-1})$, respectively.

Similar to (3.8), $\hat{C}_u^{FS}$ and $\hat{D}_u^{FS}$ are expected to give better approximation for the mean and variance of $h^{\frac{1}{2}} \sum_{t=1}^{T} \frac{1}{T} \sum_{s=t+1}^{T} \Re \left[ Z_{1,s}(u_1, \hat{\theta}) Z_{1,r}(u_1, \hat{\theta}) \right] dW(u_1) \hat{W} \left( \frac{Y_{s-1} - Y_{t-1}}{h} \right)^2 a(Y_{t-1})$ in finite samples. Consequently, $\hat{M}_u^{FS}$ is expected to deliver better sizes in finite samples. We now derive the asymptotic distribution of $\hat{M}_u$ under $\mathbb{H}_0$ and examine its asymptotic power property under $\mathbb{H}_A$ respectively.

**Theorem 4:** Suppose Assumptions B.1–B.5 given in the appendix hold, and $h = cT^{-\lambda}$ for $0 < \lambda < \frac{1}{2d_1}$ and $0 < c < \infty$. Then $\hat{M}_u \xrightarrow{d} N(0,1)$ under $\mathbb{H}_0$ as $T \to \infty$.

Assumptions B.1–B.5, given in the appendix, are simple modifications of Assumptions A.1–A.5. Like $\hat{M}$, the use of the estimated generalized residuals $\{Z_{1,t}(u_1, \hat{\theta})\}$ in place of the true unobservable residuals $\{Z_{1,t}(u_1, \theta_0)\}$ has no impact on the limit distribution of $\hat{M}_u$. One can proceed as if the true parameter value $\theta_0$ were known and equal to $\hat{\theta}$.

**Theorem 5:** Suppose Assumptions B.1–B.5 given in the appendix hold, and $h = cT^{-\lambda}$ for $0 < \lambda < \frac{2}{3d_1}$ and $0 < c < \infty$. Then as $T \to \infty$,

$$T^{-1} h^{\frac{1}{2}} \to \hat{M}_u \xrightarrow{d} \int \int |m(u_1, y, \theta^*)|^2 a(y) f(y) dy dW(u_1),$$

where $f(\cdot)$ is the stationary probability density function of $Y_t$, and the scaling factor

$$D_u = \int \int \int [E_{\theta^*} [\varphi(u_1 + v_1, t|I_{1,t-1}, \theta^*) - \varphi(u_1, t|I_{1,t-1}, \theta^*) \varphi(v_1, t|I_{1,t-1}, \theta^*) |y]|^2 a^2(y) dy dW(u_1) dW(v_1)) \times \int [\int K(\tau) K(\tau + \eta) d\eta]^2 d\eta.$$
and Mele 2006, Skaug and Tjøstheim 1996), we can use a pairwise testing approach and consider the following alternative test statistic:

\[ \hat{M}_u = \sum_{j=1}^{l} \hat{M}_u (j), \]

where

\[ \hat{M}_u (j) = h \frac{d}{2} \sum_{t=l+1}^{T} \left| \tilde{m}(u_1, X_{1,t-j}, \hat{\theta}) \right|^2 a(X_{1,t-j}) dW(u_1). \]

This avoids the "curse of dimensionality" with nonparametric estimation. By a similar but more tedious proof, we could derive the asymptotic normality of \( \hat{M}_u \) under \( H_0 \). Alternatively, a parametric bootstrap can be used, which in fact gives better size in finite samples.

7. MONTE CARLO EVIDENCE

We now study the finite sample performance of the proposed tests, in comparison with Hong and Li’s (2005, HL) test. We consider both univariate and bivariate continuous-time models.

7.1 Univariate models

7.1.1 Size of the \( \hat{M} \) tests

To examine the size of \( \hat{M} \) for univariate continuous-time Markov models, we simulate data from Vasicek’s (1977) model (DGP A0):

\[ dX_t = \kappa (\alpha - X_t) dt + \sigma dW_t, \quad (7.1) \]

where \( \alpha \) is the long run mean and \( \kappa \) is the speed of mean reversion. The smaller \( \kappa \) is, the stronger the serial dependence in \( \{X_t\} \), and consequently, the slower the convergence to the long run mean. We are particularly interested in the possible impact of dependent persistence in \( \{X_t\} \) on the size of \( \hat{M} \). Since the finite sample performance of \( \hat{M} \) may depend on both the marginal density and dependent persistence of \( \{X_t\} \), we follow HL and Pritsker (1998) to change \( \kappa \) and \( \sigma^2 \) in the same proportion so that the marginal density of \( X_t \) is unchanged; namely,

\[ p(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma_s^2} \exp \left[ -\frac{(x - \alpha)^2}{2\sigma_s^2} \right], \]

where \( \theta = (\kappa, \alpha, \sigma^2)' \) and \( \sigma_s^2 = \sigma^2/(2\kappa) = 0.01226 \). In this way, we can focus on the impact of dependent persistence. We consider both low and high levels of dependent persistence and use the same parameter values as HL and Pritsker (1998): \((\kappa, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185) \) and \((0.214592, 0.089102, 0.000546) \) for the low and high persistent dependence cases respectively.

For each parameterization, we simulate 1,000 data sets of a random sample \( \{X_t\}_{t=\Delta}^{T} \) at the monthly frequency for \( T = 250, 500, 1,000 \) respectively.\(^{19}\) The simulation is carried out by first generating an

---

\(^{18}\)The choice of the lag truncation order \( l \) has been a long-standing problem in the time series literature. It depends on the interest of practitioners and data availability. For example, 12 and 4 lags can be included if using monthly and quarterly data respectively.

\(^{19}\)We simulate data sets at the weekly and daily frequencies as well and simulation patterns are similar. These results are available from the authors upon request.
initial value $X_0$ from the marginal density $p(x, \theta)$, and given a value $X_t$, generating $X_{t+1}$ from the transitional normal density with the mean

$$\mu_t = X_t \exp(-\kappa \Delta) + \alpha [1 - \exp(-\kappa \Delta)] \quad (7.2)$$

and the variance

$$\sigma_t^2 = \sigma^2 [1 - \exp(-2\kappa \Delta)]/(2\kappa). \quad (7.3)$$

The sample sizes of $T = 250, 500, 1, 000$ correspond to about 20 to 100 years of monthly data. For each data set, we estimate a Vasicek model via MLE and compute the $\hat{M}$ statistic. We consider the empirical rejection rates using the asymptotic critical values at the 10% and 5% significance levels respectively. For $T = 250$, we also consider a parametric bootstrap method.

Following Ait-Sahalia et al. (2001), we use the Gaussian kernel $k(\cdot)$ and the truncated weighting $a(x) = 1(|x| \leq 1.5)$, where $1(\cdot)$ is the indicator function and $X_t$ has been standardized. We choose the $N(0,1)$ CDF for $W(\cdot)$. Our simulation experience suggests that the choices of $k(\cdot)$, $W(\cdot)$ and $a(\cdot)$ have little impact on the size performance of the tests. For simplicity, we choose $h = T^{-\frac{1}{5}}$. This simple bandwidth rule attains the optimal rate for the local linear fitting.

Table 1 reports the empirical sizes of $\hat{M}$ and $\hat{M}^{(v)}$ at the 10% and 5% levels under a correct Vasicek model with low and high persistence of dependence respectively. Both the asymptotic version in (3.7) and the finite sample version in (3.8) of our omnibus test tend to overreject when $T = 250$, but they improve as $T$ increases. As expected, the finite sample version $\hat{M}_{FS}$ has better sizes. The tests display a bit more overrejections under high persistence than under low persistence, but the difference becomes smaller as $T$ increases. For comparison, Table 2 reports the empirical sizes of the HL test under the same DGPs. Similarly, HL has some overrejection which is more severe than that of the $\hat{M}$ tests. We also consider the diagnostic tests $\hat{M}^{(v)}$ for $v = 1, 2$, which check model misspecifications in the conditional mean and conditional variance of the state variable. The $\hat{M}^{(v)}$ tests have similar size patterns as $\hat{M}$ except the overrejection is more severe in small samples.

Because the sizes of our tests using asymptotic theory differ significantly from the asymptotic significance level in small samples, we also consider the following parametric bootstrap procedure:

- Step (i): Use (e.g.) the Euler scheme or the generalized Milstein scheme to obtain a bootstrap sample $\mathcal{X}^b \equiv \{X_t^b\}_{t=\Delta}$ from the estimated null model

$$dX_t = \mu \left( X_t, \tilde{\theta} \right) dt + \sigma \left( X_t, \tilde{\theta} \right) dW_t + dJ_t(\tilde{\theta});$$

- Step (ii): Estimate the null model using the bootstrap sample $\mathcal{X}^b$, and compute a bootstrap statistic $\hat{M}^b$ in the same way as $\hat{M}$, with $\mathcal{X}^b$ replacing the original sample $\mathcal{X} = \{X_t\}_{t=\Delta}$;

- Step (iii): Repeat steps (i) and (ii) $B$ times to obtain $B$ bootstrap test statistics $\{\hat{M}_t^b\}_{t=1}^B$;

- Step (iv): Compute the bootstrap $p$-value $p_b \equiv B^{-1} \sum_{i=1}^B 1(\hat{M}_t^b > \hat{M})$. To obtain an accurate bootstrap $p$-value, $B$ must be sufficiently large.
Due to the high computational cost, we only consider bootstraps for $T = 250$. We generate 500 data sets of random sample $\{X_t\}_{t=1}^{T}$ and use $B = 100$ bootstrap iterations for each simulated data set. Table 1 shows that the bootstrap indeed approximates the finite sample distribution of test statistics more accurately. In particular, the bootstrap significantly reduces the overrejection of the asymptotic version of our derivative tests for conditional moments. The improvement of the finite sample version $\hat{M}_F$ is less significant since $\hat{M}_F$ has achieved reasonable sizes using the asymptotic theory.

7.1.2 Power of the $\hat{M}$ tests

To investigate the power of $\hat{M}$ in differentiating Vasicek’s (1977) model from other diffusion models, we simulate data from five popular diffusion models respectively and test the null hypothesis that data are generated from a Vasicek model. The first four models and the last model have been considered in HL and Ait-Sahalia et al. (2006) respectively:

- **DGP A1 [CIR Model]:**
  \[ dX_t = \kappa (\alpha - X_t) \, dt + \sigma \sqrt{X_t} \, dW_t, \]  
  where $(\kappa, \alpha, \sigma^2) = (0.89218, 0.090495, 0.032742)$.

- **DGP A2 [Ahn and Gao’s (1999) Inverse-Feller Model]:**
  \[ dX_t = X_t [\kappa - (\sigma^2 - \kappa \alpha) \, X_t] \, dt + \sigma X_t^{3/2} \, dW_t, \]  
  where $(\kappa, \alpha, \sigma^2) = (3.4387, 0.0828, 1.420864)$.\(^\text{20}\)

- **DGP A3 [CKLS (Chan, Karolyi, Longstaff and Sanders, 1992) Model]:**
  \[ dX_t = \kappa (\alpha - X_t) \, dt + \sigma X_t^p \, dW_t, \]  
  where $(\kappa, \alpha, \sigma^2, \rho) = (0.0972, 0.0808, 0.52186, 1.46)$.

- **DGP A4 [Ait-Sahalia’s (1996a) Nonlinear Drift Model]:**
  \[ dX_t = (\alpha_{-1} X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2) \, dt + \sigma X_t^p \, dW_t, \]  
  where $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma^2, \rho) = (0.00107, -0.0517, 0.877, -4.604, 0.64754, 1.50)$.

- **DGP A5 [Jump Diffusion Model]:**
  \[ dX_t = \kappa (\alpha - X_t) \, dt + \sigma dW_t + J_t dN_t, \]  
  where $(\kappa, \alpha, \sigma^2) = (0.214592, 0.089102, 0.001986)$, $N_t$ is a Poisson process with the intensity $\lambda = 0.05$, and $J_t$ is the jump size, which is independent of $N_t$ and has a normal distribution $N(0, \eta^2 = 0.003973)$.

\(^{20}\)There are some typos in the parameter values of Ahn and Gao’s (1999) inverse-feller model used in Hong and Li (2005). We have corrected them correspondingly.
Like HL, the parameter values for the CIR model are taken from Pritsker (1998), and the parameter values for Ahn and Gao’s inverse-Feller model are taken from Ahn and Gao (1999). For DGPs A3 and A4, the parameter values are taken from Ait-Sahalia’s (1999) estimates of real interest rate data. For the jump diffusion model, the parameter values are calculated from model (7.1) using Ait-Sahalia et al.’s (2006) method in their Example 3. For each of these four alternatives, we generate 500 data sets of the random sample \( \{X_t\}_{t=1,000} \), where \( T = 250, 500 \) and \( 1,000 \) respectively at the monthly sample frequency.

For the CIR, Ahn and Gao’s model and jump diffusion models, we simulate data from model transition densities, which have closed forms. For the CKLS and Ait-Sahalia’s nonlinear drift models, whose transition densities have no closed-form, we simulate data by the Euler scheme. 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 (7.1). Table 3 reports the rejection rates of \( \hat{M} \) and \( \hat{M}^{(2)} \) at the 10% and 5% levels using empirical critical values, which are obtained under \( H_0 \) and provide fair comparison on an equal ground. We include tests using bootstrap critical values when \( T = 250 \). Under DGP A1, model (7.1) is correctly specified for the drift function but is misspecified for the diffusion function because it fails to capture the "level effect". Both asymptotic and finite sample versions of the omnibus test have reasonable power under DGP A1, with rejection rates around 70% at the 5% level when \( T = 1,000 \). The finite sample \( \hat{M}_{FS} \) has a bit higher rejection rates than the asymptotic version of the \( \hat{M} \) test. The HL test is less powerful than \( \hat{M} \) tests, with rejection rates around 45% at the 5% level when \( T = 1,000 \). The variance test \( \hat{M}^{(2)} \) has good power and the rejection rates increase with \( T \). Interestingly, the mean test \( \hat{M}^{(1)} \) has no power, indicating that these diagnostic tests do not overreject correctly specified conditional mean dynamics.

Under DGP A2, model (7.1) is misspecified for both the conditional mean and conditional variance because it ignores the nonlinear drift and diffusion. As expected, both asymptotic and finite sample versions of the omnibus \( \hat{M} \) test have good power when model (7.1) is used to fit the data generated from DGP A2. The power of \( \hat{M} \) increases significantly with \( T \) and approaches unity when \( T=1,000 \). The HL test is more powerful than the \( \hat{M} \) tests in small samples but the difference becomes smaller as \( T \) increases. Both the mean test \( \hat{M}^{(1)} \) and the variance test \( \hat{M}^{(2)} \) have power and the rejection rates increase with \( T \).

Under DGP A3, the diffusion is no longer a linear function of \( X_t \). Consequently, model (7.1) is misspecified for the conditional mean and conditional variance. Both the asymptotic and finite sample versions of the omnibus \( \hat{M} \) test have good power when model (7.1) is used to fit the data generated from DGP A3. The rejection rates increase with \( T \) and approach unity when \( T = 1,000 \). However, the mean test \( \hat{M}^{(1)} \) has little power in detecting mean misspecification. One conjecture is that the difference between the true conditional mean under DGP A3 and the model (7.1)-implied conditional mean is small. This can be seen from a discretized version of DGP A3 by the Milstein scheme; namely,

\[
X_{t+1} = X_t + \left[ \kappa (\alpha - X_t) - \frac{1}{2} \sigma^2 X_t^{2\beta} \right] \Delta_t + \sigma X_t^{\beta} \Delta W_t + \frac{1}{2} \sigma^2 X_t^{2\beta} (\Delta W_t)^2, \\
\]

where \( \Delta_t \) is the length of the time discretization subinterval and \( \Delta W_t \) is the increment of the Brownian
motion. With small \( X_t \) (so that \( X_t^{2\rho} \) is negligible for \( \rho = 1.46 \), the leading term that determines the true condition mean under DGP A3 is \( \kappa (\alpha - X_t) \), which coincides with the drift of model (7.1). Thus, our mean test \( \hat{M}^{(1)} \) has litter power in detecting the small differences between the null and alternative models. The variance test \( \hat{M}^{(2)} \), however, has power and the rejection rates increase with \( T \). The HL test is more powerful than the \( \hat{M} \) tests in small samples, but the difference becomes smaller as \( T \) increases.

Under DGP A4, model (7.1) is misspecified for the conditional mean and conditional variance because it ignores the nonlinearity in both drift and diffusion. The patterns of our omnibus and diagnostic tests are similar to those under DGP A3. The HL test is more powerful than the \( \hat{M} \) tests when \( T = 250 \), but the rejection rates of the new tests increase quickly with \( T \) and approach unity when \( T = 1,000 \).

Under DGP A5, as shown in Ait-Sahalia et al. (2006), the transition density is, at the first order in \( \Delta \), a mixture of normal distributions: \( (1 - \lambda \Delta) N (\mu_1, \sigma_1^2) + \lambda \Delta N (\mu_2, \sigma_2^2) \), where \( \mu_t \) and \( \sigma_t^2 \) are given in (7.2) and (7.3). Under DGP A5, the conditional mean is correctly specified but the conditional variance is misspecified. Both the \( \hat{M} \) tests and the HL test have good power against this jump alternative and the HL test is more powerful. We note that the asymptotic version of our variance test \( \hat{M}^{(2)}_{AS} \) has good power in detecting variance misspecification but the finite sample version \( \hat{M}^{(2)}_{FS} \) has puzzlingly little power. In most cases, tests using bootstrap critical values have better power than tests using empirical critical values.

7.2 Bivariate models
7.2.1 Size of the \( \hat{M} \) tests

To examine the size of \( \hat{M} \) for bivariate models, we consider the following DGP:

- **DGP B0 (Bivariate Uncorrelated Gaussian Diffusion):**

\[
\begin{align*}
\frac{d}{dt} \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} &= \begin{pmatrix} \kappa_{11} & 0 \\ 0 & \kappa_{22} \end{pmatrix} \begin{pmatrix} \theta_1 - X_{1,t} \\ \theta_2 - X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} d \begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix},
\end{align*}
\] (7.9)

We set \((\kappa_{11}, \kappa_{22}, \theta_1, \theta_2, \sigma_{11}, \sigma_{22}) = (0.2, 0.8, 0, 0, 1, 1)\).\(^{21}\) With a diagonal matrix \( \kappa = \text{diag}(\kappa_{11}, \kappa_{22}) \), DGP B0 is an uncorrelated 2-factor Gaussian diffusion process. As shown in Duffee (2002), the Gaussian diffusion model has analytic expressions for the conditional mean and conditional variance respectively:

\[
\begin{align*}
E(X_t | X_s) &= \left[ I - e^{-\kappa(t-s)} \right] \theta + e^{-\kappa(t-s)} X_s, \\
\text{var}(X_t | X_s) &= \text{diag} \left\{ \frac{\sigma_{11}^2}{2\kappa_{11}} \left[ 1 - e^{2\kappa_{11}(s-t)} \right], \frac{\sigma_{22}^2}{2\kappa_{22}} \left[ 1 - e^{2\kappa_{22}(s-t)} \right] \right\},
\end{align*}
\] (7.10) (7.11)

where \( \theta = (\theta_1, \theta_2)' \) and \( s < t \). We simulate 1,000 data sets of the random sample \( \{X_t\}_{t=1}^{T} \) at the monthly frequency for \( T = 250, 500, 1,000 \) and 2,500 respectively from a bivariate normal distribution. For each data set, we use MLE to estimate model (7.9), with no restrictions on the intercept coefficients and compute the \( \hat{M} \) and HL test statistics.

\(^{21}\)We also try different parameters controlling the degree of persistence and simulation results show that our tests are not very sensitive to persistence of serial dependence in observations for two-dimensional case as well.
We focus on the finite sample version of the omnibus test in the bivariate case, which gives better sizes in finite samples than the asymptotic version of the omnibus test. To reduce computational costs, we generate \( \hat{\mathbf{u}} \) from a \( N(\mathbf{0}, \mathbf{I}_2) \) distribution, with each \( \hat{\mathbf{u}} \) having 15 grid points in \( \mathbb{R}^2 \) and let \( \mathbf{u} = (\hat{\mathbf{u}}', -\hat{\mathbf{u}}')' \) to ensure its symmetry. We standardize each component of \( \mathbf{X}_t \) and choose \( h = T^{-\frac{1}{6}} \), which attains the optimal rate for bivariate local linear fitting. The calculation of PITs for the bivariate model (7.9) used in HL is described in Section 2 (see also (18) and (19) of HL).

Table 4 reports the rejection rates of \( \hat{M}, \hat{M}^{(\nu)} \) and HL under DGP B0 at the 10% and 5% levels, using asymptotic theory. The \( \hat{M} \) test tends to underreject a bit and HL tends to overreject a bit. With \( |\nu| = 1, 2 \), the diagnostic tests \( \hat{M}^{(\nu)} \) check model misspecifications in conditional means, conditional variances and conditional correlation of state variables. The \( \hat{M}^{(\nu)} \) tests tend to overreject a bit, but not excessively. Overall speaking, both omnibus and diagnostic tests have reasonable sizes at both the 10% and 5% levels for sample sizes as small as \( T = 250 \) (i.e., about 20 years of monthly data). Our results show that the reasonable size performance of \( \hat{M} \) and \( \hat{M}^{(\nu)} \) in the univariate models carries over to the bivariate models. We also consider tests using bootstrap critical values for \( T = 250 \), which provide better sizes than asymptotic theory.

### 7.2.2 Power of the \( \hat{M} \) tests

To investigate the power of \( \hat{M} \) and \( \hat{M}^{(\nu)} \) in distinguishing model (7.9) from alternative models, we also generate data from four bivariate affine diffusion models respectively:

- **DGP B1** [Bivariate Correlated Gaussian Diffusion, with Constant Correlation in Diffusion]

  \[
  d \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.8 \end{pmatrix} \begin{pmatrix} -X_{1,t} \\ -X_{2,t} \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0.8 & 1 \end{pmatrix} d \begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}; \tag{7.12}
  \]

- **DGP B2** [Bivariate Correlated Gaussian Diffusion, with Constant Correlation in Drift]

  \[
  d \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0.2 & 0 \\ 0.8 & 0.8 \end{pmatrix} \begin{pmatrix} -X_{1,t} \\ -X_{2,t} \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} d \begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}; \tag{7.13}
  \]

- **DGP B3** [Bivariate Dai and Singleton’s (2000) A1 (2) process]

  \[
  d \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.8 \end{pmatrix} \begin{pmatrix} 2 - X_{1,t} \\ -X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \sqrt{X_{1,t}} & 0 \\ 0 & 1 \end{pmatrix} d \begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}. \tag{7.14}
  \]

- **DGP B4** [Bivariate Correlated Diffusion, with Time-varying Correlation in Diffusion]

  \[
  d \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.8 \end{pmatrix} \begin{pmatrix} 2 - X_{1,t} \\ 1 - X_{2,t} \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0.5 \sqrt{X_{1,t}} & 1 \end{pmatrix} d \begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}. \tag{7.15}
  \]

We use the Euler scheme to simulate 500 data sets of the random sample \( \{X_t\}_{t=\Delta}^{T} \) at the monthly frequency for \( T = 250, 500 \) and 1,000 respectively. For each data set, we use MLE to estimate model (7.9), with no restrictions on intercept coefficients. Table 5 reports the rejection rates of \( \hat{M}_{FS}, \hat{M}^{(\nu)}_{FS} \) and
HL at the 10% and 5% levels using empirical critical values. The empirical critical values are obtained under DGP B0.

With a nondiagonal matrix $\Sigma$, DGP B1 is a bivariate correlated Gaussian diffusion process with constant correlation in diffusion. Under DGP B1, model (7.9) ignores the nonzero constant correlation between state variables. The $\hat{M}_{FS}$ test has good power in detecting misspecification in the joint dynamics, with its rejection rate around 43% at the 5% level when $T = 1,000$. Interestingly, HL has no power with rejection rates around significance levels. This is not surprising because the conditional densities of individual variables $p(X_{1,t}, t|I_{t-1}, X_{2,t}, \theta)$ and $p(X_{2,t}, t|I_{t-1}, \theta)$ are correctly specified despite the joint dynamics is misspecified. Our correlation test $\hat{M}^{(1,1)}_{FS}$ has good power against correlation misspecification. Its rejection rate is about 75% at the 5% level when $T = 1,000$.

DGP B2 is another bivariate correlated Gaussian diffusion process, where the correlation between state variables comes from drifts rather than diffusions. Under DGP B2, model (7.9) is correctly specified for the diffusion function but is misspecified for the drift function. The power patterns of the $\hat{M}_{FS}$ and HL tests against the bivariate Vasicek model (7.9) are very similar to those under DGP B1. The rejection rate of $\hat{M}_{FS}$ increases with $T$ and approaches unity when $T = 1,000$, while the power of HL is close to 5% at the 5% level. The conditional mean and variance of $X_{2,t}$ and the conditional correlation between $X_{1,t}$ and $X_{2,t}$ are misspecified and our diagnostic tests are able to detect them.

DGP B3 is Dai and Singleton’s (2000) $A_1(2)$ model, where the first factor affects the instantaneous variance of $X_t$. Under DGP B3, model (7.9) is correctly specified for the drift function but is misspecified for the diffusion function because it fails to capture the "level effect" of $X_{1,t}$. Both the $\hat{M}_{FS}$ and HL tests have excellent power under DGP B3. In small samples, $\hat{M}_{FS}$ is more powerful than HL, but their rejection rates become very close when $T = 1,000$. The variance test $\hat{M}^{(2,0)}_{FS}$ has power against the misspecification in conditional variance of $X_{1,t}$. The mean test $\hat{M}^{(1,0)}_{FS}$ tends to overreject a bit although the conditional mean of $X_{1,t}$ is correctly specified. Nevertheless, the overrejection is not severe. The mean test $\hat{M}^{(0,1)}_{FS}$, the variance test $\hat{M}^{(0,2)}_{FS}$ and the correlation test $\hat{M}^{(1,1)}_{FS}$ do not overreject correctly specified conditional moments.

DGP B4 is a bivariate time varying correlated Gaussian diffusion process, where the correlation depends on $X_{1,t}$. If we use model (7.9) to fit data generated from DGP B4, there exists dynamic misspecification in conditional covariance between state variables. The $\hat{M}_{FS}$ test has good power when (7.9) is used to fit data generated from DGP B4. The rejection rate of the $\hat{M}_{FS}$ test increases to 95.8% at the 5% level when $T = 1,000$. The power of HL is still close to 5% at the 5% level. The correlation test $\hat{M}^{(1,1)}_{FS}$ has good power against this dynamic correlation misspecification, as the rejection rate is about 82% at the 5% level when $T = 1,000$.

To sum up, we observe:

- For both univariate and bivariate models, the $\hat{M}$ and $\hat{M}^{(v)}$ tests have reasonable sizes in finite samples, particularly when the parametric bootstrap is used. Although they tend to overreject a bit when $T=250$, they improve as the sample size $T$ increases. The finite sample versions of the proposed tests have better sizes than the asymptotic versions of the tests.
• The omnibus test $\tilde{M}$ has reasonable omnibus power in detecting various model misspecifications. It has reasonable power even when the sample size $T$ is as small as 250. It has some advantages in a multivariate framework. Particularly, it has good power in detecting misspecification in the joint dynamics even when the dynamics of individual components is correctly specified. This feature is not attainable by the HL test.

• The directional diagnostic tests $\tilde{M}^{(\nu)}$ can check various specific aspects of model misspecifications. Generally speaking, the mean test $\tilde{M}^{(\nu)}$, with $|\nu| = 1$, can detect misspecification in drifts; the variance test $\tilde{M}^{(\nu)}$, with $|\nu| = 2$, can check misspecifications in variances and correlations respectively. However, the mean test may fail to detect mean misspecification if the discrepancy between the data-implied conditional mean and the model-implied conditional mean is small.

8. CONCLUSION

The CCF-based estimation of continuous-time multivariate Markov models has attracted an increasing attention in financial econometrics. We have complemented this literature by proposing a CCF-based nonparametric regression omnibus test for the adequacy of a continuous-time multivariate Markov model. Our omnibus test fully exploits the information in the joint dynamics of state variables and thus can capture misspecification in modelling the joint dynamics, which may be easily missed by existing procedures. In addition, our omnibus test exploits the Markov property under both the null and alternative hypotheses and is an efficient approach when the DGP is indeed Markov. A class of diagnostic procedures is supplemented to gauge possible sources of model misspecifications. All test statistics follow an asymptotic null $N(0, 1)$ distribution, and they are applicable to various estimation methods, including suboptimal but consistent estimators. Simulation studies show that the proposed tests perform reasonably in finite samples for both univariate and bivariate continuous-time models, which demonstrate some nice merits of the CCF-based tests.

REFERENCES


### Table 1. Sizes of specification tests under DGP A0

<table>
<thead>
<tr>
<th>Significance level</th>
<th>.10</th>
<th>.05</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BCV-AS</td>
<td>BCV-FS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>T=250</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low persistence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{M} )</td>
<td>.126</td>
<td>.128</td>
</tr>
<tr>
<td>( \hat{M}^{(1)} )</td>
<td>.134</td>
<td>.140</td>
</tr>
<tr>
<td>( \hat{M}^{(2)} )</td>
<td>.110</td>
<td>.104</td>
</tr>
<tr>
<td>High persistence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{M} )</td>
<td>.112</td>
<td>.112</td>
</tr>
<tr>
<td>( \hat{M}^{(1)} )</td>
<td>.116</td>
<td>.114</td>
</tr>
<tr>
<td>( \hat{M}^{(2)} )</td>
<td>.098</td>
<td>.094</td>
</tr>
<tr>
<td></td>
<td>ACV-AS</td>
<td>ACV-FS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>T=500</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low persistence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{M} )</td>
<td>.159</td>
<td>.108</td>
</tr>
<tr>
<td>( \hat{M}^{(1)} )</td>
<td>.166</td>
<td>.107</td>
</tr>
<tr>
<td>( \hat{M}^{(2)} )</td>
<td>.165</td>
<td>.123</td>
</tr>
<tr>
<td>High persistence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{M} )</td>
<td>.191</td>
<td>.127</td>
</tr>
<tr>
<td>( \hat{M}^{(1)} )</td>
<td>.209</td>
<td>.146</td>
</tr>
<tr>
<td>( \hat{M}^{(2)} )</td>
<td>.238</td>
<td>.189</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>T=1000</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low persistence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{M} )</td>
<td>.136</td>
<td>.098</td>
</tr>
<tr>
<td>( \hat{M}^{(1)} )</td>
<td>.142</td>
<td>.113</td>
</tr>
<tr>
<td>( \hat{M}^{(2)} )</td>
<td>.139</td>
<td>.099</td>
</tr>
<tr>
<td>High persistence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{M} )</td>
<td>.136</td>
<td>.100</td>
</tr>
<tr>
<td>( \hat{M}^{(1)} )</td>
<td>.157</td>
<td>.124</td>
</tr>
<tr>
<td>( \hat{M}^{(2)} )</td>
<td>.185</td>
<td>.148</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: (i) DGP A0 is the Vasicek’s (1977) model, given in Eq. (7.1); (ii) Low persistence and high persistence correspond to \((\kappa, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)\) and \((0.214592, 0.089102, 0.000546)\) respectively; (iii) \(\hat{M}, \hat{M}^{(1)}\) and \(\hat{M}^{(2)}\) are the omnibus test, the conditional mean test and the conditional variance test respectively; (iv) BCV-AS and BCV-FS denote the asymptotic version and the finite-sample version using bootstrap critical values respectively; ACV-AS and ACV-FS denote the asymptotic version and the finite-sample version using asymptotic critical values respectively; (v) The \(p\) values of ACV-AS and ACV-FS are based on the results of 1,000 iterations; the \(p\) values of BCV-AS and BCV-FS are based on the results of 500 iterations.
Table 2. Sizes and Powers of Hong and Li’s (2005) tests under DGPs A0-A5

<table>
<thead>
<tr>
<th>Sample size</th>
<th>250</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significance level</td>
<td>.10</td>
<td>.05</td>
<td>.10</td>
</tr>
<tr>
<td>Lag p</td>
<td>10</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

Sizes

<table>
<thead>
<tr>
<th></th>
<th>.155</th>
<th>.146</th>
<th>.104</th>
<th>.103</th>
<th>.140</th>
<th>.136</th>
<th>.079</th>
<th>.094</th>
<th>.137</th>
<th>.151</th>
<th>.082</th>
<th>.095</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vasicek (low persistence)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vasicek (high persistence)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Powers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CIR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ahn &amp; Gao</td>
<td>.824</td>
<td>.782</td>
<td>.770</td>
<td>.728</td>
<td>.990</td>
<td>.994</td>
<td>.990</td>
<td>.984</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CKLS</td>
<td>.674</td>
<td>.660</td>
<td>.634</td>
<td>.612</td>
<td>.938</td>
<td>.930</td>
<td>.922</td>
<td>.914</td>
<td>.1.00</td>
<td>.998</td>
<td>.996</td>
<td>.992</td>
</tr>
<tr>
<td>Ait-Sahalia</td>
<td>.956</td>
<td>.954</td>
<td>.944</td>
<td>.932</td>
<td>.992</td>
<td>.996</td>
<td>.992</td>
<td>.996</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Jump Diffusion</td>
<td>.780</td>
<td>.766</td>
<td>.754</td>
<td>.738</td>
<td>.942</td>
<td>.960</td>
<td>.924</td>
<td>.938</td>
<td>.998</td>
<td>.996</td>
<td>.998</td>
<td>.996</td>
</tr>
</tbody>
</table>

Notes: (i) DGP A0 is the Vasicek’s (1977) model, given in Eq. (7.1); low persistence and high persistence correspond to $(\kappa, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)$ and $(0.214592, 0.089102, 0.000546)$ respectively; DGPs A1-A5 are CIR model, Ahn and Gao’s (1997) inverse-feller model, CKLS model, Ait-Sahalia’s (1996) nonlinear drift model and jump diffusion model, given in Eqs. (7.4)-(7.8) respectively; (ii) Results are based on Hong and Li’s (2005) test; (iii) The $p$ values of sizes are based on the results of 1000 iterations using asymptotic critical values; the $p$ values of powers are based on the results of 500 iterations using empirical critical values.
Table 3. Powers of specification tests under DGPs A1-A5

<table>
<thead>
<tr>
<th>Significance level</th>
<th>.10</th>
<th>.05</th>
<th>BCV-AS</th>
<th>BCV-FS</th>
<th>BCV-AS</th>
<th>BCV-FS</th>
</tr>
</thead>
<tbody>
<tr>
<td>T=250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M} )</td>
<td>.428</td>
<td>.454</td>
<td>.270</td>
<td>.266</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(1)} )</td>
<td>.134</td>
<td>.134</td>
<td>.076</td>
<td>.070</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(2)} )</td>
<td>.326</td>
<td>.314</td>
<td>.180</td>
<td>.132</td>
<td></td>
</tr>
<tr>
<td>Ahn &amp; Gao</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M} )</td>
<td>.848</td>
<td>.888</td>
<td>.688</td>
<td>.740</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(1)} )</td>
<td>.306</td>
<td>.336</td>
<td>.202</td>
<td>.218</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(2)} )</td>
<td>.544</td>
<td>.438</td>
<td>.424</td>
<td>.344</td>
<td></td>
</tr>
<tr>
<td>CKLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M} )</td>
<td>.506</td>
<td>.554</td>
<td>.324</td>
<td>.370</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(1)} )</td>
<td>.196</td>
<td>.214</td>
<td>.110</td>
<td>.120</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(2)} )</td>
<td>.356</td>
<td>.294</td>
<td>.240</td>
<td>.200</td>
<td></td>
</tr>
<tr>
<td>Ait-Sahalia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M} )</td>
<td>.818</td>
<td>.906</td>
<td>.612</td>
<td>.754</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(1)} )</td>
<td>.222</td>
<td>.316</td>
<td>.150</td>
<td>.184</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(2)} )</td>
<td>.318</td>
<td>.372</td>
<td>.216</td>
<td>.260</td>
<td></td>
</tr>
<tr>
<td>Jump diffusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M} )</td>
<td>.580</td>
<td>.590</td>
<td>.538</td>
<td>.564</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(1)} )</td>
<td>.130</td>
<td>.068</td>
<td>.076</td>
<td>.042</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(2)} )</td>
<td>.374</td>
<td>.076</td>
<td>.274</td>
<td>.044</td>
<td></td>
</tr>
<tr>
<td>ECV-AS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M} )</td>
<td>.122</td>
<td>.126</td>
<td>.048</td>
<td>.058</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(1)} )</td>
<td>.016</td>
<td>.018</td>
<td>.004</td>
<td>.006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(2)} )</td>
<td>.106</td>
<td>.084</td>
<td>.058</td>
<td>.044</td>
<td></td>
</tr>
<tr>
<td>Ahn &amp; Gao</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M} )</td>
<td>.548</td>
<td>.594</td>
<td>.378</td>
<td>.408</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(1)} )</td>
<td>.078</td>
<td>.070</td>
<td>.032</td>
<td>.034</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(2)} )</td>
<td>.294</td>
<td>.218</td>
<td>.216</td>
<td>.154</td>
<td></td>
</tr>
<tr>
<td>CKLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M} )</td>
<td>.204</td>
<td>.248</td>
<td>.108</td>
<td>.134</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(1)} )</td>
<td>.034</td>
<td>.040</td>
<td>.014</td>
<td>.020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(2)} )</td>
<td>.174</td>
<td>.132</td>
<td>.104</td>
<td>.076</td>
<td></td>
</tr>
<tr>
<td>Ait-Sahalia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M} )</td>
<td>.632</td>
<td>.732</td>
<td>.482</td>
<td>.578</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(1)} )</td>
<td>.176</td>
<td>.214</td>
<td>.112</td>
<td>.128</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(2)} )</td>
<td>.286</td>
<td>.318</td>
<td>.192</td>
<td>.212</td>
<td></td>
</tr>
<tr>
<td>Jump diffusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M} )</td>
<td>.604</td>
<td>.622</td>
<td>.536</td>
<td>.592</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(1)} )</td>
<td>.116</td>
<td>.074</td>
<td>.084</td>
<td>.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{M}^{(2)} )</td>
<td>.372</td>
<td>.100</td>
<td>.274</td>
<td>.046</td>
<td></td>
</tr>
</tbody>
</table>

Notes: (i) \( \hat{M} \), \( \hat{M}^{(1)} \) and \( \hat{M}^{(2)} \) are the omnibus test, the conditional mean test and the conditional variance test respectively; (ii) BCV-AS and BCV-FS denote the asymptotic version and the finite-sample version using bootstrap critical values respectively; ECV-AS and ECV-FS denote the asymptotic version and the finite-sample version using empirical critical values respectively; (iii) 500 iterations.
Table 3 (continued). Powers of specification tests under DGPs A1-A5

<table>
<thead>
<tr>
<th>Significance level</th>
<th>ECV-AS</th>
<th>ECV-FS</th>
<th>ECV-AS</th>
<th>ECV-FS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.10</td>
<td>.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=500</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>.354</td>
<td>.368</td>
<td>.206</td>
<td>.214</td>
</tr>
<tr>
<td>$\hat{M}^{(1)}$</td>
<td>.012</td>
<td>.014</td>
<td>.006</td>
<td>.008</td>
</tr>
<tr>
<td>$\hat{M}^{(2)}$</td>
<td>.278</td>
<td>.240</td>
<td>.162</td>
<td>.118</td>
</tr>
<tr>
<td>Ahn &amp; Gao</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>.974</td>
<td>.980</td>
<td>.912</td>
<td>.944</td>
</tr>
<tr>
<td>$\hat{M}^{(1)}$</td>
<td>.138</td>
<td>.170</td>
<td>.076</td>
<td>.084</td>
</tr>
<tr>
<td>$\hat{M}^{(2)}$</td>
<td>.532</td>
<td>.432</td>
<td>.396</td>
<td>.318</td>
</tr>
<tr>
<td>CKLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>.666</td>
<td>.706</td>
<td>.434</td>
<td>.504</td>
</tr>
<tr>
<td>$\hat{M}^{(1)}$</td>
<td>.096</td>
<td>.116</td>
<td>.048</td>
<td>.062</td>
</tr>
<tr>
<td>$\hat{M}^{(2)}$</td>
<td>.174</td>
<td>.132</td>
<td>.104</td>
<td>.076</td>
</tr>
<tr>
<td>Ait-Sahalia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>.922</td>
<td>.952</td>
<td>.788</td>
<td>.870</td>
</tr>
<tr>
<td>$\hat{M}^{(1)}$</td>
<td>.142</td>
<td>.196</td>
<td>.100</td>
<td>.126</td>
</tr>
<tr>
<td>$\hat{M}^{(2)}$</td>
<td>.286</td>
<td>.318</td>
<td>.192</td>
<td>.212</td>
</tr>
<tr>
<td>Jump diffusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>.754</td>
<td>.752</td>
<td>.738</td>
<td>.738</td>
</tr>
<tr>
<td>$\hat{M}^{(1)}$</td>
<td>.148</td>
<td>.110</td>
<td>.110</td>
<td>.054</td>
</tr>
<tr>
<td>$\hat{M}^{(2)}$</td>
<td>.712</td>
<td>.106</td>
<td>.612</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>.860</td>
<td>.890</td>
<td>.710</td>
<td>.726</td>
</tr>
<tr>
<td>$\hat{M}^{(1)}$</td>
<td>.026</td>
<td>.028</td>
<td>.004</td>
<td>.004</td>
</tr>
<tr>
<td>$\hat{M}^{(2)}$</td>
<td>.440</td>
<td>.406</td>
<td>.312</td>
<td>.268</td>
</tr>
<tr>
<td>Ahn &amp; Gao</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{M}^{(1)}$</td>
<td>.252</td>
<td>.286</td>
<td>.126</td>
<td>.180</td>
</tr>
<tr>
<td>$\hat{M}^{(2)}$</td>
<td>.816</td>
<td>.764</td>
<td>.744</td>
<td>.682</td>
</tr>
<tr>
<td>CKLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>.986</td>
<td>.996</td>
<td>.936</td>
<td>.952</td>
</tr>
<tr>
<td>$\hat{M}^{(1)}$</td>
<td>.134</td>
<td>.166</td>
<td>.084</td>
<td>.104</td>
</tr>
<tr>
<td>$\hat{M}^{(2)}$</td>
<td>.582</td>
<td>.526</td>
<td>.470</td>
<td>.422</td>
</tr>
<tr>
<td>Ait-Sahalia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>.998</td>
<td>1.00</td>
<td>.992</td>
<td>.996</td>
</tr>
<tr>
<td>$\hat{M}^{(1)}$</td>
<td>.106</td>
<td>.142</td>
<td>.052</td>
<td>.078</td>
</tr>
<tr>
<td>$\hat{M}^{(2)}$</td>
<td>.484</td>
<td>.456</td>
<td>.348</td>
<td>.334</td>
</tr>
<tr>
<td>Jump diffusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>.840</td>
<td>.840</td>
<td>.832</td>
<td>.832</td>
</tr>
<tr>
<td>$\hat{M}^{(1)}$</td>
<td>.160</td>
<td>.094</td>
<td>.082</td>
<td>.048</td>
</tr>
<tr>
<td>$\hat{M}^{(2)}$</td>
<td>.918</td>
<td>.168</td>
<td>.872</td>
<td>.128</td>
</tr>
</tbody>
</table>
Table 4. Sizes of specification tests under DGP B0

<table>
<thead>
<tr>
<th>T</th>
<th>250</th>
<th>.10</th>
<th>.05</th>
<th>500</th>
<th>.10</th>
<th>.05</th>
<th>1000</th>
<th>.10</th>
<th>.05</th>
<th>2500</th>
<th>.10</th>
<th>.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>.132</td>
<td>.098</td>
<td>.076</td>
<td>.051</td>
<td>.075</td>
<td>.046</td>
<td>.064</td>
<td>.043</td>
<td>.079</td>
<td>.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}^1$</td>
<td>$\hat{M}^{(1,0)}$</td>
<td>.110</td>
<td>.172</td>
<td>.056</td>
<td>.113</td>
<td>.083</td>
<td>.051</td>
<td>.125</td>
<td>.077</td>
<td>.107</td>
<td>.063</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{M}^{(0,1)}$</td>
<td>.146</td>
<td>.169</td>
<td>.090</td>
<td>.118</td>
<td>.075</td>
<td>.046</td>
<td>.131</td>
<td>.085</td>
<td>.111</td>
<td>.077</td>
<td></td>
</tr>
<tr>
<td>$\hat{M}^2$</td>
<td>$\hat{M}^{(2,0)}$</td>
<td>.118</td>
<td>.095</td>
<td>.052</td>
<td>.053</td>
<td>.084</td>
<td>.060</td>
<td>.130</td>
<td>.089</td>
<td>.119</td>
<td>.079</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{M}^{(0,2)}$</td>
<td>.130</td>
<td>.090</td>
<td>.072</td>
<td>.056</td>
<td>.084</td>
<td>.047</td>
<td>.102</td>
<td>.071</td>
<td>.114</td>
<td>.073</td>
<td></td>
</tr>
<tr>
<td>$\hat{M}^3$</td>
<td>$\hat{M}^{(1,1)}$</td>
<td>.098</td>
<td>.107</td>
<td>.052</td>
<td>.066</td>
<td>.105</td>
<td>.068</td>
<td>.129</td>
<td>.093</td>
<td>.129</td>
<td>.083</td>
<td></td>
</tr>
<tr>
<td>ACV</td>
<td>.167</td>
<td>.105</td>
<td>.176</td>
<td>.110</td>
<td>.144</td>
<td>.092</td>
<td>.146</td>
<td>.106</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: (i) DGP B0 is a bivariate uncorrelated Gaussian diffusion process, given in Eq. (7.9);
(ii) $\hat{M}$ is the omnibus test; HL is Hong and Li’s (2005) test; $\hat{M}^1$, $\hat{M}^2$, $\hat{M}^3$ are conditional mean tests, conditional variance tests and conditional correlation test respectively;
(iii) BCV-FS and ACV-FS denote the finite-sample version tests using bootstrap and asymptotic critical values respectively;
(iv) The $p$ values of BCV-FS are based on the results of 500 iterations; the $p$ values of ACV-FS and HL are based on the results of 1,000 iterations.
Table 5. Powers of specification tests under DGPs B1-B4

<table>
<thead>
<tr>
<th>Sample size</th>
<th>250</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significance level</td>
<td>.10</td>
<td>.05</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>BCV-FS</td>
<td>ECV-FS</td>
<td>BCV-FS</td>
</tr>
<tr>
<td>DGP B1 (Bivariate Correlated Gaussian Diffusion Process, with Constant Correlation in Diffusion)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{M} )</td>
<td>.278</td>
<td>.210</td>
<td>.184</td>
</tr>
<tr>
<td>HL</td>
<td>.080</td>
<td></td>
<td>.052</td>
</tr>
<tr>
<td>( \hat{M}^{(1,0)} )</td>
<td>.110</td>
<td>.098</td>
<td>.054</td>
</tr>
<tr>
<td>( \hat{M}^{(0,1)} )</td>
<td>.100</td>
<td>.082</td>
<td>.042</td>
</tr>
<tr>
<td>DGP B2 (Bivariate Correlated Gaussian Diffusion Process, with Constant Correlation in Drift)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{M} )</td>
<td>.454</td>
<td>.489</td>
<td>.342</td>
</tr>
<tr>
<td>HL</td>
<td>.074</td>
<td></td>
<td>.046</td>
</tr>
<tr>
<td>( \hat{M}^{(1,0)} )</td>
<td>.080</td>
<td>.054</td>
<td>.048</td>
</tr>
<tr>
<td>( \hat{M}^{(0,1)} )</td>
<td>.660</td>
<td>.637</td>
<td>.544</td>
</tr>
</tbody>
</table>

Notes: (i) DGP B1 is a bivariate correlated Gaussian diffusion model, with constant correlation in diffusion, given in Eq. (7.12); DGP B2 is a bivariate correlated Gaussian diffusion model, with constant correlation in drift, given in Eq. (7.13); DGP B3 is Dai and Singleton’s (2000) \( A_1(2) \), given in Eq. (7.14); DGP B4 is a bivariate correlated diffusion model, with time-varying correlation in diffusion, given in Eq. (7.15); (ii) The p values are based on the results of 500 iterations; (iii) BCV-FS and ECV-FS denote the finite-sample version tests using bootstrap and empirical critical values respectively.
Table 5 (continued). Powers of specification tests under DGPs B1-B4

<table>
<thead>
<tr>
<th>Sample size</th>
<th>250</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significance level</td>
<td>.10</td>
<td>.05</td>
<td>.10</td>
</tr>
<tr>
<td>BCV-FS</td>
<td>ECV-FS</td>
<td>BCV-FS</td>
<td>ECV-FS</td>
</tr>
</tbody>
</table>

**DGP B3 (Dai and Singleton's (2000) $A_1(2)$ process)**

| $\hat{M}$ | .720 | .816 | .618 | .754 | .946 | .900 | .998 | .990 |
| HL | .728 | .688 | .946 | .936 | 1.00 | 1.00 |

| $\hat{M}^{(1,0)}$ | .166 | .216 | .102 | .136 | .164 | .096 | .196 | .110 |
| $\hat{M}^{(0,1)}$ | .122 | .114 | .056 | .050 | .114 | .064 | .092 | .054 |

| $\hat{M}^{(2,0)}$ | .268 | .252 | .184 | .190 | .373 | .261 | .459 | .285 |
| $\hat{M}^{(0,2)}$ | .102 | .102 | .068 | .050 | .130 | .062 | .108 | .046 |

| $\hat{M}^{(1,1)}$ | .110 | .098 | .044 | .058 | .092 | .046 | .064 | .028 |

**DGP B4 (Bivariate Correlated Diffusion Process, with Time-varying Correlation in Diffusion)**

| $\hat{M}$ | .106 | .214 | .060 | .144 | .590 | .446 | .982 | .958 |
| HL | .130 | .094 | .166 | .110 | .122 | .076 |

| $\hat{M}^{(1,0)}$ | .114 | .060 | .054 | .034 | .196 | .132 | .254 | .162 |
| $\hat{M}^{(0,1)}$ | .126 | .044 | .068 | .012 | .106 | .054 | .142 | .084 |

| $\hat{M}^{(2,0)}$ | .138 | .134 | .076 | .070 | .230 | .158 | .286 | .190 |
| $\hat{M}^{(0,2)}$ | .126 | .098 | .052 | .052 | .114 | .056 | .152 | .080 |

| $\hat{M}^{(1,1)}$ | .454 | .474 | .328 | .344 | .724 | .594 | .912 | .818 |