BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2009-34

A Consistent Test for Multivariate Conditional Distributions

by Fuchun Li and Greg Tkacz

www.bank-banque-canada.ca

Bank of Canada Working Paper 2009-34

December 2009

A Consistent Test for Multivariate Conditional Distributions

by

Fuchun Li[1] and Greg Tkacz[2]

[1]Financial Stability Department
[2]Canadian Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
fuchunli@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

We thank seminar participants at the Bank of Canada and at the 2007 annual meeting of the Canadian Economics Association for useful comments and suggestions.

Abstract

We propose a new test for a multivariate parametric conditional distribution of a vector of variables y_t given a conditional vector x_t. The proposed test is shown to have an asymptotic normal distribution under the null hypothesis, while being consistent for all fixed alternatives, and having non-trivial power against a sequence of local alternatives. Monte Carlo simulations show that our test has reasonable size and good power for both univariate and multivariate models, even for highly persistent dependent data with sample sizes often encountered in empirical finance.

JEL classification: C12, C22
Bank classification: Econometric and statistical methods

Résumé

Les auteurs proposent un nouveau test en vue de vérifier la validité de la distribution conditionnelle paramétrique multivariée d'un vecteur de variables y_t étant donné un vecteur conditionnel x_t. Ils montrent que la statistique du test suit asymptotiquement une loi normale sous l'hypothèse nulle et que le test est convergent pour toutes les hypothèses alternatives spécifiées et relativement puissant sous une suite d'alternatives locales. D'après les résultats de simulations de Monte-Carlo, le niveau et la puissance du test sont raisonnables, que les modèles considérés soient univariés ou multivariés, même en cas de forte persistance des données dépendantes et d'échantillons de taille usuelle en finance empirique.

Classification JEL : C12, C22
Classification de la Banque : Méthodes économétriques et statistiques

1. Introduction

The forecast of a probability distribution and its associated aspects, such as value-at-risk and expected shortfall probabilities, have been widely used in economics and finance. For instance, in the explosively growing field of financial risk management, much effort has been put into providing forecasts of probability distributions of credit loss, asset and portfolio returns, etc., to capture a complete characterization of the uncertainty associated with these financial variables (Okhrin and Schmid 2006; Berkowitz 2001; Duffie and Pan 1997). In macroeconomics, monetary authorities in the United States (the Federal Reserve Bank of Philadelphia) and United Kingdom (the Bank of England) have been conducting quarterly surveys on the distribution forecasts for inflation and output growth to help set their policy instruments (e.g., inflation target, Tay and Wallis 2000). The validity of the forecast of a probability distribution, and its resulting inferences, however, are conditional on the hypothesis that the model used to produce the probability distribution is correctly specified. Obviously, a possible serious problem with the forecast of a probability distribution is model misspecification. A misspecified model can yield large errors in pricing, hedging, and risk management. A test is thus required to determine whether the forecast of a probability distribution implied by the model corresponds to the one implied by the data.

The work on testing whether a random variable originates from a stipulated unconditional distribution dates from as early as Pearson's chi-square test, the Kolmogorov-Smirnov test, and the Cramér-von Mises test (Darling 1957). Since then, many consistent specification tests have been developed for unconditional distribution functions (Fan, Li, and Min, 2006, and the references therein). Andrews (1997) extended the three tests (Pearson, Kolmogorov-Smirnov, and Cramér-von Mises) to the conditional distribution case. Stinchcombe and White (1998) provided consistent nonparametric tests for conditional distributions. Zheng (2000) then provided a consistent test of conditional density functions based on the first-order linear expansion of the Kullback-Leibler information criterion. More recently, Fan, Li and Min (2006) proposed a bootstrap test for conditional distributions, in which the conditional variables can be both discrete and continuous. A limitation of all these tests of conditional distributions is that the data must be independently and identically distributed (i.i.d.); clearly, this rules out time-series applications.

In a time-series context, Diebold, Gunther and Tay (1998) developed a variety of graphical approaches for evaluating conditional distribution functions based on a probability integral transform of the conditional density function. The rationale behind their approach is an early result of Rosenblatt (1952), who showed that the probability integral transform would be distributed as an independent and identical uniform distribution under the correct specification of a distribution function. Recently, standard statistical techniques have been used to carry out the test of the independent and identical uniformity of the transformed data. For instance, Berkowitz (2001) developed a test based on an extension of the Rosenblatt transformation in which the data can be transformed to independent and identical standard normal distribution under the correct specification of a distribution function. Berkowitz applied the likelihood ratio test to test the independence and normality of the transformed data in a linear first-order autoregressive model. Since the model used in Berkowitz's test captures only a specific sort of serial dependence in the data, he showed how to expand the model and associated tests to higher-order autoregressive models. However, this results in an increasing number of model parameters and reduces the power of the test.

Bai (2003) and Corradi and Swanson (2006) considered tests for the parametric conditional distribution in a dynamic model using an empirical distribution function. Bai's test can be applicable for various dynamic models, where the conditioning event allows for an infinite past history of information. The Corradi and Swanson (2006) test allows for dynamic misspecification under null hypothesis. Both the Bai and Corradi and Swanson tests have power against violations of uniformity but not against violations of independence of the transformed data.

Hong and Li (2005) developed an omnibus nonparametric specification test for independent and identical uniformity of the transformed data by comparing unity with a nonparametric kernel estimator for the joint density of the transformed data. Their test can be used for a wide variety of continuous-time and discrete-time dynamic models. However, like all above-mentioned tests based on the probability integral transform of the conditional density function, the Hong and Li test cannot be extended to a multivariate conditional density function because it is well-known that the probability integral transform of data with respect to a multivariate conditional density function is no longer i.i.d. uniformity even if the model is correctly specified.

Alternative tests for conditional density (distribution) functions have recently been suggested. Li and Tkacz (2006) built a consistent bootstrap test for conditional density functions with time-series. Aït-Sahalia *et al*. (2009) developed a nonparametric specification test for the conditional density function of a Markovian process. Both the Li and Tkacz and Aït-Sahalia *et al*. tests can only be used to test the conditional density (distribution) function with compact support (Assumption A.4 in Li and Tkacz 2006, Condition 2 in Aït-Sahalia *et al*. 2009). Consequently, their tests cannot be used to test the whole conditional density (distribution) functions with an unlimited domain.

In this paper we propose a consistent specification test for a parametric model which specifies a multivariate joint conditional distribution of a vector of variables y_t given a conditional vector x_t, where the conditional vector x_t may contain lags of y_t and (or) lags of some other variables[1]. Many models used in economics and finance are of this type; for instance, a (possibly multivariate) regression model with a given conditional distribution function of a ARCH process for the error terms, and a (possibly multivariate) continuous-time diffusion model with a given specification of its transitional distribution function. These models are popular analytic tools to model the stochastic dynamics of economic and financial variables, such as asset prices, interest rates, exchange rates, and macroeconomic factors, etc. It is shown that our test statistic follows an asymptotic normal distribution under the null hypothesis, while being consistent for all fixed alternatives and powerful against a sequence of local alternatives to the null hypothesis.

The rest of this paper is organized as follows. In the next section we introduce the test statistic for the conditional distributions in time-series models. The asymptotic null distribution, the consistency, and local power properties of the test statistic are then established. Section 3 presents a Monte Carlo simulation study to investigate the performance of the test in finite samples. Section 4 concludes, and the proofs are in the Appendix.

2. A Consistent Test For Parametric Conditional Distributions

Let the observations consist of $\{Z_t\}_{t=1}^{n}$, where $Z_t = (x_t, y_t)$, with unknown conditional distribution function $F(y|x)$ of y_t given $x_t = x$ and distribution function $F_y(y)$ of y_t, with $x_t \equiv (x_t^1, x_t^2, \ldots, x_t^p)$ and $y_t \equiv (y_t^1, y_t^2, \ldots, y_t^q)$ being vectors of dimension p and q respectively.

---

1. We note that, based on the Khmaladze martingale transformation, Bai and Chen (2008) propose a test for multivariate distributions with a focus on the multivariate unconditional normal and unconditional t-distributions.

$F_0(y|x, \theta)$ is a parametric family of conditional distribution functions with $\theta \in \Theta$ being a subset of R^d. For notational simplicity, throughout the rest of this paper, we use the following notation: $\sum_{s \neq t} = \sum_{s \neq t, s = 1}^{n}$, $\sum_{l, s \neq t} = \sum_{l = 1}^{n}\left(\sum_{s \neq t}\right)$, and all the limits are taken as $n \to \infty$.

We assume the sample $\{Z_t\}_{t=1}^{n}$ comes from a random sequence that is a strictly stationary and absolutely regular process with coefficient β_s, which is defined as $\beta_s = sup_{t \in N} E[sup_{A \in M_{t+s}^{\infty}(Z)}\{\mathrm{P}(\mathrm{A}|\mathrm{M}_{-\infty}^{t}(Z)) - P(A)\}]$, where $M_a^b(Z)$ denotes the Borel σ-algebra of events generated by $(Z_a, \dots ,Z_b)$ for $a \leq b$.

The null hypothesis to be tested is that the conditional distribution function is correctly specified:

$$\mathrm{H}_0\text{: } F(y|x) = F_0(y|x, \theta_0) \quad \text{for some } \theta_0 \in \Theta, \text{ almost everywhere in } (x, y)\,; \tag{1}$$

the alternative hypothesis H_1 of interest is the negation of H_0.

For any $y \equiv (y^1, y^2, \dots, y^q) \in R^q$, denote $I(y_t \leq y)$ as the indicator function of the event $\{y_t^1 \leq y^1, y_t^2 \leq y^2, \dots, y_t^q \leq y^q\}$. Let $\varepsilon_t(y, \theta) \equiv I(y_t \leq y) - F_0(y|x_t, \theta)$. Then H_0 holds if and only if there exists $\theta_0 \in \Theta$ such that $E(\varepsilon_t(y, \theta_0)|x_t) = F(y|x_t) - F_0(y|x_t, \theta_0) = 0$ for almost everywhere in y with probability one.

Let $\pi(x)$ be the unknown marginal density function of x_t. Then we define:

$$J(\theta) \equiv \int E\{\varepsilon_t(y, \theta) E[\varepsilon_t(y, \theta)|x_t]\pi(x_t)\} dF_y(y)$$

$$= \int E\{[F(y|x_t) - F_0(y|x_t, \theta)]^2 \pi(x_t)\} dF_y(y) \ . \tag{2}$$

Under the null hypothesis, we have $J(\theta_0) = 0$ for some $\theta_0 \in \Theta$, and under the alternative, we have $J(\theta) > 0$ for all $\theta \in \Theta$. Hence, $J(\theta)$, as a measure of departure from the null hypothesis, can be used as an indicator for constructing a consistent test for parametric conditional distributions. Our test will be based on the estimator of $J(\theta_0)$. As in Li (1999) and Powell, Stock, and Stoker (1989), the density weighting function $\pi(x)$ in (2) is introduced to avoid the random denominator problem associated with kernel estimation.

Let $\hat{\theta}_n$ be an estimator of θ_0, and $\hat{\pi}(x_t)$ and $\hat{E}[\varepsilon_t(y, \hat{\theta}_n)|x_t]$ the leave-one-out kernel estimators of $\pi(x_t)$ and $E[\varepsilon_t(y, \hat{\theta}_n)|x_t]$, respectively. Then, the parametric conditional distribution function $F_0(y|x, \theta_0)$, and $E[\varepsilon_t(y, \theta_0)|x_t]\pi(x_t)$ can be respectively estimated by

$F_0(y|x, \hat{\theta}_n)$ and $\hat{E}[\varepsilon_t(y, \hat{\theta}_n)|x_t]\hat{\pi}(x_t)$ which is,

$$\hat{E}[\varepsilon_t(y, \hat{\theta}_n)|x_t]\hat{\pi}(x_t) \equiv \frac{1}{nh^p}\sum_{s \neq t} K\left(\frac{x_s - x_t}{h}\right)\varepsilon_s(y, \hat{\theta}_n) , \tag{3}$$

where $K(\cdot)$ is a product kernel function $K\left(\frac{x_s - x_t}{h}\right) \equiv \prod_{l=1}^{p} k\left(\frac{x_s^l - x_t^l}{h}\right)$, and we assume that each of the window widths in the product kernel function is equal to h [2].

2. In general, the left-hand-side in equation (3) can be estimated by $(n\det(\mathrm{H}))^{-1}\sum_{s \neq t} K\left(H^{-1}(x_s - x_t)\right)\varepsilon_s(y, \hat{\theta}_n)$, where H is a non-singular window-width matrix and $K(\cdot)$ is the multivariate kernel function. The right-hand - side in equation (3) is obtained by assuming that each of the window widths is equal to h (h is scalar and $H = hI$) and $K(\cdot)$ is a product kernel function.

Let $\hat{F}_y(y)$ be the empirical cumulative distribution estimator of $F_y(y)$. Inserting these estimators above into the definition of $J(\theta_0)$, given by (2), yields the following estimator of $J(\theta_0)$:

$$J_n \equiv n^{-3}h^{-p} \sum_{l,\, s \neq t} K\left(\frac{x_t - x_s}{h}\right)\varepsilon_t(y_l, \hat{\theta}_n)\varepsilon_s(y_l, \hat{\theta}_n) \,. \tag{4}$$

Based on J_n, a feasible test statistic for H_0 is obtained:

$$T_n \equiv nh^{p/2}J_n/\hat{\sigma}_n \quad , \tag{5}$$

where $\hat{\sigma}_n$ will be defined in Theorem 1. The test statistic T_n does not contain an asymptotic bias term, because the asymptotic bias term is removed by using the "leave-one-out" estimator of $E[\varepsilon_t(y, \theta_0)|x_t]\pi(x_t)$.

In order to establish the asymptotic validity of this test statistic, we require the following assumptions.

Assumption 1. *$K(\cdot)$ is bounded and symmetric with* $\int K(\mathrm{u})du = 1$ *and* $\int |u|^2 K(u)du < \infty$.

Assumption 2. *The process* $\{Z_t = (x_t, y_t)\}$ *is strictly stationary absolutely regular with mixing coefficient* $\beta_s = O(\lambda^s)$ *for some* $0 < \lambda < 1$.

Assumption 3. *The parametric space* Θ *is compact and convex subset of* R^d. *Let* $\|\cdot\|$ *denote the Euclidean norm,* $(\partial F(y|x, \theta)/\partial y)^{-1}$, $\|\partial(\partial F(y|x, \theta)/\partial y)/\partial\theta\|$, $\|\partial^2(\partial F(y|x, \theta)/\partial y)/\partial\theta\partial\theta'\|$, *and* $\|\partial(\partial F(y|x, \theta)/\partial y)/\partial\theta \times \partial(\partial F(y|x, \theta)/\partial y)/\partial\theta'\|$ *are all bounded by a non-negative integrable function* $M(x, y)$. *For all* θ *in a neighbourhood of* θ_0 *of* Θ, $F(y|x, \theta)$, $||\partial F(y|x, \theta)/\partial\theta||$, $\|\partial^2 F(y|x, \theta)/\partial\theta\partial\theta'\|$ *are bounded by* $M(x, y)$.

Assumption 4. *There exists an estimator* $\hat{\theta}_n$ *of* θ_0 *such that* $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$ *under the null, whereas under the alternative,* $\hat{\theta}_n \to \theta^*$ *in probability where* $\theta^* \in \Theta$.

Assumption 1 is a standard regularity condition imposed on a kernel function. Assumption 2 requires that the process $\{Z_t\}$ be stationary and absolutely regular with geometric decay rate. The stationary absolutely regular property of process $\{Z_t\}$ is to ensure that a central limit theorem for second order degenerate U-statistics of absolutely regular processes can be used. The geometric decay rate is needed to derive some inequalities for asymptotic results. Absolutely regular processes with geometric decay rate have been used in different contexts by various authors, including Aït-Sahalia, Bickel, and Stocker (2001), Li (1999), and Fan and Li (1999), to make possible a satisfactory asymptotic theory of inference and estimation. Assumption 3 ensures the consistency and asymptotic normality of the Quasi-maximum likelihood estimator of White (1982). Assumption 4 is known to hold for many economic and finance models including some general regression models involving time series (Fuller 1996 and White 1994), conditional heteroskedasticity models (Newey and Steigerwald 1997), and continuous-time parametric models.

The asymptotic null distribution and consistency of T_n is provided in the following theorem.

Theorem 1. *Given Assumptions 1-4, if* $h \to 0$ *and* $nh^p \to \infty$, *we have*

(a) *Under* H_0, $T_n \equiv nh^{p/2} J_n / \hat{\sigma}_n \to N(0,1)$ *in distribution, where*

$$\hat{\sigma}_n^2 = \frac{2}{n^2 h^p} \sum_{s \neq t} K^2\left(\frac{x_t - x_s}{h}\right) \left[\int \varepsilon_t(y, \hat{\theta}_n) \varepsilon_s(y, \hat{\theta}_n) d\hat{F}_y(y)\right]^2 , \qquad (6)$$

which is a consistent estimator of

$$\sigma^2 \equiv 2\int K^2(u)du\int\int E\{[F(y \wedge \bar{y}|x_1) - F_0(y|x_1)F_0(\bar{y}|x_1)]^2 \pi(x_1)\} dF_y(y)dF_y(\bar{y}) \cdot \quad (7)$$

(b) *Under* H_1, $Pr(T_n \geq B_n) \to 1$ *, for any non-stochastic sequence* $B_n = o(nh^{p/2})$.

Proof: *See Appendix.*

Because our test is a centered statistic (by the leave-one-out estimation approach), it has a zero-mean limiting distribution. We next determine the power of our test against continuous local alternatives to the null hypothesis.

We define the following sequence of local alternative conditional distribution functions of y_t given x_t:

$$F_n(y|x) \equiv (1 - \delta_n)F_0(y|x, \theta_0) + \delta_n \times H(y, x) , \quad (8)$$

where δ_n is a sequence of positive real numbers tending to zero, and both $F_0(y|x, \theta_0)$ and $H(y, x)$ are conditional distribution functions. The null hypothesis states that the conditional distribution of y_t given x_t is $F_0(y|x, \theta_0)$, whereas under the alternative hypothesis the conditional distribution is $F_n(y|x)$. The asymptotic distribution of our test under the local alternative (8) is given in the following theorem.

Theorem 2: *Given Assumptions 1-4, if* $h \to 0$, $nh^p \to \infty$, *and* $\delta_n = n^{-1/2}h^{-p/4}$, *then under the local alternative (8), we have* $T_n \to N(\mu, 1)$ *in distribution, where*

$$\mu = E[\int(F_0(y|x_1, \theta_0) - H(y, x_1))^2 dF_y(y)\pi(x_1)]/\sigma.$$

Proof: *See Appendix.*

Let Z_{α} express the quantile at level α of the standard normal distribution, then the asymptotic local power of our test is $Pr(T_n > Z_{\alpha}) \to 1 - \Phi(Z_{\alpha} - \mu)$. Hence, our test has nontrivial power against the local alternatives in (8) because of $\mu > 0$.

In practice, the smoothing parameter h can be selected by several commonly used procedures, including the cross-validation method, the plug-in method, and some ad hoc methods. For the cross-validation method, we select the bandwidth h to minimize the integrated squared error function:

$$ISE(h) \equiv \int \{F(y|x) - F_0(y|x, \theta_0) - \hat{E}[\varepsilon_t(y, \theta_0)|x_t = x]\}^2 dF_y(y) dF_x(x), \tag{9}$$

where $F_x(.)$ is the distribution of x_t. A discrete approximation to $ISE(h)$ is the average squared error function:

$$ASE(h) \equiv n^{-2} \sum_{s,\, t\, =\, 1}^{n} [F(y_s|x_t) - F_0(y_s|x_t, \hat{\theta}) - \hat{g}_h(x_t, y_s)]^2,$$

where $\hat{g}_h(x_t, y_s)$ is the estimator of the regression function $E[\varepsilon_t(y, \theta)|x_t = x] = F(y|x) - F_0(y|x, \theta_0)$ for every $x = x_t, y = y_s$, that is:

$$\hat{g}_h(x_t, y_s) = \frac{\sum_{l \neq t} K\left(\frac{x_l - x_t}{h}\right)[I(y_l \le y_s) - F_0(y_s|x_l, \hat{\theta})]}{\sum_{l \neq t} K\left(\frac{x_l - x_t}{h}\right)}.$$

Minimizing $ASE(h)$ will yield an asymptotically optimal bandwidth that is proportional to $n^{-\frac{1}{4+p}}$. Hence we can choose the bandwidth h to be $cn^{-\frac{1}{4+p}}$, where c is a constant. Following Hardle[3] (1990), we can use the grid search method to find the optimal c that minimizes $CV(cn^{-1/(4+p)})$, where $CV(h) = n^{-2} \sum_{s,\, t\, =\, 1}^{n} \{I(y_t \le y_s) - F_0(y_s|x_t, \hat{\theta}) - \hat{g}_h(x_t, y_s)\}^2$.

3. Monte Carlo Study

In this section we present some Monte Carlo simulation results to investigate the performance of our test for both univariate and multivariate models. In general, the data-generating processes will be simulated from the continuous-time model represented by a stochastic differential equation. For univariate models, we simulate data from four popular one-factor term structural models examined in Aït-Sahalia (1999). For multivariate models we focus on affine diffusion models, given their importance in the existing financial literature (Duffie, Pedersen, and Singleton (2003)).

3.1 Univariate continuous-time models

To examine the test's size performance, we simulate data from the Vasicek (1977) model:

$$dx_t = \beta(\alpha - x_t)dt + \sigma dw_t, \tag{10}$$

where α is long-run mean of x_t, and β is the speed at which the process returns to the long-run mean. The β determines the dependent persistence of the process, i.e., the smaller β is, the stronger the serial dependence of x_t, and consequently, the slower the convergence to the long-run mean. As with Pritsker (1998), to examine the impact of dependent persistence of x_t on the size of our test, we consider both low and high levels of persistent dependence and adopt the same parameter values as Pritsker (1998). The parameter values for low and high levels of persistent dependence are, respectively,

---

3. Let $h_{cv} = \arg min\ CV(h)$ and $h_{ASE} = \arg min\ ASE(h)$, then $ASE(h_{cv})/h_{ASE} \to 1$ in probability (Theorem 7.1.1, Hardle, 1990).

$(\beta, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)$, and $(\beta, \alpha, \sigma^2) = (0.214592, 0.089102, 0.000546)$. The null hypothesis H_0 is:

$$F(x_t \le y | x_{t-\Delta} = x) = \left[2\pi V_E(1-e^{-\beta\Delta})\right]^{-1/2} \int_{-\infty}^{y} \exp\left[-\frac{1}{2}(u-(\alpha+(x-\alpha)e^{-\beta\Delta}))^2 / V_E(1-e^{-\beta\Delta})\right] du, \quad (11)$$

where $V_E = \sigma^2/(2\beta)$, and Δ is the sample interval.

Since Vasicek's model has a closed-form transition density and marginal density functions (Pritsker, 1998), the simulated sample path can be constructed by its transition density. The initial values are drawn from its marginal density.

To study the test's power performance, we simulate data from three diffusion processes and test the null hypothesis that the data is generated from the Vasicek model. The three diffusion processes are:

- Cox, Ingersoll and Ross (1985) model, henceforth CIR:

$$dx_t = \beta(\alpha - x_t)dt + \sigma\sqrt{x_t}dw_t, \quad (12)$$

where $(\beta, \alpha, \sigma^2) = (0.89218, 0.090495, 0.032742)$.

- Chan *et al*. (1992) model, henceforth CKLS:

$$dx_t = \beta(\alpha - x_t)dt + \sigma x_t^{\rho} dw_t, \quad (13)$$

where $(\beta, \alpha, \sigma^2, \rho) = (0.0972, 0.0808, 0.52186, 1.46)$.

- Aït-Sahalia (1996) nonlinear drift model:

$$dx_t = (\alpha_{-1}x_t^{-1} + \alpha_0 + \alpha_1 x_t + \alpha_2 x_t^2)dt + \sigma x_t^{\rho} dw_t, \quad (14)$$

where $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma^2, \rho) = (0.00107, -0.0517, 0.877, -4.604, 0.64754, 1.50)$.

The parameter values for the CIR model are taken from Pritsker (1998). For the CKLS and Aït-Sahalia models the parameter values are taken from Aït-Sahalia (1999). For the CIR model we simulate data from the model transition and marginal density functions. For the CKLS and Aït-Sahalia models, whose transition densities have no closed form, we simulate data using the Milstein scheme. We simulate 1000 data sets of a random sample $\{x_{t\Delta_n}\}_{t=1}^{n}$ with the same sampling interval $\Delta = 1/250$, that is, we sample the data at daily frequency. The sample sizes are $n = 250, 500, 1000, 2500$, which correspond to 1 year, 2 years, 4 years and 10 years of daily data, respectively. The kernel function is chosen to be the standard normal density function. The smoothing parameter h is selected to minimize $ASE(h)$. This yields an asymptotically optimal smoothing parameter, $h = cn^{-1/5}$, where c is a positive constant. We use the grid search method to find the optimal c that minimizes the $CV(ch^{-1/5})$. We let $0.2 \le c \le 2$, and the grid points start from 0.2 to 2 with an increment of 0.04.

The simulation results are reported in Table 1. We find that the T_n test has satisfactory size performance at all three levels for sample sizes as small as $n = 250$. The impact of the level of the persistent dependence on the size of our test is minimal, which suggests that our test achieves robustness to the persistent dependence. This can be explained by the fact that to test the null hypothesis at the level α_0, our test would use the following critical region: reject H_0 when $nh^{1/2}J_n \ge \hat{\sigma}_n z_{1-\alpha_0}$, where $\hat{\sigma}_n$ is the estimator of σ. Equations (6) and (7) indicate that the value of $\hat{\sigma}_n$ will change with respect to the different value of the persistence parameter β. Therefore, the critical values of our test statistic can be automatically adjusted for different values of the persistence parameter β[4]. From Table 1 it is also observed that our test has good power in detecting misspecification of the Vasicek model against its three alternatives. For a given alternative, the test's power always increases rapidly with respect to the sample size, in line with the test's consistency property.

3.2 Multivariate Continuous-Time Models

To examine the size of our test for multivariate diffusion processes, we simulate data from the affine two-factor Brennan-Schwartz model (Hsin, 1995):

$$d\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\left\{\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} - \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right\}dt + \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} d\begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix}, \tag{15}$$

where we set the parameter values as:

$$(b_{11}, b_{12}, b_{21}, b_{22}, a_1, a_2, \sigma_{11}, \sigma_{22}) = (-3.5804, 3.5804, 0, -0.3378, -2.343, 0.5760, 0.1455),$$

which are from Hsin (1995). The null hypothesis is that the data is generated from the process $x_t = (x_{1t}, x_{2t})'$ with the two-dimension transitional distribution function as follows:

$$H_0 : F(y_1, y_2 | x_1, x_2) = \int_{-\infty}^{y_2}\int_{-\infty}^{y_1} \frac{1}{\sqrt{2\pi}|v|^{1/2}} \exp[-(1/2)(u-\mu-\phi(x-\mu))'v^{-1}(u-\mu-\phi(x-\mu))]du_1 du_2, \tag{16}$$

where

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \mu = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \phi = \begin{bmatrix} \exp(b_{11}\Delta) & b_{11}/(b_{11}-b_{22})(\exp(b_{22}\Delta) - \exp(b_{11}\Delta)) \\ 0 & \exp(b_{22}\Delta) \end{bmatrix}, v = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix}$$

and

---

4. As pointed out by Pritsker (1998), the critical values in Aït-Sahalia's test (1996) are invariant to proportional changes in the variance and the persistent parameter β.

$$v_{11} = \sigma_{11}^2(1-\exp(-2b_{11}\Delta))/(2b_{11}) + \sigma_{22}^2(b_{11}/(b_{11}-b_{22}))^2[(\exp(-2b_{22}\Delta)-1)/(2b$$

$$-2(\exp((b_{11}+b_{22})\Delta)-1)/(b_{11}+b_{22}) + (\exp(-2b_{11}\Delta)-1)/(2b_{11})$$

$$v_{12} = \sigma_{11}^2 b_{11}/(b_{11}-b_{22})\{(\exp(2b_{22}\Delta)-1))/(2b_{22}) - (\exp((b_{11}+b_{22})\Delta)-1)/(b_{11}-b_{22})\}$$

$$v_{21} = v_{12},\ \ v_{22} = \sigma_{22}^2(\exp(2b_{22}\Delta)-1)/(2b_{22}),\ \Delta = 1/250\ .$$

To investigate the power of our test, we simulate data from two other affine term structure models and test the null hypothesis that the data is generated from the two-factor Brennan-Schwartz model in (15). We set the parameter values as in Aït-Sahalia and Kimmel (2010) in the following two affine models:

$$A_2(1):\qquad d\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} a_1 \\ 0 \end{bmatrix} dt + \begin{bmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_{11}\sqrt{x_{1t}} & 0 \\ 0 & \sigma_{22}\sqrt{1+\beta_{21}x_{2t}} \end{bmatrix} d\begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix}, \qquad (17)$$

where $(a_1, b_{11}, b_{21}, b_{22}, \beta_{21}, \sigma_{11}, \sigma_{22}) = (0.0075, -0.1, 0, -3, 0, 0.01, 0.01)$.

$$A_2(2):\qquad d\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} dt + \begin{bmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_{11}\sqrt{x_{1t}} & 0 \\ 0 & \sigma_{22}\sqrt{x_{2t}} \end{bmatrix} d\begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix}, \qquad (18)$$

where $(a_1, a_2, b_{11}, b_{21}, b_{22}, \sigma_{11}, \sigma_{22}) = (0.0075, 0.2325, -0.03, 0, -2.5, 0.01, 0.01)$.

We simulate 1000 data sets of a random sample $\{x_{t\Delta_n}\}_{t=1}^n$ with the same sampling interval $\Delta = 1/250$, that is, we sample the data at daily frequency. The sample sizes are $n = 250, 500, 1000$. The kernel function is chosen to be the bivariate standard normal density function. The smoothing parameter h is selected to minimize $ASE(h)$. This yields an asymptotically optimal smoothing parameter, $h = cn^{-1/6}$, where c is a positive constant. We use the grid search method to find the optimal c that minimizes the $CV(ch^{-1/6})$. We let $0.2 \le c \le 2$, and the grid points start from 0.2 to 2 with an increment of 0.04.

Table 2 reports the simulation results. We observe that the estimated sizes are close to their nominal sizes, and T_n can powerfully detect the bivariate Brennan-Schwartz model from the two alternative models with the misspecification of diffusion terms. It is noted that the test has higher power against $A_2(2)$ than $A_2(1)$. This is apparently due to the fact that the transition distribution of $A_2(2)$, which is a two-dimension non-central Chi-square, deviates more significantly from the two-dimension Guassian distribution and non-central Chi-square distribution. Overall, our simulation results reveal that our test performs rather well in finite samples for multivariate models, which suggests that the good finite-sample performance of our test in the univariate continuous time models can carry over to the multivariate models as well.

4. Conclusion

In this paper we propose a new test for a multivariate parametric conditional distribution of a vector of variables y_t given a conditional vector x_t. Under appropriate conditions, the proposed test statistic has been shown to follow a standard normal distribution under the null hypothesis and to be consistent against all possible fixed alternatives and $n^{-1/2}h^{-p/4}$ local alternatives. Simulation studies have shown that the test has reasonable size and good power in finite samples.

The test can be applied to evaluate a variety of univariate spot rate models and multivariate term structure models. We are currently investigating these issues.

APPENDIX: Proofs

Lemma 1. *Under assumptions 1-4 and the null hypothesis, we have:*

$$J_n = \bar{J}_n + o_p((nh^{p/2})^{-1}), \text{ where } \bar{J}_n = n^{-2}h^{-p}\sum_{s \neq t} K\left(\frac{x_s - x_t}{h}\right)\int \varepsilon_s(y, \hat{\theta}_n)\varepsilon_t(y, \hat{\theta}_n)dF_y(y) \ .$$

Proof of Lemma 1: Denoting $K_{st} \equiv K\left(\frac{x_s - x_t}{h}\right)$, $F(y) \equiv F_y(y)$, and $\hat{F}(y) \equiv \hat{F}_y(y)$, we need to prove that: $\Delta_n = n^{-2}h^{-p}\sum_{s \neq t} K_{st}\int \varepsilon_s(y, \hat{\theta}_n)\varepsilon_t(y, \hat{\theta}_n)d(\hat{F}(y) - F(y)) = o_p((nh^{p/2})^{-1}$.

Denoting $F_0(y|x_t, \theta_0) - F_0(y|x_t, \hat{\theta}_n)$ by $D_t(y, \theta_0, \hat{\theta}_n)$, we can write Δ_n as follows:

$$\Delta_n = n^{-3}h^{-p}\sum_{l, s \neq t} K_{st}[\varepsilon_s(y_l, \theta_0)\varepsilon_t(y_l, \theta_0) - \int \varepsilon_s(y, \theta_0)\varepsilon_t(y, \theta_0)dF(y)]$$

$$-2n^{-3}h^{-p}\sum_{l, s \neq t} K_{st}\{\varepsilon_s(y_l, \theta_0)D_t(y_l, \theta_0, \hat{\theta}_n) - \int \varepsilon_s(y, \theta_0)D_t(y, \theta_0, \hat{\theta}_n)dF(y)\}$$

$$+ n^{-3}h^{-p}\sum_{l, s \neq t} K_{st}\{D_s(y_l, \theta_0, \hat{\theta}_n)D_t(y_l, \theta_0, \hat{\theta}_n) - \int D_s(y, \theta_0, \hat{\theta}_n)D_t(y, \theta_0, \hat{\theta}_n)dF(y)\}$$

$$= \Delta_{n1} - \Delta_{n2} + \Delta_{n3}.$$

We shall show that $\Delta_{ni} = o((nh^{p/2})^{-1})$ for $i = 1, 2, 3$. We define $m = [C\log n]$, where C is a large positive constant. By Assumption 2, we have $\beta_m = O(\lambda^{-C\rho\log_\lambda n}) = O(n^{-C\rho})$, where $\rho = -\log\lambda > 0$. We first show that $\Delta_{n1} = o_p((nh^{p/2})^{-1})$. It is sufficient to show that $E(nh^{p/2}\Delta_{n1})^2 = o(1)$. Let $W_{s,t,l} \equiv K_{st}[\varepsilon_s(y_l, \theta_0)\varepsilon_t(y_l, \theta_0) - \int \varepsilon_s(y, \theta_0)\varepsilon_t(y, \theta_0)dF(y)]$. Then we have $E(nh^{p/2}\Delta_{n1})^2 = n^{-4}h^{-p}\sum_{l, s \neq t}\ \sum_{l', s' \neq t'} E[W_{s,t,l}W_{s',t',l'}]$.

We consider four different cases. (a): For any two summation indices k and L from l, s, t, l', s', t', we have $|k - L| > m$ for all $L \neq k$; (b): There exist exactly four different summation indices such that any index k from these four indices, we have $|k - L| > m$ for all $L \neq k$.

(c): There exist exactly three different summation indices such that any index k, from these three indices, we have $|k - L| > m$ for all $L \neq k$. (d) All the other remaining cases. We will use $EA_{(s)}$ to denote these cases $(s = a, b, c, d)$. Let $a(.)$ be a Borel measurable function. We denote:

$$E_s[a(Z_s, Z_t, Z_l)] \equiv \int a(Z, Z_t, Z_l) dF(Z), \ E_{s,t}[a(Z_s, Z_t, Z_l)] \equiv \iiint a(Z, \bar{Z}, Z_l) dF(Z, \bar{Z}),$$

where, $F(Z)$ and $F(Z, \bar{Z})$ are the marginal distribution function for Z_t and joint distribution function for (Z_s, Z_t), respectively.

For case (a), using Lemma 1 in Yoshihara (1976), $E_l[W_{s,t,l}] = 0$ or $E_{l'}[W_{s',t',l'}] = 0$, and choosing $C > 3(1+\delta)/\rho\delta$ in $m = [C\log n]$, we have:

$$EA_{(a)} \leq 0 + cn^2 h^{-p} M_n^{\delta/(1+\delta)} \beta_m^{\delta/(1+\delta)} = o(1),$$

where we used the fact that $M_n = O(h^p)$. For case (b), we only need to consider the case $|l - l'| \leq m$, since otherwise we will have l or l' is at least m period away from any other indices and by Lemma 1 in Yoshihara (1976), we know it is bounded by $cn^3 \beta_m^{\delta/(1+\delta)}$. With $|l - l'| \leq m$, we must have one index k at least m periods away from any other indices for $k = s, t, s', t'$. Hence, repeating application of Lemma 1 in Yoshihara (1976), we have:

$$EA_{(b)} \leq n^{-4} h^{-p} \sum_{case(b)} E\{[E_t E_s[K_{st} \varepsilon_s(y_l, \theta_0) \varepsilon_t(y_l, \theta_0)] - E_t E_s[K_{st} \int \varepsilon_s(y, \theta_0) \varepsilon_t(y, \theta_0) dF]$$

$$\times [E_{t'} E_{s'}[K_{s't'} \varepsilon_{s'}(y_l, \theta_0) \varepsilon_{t'}(y_l, \theta_0)] - E_{t'} E_{s'}[K_{s't'} \int \varepsilon_{s'}(y, \theta_0) \varepsilon_{t'}(y, \theta_0)] dF(y)] \ \}$$

$$+ cn^3 \beta_m^{\delta/(1+\delta)} = cn^3 \beta_m^{\delta/(1+\delta)},$$

because $E_s[K_{st} \varepsilon_s(y_l, \theta_0) \varepsilon_t(y_l, \theta_0)] = E_s[K_{st} \varepsilon_s(y_l, \theta_0)] \varepsilon_t(y_l, \theta_0)$, and

$$E_s K_{st} \varepsilon_s(y_l, \theta_0) = \iint K_{st}[I(y_s \leq y_l) - F(y_l | x_s, \theta_0)] f(y_s | x_s) \pi(x_s) dy_s dx_s$$

$$= \int [\int K_{st} I(y_s \le y_l) f(y_s | x_s) dy_s - \int K_{st} F(y_l | x_s, \theta_0) f(y_s | x_s) dy_s] \pi(x_s) dx_s$$

$$= \int [K_{st} F(y_k | x_s, \theta_0) - K_{st} F(y_l | x_s, \theta_0)] \pi(x_s) dx_s = 0 .$$

Similarly, $E_s K_{st} \varepsilon_s(y, \theta_0) = E_{s'} K_{s't'} \varepsilon_{s'}(y_{l'}, \theta_0) = E_{s'} K_{s't'} \varepsilon_s(y, \theta_0) = 0$.

For case (c), we only need to consider $|l - l'| \le m$, $|k - l| \le m$ (or $|k - l'| \le m$) for exactly one index $k \in \{s, t, s', t'\}$ since otherwise it will be bounded by $cn^3 \beta_m^{\delta/(1+\delta)}$. By symmetry we only need to consider $k = s$, repeating application of lemma 1 in Yoshihara (1976), we have:

$$EA_{(c)} \le n^{-4} h^{-p} \sum_{case(c)} E\{[[K_{st} \varepsilon_s(y_l, \theta_0) \varepsilon_t(y_l, \theta_0)] - [K_{st} \int \varepsilon_s(y, \theta_0) \varepsilon_t(y, \theta_0) dF(y)]$$

$$\times [E_{t'} E_{s'} [K_{s't'} \varepsilon_{s'}(y_l, \theta_0) \varepsilon_{t'}(y_l, \theta_0)] - E_{t'} E_{s'} [K_{s't'} \int \varepsilon_{s'}(y, \theta_0) \varepsilon_{t'}(y, \theta_0)] dF(y)] \}$$

$$+ cn^3 \beta_m^{\delta/(1+\delta)} = 0 + cn^3 \beta_m^{\delta/(1+\delta)} ,$$

because $E_{s'}[K_{s't'} \varepsilon_{s'}(y_l, \theta_0) \varepsilon_{t'}(y_l, \theta_0)] = E_{s'}[K_{s't'} \int \varepsilon_{s'}(y, \theta_0) \varepsilon_{t'}(y, \theta_0)] dF(y)] = 0$.

For case (d), for any three different indices, we has at most $m^3 n^3$ terms. Hence, we have $EA_{(d)} \le cm^3 n^3 (n^4 h^p)^{-1} h^p = o(1)$.

To prove $\Delta_{n2} = o_p((nh^{p/2})^{-1})$, we expand $F_0(y | x_t, \hat{\theta}_n)$ around θ_0 to obtain

$$F_0(y | x_s, \hat{\theta}_n) - F_0(y | x_s, \theta_0)$$

$$= \nabla' F_0(y | x_s, \theta_0)(\hat{\theta}_n - \theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)' \nabla^2 F_0(y | x_s, \bar{\theta})((\hat{\theta}_n - \theta_0)) . \qquad \text{(A.1)}$$

where $\bar{\theta}$ is equal to some convex combination of θ_0 and $\hat{\theta}_n$. Using (A.1), we have:

$$\Delta_{n2} = 2(\hat{\theta}_n - \theta_0)' n^{-3} h^{-p} \sum_{l, s \ne t} K_{st} \{\varepsilon_s(y_l, \theta_0) \nabla' F_0(y_l | x_t, \theta_0)$$

$$-\int \varepsilon_t(y, \theta_0)\nabla' F_0(y|x_t, \theta_0)dF(y)\}$$

$$+ (\hat{\theta}_n - \theta_0)' n^{-3}h^{-p}\sum_{l, s \neq t} K_{st}\{\varepsilon_s(y_l, \theta_0)\nabla^2 F_0(y_l|x_t, \bar{\theta})$$

$$-\int \varepsilon_t(y, \theta_0)\nabla^2 F_0(y|x_t, \bar{\theta})dF(y)\}(\hat{\theta}_n - \theta_0)$$

$$= 2(\hat{\theta}_n - \theta_0)'\Delta_{n21} + (\hat{\theta}_n - \theta_0)'\Delta_{n22}(\hat{\theta}_n - \theta_0)\,.$$

Using the same approach to prove that $\Delta_{n1} = o_p((nh^{p/2})^{-1})$, we can prove that $\Delta_{n21} = o_p((nh^{p/2})^{-1})$. By Assumption 3 and $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$, we have $\Delta_{n22} = O(1)$. Hence, we have $\Delta_{n2} = o_p((nh^{p/2})^{-1})$. By the mean value theorem and Assumption (3), we have $\left|D_s(y_l, \theta_0, \hat{\theta}_0)\right| \leq M(x_s, y_l)\left\|\hat{\theta}_0 - \theta_0\right\|$. Hence,

$$\Delta_{n3} \leq n^{-3}h^{-p}\sum_{l, s \neq t} K_{st}\{M(x_s, y_l)M(x_t, y_l) + \int M(x_s, y)M(x_t, y)dF(y)\}\left\|\hat{\theta}_0 - \theta_0\right\|^2 = O_p(n^{-1})$$

because $E\left|n^{-3}h^{-p}\sum_{l, s \neq t} K_{st}\{M(x_s, y_l)M(x_t, y_l) + \int M(x_s, y)M(x_t, y)dF(y)\}\right| = O(1)$ and $\left\|\hat{\theta}_0 - \theta_0\right\|^2 = O_p(n^{-1})$.

Proof of (a) of Theorem 1: We decompose $\bar{J}_n$ into the following three terms,

$$\bar{J}_n = \frac{2}{n^2h^p}\sum_{1 \leq s < t \leq n} K_{st}\int \varepsilon_s(y, \theta_0)\varepsilon_t(y, \theta_0)dF(y)$$

$$+ \frac{2}{n^2h^p}\sum_{s \neq t} K_{st}\int [I(y_s \leq y) - F_0(y|x_s, \theta_0)][F_0(y|x_t, \theta_0) - F_0(y|x_t, \hat{\theta}_n)]dF(y)$$

$$+ \frac{1}{n^2h^p}\sum_{s \neq t} K_{st}\int [F_0(y|x_s, \theta_0) - F_0(y|x_s, \hat{\theta}_n)][F_0(y|x_t, \theta_0) - F_0(y|x_t, \hat{\theta}_n)]dF(y)$$

$$= \bar{J}_{n1} + 2\bar{J}_{n2} + \bar{J}_{n3}\,.$$

From Lemma 1, we will complete the proof of Theorem 1 by showing:

(i) $nh^{p/2}\bar{J}_{n1} \to N(0,\sigma^2)$ in distribution, (ii)$\bar{J}_{n2} = o_p((nh^{p/2})^{-1})$, (iii) $\bar{J}_{n3} = o_p(n^{-1})$,

(iv)$\hat{\sigma}^2 = \sigma^2 + o_p(1)$.

Proof of (i): $nh^{p/2}\bar{J}_{n1} \to N(0,\sigma^2)$.

let $Z_s = (x_s, y_s)$, $Z_t = (x_t, y_t)$, $H_n(Z_s, Z_t) = K_{st}\int \varepsilon_s(y, \theta_0)\varepsilon_t(y, \theta_0)dF(y)$ and $U_n = \sum_{1 \le s < t \le n} H_n(Z_s, Z_t)$. Because $E[H_n(Z_s, Z_t)|Z_s = z] = 0$ under null hypothesis, U_n is degenerate. Let $\bar{r} = [n^{1/4}]$ and $m = [C \ \log n]$, where C is a positive constant.

We will use a central limit theorem for degenerate U-statistics from Fan and Li (1999) to prove (i). We now verify that Assumptions (1)-(4) in Fan and Li (1999) are satisfied under Assumptions 1-4. We express $\varepsilon_s(y, \theta_0)$ by $\varepsilon_s(y)$. We have:

$$
\begin{aligned}
\sigma_n^2 &\equiv E[H^2(\tilde{Z}_1, \tilde{Z}_2)] \\
&= \iint E\left\{K^2\left(\frac{x_1 - x_2}{h}\right)E[\varepsilon_1(y)\varepsilon_1(\bar{y})|x_1]E[\varepsilon_2(y)\varepsilon_2(\bar{y})|x_2]\right\}dF(y)dF(\bar{y}) \\
&= \iint E\left\{K^2\left(\frac{x_1 - x_2}{h}\right)[F(y \wedge \bar{y}|x_1) - F_0(y|x_1)F_0(\bar{y}|x_1)][F(y \wedge \bar{y}|x_2) - F_0(y|x_2)F_0(\bar{y}|x_2)]\right\}
\end{aligned}
$$

$\times dF(y)dF(\bar{y})$. Changing variables to $\frac{x_1 - x_2}{h} = u$, $x_1 = x_1$, $y = y$ and $\bar{y} = \bar{y}$, we obtain:

$$
\begin{aligned}
\sigma_n^2 = \ & h^p \iiiint K^2(u)f(x_1)f(x_1 - hu)[F(y \wedge \bar{y}|x_1) - F_0(y|x_1)F_0(\bar{y}|x_1)] \\
& \times [F(y \wedge \bar{y}|x_1 - hu) - F_0(y|x_1 - hu)F_0(\bar{y}|x_1 - hu)]dudx_1dF(y)dF(\bar{y}) \\
= \ & h^p \int K^2(u)du \iint E\{[F(y \wedge \bar{y}|x_1) - F_0(y|x_1)F_0(\bar{y}|x_1)]^2 f(x_1)\}dF(y)dF(\bar{y}) + o(h^p) \\
= \ & O(h^p).
\end{aligned}
\tag{A.2}
$$

Similarly we have:

$$\mu_{n4} = E[H^4(Z_1, Z_2)]$$

$$= \iiiint E\left\{K^4\left(\frac{x_1 - x_2}{h}\right)E[\varepsilon_1(y_1)\varepsilon_1(y_2)\varepsilon_1(y_3)\varepsilon_1(y_4)|x_1]E[\varepsilon_2(y_1)\varepsilon_2(y_2)\varepsilon_2(y_3)\varepsilon_2(y_4)|x_2]\right\}$$

$$\times dF(y_1)dF(y_2)dF(y_3)dF(y_4) = O(h^p) . \tag{A.3}$$

For $(i, j) = (1, 1), (2, 2)$ or $(3, 3)$, we have:

$$\gamma_{nij} \sim E[H^i(Z_t, Z_s)H^j(Z_{t'}, Z_{s'})]$$

$$= E\left\{K^i_{ts}K^j_{t's'}[\int\varepsilon_t(y)\varepsilon_s(y)dF(y)]^i[\int\varepsilon_{t'}(y)\varepsilon_{s'}(y)dF(y)]^j\right\} = O(h^{2p}) . \tag{A.4}$$

$$\tilde{\gamma}_{n14} \sim \int\{E[H(z, Z_t)H(z, Z_s)]\}^2 dF(z)$$

$$= \int\left\{E\left[K\left(\frac{z_1 - Z_t^1}{h}\right)\left(K\left(\frac{z_1 - Z_s^1}{h}\right)\right)\iint\varepsilon_1(y)\varepsilon_t(y)\varepsilon_1(\bar{y})\varepsilon_s(\bar{y})dF(y)dF(\bar{y})\right]\right\}^2 dF(z) = O(h^{4p}).$$

(A.5)

$$\tilde{\gamma}_{n22} = E[H^2(\tilde{Z}_1, \tilde{Z}_2)H^2(\tilde{Z}_1, \tilde{Z}_3)] = O(h^{2p}) . \tag{A.6}$$

From equations (A.2)-(A.6), we have shown $\sigma_n^2 = O(h^p)$, $\sigma_n^2 = O(h^p)$, $\mu_{n4} = O(h^p)$, $\gamma_n = max\{\gamma_{n11}, \tilde{\gamma}_{n22}, \tilde{\gamma}_{n14}\} = O(h^{2p})$, $v_n = max\{\gamma_{n22}, \gamma_{n13}\} = O(h^{2p})$. These results imply (A1) (i)-(iii) in Fan and Li (1999). Next, we prove that (A2) in Fan and Li (1999) is satisfied. We consider:

$$G(z_t, z_s) = E[H(Z, z_t)H(Z, z_s)]$$

$$= \int\left\{K\left(\frac{z - x_t}{h}\right)K\left(\frac{z - x_s}{h}\right)\int\varepsilon_t(y)\varepsilon_z(y)dF(y)\int\varepsilon_s(\bar{y})\varepsilon_z(\bar{y})dF(\bar{y})\right\}\pi(z)dz$$

$$= h^p \int \left\{ K(u) K\left(\frac{x_t - x_s}{h} + u\right) \int \varepsilon_t(y) \varepsilon_z(y) dF(y) \int \varepsilon_s(\bar{y}) \varepsilon_z(\bar{y}) dF(\bar{y}) \right\} \pi(x_t + hu) du \quad . \text{(A.7)}$$

Noting that $G(z_t, z_s)$ given in (A.7) already has a h^p factor, then changing variables, it is straightforward to show that $\sigma_G^2 = E[G^2(Z_t, Z_t)] = O(h^{2p})$, $\gamma_{nG11} = O(h^{3p})$ and $\mu_{nG2} = max_{s \neq t} E[G^2(Z_t, Z_s)] = O(h^{3p})$. Thus (A2) in Fan and Li (1999) is satisfied.

Finally, $n^2 \beta_m^{\delta/(1+\delta)} (m^2 / \sigma_n^4) = n^2 m^2 O(n^{\delta C \log(\lambda)/(1+\delta)} h^{2p}) = o(1)$, which implies $n^2 \beta_m^{\delta/(1+\delta)} / \sigma_n^4 = o(1)$ provided we choose C sufficiently large. Also it is easy to check that M_n is bounded by some positive constant. Hence (A3) in Fan and Li (1999) is satisfied.

Proof of (ii): $\bar{J}_{n2} = o_p((nh^{p/2})^{-1})$. Using (A.1), we have:

$$\bar{J}_{n2} = (\hat{\theta}_n - \theta_0)' A_{n1} + (\hat{\theta}_n - \theta_0)' A_{n2} (\hat{\theta}_n - \theta_0),$$

where
$$A_{n1} = (n^2 h^p)^{-1} \sum_{s \neq t} K_{st} \int \nabla F_0(y | x_s, \theta_0) \varepsilon_s(y) dF(y) \ ,$$
$$A_{n2} = (n^2 h^p)^{-1} \sum_{s \neq t} K_{st} \int \nabla^2 F_0(y | x_s, \theta_0) \varepsilon_s(y) dF(y).$$

We consider two different cases for A_{n1}: (a) $min\{|s - s'|, |s - t|, |s - t'|\} > m$ and (b) $min\{|s - s'|, |s - t|, |s - t'|\} \le m$. We use A_{n1a} and A_{n1b} to denote these two cases. Let $V(y, x_t) = \nabla' F_0(y | x_t, \theta_0)$, we have:

$$E(\|A_{n1}\|^2) = (n^2 h^p)^{-2} \sum_{s \neq t} \sum_{s' \neq t'} E[K_{st} K_{s't'} \int\int \varepsilon_s(y) \varepsilon_{s'}(\bar{y}) V(y, x_s) V'(\bar{y}, x_{s'}) dF(\bar{y}) dF(\bar{y})]$$
$$= (n^2 h^p)^{-2} [EA_{n1a} + EA_{n1b}] \ . \qquad \text{(A.8)}$$

Using Lemma 1 in Yoshihara (1976) and Buniakowsky-Schwarz inequality, we have:

$$A_{n1} = 0_p((n^2 h^p)^{-1}) + O_p(m(nh^{2p(\eta - 1)/\eta})^{-1/2}), \qquad \text{(A.9)}$$

$$A_{n2} = O(h^{(1-\eta)p/\eta}), \tag{A.10}$$

where $1 < \eta < 2$. (A.9) and (A.10) lead to $(\hat{\theta}_n - \theta_0)' A_{n1} = o_p((nh^{p/2})^{-1})$ and

$(\hat{\theta} - \theta_0)' A_{n2} (\hat{\theta} - \theta_0) = o_p((nh^{p/2})^{-1})$ respectively.

Proof of (iii): $\bar{J}_{n3} = o_p(n^{-1})$.

$$J_{n3} = \frac{1}{n^2 h^p}(\hat{\theta}_n - \theta_0)' \sum_{s \neq t} K_{st} \int \nabla_\theta F_0(y|x_s, \bar{\theta}_1) \nabla_\theta' F_0(y|x_t, \bar{\theta}_2) dF(y) (\hat{\theta}_n - \theta_0) = O_p(n^{-1}),$$

because $\sum_{s \neq t} K_{st} \int \nabla_\theta F_0(y|x_s, \bar{\theta}_1) \nabla_\theta' F_0(y|x_t, \bar{\theta}_2) dF(y) = O_p(n^2 h^p)$ by Assumption 3.

Proof of (iv): $\hat{\sigma}^2 = \sigma^2 + o_p(1)$.

The proof for (iv) is similar to that (iii). Hence, we will provide a sketch proof here.

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{2}{n^2 h^p} \sum_{s \neq t} K_{st}^2 [\int \varepsilon_s(y, \hat{\theta}_n) \varepsilon_t(y, \hat{\theta}_n) d\hat{F}(y)]^2 \\
&= \frac{2}{n(n-1)h^p} \sum_{s \neq t} K_{st}^2 [\int \varepsilon_s(y, \theta_0) \varepsilon_t(y, \theta_0) dF(y)]^2 + o_p(1) \equiv \hat{\sigma}_1^2 + o_p(1).
\end{aligned}$$

By Lemma 1 we have:

$$\begin{aligned}
E\hat{\sigma}^2 &= \frac{4}{n(n-1)h^p} \sum_{1 \le s < t \le n} E[K_{st}^2 [\int \varepsilon_s(y) \varepsilon_t(y) dF(y)]^2] \\
&= \frac{4}{n(n-1)h^p} \sum_{1 \le s < t \le n} E[K_{st}^2 \int\int E[\varepsilon_s(y)\varepsilon_s(\bar{y}) | x_s] E[\varepsilon_t(y)\varepsilon_t(\bar{y}) | x_t] dF(y) dF(\bar{y})] \\
&\quad + O\left((nh^p)^{-1} \sum_{k=1}^{n} \beta_k^{\delta/(1+\delta)}\right) = \sigma^2 + O((nh^p)^{-1}) .
\end{aligned} \tag{A.11}$$

Hence from (A.11) and $Var(\hat{\sigma}^2) = o(1)$, we have $\hat{\sigma}^2 = \sigma^2 + o_p(1)$.

Proof of (b) of Theorem 1: Using the similar arguments as those in the proof of (a) of Theorem 1, we can prove:

$$J_n = \bar{J}_n + o_p(1) = \int E\{[F(y|x_s) - F_0(y|x_s, \theta^*)]^2 \pi(x_s)\} dF(y) + o_p(1) .$$

Under H_1, $\int E\{[F(y|x_s) - F_0(y|x_s, \theta^*)]^2 \pi(x_s)\} dF(y) > 0$, and $\sigma_n^2 = O(1)$, these two results complete the proof for (b) of Theorem 1.

Proof of Theorem 2:

Following the same approach to prove Lemma 2 and (a) of Theorem 1, we can show that $J_n = \bar{J}_{n1} + o_p((nh^{p/2})^{-1})$. Let $\Delta(x, y) = H(y, x) - F_0(y|x, \theta_0)$. We can decompose $\bar{J}_{n1}$ as follows:

$$\bar{J}_{n1} = n^{-2}h^{-p}\sum_{s \neq t} K_{st}\int[\varepsilon_s(y) - \delta_n \times \Delta(y, x_s)][\varepsilon_t(y) - \delta_n \times \Delta(y, x_t)]dF(y)$$

$$-2\delta_n n^{-2}h^{-p}\sum_{s \neq t} K_{st}\int[\varepsilon_s(y) - \delta_n \times \Delta(y, x_s)]\Delta(y, x_t)dF(y)$$

$$+ n^{-2}h^{-p}\delta_n^2\sum_{s \neq t} K_{st}\int\Delta(y, x_s)\Delta(y, x_t)dF(y) = \bar{J}_{n1, 1} - 2\delta_n\bar{J}_{n1, 2} + \delta_n^2\bar{J}_{n(1, 3)} .$$

Noting $E[\varepsilon_s(y)|x_s] = \delta_n\Delta(y, x_s)$ under the local alternative, from Theorem 1 we have proved that $nh^{p/2}\bar{J}_{n1, 1} \to N(0, \sigma^2)$ in distribution. $E(nh^{p/2}\, \bar{J}_{n1, 3})$ converges to $E[\int\Delta^2(y, x_1)\pi(x_1)dF(y)]$, and by using Lemma 1 we can prove that $Var(\bar{J}_{n1, 3}) = o(1)$. Hence, by the Chebychev inequality, we have $\bar{J}_{n1, 3}$ converges to $E[\int\Delta^2(y, x_1)\pi(x_1)dF(y)]$ in probability. Hence, if we choose $\delta_n = n^{-1/2}h^{-p/4}$, then $nh^{p/2}\bar{J}_{n1} \to N[\bar{\mu}, \sigma^2]$, where $\bar{\mu} = \int H^2(y, x_1)\pi(x_1)dF(y)/\sigma$.

References

Aït-Sahalia, Y., 1996, Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 12, 721-765.

Aït-Sahalia, Y., 1999, Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance,* 54, 1361-1395

Aït-Sahalia, Y., P.J.Bickel and T.M.Stoker, 2001, Goodness-of-fit tests for regression using kernel methods. *Journal of Econometrics* 105, 363-412.

Aït-Sahalia, Y., and R. Kimmel, 2010, Estimating affine multifactor term structure models using closed-form likelihood expansions. *Journal of Financial Economics,* forthcoming.

Aït-Sahalia, Y., J. Fan and H.Peng, 2009, Nonparametric transition-based tests for jump diffusions. *Journal of the American Statistical Association* 104, 1102-1116.

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

Bai, J., 2003, Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics* 85, 531-549.

Bai, J., and Z. Chen, 2008, Testing multivariate distributions in GARCH models. *Journal of Econometrics* 143, 19-36.

Berkowitz, J., 2001, Testing density forecasts, with application to risk management. *Journal of Business & Economic Statistics* 19, 465-474.

Chan, K.C., G.A.Karolyi, F.A. Longstaff, and A.B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 47, 1209-1227.

Corradi, V. and N.R. Swanson, 2006, Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics* 133, 779-806.

Cox, J.C., J.E.Ingersoll, and S.A.Ross, 1985, A theory of the term structure of interest rates. *Econometrica* 53, 385-407.

Darling, D.A., 1957, The Kolmogorov-Smirnov, Cramer-Von Mises tests. *Annals of Mathematical Statistics* 28, 823-838.

Diebold, F. X., T. Gunther and A. S. Tay, 1998, Evaluating density forecasts, with applications to financial risk management. *International Economic Review* 39, 863-883.

Duffie, D. and J. Pan, 1997, An Overview of Value at Risk, *The Journal of Derivatives,* Spring, 7-49.

Duffie, D., L. Pedersen, and K. Singleton, 2003, Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343-1376.

Fan, Y. and Q. Li, 1999, Central limit theorem for degenerate U-statistic of absolute regular processes with application to model specification testing. *Journal of Nonparametric Statistics* 10, 245-271.

Fan, Y., Q. Li and I. Min, 2006, A nonparametric bootstrap test of conditional distributions. *Econometric Theory* 22, 587-613.

Fuller, W.A., 1996, *Introduction to Statistical Time Series*, 2nd Edition. John Wiley & Sons, Inc., New York.

Hardle, W., 1990, *Applied Nonparametric Regression*. Cambridge University Press, New York.

Hong, Y., and H. Li, 2005, Nonparametric specification testing for continuous-time models with application to spot interest rates. *The Review of Financial Studies* 18, 38-84.

Hsin, C.W., 1995, An empirical investigation of the two-factor Brennan-Schwartz term structure model. *Review of Quantitative Finance and Accounting* 5, 71-92.

Li, F. and G.Tkacz., 2006, A consistent bootstrap test for conditional density functions with time series data. *Journal of Econometrics* 133, 841-862.

Li, Q., 1999, Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101-147.

Newey, W.K. and D.G. Steigerwald, 1997, Asymptotic bias for quasi-maximum-likelihood estimations in conditional heteroskedasticity models. *Econometrica* 65, 587-599.

Okhrin, Y. and W. Schmid, 2006, Distributional properties of portfolio weights. *Journal of Econometrics* 134, 235-256.

Powell, J.L., J.H. Stock, &T.M. Stocker, 1989, Semiparametric estimation of index coefficients. *Econometrica* 57, 1403-1430.

Pritsker, M., 1998, Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies* 11, 449-487.

Rosenblatt, M., 1952, Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23, 470-472.

Stinchcombe, M.B., and H. White, 1998, Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295-325.

Tay, A.S. and K.F. Wallis, 2000, Density forecasting: A survey, *Journal of Forecasting* 19, 235-254.

Vasicek, O., 1977, An Equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177-188.

White, H., 1982, Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1-25.

White, H., 1994, *Estimation, inference and specification analysis.* Econometric society monographs No. 22, Cambridge university press.

White, H. and I. Domowitz, Nonlinear regression dependent observations. *Econometrica* 52, 143-161.

Yoshihara, K., 1976, Limiting behavior of U-statistic for stationary, absolutely regular processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 35, 237-252.

Zheng, J.X., 2000, A consistent test of conditional parametric distributions. *Econometric Theory* 16, 667-691.

Table 1: Percentage of Rejections of the H_0 for Univariate Diffusions

		T_n	
n	1%	5%	10%
		Vasicek (1977) Model with Low Level of Dependent Persistence	
250	0.006	0.033	0.041
500	0.005	0.023	0.064
1000	0.006	0.042	0.077
2500	0.008	0.047	0.085
		Vasicek (1977) Model with High Level of Dependent Persistence	
250	0.002	0.020	0.039
500	0.007	0.031	0.063
1000	0.005	0.043	0.076
2500	0.009	0.048	0.086
		Cox, Ingersoll, and Ross (1985) Model	
250	0.512	0.544	0.615
500	0.627	0.746	0.882
1000	0.923	0.945	0.961
2500	1.000	1.000	1.000
		Chan *et al*. (1992) Model	
250	0.705	0.759	0.814
500	0.873	0.938	0.982
1000	1.000	1.000	1.000
2500	1.000	1.000	1.000
		Aït-Sahalia (1996) Nonlinear Drift Model	
250	0.814	0.827	0.841
500	0.905	0.943	0.983
1000	1.000	1.000	1.000
2500	1.000	1.000	1.000

Table 2: Percentage of Rejections of the H_0 For Multivariate Diffusions

T_n			
Multivariate Affine Diffusion Process			
n	1%	5%	10%
Two-Factor Brennan-Schwartz Model			
250	0.004	0.040	0.063
500	0.002	0.045	0.081
1000	0.007	0.051	0.087
2500	0.008	0.049	0.091
Two-Factor Affine Model $A_2(1)$			
250	0.537	0.645	0.689
500	0.745	0.798	0.898
1000	0.934	0.985	0.991
2500	1.000	1.000	1.000
Two-Factor Affine Model $A_2(2)$			
250	0.713	0.788	0.823
500	0.890	0.922	0.975
1000	0.993	1.000	1.000
2500	1.000	1.000	1.000