Bank of Canada Banque du Canada

Working Paper 2001-21 / Document de travail 2001-21

A Consistent Bootstrap Test for Conditional Density Functions with Time-Dependent Data

by

Fuchun Li and Greg Tkacz

ISSN 1192-5434

Printed in Canada on recycled paper

Bank of Canada Working Paper 2001-21

December 2001

A Consistent Bootstrap Test for Conditional Density Functions with Time-Dependent Data

by

Fuchun Li and Greg Tkacz

Monetary and Financial Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
gtkacz@bankofcanada.ca

Contents

Acknowledgements

Thanks to Jean-Marie Dufour, Yanqin Fan, John W. Galbraith, Benoît Perron, and Pierre St-Amant, as well as seminar participants at the Centre de recherche et développement en économique (Université de Montréal) conference, “Resampling Methods in Econometrics,” the Bank of Canada, and the 2001 meetings of the Canadian Economics Association, for several valuable comments.

Abstract

This paper describes a new test for evaluating conditional density functions that remains valid when the data are time-dependent and that is therefore applicable to forecasting problems. We show that the test statistic is asymptotically distributed standard normal under the null hypothesis, and diverges to infinity when the null hypothesis is false. We use a bootstrap algorithm to approximate the distribution of the test statistic in finite samples, and show that the bootstrapped distribution converges to the asymptotic distribution in probability. A Monte Carlo simulation study reveals that the bootstrap test works well and is highly robust to the value of the smoothing parameter in the kernel density estimator. An application to inflation forecasting is also presented to demonstrate the use of the test.

JEL classification: C12, C15, E37
Bank classification: Econometric and statistical methods

Résumé

Les auteurs décrivent un nouveau test qui permet d'évaluer les densités de probabilité conditionnelles dans le cas de séries temporelles et qui se révèle par conséquent utile pour la prévision. Ils montrent que la statistique du test a pour loi asymptotique une loi normale centrée réduite si l'hypothèse nulle est vraie, mais qu'elle diverge vers l'infini si celle-ci est fausse. Lorsqu'ils se servent d'un algorithme de rééchantillonnage *bootstrap* pour représenter la distribution de la statistique du test sur de petits échantillons, ils constatent que la distribution ainsi obtenue converge vers la loi asymptotique en probabilité. Une simulation de Monte-Carlo révèle que le niveau et la puissance du test *bootstrap* sont satisfaisants et qu'ils ne sont pas sensibles à la valeur prise par le paramètre de lissage dans l'estimateur à noyau de la densité. Enfin, les auteurs appliquent leur test à la prévision de l'inflation pour en démontrer l'utilité.

Classification JEL : C12, C15, E37
Classification de la Banque : Méthodes économétriques et statistiques

1. Introduction

Conditional density functions arise in a variety of areas. One of the more useful applications involves the production of density forecasts, where the probability density of the forecast of a time series, such as the rate of inflation, can be used to make probability statements regarding the future course of that series.[1] The probability density, however, and its resulting interpretation, is conditional on the hypothesis that the model used to produce the forecasts is correctly specified. A test is thus required to determine whether the conditional density function implied by the model corresponds to the one implied by the data.

There is a long history of testing to determine whether a random variable originates from a stipulated distribution. For example, Kolmogorov (1933) introduced a test for evaluating whether an independent and identically distributed sample of random variables comes from a given continuous univariate distribution function. Much later, Bickel and Rosenblatt (1973) proposed a test for density functions based on estimation of kernel densities. Their test, however, requires that the distribution under the null hypothesis be completely specified; i.e., that there be no unknown parameters under the null. Fan (1994, 1995) extended the Bickel and Rosenblatt test to allow for the presence of unknown parameters, deriving both asymptotic and bootstrapped versions of the test. Fan and Ullah (1999) proposed a further extension that allows for weakly dependent data. All of the above tests, however, were designed only for *unconditional* density functions.

1. In its *Inflation Report*, the Bank of England regularly displays density forecasts for inflation over the next two years through the use of fan charts. For details on the construction of fan charts, see page 52 of the February 1999 *Inflation Report*.

Tests for evaluating conditional densities have a more recent history. Andrews (1997) proposed a conditional Kolmogorov test of conditional distribution functions. The asymptotic null distribution of Andrews' test is dependent upon a nuisance parameter, so the critical values for his test are obtained by a bootstrap procedure. Zheng (2000) used the Kullback-Leibler information criterion as a basis for testing conditional density functions. Zheng's test is consistent against all alternatives to the null, but the simulation results revealed that the power and size are sensitive to the smoothing parameters, and that, in particular, there are large differences between the test's true and nominal sizes. A limitation of the above tests is that the data must be independently and identically distributed. Clearly, this rules out time-series applications.

In the area of time series, Diebold, Gunther, and Tay (1998) proposed an exact test for evaluating conditional density forecasts based on an integral transform of the conditional density function for dependent observations. Their approach requires a visual assessment of histograms, and is limited by the requirement that there be no unknown parameters under the null hypothesis.

In this paper, we propose a new, computationally convenient test for evaluating conditional density functions that remains valid for time-series data. The test statistic is based on the integrated squared difference between the conditional density function implied by the parametric model and a nonparametric estimate of the true conditional density function. Our test can be used to determine whether a sample of random variables originates from a given conditional distribution, which can be applied to the evaluation of density forecasts, and it can also be used to perform specification tests of parametric models.

Recognizing that the standard normal distribution may not provide an accurate approximation of the small sample distribution of our test statistic under the null hypothesis, and that the true size of the test is very sensitive to the amount of smoothing used in the kernel estimate of the conditional density function, we use the bootstrap to approximate the finite-sample distribution of our test statistic. As in Horowitz (1994) and Fan (1995), given the original sample, our bootstrap sample always satisfies the null hypothesis. Thus, the distribution of the bootstrap test statistic always approximates the true distribution of the test statistic even when the null hypothesis is false.

This paper is organized as follows. In section 2 we introduce the test statistic for the conditional density function and establish its asymptotic distribution. In section 3 we explain how the bootstrap can be used to approximate the finite-sample distribution of the test statistic. Section 4 presents a Monte Carlo simulation study to investigate the performance of the test in finite samples. Section 5 describes an application to two inflation-forecasting models, and section 6 concludes. Appendix 1 presents technical assumptions, Appendix 2 gives proofs of our results, and Appendix 3 presents the central limit theorem for degenerate U-statistics.

2. An Asymptotic Test for Evaluating Conditional Density Functions

Let the observations consist of $\{Z_t\}_{t=1}^{n}$, where $Z_t = (x_t, y_t)$, with unknown conditional density function $\pi(y|x)$ of y_t, given $x_t = x$, and marginal density function $\pi(x)$ of x_t, with y_t and x_t being vectors of dimension p and q, respectively. $\pi_0(y|x, \theta)$ is a parametric family of conditional density functions, with $\theta \in \Theta$ being a compact subset of R^d. We assume that the sample $\{Z_t\}_{t=1}^n$ comes from a random sequence that is a strictly stationary and absolutely regular process with coefficient β_l, which is defined as $\beta_l = sup_{t \in N} E[sup_{A \in M_{t+l}^{\infty}(Z)} \{P(A|M_{-\infty}^t(Z)) - P(A)\}]$, where $M_s^t(Z)$ denotes the σ-algebra generated by $(Z_s, \ldots, Z_t)$ for $s \le t$.

We are interested in testing

$$\begin{aligned} &H_0: \pi(y|x) = \pi_0(y|x, \theta_0), \quad \text{for some } \theta_0 \\ &H_1: \pi(y|x) \neq \pi_0(y|x, \theta), \quad \text{for all } \theta. \end{aligned} \tag{1}$$

The identifying information in our test is the conditional density $\pi_0(y|x, \theta_0)$ of y given x. The null hypothesis restricts the conditional density function to be of the form $\pi_0(y|x, \theta_0)$. We use this restriction to test the null hypothesis by basing our test on the following distance measure between the conditional density functions:

$$T = \int\int[(\pi(y|x) - \pi_0(y|x, \theta_0))\pi(x)]^2 w(x, y) dF(x, y), \tag{2}$$

where $w(x, y)$ is a weight function and $F(x, y)$ is the cumulative distribution function of $\{Z_i\}$. Distance measures similar to (2) were used by, for example, Bickell and Rosenblatt (1973), Hall (1984), and Fan (1994, 1995). This measure is a useful indicator of model misspecification, since $T = 0$ if and only if the model is correctly specified; that is, if H_0 is true. The weight function allows us to focus on the relevant issue: namely, how the model performs in a particular range of the conditional distribution. The cumulative distribution function $F(x, y)$ is introduced to make our test statistics computationally simpler. In this paper, we assume that $w(x, y) = 1$; $w(x, y) \neq 1$ can be treated similarly.

The true conditional density function $\pi(y|x)$ can be consistently estimated by the kernel estimator whether or not the model is correctly specified, where

$$\hat{\pi}(y|x) = \frac{\hat{\pi}(x, y)}{\hat{\pi}(x)} \tag{3}$$

and

$$\hat{\pi}(x, y) = \left(\frac{1}{n \times h^{p+q}}\right) \sum_{t=1}^{n} K\left(\frac{x - x_t}{h}\right) K\left(\frac{y - y_t}{h}\right)$$

$$\hat{\pi}(x) = \left(\frac{1}{n \times h^{q}}\right) \sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right).$$

In the above equations, $K(\cdot)$ is a kernel function and $h \equiv h_n$ is a sequence of smoothing parameters used in the nonparametric estimation of the conditional density function. We use the product kernel, for example $K(y_t) = \Pi_{i=1}^{p} k(y_{t,i})$, $K(x_t) = \Pi_{i=1}^{q} k(x_{t,i})$, and $K(y_t, x_t) = \Pi_{i=1}^{p} k(y_{t,i}) \times \Pi_{i=1}^{q} k(x_{t,i})$, where $K(\cdot)$ is a univariate kernel function and $y_{t,i}$ and $x_{t,i}$ are the i^{th} components of y_t and x_t, respectively.

Suppose that $\hat{\theta}_n$ is an asymptotically normal estimator of θ_0 under the null. The parametric conditional density function is estimated by $\pi_0(y|x, \hat{\theta}_n)$. Let $\hat{F}_n$ be the empirical cumulative distribution estimate. Inserting these estimates into the definition of T, given by (2), yields the following estimator of T:

$$T_n(\hat{F}_n, \hat{\theta}_n) = \frac{1}{n} \sum_{t=1}^{n} [(\hat{\pi}(y_t|x_t) - \pi_0(y_t|x_t, \hat{\theta}_n)) \times \hat{\pi}(x_t)]^2 . \tag{4}$$

Under the null hypothesis, the parametric and nonparametric conditional density function estimators are consistent estimators of the true conditional density function. Thus, their distance will converge to zero as the sample size increases. If the null hypothesis is false, the distance will not converge to zero, because the nonparametric conditional density estimate will converge to a conditional density function outside the parametric class of $\pi_0(y_t|x_t, \theta)$. The following result describes the asymptotic properties of the test statistic.

Theorem 1: *Let*

$$J_n = \hat{\sigma}^{-1} n h^{(p+q)/2} [T_n(\hat{F}_n, \hat{\theta}_n) - \hat{\sigma}_1 n^{-1} h^{-(p+q)} + \hat{\sigma}_2 n^{-1} h^{-q}] .$$

Under Assumptions 1–5 in Appendix 1, we have

(a) Under H_0 , $J_n \to N(0,1)$ *in distribution.*

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

where $\hat{\sigma}, \hat{\sigma}_1$ *, and* $\hat{\sigma}_2$ *are consistent estimators of* σ, σ_1 *, and* σ_2 *, respectively. These are*

$$\sigma^2 = 2\int\int\{\int\int K(u_1)K(u_2)K(u_1+v_1)K(u_2+v_2)du_1du_2\}^2 dv_1dv_2 \times \int\int\{\pi(x,y)\}^4 dxdy \quad (5)$$

$$\sigma_1 = \int\int K^2(x,y)dxdy \times \int\int\{\pi(x,y)\}^2 dxdy \quad (6)$$

$$\sigma_2 = \int K^2(x)dx \times \int\int \pi^2(y|x)\pi(x)\pi(x,y)dxdy \quad (7)$$

$$\hat{\sigma}^2 = 2\int\int\{\int\int K(u_1,u_2)K(u_1+v_1,u_2+v_2)du_1du_2\}^2 dv_1dv_2 \times n^{-1}\sum_{i=1}^{n}[\hat{\pi}(x_i,y_i)]^3 \quad (8)$$

$$\hat{\sigma}_1 = \int\int K^2(x,y)dxdy \times n^{-1}\sum_{i=1}^{n}(\hat{\pi}(x_i,y_i)) \quad (9)$$

$$\hat{\sigma}_2 = \int K^2(y)dy \times n^{-1}\sum_{i=1}^{n}\{[\hat{\pi}(x_i, y_i)]^2/\hat{\pi}(x_i)\} . \tag{10}$$

Proof: *See Appendix 2.*[2]

3. A Bootstrap Test for Evaluating Conditional Density Functions

The test statistic in Theorem 1 has an asymptotic standard normal distribution under the null hypothesis. Our simulation studies reveal that there are large differences between the true and nominal sizes of the test in small samples (see section 4.1), which indicates that the finite-sample distribution of the test statistic cannot be well approximated by the standard normal distribution. Consequently, there is a need to provide a better approximation of the finite-sample distribution under the null hypothesis. In this section, we propose that the bootstrap be used to approximate the finite-sample distribution to conduct inference in small samples.

We propose the following bootstrapping algorithm to obtain the null distribution of the test:

1. Estimate the unknown parameter θ_0 using the sample $\{Z_i = (x_i, y_i)\}_{i=1}^{n}$, with the estimate denoted as $\hat{\theta}_0$.

2. For each x_i in the sample $\{Z_i = (x_i, y_i)\}_{i=1}^{n}$, generate a value of y_i* by random sampling from the conditional density function $\pi_0(y|x_i, \hat{\theta}_0)$. Use this simulated sample, $\{Z_i* = (x_i, y_i*)\}_{i=1}^{n}$, as the bootstrap sample.

3. Use the sample $\{Z*_t\}_{t=1}^{n}$ to compute J_n, given in Theorem 1, and denote it by J_n*,

2. Note that the integrals of squared kernels are generally known for some popular kernels, such as the Gaussian kernel, and do not need to be computed. See, for example, Pagan and Ullah (1999).

$$J_n^* = \hat{\sigma}^{*-1} n h^{(p+q)/2}[T_n^*(\hat{F}_n^*, \hat{\theta}_n^*) - \hat{\sigma}_1^* n^{-1} h^{-(p+q)} + \hat{\sigma}_2^* n^{-1} h^{-q}],$$

where $\hat{\sigma}^*, \hat{F}_n^*, \hat{\sigma}_1^*, \hat{\sigma}_2^*$ are estimators of $\sigma, F, \sigma_1, \sigma_2$, respectively, from the bootstrap sample. Also, $\hat{\theta}_n^*$ is the estimator of $\hat{\theta}_0$ from the bootstrap sample.

4. Repeat Steps 2 and 3 m times, yielding bootstrap replications $J_n^{*1}, \ldots, J_n^{*m}$. The empirical distribution of $J_n^{*1}, \ldots, J_n^{*m}$ is used to approximate the finite sample distribution of J_n under the null.

Having computed the test statistic, J_n, and obtained the empirical sampling distribution of the bootstrap test statistic, J_n^*, under the null, the α-level bootstrap critical values are given by $J_n^{*(1-\alpha)\times m}$. A test of size α can therefore be conducted by obtaining the null distribution for J_n and determining whether $J_n > J_n^{*(1-\alpha)\times m}$. If so, the null is rejected; otherwise, we fail to reject the null.

The following result shows that the distribution of the bootstrap test also tends to the standard normal in probability.

Theorem 2: *Under Assumptions 1–6 in Appendix 1, conditional on* $X_n = \{(x_i, y_i)\}_{i=1}^{n}$, $J_n^* \to N(0, 1)$ *in distribution.*

Proof of Theorem 2: *See Appendix 2.*

4. Monte Carlo Study

In this section, we report some simulation results to examine the finite-sample size and power of the asymptotic and bootstrap tests.

4.1 Size

To study the size performance, we simulate data from the Ornstein-Uhlenbeck process, which Vasicek (1977) used to model the dynamic movement of interest rates. The Ornstein-Uhlenbeck process is $dx_t = (\beta - x_t)dt + \sigma dW_t$, where W_t is a Wiener process, β and σ are constants, with the normal transition (conditional) density function given by

$$\pi_0(y|x, \theta) = p(x_s = y|x_t = x) = \frac{1}{\sqrt{2\pi s^2(\tau)}}\exp\left\{-\frac{(y-\beta-(x-\beta)e^{-(\tau)})^2}{2s^2(\tau)}\right\}, \tag{11}$$

where $s^2(\tau) = \frac{\sigma^2}{2}[1-e^{-2(\tau)}]$, $\tau = s-t$, and $\theta = (\beta, \sigma)$. The marginal density function is $\pi(x) = \frac{1}{\sqrt{2\pi v}}\exp\left\{-\frac{1}{2}\left(\frac{x-\beta}{\sqrt{v}}\right)^2\right\}$, where $v = \frac{\sigma^2}{2}$. The discrete sampling observations along the continuous sampling interval are observed over equispaced intervals with sampling interval $\tau = 1$. The starting values of the data-generating process are directly drawn from its marginal density. We choose the parameter space $\Theta = \left\{(\beta, \sigma), \frac{1}{2} \le \beta \le 10, \frac{1}{2} \le \sigma \le 10\right\}$. The parameters $(\theta_0 = (\beta_0, \sigma_0))$ are set to $\beta_0 = 1$ and $\sigma_0 = 1$.

Since the transition (conditional) density function has an explicit functional form, the log-likelihood function conditioned on the first observation of the Ornstein-Uhlenbeck process can be calculated as

$$\begin{aligned} l_n(\theta) &= \sum_{t=1}^{n} \ln[\pi_0(x_{t+1}|x_t, \theta)] \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(s^2(1)) - \sum_{t=1}^{n}\left[\frac{(x_{t+1}-\beta-(x_t-\beta)e^{-1})^2}{2s^2(1)}\right]. \end{aligned}$$

Under H_0, the conditional maximum-likelihood estimator of $\theta = (\beta, \sigma^2)$ is given by

$$\hat{\beta}_n = -\frac{\bar{x}_n - e^{-1}\bar{x}_{n-1}}{e^{-1} - 1} \tag{12}$$

and

$$\hat{\sigma}_n^2 = \frac{2}{1 - e^{-2}} \sum_{t=1}^{n} \left[\frac{(x_t - (\hat{\beta}_n - x_{t-1}))^2}{n} \right]. \tag{13}$$

Thus, $\pi_0(y|x, \hat{\theta}_n)$ can be obtained.

It is easy to check that $\lim_{x \to \infty} \sigma\pi(x) = 0, \lim_{x \to \infty} \left| \frac{\sigma}{(2(\beta - x))} \right| = 0$. From Hansen and Scheinkman (1995), the integral operator is a strong contraction. Therefore, there exists λ such that $\beta_l = O(\lambda^l)$, where $0 < \lambda < 1$. It follows that Assumption 3 in Appendix 1 is satisfied. $l_n(\theta)$ is the quasi-log-likelihood function and $\hat{\theta}_n = (\hat{\beta}_n, \hat{\sigma}_n^2)$ the quasi-maximum-likelihood estimator. In Assumption 5 (Appendix 1), $A_0(\theta) = \{E_x(\partial^2(\log\pi_0(x_t|x_{i-1}, \theta))/\partial\beta\partial\sigma^2)\}^{-1}$, where E_x expresses the conditional expectation of x_t given x_{t-1} and $\varphi(Z_i, \theta) = (\partial\log\pi_0(x_i|x_{i-1}, \theta)/\partial\theta)$.

We use the standard normal density, $K(x) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}$ as the kernel function. The smoothing parameter h is chosen according to $h = cn^{-1/2.25}$, where c is a positive constant. To see how the bootstrap approximation performs, we choose values of c in the interval [0.4, 1.4]; similar results are found for values outside this range.

Throughout the experiment, we discard the first 500 observations to eliminate any start-up effects. The number of Monte Carlo replications is fixed at 500, and the number of bootstrap replications is 100 for sample size n = 50, 100. See Hall (1986) for a theoretical explanation of the ability of the bootstrap to produce satisfactory results with few replications of the bootstrap sampling process. The bootstrap critical values are given by

$J_n{}^{*99}$ for the nominal level $\alpha = 1$ per cent, $J_n{}^{*95}$ for $\alpha = 5$ per cent, and $J_n{}^{*90}$ for $\alpha = 10$ per cent. The bootstrap test will reject the null hypothesis at 1 per cent (5 per cent, 10 per cent), if $J_n > J_n{}^{*99}(J_n{}^{*95}, J_n{}^{*90})$.

Table 1 reports the true sizes of the tests (J_n and $J_n{}^*$) based on both the asymptotic critical values and bootstrapped critical values. We find that the empirical sizes vary enormously from the nominal sizes when asymptotic critical values are used. By contrast, the use of bootstrapped critical values dramatically reduces the differences between the empirical and nominal sizes. In addition, empirical sizes of the asymptotic test change drastically with respect to the smoothing parameters. However, the empirical sizes of the bootstrap test are very close to the nominal sizes at 1 per cent, 5 per cent, and 10 per cent significant levels for all values of the smoothing parameters considered.

4.2 Power

To study the power performances, we simulate data from a square-root process, which Cox, Ingersoll, and Ross (1985) used to model the dynamic movement of spot interest rates. The square-root process, $dx_t = (\beta - x_t)dt + \sigma x_t^{1/2} dW_t$, where β and σ are constants, has the transition density function given by

$$\pi_0(y|x, \theta) = P(X_t = y|X_s = x) = \bar{c}e^{-u-v}\left(\frac{v}{u}\right)^{\frac{\bar{q}}{2}} I_{\bar{q}}\left((2uv)^{\frac{1}{2}}\right),$$

with X_t taking non-negative values, where $\theta = (\beta, \sigma)$, $\bar{c} = \frac{2}{\sigma^2(1 - e^{-(t-s)})}$, $u = \bar{c}ye^{-(t-s)}$, $v = \bar{c}x$, $\bar{q} = \frac{2\beta}{\sigma^2} - 1$, and $I_{\bar{q}}(.)$ is the modified Bessel function of the first kind of order q. The marginal density function is a gamma density function; i.e., $\pi(y) = \frac{w^v}{\Gamma(v)}\pi^{v-1}e^{-wv}$, where $w = 2\frac{1}{\sigma^2}$ and $v = 2\frac{\beta}{\sigma^2}$, with parameters being set to $\beta = 1$, $\sigma = 0.5$.

In our study of the test's power, the null hypothesis continues to stipulate that the data are generated by the Ornstein-Uhlenbeck process, while the data are in fact generated by the square-root process. Good power therefore requires a high percentage of rejections of the null. Table 2 indicates that for any given choice of smoothing parameter, the power of the test J_n increases rapidly with respect to the sample size and is very close to 100 per cent rejection of the null at $n = 400$, in line with the consistency property of the test. Similar results can be found for the bootstrap test, $J_n{}^*$.

5. Application: Inflation Forecasting

To illustrate the use of the bootstrap test developed in this paper, we apply it to two Canadian inflation-forecasting models: a simple AR(1) and a vector autoregression (VAR) proposed by McPhail (2000). We empirically test the conditional density functions implied by these two models, given the specifications of their disturbance terms. The AR(1) model is

$$\dot{P}_t = \alpha + \beta \dot{P}_{t-1} + \varepsilon_t, \qquad (14)$$

where $\dot{P}_t = 100 \times \log(P_t / P_{t-12})$ is the year-over-year growth rate of the core consumer price index (i.e., the CPI excluding food, energy, and the effect of changes in indirect taxes) and ε_t is assumed to be a normally distributed disturbance term with mean zero and variance σ^2. This measure of CPI is closely monitored by the Bank for evidence of underlying price pressures in the economy. The conditional density function of $\dot{P}_t$, given $\dot{P}_{t-1}$, is therefore $\pi_0(\dot{P}_t | \dot{P}_{t-1}, \theta) = \phi[(\dot{P}_t - \alpha - \beta \dot{P}_{t-1})/\sigma]/\sigma$, where $\theta = (\alpha, \beta, \sigma^2)$ and ϕ is the standard normal density function.

The VAR is based on a view that broad money contains useful information on the future path of inflation, especially in the long run. Let *P*, *M*, *Y*, and *S* express, respectively, the core CPI, broad money (the sum of $M2_+$, Canada Savings Bonds, and non-money market mutual funds sponsored by deposit-taking institutions), GDP, and the log of the spread between rates on medium-term (3- to 5-year) government bonds and short-term (90-day) commercial paper. The general form of the VAR is

$$\begin{pmatrix} \dot{P}_t \\ \dot{M}_t \\ \dot{Y}_t \\ S_t \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} + \begin{pmatrix} \alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{14} \\ \alpha_{21}, \alpha_{12}, \alpha_{23}, \alpha_{24} \\ \alpha_{31}, \alpha_{32}, \alpha_{33}, \alpha_{34} \\ \alpha_{41}, \alpha_{42}, \alpha_{43}, \alpha_{44} \end{pmatrix} \times \begin{pmatrix} \dot{P}_{t-1} \\ \dot{M}_{t-1} \\ \dot{Y}_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \\ \varepsilon_{t3} \\ \varepsilon_{t4} \end{pmatrix}, \tag{15}$$

where $\dot{k} = 100 \times \log(k_t / k_{t-4})$, for $k = \{P, M, Y\}$. Assume that ε_{t1} is normally distributed with mean zero and variance σ^2. The conditional density function of $\dot{P}_t$, given $\dot{P}_{t-1}, \dot{M}_{t-1}, \dot{Y}_{t-1}, S_{t-1}$, $\pi_0(\dot{P}_t | \dot{P}_{t-1}, \dot{M}_{t-1}, \dot{Y}_{t-1}, S_{t-1}, \theta)$, is

$$\phi[(\dot{P}_t - \alpha_{11}\dot{P}_{t-1} - \alpha_{12}\dot{M}_{t-1} - \alpha_{13}\dot{Y}_{t-1} - \alpha_{14}S_{t-1})/\sigma]/\sigma ,$$

where $\theta = (\alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{14}, \sigma)$. For the AR(1) model we use monthly observations of core CPI from January 1984 to March 2001. For the VAR model we are constrained by the frequency of national accounts data, and therefore use quarterly observations from 1968Q1 to 2001Q1.

The number of bootstrap replications, *m*, is 100. Bandwidths are chosen via leave-one-out cross-validation. Table 3 gives the results of the tests. At the 5 per cent significance level, we reject the null of equality of the conditional densities for both models, indicating that we should not use the conditional densities implied by the parametric models to make probability statements about future inflation. Figures 1 and 2 plot graphs of both the nonparametric and model-implied conditional density functions. Figure 1 shows that the nonparametric density is somewhat skewed relative to the parametric density, and that a higher probability should be given to lower levels of inflation. In Figure 2, the nonparametric conditional density has thinner tails than the parametric density, with most probability located around an inflation rate of 2 per cent.

Based on the nonparametric estimates of $\pi(\dot{P}_t|\dot{P}_{t-1})$ and $\pi(\dot{P}_t|\dot{P}_{t-1}, \dot{M}_{t-1}, \dot{Y}_{t-1}, S_{t-1})$ using (3), and conditional density functions implied by the AR(1) and VAR models, Table 4 computes the probability that next period's (April 2001 for the AR(1) and 2001Q2 for the VAR) inflation rate will lie in various ranges. For the AR(1) model, the parametric density produces a probability of 0.41 that next period's inflation will be between 1 and 2 per cent, while this probability rises to 0.60 for inflation being between 2 and 3 per cent. In contrast, the nonparametric density places a larger weight on lower inflation, namely 0.71 between 1 and 2 per cent, and 0.29 between 2 and 3 per cent. Both densities agree, however, that it is virtually certain that inflation will be in the 1 to 3 per cent official target range set by the Bank. For the VAR model, the densities are less agreeable. In particular, the parametric density implies that there is about a 20 per cent chance of witnessing an inflation rate above 3 per cent, while the nonparametric density views this possibility as being virtually non-existent. In the conduct of monetary policy, if the erroneous parametric density were used to set interest rates, policy-makers would be presented with a scenario in which the risk of higher inflation would be overstated. This could lead to interest rates being set at a higher level than would be necessary had the correct conditional density been used.

6. Conclusion

This paper has proposed asymptotic and bootstrap tests for testing parametric conditional density functions. These tests are based on comparing kernel density estimates of the true unknown conditional density function with the conditional density function implied by the model. The test statistics are more convenient to compute than others in the literature. Under appropriate conditions, the proposed tests are shown to have standard normal distributions under the null, and to be consistent against all possible fixed alternatives. The Monte Carlo simulation study illustrates that the bootstrap test provides a reliable approximation to the null distribution of T_n for sample sizes as small as 50, with the bootstrap test being robust to the choice of the smoothing parameter.

In an application to inflation forecasting, we show that the probabilities of inflation being in different ranges can vary considerably when they are extracted from either the parametric or nonparametric conditional densities, and such differences in probabilities can influence policy decisions. In particular, the VAR model considered in this paper appears to overstate the risk of high inflation rates relative to the nonparametric conditional density. Such miscalculations of risk can lead to less-informed policy decisions.

In future work, as another empirical exercise, we could apply the bootstrap test proposed in this paper to test the specification of continuous-time models of the spot interest rate. Furthermore, these tests could be extended to test for conditional symmetry (Bai and Ng 2001) and two-sample goodness-of-fit tests for conditional density functions.

References

Andrews, D.W.K. 1997. “A Conditional Kolmogorov Test.” *Econometrica* 65: 1097-1128.

Bai, J. and S. Ng. 2001. “A Consistent Test for Conditional Symmetry in Time Series Model.” *Journal of Econometrics* 103: 183–224.

Bickel, P.J. and M. Rosenblatt. 1973. “On Some Global Measures of the Deviations of Density Function Estimates.” *Annals of Statistics* 1: 1071–95.

Bierens, H.J. 1983. “Uniform Consistency of Kernel Estimators of a Regression Function under Generalized Conditions.” *Journal of the American Statistical Association (*September): 699–707.

Cox, J.C., J.E. Ingersoll, and S.A. Ross. 1985. “A Theory of the Term Structure of Interest Rates.” *Econometrica* 53: 385–407.

Denker, M. and G. Keller. 1983. “On U-Statistic and v. Mises’ Statistics for Weakly Dependent Processes.” *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* 64: 505–22.

Diebold, F.X., T. Gunther, and A.S. Tay. 1998. “Evaluating Density Forecasts, with Applications to Financial Risk Management.” *International Economic Review* 39: 863–83.

Fan, Y. 1994. “Testing the Goodness-of-Fit of a Parametric Density Function by Kernel Method.” *Econometric Theory* 10: 316–56.

———. 1995. “Bootstrapping a Consistent Nonparametric Goodness-of-Fit Test.” *Econometric Reviews* 14: 367–82.

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: 245–71.

Fan, Y. and A. Ullah. 1999. “On Goodness-of-Fit Tests for Weakly Dependent Processes using Kernel Methods.” *Journal of Nonparametric Statistics* 10: 337–60.

Hall, P. 1984. “Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators.” *Journal of Multivariate Analysis* 14: 1–16.

———. 1986. “On the Number of Bootstrap Simulations Needed to Construct a Confidence Interval.” *Annals of Statistics* 14: 1453–62.

Hansen, L.P. and J.A. Scheinkman. 1995. “Back to the Future: Generating Moment Implications for Continuous Time Markov Processes.” *Econometrica* 63: 767–804.

Horowitz, J.L. 1994. “Bootstrap-Based Critical Values for the Information Matrix Test.” *Journal of Econometrics* 61: 395–411.

Kolmogorov, A.N. 1933. “Sulla determinazione empirica di una legge di distribuzione.” *Giornale Dell’ Instituto Ital. Degli Attuari* 4: 83–91.

Masry, E. and D. Tjostheim. 1995. “Nonparametric Estimation and Identification of Nonlinear ARCH Time Series: Strong Convergence and Asymptotic Normality.” *Econometric Theory* 11: 258–89.

McPhail, K. 2000. “Broad Money: A Guide for Monetary Policy.” In *Money, Monetary Policy and Transmission Mechanisms*, 51–84. Proceedings of a conference held by the Bank of Canada, November 1999. Ottawa: Bank of Canada.

Pagan, A. and A. Ullah. 1999. “Nonparametric Econometrics.” Cambridge: Cambridge University Press.

Vasicek, O. 1977. “An Equilibrium Characterization of the Term Structure.” *Journal of Financial Economics* 5: 177–88.

White, H. 1982. “Maximum Likelihood Estimation of Misspecified Models.” *Econometrica* 50: 1–25.

Yoshihara, K. 1976. “Limiting Behavior of U-Statistic for Stationary, Absolutely Regular Processes.” *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* 35: 237–52.

Zheng, J.X. 2000. “A Consistent Test of Conditional Parametric Distributions.” *Econometric Theory* 16: 667–91.

Table 1: Percentage Rejections of the True Null (Size) with Asymptotic and Bootstrap Critical Values

1 per cent significance level (n=50)						
c	0.4	0.6	0.8	1	1.2	1.4
J_n	100	74.2	18.6	5.2	2.0	0.2
J_n*	1.4	1.2	0.8	1.2	1.0	0.6
5 per cent significance level (n=50)						
J_n	100	92.8	53.0	18.8	9.0	2.8
J_n*	5.4	4.6	4.2	4.0	4.8	4.4
10 per cent significance level (n=50)						
J_n	100	98.8	71.0	33.8	19.8	8.0
J_n*	10.0	10.0	8.0	8.2	9.2	9.8
1 per cent significance level (n=100)						
J_n	100	76.0	17.0	8.2	3.2	1.8
J_n*	1.2	1.0	1.2	0.6	0.8	1.0
5 per cent significance level (n=100)						
J_n	100	91.8	60.2	18.2	8.4	7.6
J_n*	5.8	4.8	5.2	4.2	4.6	4.6
10 per cent significance level (n=100)						
J_n	100	99.4	79.0	32.8	19.4	8.4
J_n*	11.0	9.4	8.2	10.2	9.8	8.6

Notes: In each case, the number of bootstrap replications is m = 100, with 500 Monte Carlo replications performed. Data are generated from the Ornstein-Uhlenbeck process. J_n is a test performed with asymptotic critical values, and J_n* with bootstrapped critical values. c is used to select the smoothing parameter $h_n = cn^{-1/2.25}$ of the conditional nonparametric kernel density function. n is the size of the simulated sample.

Table 2: Percentage Rejections of the False Null (Power)

c	0.4	0.6	0.8	1	1.2	1.4
1 per cent significance level						
n=50	94.8	39.8	17.2	10.6	4.2	4.0
n=100	100	72.6	44.2	28.2	21.2	15.2
n=200	100	98.6	81.8	62.2	50.4	44.6
n=400	100	99.8	98.2	92.4	93.0	88.8
5 per cent significance level						
n=50	99.0	67.2	41.8	28.2	16.6	13.4
n=100	100	92.2	71.2	55.0	44.8	33.4
n=200	100	99.4	96.2	84.4	76.0	72.6
n=400	100	100	99.8	97.8	98.6	99.4
10 per cent significance level						
n=50	100	83.6	59.4	42.8	31.2	22.8
n=100	100	96.0	82.8	71.6	58.8	50.8
n=200	100	100	98.4	91.6	86.2	87.0
n=400	100	100	100	98.8	99.6	99.4

Notes: Data generated by the square-root process, with the false null stipulating the Ornstein-Uhlenbeck process. Results shown are for the asymptotic test, J_n. c is used to select the smoothing parameter $h_n = cn^{-1/2.25}$ of the conditional nonparametric kernel density function. n is the size of the simulated sample. 500 Monte Carlo replications were performed.

Table 3: Bootstrap Test Results for Inflation Forecasts

Model	Bootstrap critical values (5% level)	Estimated test statistic	Result
AR(1)	2.4145	2.9244	Reject H_0
VAR	45.8011	64.7704	Reject H_0

Note: The null hypothesis states that the parametric conditional density equals the true conditional density approximated by the nonparametric kernel density; the alternative is that the conditional densities are not equal.

Table 4: Inflation Probabilities at Different Ranges

Inflation range	AR(1)		VAR	
	Nonparametric density forecast	AR(1) implied density forecast	Nonparametric density forecast	VAR implied density forecast
< 1%	0.005	0.000	0.007	0.050
1% to 2%	0.708	0.406	0.487	0.286
2% to 3%	0.287	0.594	0.485	0.453
> 3%	0.000	0.000	0.021	0.211
1 to 3%	0.995	1.000	0.971	0.739

Notes: Inflation is defined as the annual core CPI inflation rate. For the AR(1) model, the probabilities are for inflation being in the stipulated ranges in April 2001. For the VAR model, the probabilities are computed for the second quarter of 2001.

Figure 1: Conditional Density Functions, Nonparametric and AR(1)

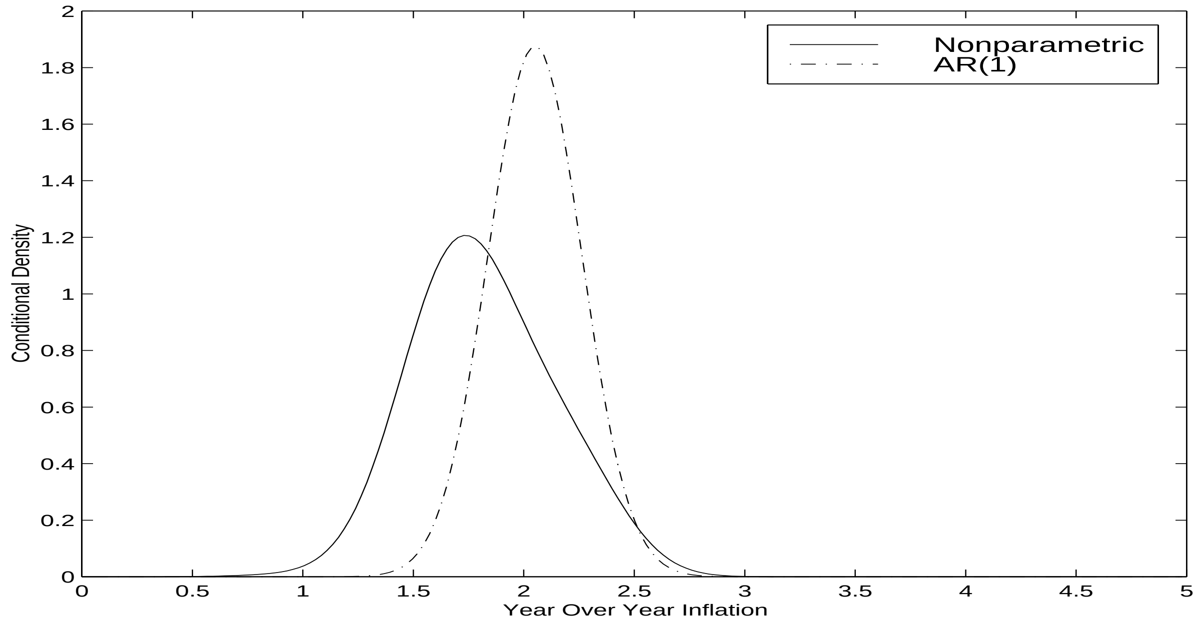
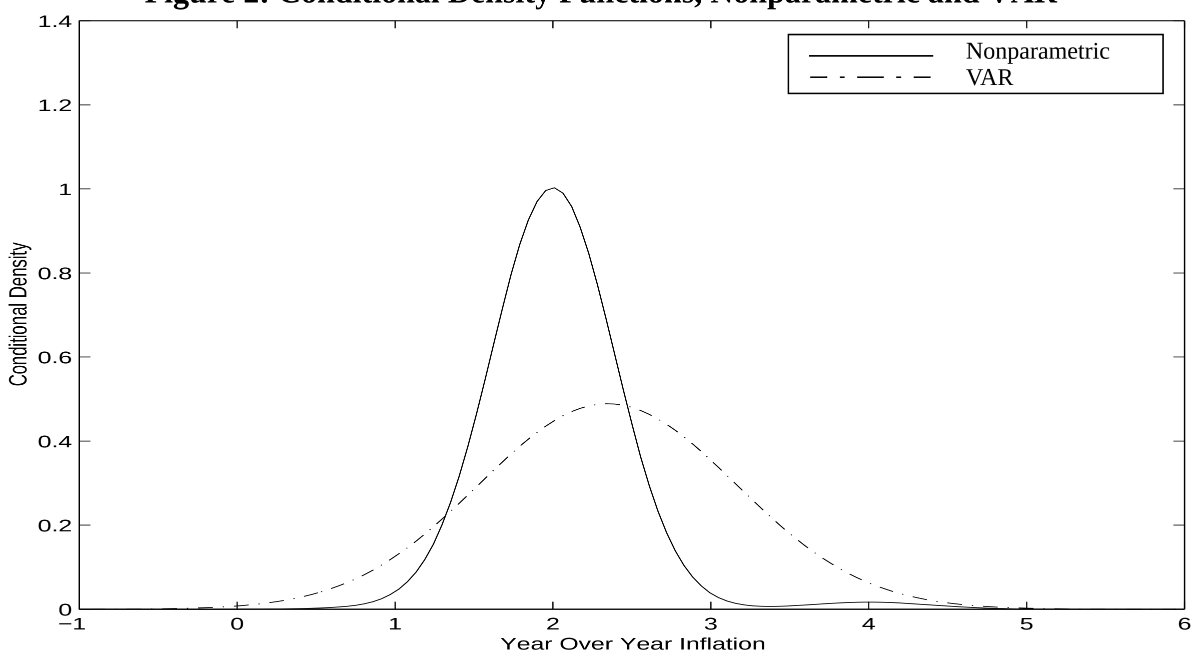


Figure 2: Conditional Density Functions, Nonparametric and VAR



Appendix 1: Assumptions

The following assumptions are used to derive the limiting distribution of J_n and J_n^*.

Assumption 1. The kernel function $k(\cdot)$ is a bounded, symmetric about 0, of order r, twice continuously differentiable function on R, and satisfies $\int_{-\infty}^{\infty} k(x)dx = 1$. $r > max\{(p+q)/4, (q-p)/2\}$.

Assumption 2. The smoothing parameters are chosen as $h \equiv h_n = cn^{-1/\alpha}$, where c is a positive constant. Further, $p + q < \alpha < (p+q)/2 + 2r$.

Assumption 3. The sequence of observations, $\{Z_t\}_{t=1}^{n} = \{x_t\ y_t\}_{t=1}^{n}$, is absolutely regular with coefficient $\beta_l = O(\lambda^l)$ for some fixed $0 < \lambda < 1$.

Assumption 4. The parameter space $\Theta \subset R^d$ is compact. The joint density function $\pi(x, y)$ and the marginal density function $\pi(x)$ of x_t are $r+1$ times continuously differentiable and uniformly bounded. The conditional density function $\pi_0(x|y, \theta)$ and its derivatives with respect to x, y, and θ are uniformly bounded and uniformly continuous on $R^p \times R^q \times \Theta$.

Assumption 5. Let $\hat{\theta}_n$ be an estimator of θ_0. There exists $\theta^* \in \Theta$ such that $\hat{\theta}_n \to \theta^*$ consistently, and $\hat{\theta}_n$ has a linear expansion of the form

$$\sqrt{n}(\hat{\theta}_n - \theta^*) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} A_0(\theta^*)\varphi(Z_i, \theta^*) + o_P(1), \tag{16}$$

where $A_0(\theta)$ is a random matrix, $\varphi(z, \theta)$ is a measurable function, and $E[\varphi(Z_t, \theta^*)|x_i] = 0$ for any $t \geq 1$. $\theta^* = \theta_0$ if the model is correctly specified.

Assumption 6. Let P^* denote the probability measure corresponding to the distribution of Z_i^* conditional on $\{Z_i\}_{i=1}^n$, and let $\hat{\theta}_n^*$ be the bootstrap estimator of $\hat{\theta}_0$. Then

$$\hat{\theta}_n^* - \hat{\theta}_0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} A_0(\hat{\theta}_0) \varphi(Z_i, \hat{\theta}_0) + o_{P*}(1),$$

where $\hat{\theta}_0$ is the estimator of θ_0 under the null hypothesis of correct specification.

From Assumption 1, there is no need to use a kernel of order greater than 2 unless $p + q > 7$ and $q \le 3p$, or $q > 3p$ and $q \ge p + 4$. Under Assumption 2, we have $nh^{(p+q)/2+2r} \to 0, nh^{p+q} \to \infty$. Assumption 3 is used to restrict the amount of dependence allowed in the observations such that asymptotic theory can be applied. It requires that the underlying process $\{Z_t\}$ be absolutely regular with a geometric decay rate. This is not a very restrictive assumption, because many well-known processes satisfy Assumption 3. For example, a continuous-time parametric diffusion process, $dx_t = \mu(x_t, \theta)dt + \sigma(x_t, \theta)dW_t$, which is widely used in theoretical financial models to represent the stochastic dynamics of asset prices, interest rates, macroeconomic factors, etc., satisfies this assumption provided the marginal density function $\pi(x_t)$, the drift function $\mu(x_t, \theta)$, and volatility $\sigma(x_t, \theta)$ satisfy $\lim_{x \to 0 \text{ or } x \to \infty} \sigma(x, \theta)\pi(x) = 0$ and $\lim_{x \to 0 \text{ or } x \to \infty} |\sigma(x, \theta)/(2\mu(x, \theta) - \sigma(x, \theta)\partial\sigma(x, \theta)/\partial x)| < \infty$. Masry and Tjostheim (1995) also established Assumption 3 for non-linear ARCH models. Assumptions 4 and 5 are required to ensure that under H_0, the effect of estimating $\pi_0(y|x, \theta_0)$ by $\pi_0(y|x, \hat{\theta}_n)$ on the asymptotic distribution of J_n is asymptotically negligible as the sample size increases to infinity. These two assumptions are also standard for ensuring the consistency and asymptotic normality of the quasi-maximum-likelihood estimator (see White 1982). In Assumption 5, in the case of the maximum-likelihood estimator, the function $\varphi(z, \theta)$ is the conditional score function $(\partial/\partial\theta)\log\pi_0(y|x, \theta)$ and the matrix $A_0(\theta) = [\int\int(\partial^2/\partial^2\theta)(\log\pi_0(y|x, \theta) \times \pi_0(y|x, \theta))dxdy]^{-1}$ is the inverse of the asymptotic information matrix. Finally, Assumption 6 ensures that the bootstrap estimator $\hat{\theta}_n{}^*$ has the same asymptotic properties as $\hat{\theta}_n$.

Appendix 2: Proofs

Proof of Theorem 1: Let

$$T_n(F, \theta_0) = \iint [(\hat{\pi}(y|x) - \pi_0(y|x,\theta_0))\hat{\pi}(x)]^2 dF$$

$$= n^{-2} \iint \left[\sum_{j=1}^{n} s_n^j(x, y) \right]^2 dF, \qquad (17)$$

where

$$s_n^j(x, y) = K_h(x - x_j) K_h(y - y_j) - \pi_0(y|x, \theta_0) K_h(x - x_j)$$

$$K_h(x - x_j) = h^{-q} K[(x - x_j)/h_n], K_h(y - y_j) = h^{-p} K[(y - y_j)/h_n],$$

denoting $\bar{s}_n^j(x, y) = s_n^j(x, y) - E(s_n^j(x, y))$.

We will prove (a) of Theorem 1 by showing that

(i) $\hat{\sigma}^{-1} n h^{(p+q)/2} [T_n(F, \theta_0) - \hat{\sigma}_1 n^{-1} h^{-(p+q)} + \hat{\sigma}_2 n^{-1} h^{-q}] \to N(0,1)$ in distribution,

(ii) $\hat{\sigma}^{-1} n h^{(p+q)/2} [T_n(F, \hat{\theta}_n) - \hat{\sigma}_1 n^{-1} h^{-(p+q)} + \hat{\sigma}_2 n^{-1} h^{-q}] \to N(0,1)$ in distribution, and

(iii) $\hat{\sigma}^{-1} n h^{(p+q)/2} [T_n(\hat{F}_n \hat{\theta}_n) - \hat{\sigma}_1 n^{-1} h^{-(p+q)} + \hat{\sigma}_2 n^{-1} h^{-q}] \to N(0,1)$ in distribution.

Proof of (i): $\hat{\sigma}^{-1} n h^{(p+q)/2} [T_n(F, \theta_0) - \hat{\sigma}_1 n^{-1} h^{-(p+q)} + \hat{\sigma}_2 n^{-1} h^{-q}] \to N(0,1)$

We decompose $T_n(F, \theta_0)$ into the following four terms,

$$T_n(F, \theta_0) = n^{-2} \sum_{j,k=1}^{n} \iint (s_n^j(x, y) s_n^k(x, y)) dF$$

$$= n^{-2} 2 \sum_{1 \le j < k \le n} \iint \bar{s}_n^j(x, y) \bar{s}_n^k(x, y) dF \quad + n^{-2} \sum_{j=1}^{n} \iint (s_n^j(x, y))^2 dF$$

$$+ 2(n^{-1} - n^{-2}) \sum_{j=1}^{n} \iint \bar{s}_n^j(x, y) E(s_n^1(x, y)) dF$$

$$+ (1 - n^{-1})\int (E(s_n^1(x, y)))^2 dF$$

$$= D_{n1}(\theta_0) + D_{n2}(\theta_0) + D_{n3}(\theta_0) + D_{n4}(\theta_0) \ . \qquad (18)$$

Noting that, under H_0, by changing variables we have

$$E(s_n^1(x, y)) = E[K_h(x - x_1)K_h(y - y_1)] - \pi(y|x)E[K_h(x - x_1)]$$
$$= \int\int K(u)K(v)\pi(x - hu, y - hv)dudv - \pi(y|x)\int K(s)\pi(x - hs)ds$$

$= O(h^r)$ uniformly in (x, y) by Assumption 1 and Assumption 4.

First, we show $nh^{(p+q)/2}[D_{n2}(\theta_0) - \sigma_1 n^{-1}h^{-(p+q)} + \sigma_2 n^{-1}h^{-q}] = o_p(1)$. For this purpose, we evaluate $E(D_{n2}(\theta_0))$ and $Var(D_{n2}(\theta_0))$ separately. By using Assumption 1 and Assumption 4, we have

$$E(nD_{n2}(\theta_0)) = \int\int E[K_h^2(x - x_1)K_h^2(y - y_1)]dF$$

$$-2\int\int \pi(y|x)E[K_h^2(x - x_1)K_h(y - y_1)]dF$$

$$+ \int\int E[\pi^2(y|x)K_h^2(x - x_1)]dF$$

$$= h^{-(p+q)}\int\int K^2(u, v)dudv\int\int \pi(x, y)dF(1 + O(h^{2r}))$$

$$-2h^{-q}\int K^2(x)dx\int\int \pi(y|x)\pi(x, y)dF(1 + O(h^r))$$

$$+ h^{-q}\int K^2(x)dx\int\int \pi^2(y|x)\pi(x)dF(1 + O(h^{2r})) \qquad (19)$$

$$Var(D_{n2}(\theta_0)) = var\left[n^{-2}\sum_{j=1}^{n}\int\int (s_n^j(x, y))^2 dF\right]$$

$$= n^{-4}\sum_{j=1}^{n} var[\int\int (s_n^j(x, y))^2 dF] +$$

$$(2n^{-4})\sum_{1\le j<k\le n} E\Big\{\Big[\iint (s_n^j(x,y))^2 dF_1 - \iint E(s_n^j(x,y))^2 dF\Big]$$
$$\times \Big[\iint (s_n^k(x,y))^2 dF - \iint E(s_n^k(x,y))^2 dF\Big]\Big\}. \tag{20}$$

The first term on the right-hand side of equation (20) equals

$$\begin{aligned} & n^{-3}var\Big(\Big[\iint (s_n^1(x,y))^2 dF\Big]\Big) \\ &= n^{-3}\Big\{E\Big[\iint (s_n^1(x,y))^2 dF\Big]^2 - \Big[E\Big[\iint (s_n^1(x,y))^2 dF\Big]\Big]^2\Big\} \\ &= O(n^{-3}h^{-2(p+q)}), \end{aligned} \tag{21}$$

and for any $0<\delta<1/3$ the absolute value of the second term on the right-hand side of equation (20) equals

$$\begin{aligned} & 2n^{-4}\sum_{1\le j<k\le n} E\Big|\Big\{\iint [(s_n^j(x,y))^2 - E[(s_n^j(x,y))^2]]dF(x,y) \\ & \qquad \times \iint [(s_n^k(\bar{x},\bar{y}))^2 - E[(s_n^k(\bar{x},\bar{y}))^2]]dF(\bar{x},\bar{y})\Big\}\Big| \\ & \le \sum_{k=1}^{n} \beta_k^{\delta/(1+\delta)} \times O(h^{-2(p+q)}n^{-3}) \\ & = O(h^{-2(p+q)}n^{-3}), \end{aligned} \tag{22}$$

by Lemma 1 (the inequality of the function of a stationary, absolutely regular process) in Yoshihara (1976). Therefore, by (21), (22), and Assumption 2,

$$\begin{aligned} Var(D_{n2}(\theta_0)) &= O(n^{-3}h^{-2(p+q)}) \\ &= o(1). \end{aligned}$$

It follows immediately by Chebyshev's inequality that

$$nh^{(p+q)/2}[D_{n2}(\theta_0) - E(D_{n2}(\theta_0))]$$

$$= nh^{(p+q)/2}[D_{n2}(\theta_0) - \sigma_1 n^{-1}h^{-(p+q)} + \sigma_2 n^{-1}h^{-q}] = o_p(1). \tag{23}$$

Now, we consider $D_{n3}(\theta_0)$ and $D_{n4}(\theta_0)$. First, because $E(D_{n3}(\theta_0)) = 0$, it is sufficient to show that $E(nh^{(p+q)/2}D_{n3}(\theta_0))^2 = o(1)$ by Chebyshev's inequality. For this purpose, we have

$$E(D_{n3}(\theta_0))^2 =$$
$$4(n^{-1} - n^{-2})^2 \sum_{1 \le j, k \le n} \iiiint E(\bar{s}_n^j(x, y)\bar{s}_n^k(\bar{x}, \bar{y}))E(s_n^1(x, y))E(s_n^1(\bar{x}, \bar{y}))dF_1(x, y)dF(\bar{x}, \bar{y})$$
$$= 4(n^{-1} - n^{-2})^2 \sum_{j=1}^{n} \iiiint E[\bar{s}_n^j(x, y)(\bar{s}_n^j(\bar{x}, \bar{y}))]E(s_n^1(x, y))E(s_n^1(\bar{x}, \bar{y}))dF(x, y)dF(\bar{x}, \bar{y})$$
$$= O(n^{-1}h^{2r}) + n^{-2} \sum_{1 \le j < k \le n} \beta_{k-j}^{\delta/(1+\delta)} O(h^{2r}), \tag{24}$$

where, for any $0 < \delta < 1/3$, the last equality above is obtained by using Lemma 1 in Yoshihara (1976) and the facts that $E(s_n^1(x, y)) = O(h^r), E(s_n^1(\bar{x}, \bar{y})) = O(h^r)$ uniformly in (x, y) and $(\bar{x}, \bar{y})$, and $\iiiint E[\bar{s}_n^j(x, y)(\bar{s}_n^j(\bar{x}, \bar{y}))]dF(x, y)dF(\bar{x}, \bar{y}) = O(1)$. Hence, from (24) we have $E(D_{n3}(\theta_0))^2 = O(n^{-1}h^{2r})$. Now, by using Chebyshev's inequality, we have $nh^{(p+q)/2}D_{n3}(\theta_0) = o_p(1)$, and by $\iint [E(s_n^1(x, y))]^2 dF(x, y) = O(h^{2r})$ and Assumption 2, we have $nh^{(p+q)/2}D_{n4}(\theta_0) = o_p(1)$.

It remains to show that $D_{n1}(\theta_0)$ gives the asymptotic normal distribution. For this purpose, we use a central limit theorem for degenerate U-statistics from Fan and Li (1999), which is reproduced in Appendix 3.

Let $Z_i = (x_i, y_i)$, $Z_j = (x_j, y_j)$ and $H_n(Z_i, Z_j) = 2\iint \bar{s}_n^i(x, y)\bar{s}_n^j(x, y)dF(x, y)$. Hence $D_{n1}(\theta_0) = n^{-2} \sum_{1 \le j < k \le n} H_n(Z_i, Z_j) = n^{-2}U_n$, and $E[H_n(Z_i, Z_j) | Z_i = z_i] = 0$ under H_0.

We now verify that assumptions (A1), (A2), and (A3) in Fan and Li (1999) are satisfied under Assumptions 1–5. Let $\bar{r} = [n^{1/4}]$ and $m = [C \log n]$, where C is a positive constant:

$$\sigma_n^2 = E[H_n(Z_1, Z_2)^2] = 2\iiiint E[\bar{s}_n^1(x, y)\bar{s}_n^2(x, y)\bar{s}_n^1(\bar{x}, \bar{y})\bar{s}_n^2(\bar{x}, \bar{y})]dF_1(x, y)dF(\bar{x}, \bar{y})$$

$$= 2\iint\{\iiiint\bar{s}_n^1(x, y)\bar{s}_n^2(x, y)\bar{s}_n^1(\bar{x}, \bar{y})\bar{s}_n^2(\bar{x}, \bar{y})F(x_1, y_1, x_2, y_2)dx_1dy_1dx_2dy_2\}$$
$$\times dF(x, y)dF(\bar{x}, \bar{y})$$

$$= \sigma_{n1}^2 + o(\sigma_{n1}^2) , \qquad (25)$$

where

$$\sigma_{n1}^2 = 2\iint\{\iiiint K_h(x-x_1)K_h(y-y_1)K_h(x-x_2)K_h(y-y_2)K_h(\bar{x}-x_1)K_h(\bar{y}-y_1)$$
$$K_h(\bar{x}-x_2)K_h(\bar{y}-y_2)F(x_1, y_1, x_2, y_2)dx_1dy_1dx_2dy_2\}dF(x, y)dF(\bar{x}, \bar{y}).$$

By changing variables, we have

$$\sigma_{n1}^2 = 2h^{-(p+q)}\iint\{\iint K(u_1, u_2)K(u_1 + w, u_2 + \bar{w})du_1du_2\}^2 dw d\bar{w} \times \iint \pi^4(x, y)dxdy. \qquad (26)$$

By (25) and (26), we have $\sigma_n^2 = h^{-(p+q)}\sigma^2 + o(h^{-(p+q)})$. Similarly, one can show that $E[H^2(Z_1, Z_{1+\tau})] = O(h^{-3(p+q)})$. Hence $\mu_{n2} = O(h^{-(p+q)})$. By definition,

$$\mu_{n4} \sim \iiiint\{\iint K_h(x-x_1)K_h(y-y_1)K_h(x-x_2)K_h(y-y_2)dF_1(x, y)\}^4 F(x_1, y_1, x_2, y_2)$$

$= O(h^{-3(p+q)})$, where we use the notation $a_n \sim b_n$ to denote that a_n and b_n have the same order of magnitude:

$$\tilde{\gamma}_{n22} = E\iiiiiiiint\bar{s}_n^1(x^1, y^1)\bar{s}_n^1(x^2, y^2)\bar{s}_n^1(x^3, y^3)\bar{s}_n^2(x^1, y^1)\bar{s}_n^1(x^2, y^2)\bar{s}_n^3(x^3, y^3)dF(x^1, y^1)$$

$$\times dF(x^2, y^2)dF(x^3, y^3)$$

$= O(h^{-2(p+q)})$. Now we consider $\tilde{\gamma}_{n14} = max_{s \neq t} \int \{E[H(z, Z_s)H(z, Z_t)]\}^2 dF(z)$. Let $z = (z_1, z_2)$, $Z_s = (Z_s^1, Z_s^2), Z_t = (Z_t^1, Z_t^2)$, then

$$\begin{aligned}\tilde{\gamma}_{n14} \sim \iint \{ \iiiint & K(u_1, v_1)K(u_2, v_2)K_h(u_1 + (z_1 - Z_s^1)/h, v_1 + (z_2 - Z_s^2)/h) \times \\ & K_h(u_2 + (z_1 - Z_t^1)/h, v_2 + (z_2 - Z_t^2)/h)F(z_1 + hu_1, z_2 + hv_1)F(z_1 + hu_2, z_2 + hv_2) \\ & du_1 du_2 dv_1 dv_2 \}^2 F(z_1, z_2) dz_1 dz_2 \\ = O(1) \cdot \end{aligned}$$

Similarly, one can show that $\gamma_{nij} = O(h^{-(p+q)(i+j-2)})$, for $(i, j) = (1, 1), (2, 2), (1, 3)$.

Summarizing the above, we have shown that $\sigma_n^2 = O(h^{-(p+q)})$,

$$\mu_{n2} = O(h^{-(p+q)}), \mu_{n4} = O(h^{-3(p+q)}), \tilde{\gamma}_{n22} = O(h^{-2(p+q)})$$

$$\tilde{\gamma}_{n14} = O(h^{-2(p+q)}), \gamma_{n11} = O(1), \gamma_{nij} = O(h^{-(p+q)(i+j-2)}), \quad \text{where}$$

$(i, j) = (1, 1), (2, 2), (1, 3)$, thus, under Assumptions 1–5, and $m = [C\log n], \bar{r} = [n^{1/4}]$, A1 (i)–(vii) in Fan and Li (1999) are all satisfied.

Next, let $x = (x_1, x_2), y = (y_1, y_2)$, by definition,

$$G(x, y) = O\left(h^{-(p+q)} \iiiint K(u, v)K\left(\frac{x_1 - y_1}{h_n} + u + w_1, \frac{(x_2 - y_2)}{h_n} + v + w_2\right) du dv dw_1 dw_2\right),$$

using a similar argument as above, it is straightforward that

$$\sigma_G^2 = E[G^2(Z_s, Z_s)] = O(h^{-2(p+q)}), \mu_{nG1} = O(1), \mu_{nG2} = O(h^{-(p+q)})$$

$$M_{nG1} = O(h^{-(p+q)\delta}), \gamma_{nG11} = O(1), M_{nG11} = O(h^{-2(p+q)(1+\delta)})$$

$$M_{nij} = O(h^{-2(p+q)(1+\delta)(i+j)+2(p+q)}), M_{n4} = O(h^{-3(p+q)(1+(3\delta)/4)}).$$

Therefore, (A2) (i)–(iv) are satisfied. To see that (A3) (i)–(vi) are satisfied, note that $\beta_m = O(\lambda^m) = O(\lambda^{-C\Gamma\log_\rho n}) = O((n^{-C\Gamma})$, where $\Gamma = \log\lambda > 0$. Hence, we have $m^2 n^2 \beta_m^{\delta/(1+\delta)} = 0(1)$ provided we choose $C > 4/\Gamma$. Hence, (A3) (i)–(vi) are satisfied.

It follows from Theorem 2.1 in Fan and Li (1999) that $\sqrt{2}[U_n - E(U_n)]/(n\sigma_n) \to N(0, 1)$ in distribution. By noting that $E(nh^{(p+q)/2}E(D_{n1}(\theta_0))) = O\left(h^{(p+q)(1/2-2\delta/(1+\delta))}\sum_{k=1}^{n}\beta_k^{\delta/(1+\delta)}\right)$ as long as $0 < \delta < 1/3$, and $\sigma_n = h^{-(p+q)/2}\sigma(1+o(1))$, we obtain that $nh^{(p+q)/2}D_{n1}(\theta_0) \to N(0, \sigma^2)$ in distribution.

Proof of (ii): $\hat{\sigma}^{-1}nh^{(p+q)/2}[T_n(F_1, \hat{\theta}_n) - \hat{\sigma}_1 n^{-1}h^{-(p+q)} + \hat{\sigma}_2 n^{-1}h^{-q}] \to N(0,1)$

Suppose the null hypothesis holds. Now we prove that the estimator $\hat{\theta}_n$ does not affect the limiting distribution of $T_n(F, \hat{\theta}_n)$. In fact, by Assumption 4,

$$T_n(F, \hat{\theta}_n) = \iint[(\hat{\pi}(y|x) - \pi_0(y|x,\hat{\theta}_n))\hat{\pi}(x)]^2 dF$$

$$= \iint[\hat{\pi}(x, y) - \pi_0(y|x,\hat{\theta}_n)\hat{\pi}(x)]^2 dF$$

$$= \iint(\hat{\pi}(x, y) - \pi(y|x)\hat{\pi}(x))^2 dF + \iint[(\pi_0(y|x, \hat{\theta}_n) - \pi_0(y|x, \theta_0))\hat{\pi}(x)]^2 dF$$

$$-2\iint(\hat{\pi}(x, y) - \pi_0(y|x, \theta_0)\hat{\pi}(x))(\pi_0(y|x, \hat{\theta}_n) - \pi_0(y|x, \theta_0))\hat{\pi}(x)dF$$

$$= T_n(F, \theta_0) - 2L_n(\theta_0, \hat{\theta}_n) + O_p(1/n),$$

where $L_n(\theta_0, \hat{\theta}_n) = \iint(\hat{\pi}(x, y) - \pi_0(y|x, \theta_0)\hat{\pi}(x))(\pi_0(y|x, \hat{\theta}_n) - \pi_0(y|x, \theta_0))\hat{\pi}(x)dF$, which can be written as

$$L_n(\theta_0, \hat{\theta}_n) = -2\iint(\hat{\pi}(x, y) - \pi(x, y))(\hat{\pi}(x) - \pi(x))(\pi_0(y|x, \hat{\theta}_n) - \pi_0(y|x, \theta_0))dF$$

$$-2\iint \pi_0(y|x,\theta_0)(\hat{\pi}(x)-\pi(x))^2(\pi_0(y|x,\hat{\theta}_n)-\pi_0(y|x,\theta_0))dF$$

$$+\,2\iint(\hat{\pi}(x,y)-\pi(x,y))(\pi_0(x|y,\hat{\theta}_n)-\pi_0(y|x,\theta_0))\pi(x)dF$$

$$-2\iint(\hat{\pi}(x)-\pi(x))(\pi_0(y|x,\hat{\theta}_n)-\pi_0(y|x,\theta_0))\pi_0(y|x,\theta_0)\pi(x)dF$$

$$= L_{n1}(\theta_0,\hat{\theta}_n)+L_{n2}(\theta_0,\hat{\theta}_n)+L_{n3}(\theta_0,\hat{\theta}_n)+L_{n4}(\theta_0,\hat{\theta}_n). \qquad (27)$$

By following the proof of Theorem 3.2 of Fan (1994) and using the central limit theorem for degenerate U-statistics in Fan and Li (1999), one can prove that $\iint(\hat{\pi}(x)-\pi(x))^2 dF = O_p(1/(nh^q))$ and $\iint(\hat{\pi}(x,y)-\pi(x,y))^2 dF = O_p(1/(nh^{p+q}))$. Therefore, the first term and second term on the right-hand side of equation (27) satisfy $nh^{(p+q)/2}L_{n1}(\theta_0,\hat{\theta}_n) = o_p(1)$.

$nh^{(p+q)/2}L_{n2}(\theta_0,\hat{\theta}_n) = o_p(1)$. $L_{n3}(\theta_0,\hat{\theta}_n)$ and $L_{n4}(\theta_0,\hat{\theta}_n)$ can be decomposed as $L_{n31}(\theta_0,\hat{\theta}_n)+L_{n32}(\theta_0,\hat{\theta}_n)$ and $L_{n41}(\theta_0,\hat{\theta}_n)+\ I_{n42}(\theta_0,\hat{\theta}_n)$, respectively, where

$$L_{n31}(\theta_0,\hat{\theta}_n) = 2\iint(\hat{\pi}(x,y)-E\hat{\pi}(x,y))(\pi_0(y|x,\hat{\theta}_n)-\pi_0(y|x,\theta_0))\pi(x)dF$$

$$L_{n32}(\theta_0,\hat{\theta}_n) = 2\iint(E\hat{\pi}(x,y)-\pi(x,y))(\pi_0(y|x,\hat{\theta}_n)-\pi_0(y|x,\theta_0))\pi(x)dF$$

$$L_{n41}(\theta_0,\hat{\theta}_n) = -2\iint(\hat{\pi}(x)-E\hat{\pi}(x))(\pi_0(y|x,\hat{\theta}_n)-\pi_0(y|x,\theta_0))\pi(x)dF$$

$$L_{n42}(\theta_0,\hat{\theta}_n) = 2\iint(E\hat{\pi}(x)-\pi(x))(\pi_0(y|x,\hat{\theta}_n)-\pi_0(y|x,\theta_0))\pi(x)dF\,.$$

First, we consider the third term, $L_{n3}(\theta_0,\hat{\theta}_n)$. By using Assumption 5, we have

$$L_{n31}(\theta_0,\hat{\theta}_n) = (n^2h^{(p+q)})^{-1}\sum_{i,j=1}^{n}\iint(K(x_i-x,y_i-y)-E[K(x_i-x,y_i-y)]) \times(\partial\pi_0(y|x,\theta)/\partial\theta)\big|_{\theta=\theta_0}A_0(\theta_0)\varphi(Z_i,\theta_0)dxdy+o_p(\Pi_n),$$

where

$$\Pi_n = (n^2h^{(p+q)})^{-1}\sum_{i,j=1}^{n}\iint(K(x_i-x,y_i-y)-E[K(x_i-x,y_i-y)])$$
$$\times(\partial\pi_0(y|x,\theta)/\partial\theta)|_{\theta=\theta_0}A_0(\theta_0)\varphi(Z_i,\theta_0)dxdy.$$

We show that $nh^{(p+q)/2}(\Pi_n) = o_p(1)$.

$$\Pi_n = (n^2h^{(p+q)})^{-1}\sum_{i\neq j}\iint(K(x_i-x,y_i-y)-E[K(x_i-x,y_i-y)])$$
$$\times(\partial\pi(y|x,\theta)/\partial\theta)|_{\theta=\theta_0}A_0(\theta_0)\varphi(Z_j,\theta_0)dxdy$$
$$+(n^2h^{(p+q)})^{-1}\sum_{i=1}^{n}\iint(K(x_i-x,y_i-y)-E[K(x_i-x,y_i-y)])$$
$$\times(\partial\pi(y|x,\theta)/\partial\theta)|_{\theta=\theta_0}A_0(\theta_0)\varphi(Z_i,\theta_0)dxdy$$
$$= \Pi_{n1}+\Pi_{n2}.$$

Because Π_{n1} is a degenerate U-statistic, it is straightforward to show that Π_{n1} = $O_P(n^{-1+\varepsilon/2})$ by Proposition 2 in Denker and Keller (1983), where ε is any small positive number. We can take it that ε is small enough such that $nh^{(p+q)/2}\Pi_{n1} = o_p(1)$. Using a similar argument as in Lemma 2 in Hall (1984), we have $nh^{(p+q)/2}\Pi_{n2} = o_p(1)$.

Hence, $nh^{(p+q)/2}L_{n3i}(\theta_0,\hat\theta_n) = o_p(1)$, i=1, 2. Similarly, $nh^{(p+q)/2}L_{n3i}(\theta_0,\hat\theta_n) = o_p(1)$, i=1, 2. This finishes the proof of (ii).

Proof of (iii): $\hat\sigma^{-1}nh^{(p+q)/2}[T_n(\hat F\ ,\hat\theta_n)-\hat\sigma_1 n^{-1}h^{-(p+q)}+\hat\sigma_2 n^{-1}h^{-q}]\rightarrow N(0,1)$

We can express $T_n(\hat{F_n},\hat\theta_n) = T_n(F,\theta_0)+\Gamma_n(\theta_0,\hat\theta_n)$, where,

$$\Gamma_n(\theta_0,\hat\theta_n) = T_n(\hat F,\hat\theta_n)-T_n(F,\theta_0)$$
$$= n^{-2}\sum_{i,j=1}^{n}\iint(s_n^i(x,y)s_n^j(x,y))d(\hat F_n-F)$$
$$= n^{-3}\sum_{i,j,k=1}^{n}\{(s_n^i(x_k,y_k)s_n^j(x_k,y_k))-\iint(s_n^i(x,y)s_n^j(x,y))dF\}$$
$$= \Gamma_{n1}(\theta_0,\hat\theta_n)+\Gamma_{n2}(\theta_0,\hat\theta_n)+\Gamma_{n3}(\theta_0,\hat\theta_n)+\Gamma_{n4}(\theta_0,\hat\theta_n),$$

where

$$\Gamma_{n1}(\theta_0, \hat{\theta}_n) = n^{-3} \sum_{k \neq i, j}^{n} (\{s_n^i(x_i, y_i)s_n^j(x_k, y_k)\} - E[s_n^i(x_k, y_k)s_n^j(x_k, y_k) | (x_i, x_j)]\})$$

$$\Gamma_{n2}(\theta_0, \hat{\theta}_n) = 2n^{-3} \sum_{i \neq j}^{n} \{s_n^i(x_i, y_i)s_n^j(x_j, y_j)\}$$

$$\Gamma_{n3}(\theta_0, \hat{\theta}_n) = n^{-3} \sum_{j=1}^{n} (s_n^j(x_j, y_j))^2$$

$$\Gamma_{n4}(\theta_0, \hat{\theta}_n) = -n^{-3} \sum_{i, j}^{n} \iint s_n^i(x, y)s_n^j(x, y)dF.$$

By $E[(s_n^1(x, y))^2] = O(h^{-(p+q)})$ and $\Gamma_{n3}(\theta_0, \hat{\theta}_n) \geq 0)$, we have $\Gamma_{n3}(\theta_0, \hat{\theta}_n) = o_p(n^{-1}h^{-(p+q)})$. $\Gamma_{n4}(\theta_0, \hat{\theta}_n) = -n^{-1}T_n(F, \theta_0) = o_p(n^{-1}h^{-(p+q)/2})$. By straightforward computation, one can show that both $\Gamma_{n1}(\theta_0, \hat{\theta}_n)$ and $\Gamma_{n2}(\theta_0, \hat{\theta}_n)$ are $o_p(n^{-1}h^{-(p+q)/2})$. Hence,

$$\begin{aligned} &\hat{\sigma}^{-1}nh^{(p+q)/2}[T_n(\hat{F} \;, \hat{\theta}_n) - \hat{\sigma}_1 n^{-1}h^{-(p+q)} + \hat{\sigma}_2 n^{-1}h^{-q}] \\ &= \hat{\sigma}^{-1}nh^{(p+q)/2}[T_n(F, \theta_0) - \hat{\sigma}_1 n^{-1}h^{-(p+q)} + \hat{\sigma}_2 n^{-1}h^{-q}] + o_p(1) \\ &\rightarrow N(0, 1), \text{ in distribution.} \end{aligned}$$

Since the proof of (b) is almost the same as that of (a), we provide only a sketch of it. It is easy to prove that $T_n(\hat{F}_n, \hat{\theta}_n) \rightarrow \iint [(\pi(y|x) - \pi_0(y|x, \theta_0))(\pi(x))]^2 dF$ in probability. Under H_1, $\iint [(\pi(y|x) - \pi_0(y|x, \theta_0))\pi(x)]^2 dF > 0$. Therefore, we have that

$$P[nh^{(p+q)/2}T_n(\hat{F}_n, \hat{\theta}_n) > B_n] \rightarrow 1 \text{ in probability.}$$

Note that under Assumptions 3 and 4 the distribution of the (x_i, y_i) is absolutely continuous. For any (i, j) there exists a positive constant $\bar{C}$ such that

$$\left|[\hat{\pi}(x_i, y_j)]^3 - [\pi(x_i, y_j)]^3\right| \leq \bar{C} \times sup_{(x, y) \in R^{p+q}} |\hat{\pi}(x, y) - \pi(x, y)| = o_p(1) ,$$

by Assumptions 1–4 and inequality (3.9) in Bierens (1983). Hence,

$$\frac{1}{n}\sum_{i=1}^{n}[\hat{\pi}(x_i, y_j)]^3 = \frac{1}{n}\sum_{i=1}^{n}[\pi(x_i, y_j)]^3 + o_p(1) = \iint[\pi(x, y)]^4 dxdy + o_p(1),$$

where the last equality is obtained by the Law of Large Numbers. Therefore, $\hat{\sigma}^2 = \sigma^2 + o_p(1)$. The same argument will also yield $\hat{\sigma}_i = \sigma_i + o_p(1), \quad i = 1, 2$.

Proof of Theorem 2:

Before we proceed, it is important to note that Step 2 ensures that, in the bootstrap world, H_0 always holds with $\theta = \hat{\theta}_0$. Since the detailed proof is similar to the proof of Theorem 1, we give only a sketch of the proof here. It is easy to see that $T_n{*}(F, \hat{\theta}_0) = D_{n1}{*}(\hat{\theta}_0) + D_{n2}{*}(\hat{\theta}_0) + D_{n3}{*}(\hat{\theta}_0) + D_{n4}{*}(\hat{\theta}_0)$.

Let $P*$ denote the probability measure relative to the distribution of the bootstrap sample conditional on the original sample, and $\hat{\theta}^*$ is the maximum-likelihood estimator of $\hat{\theta}_0$. Then $\hat{\theta}^* - \hat{\theta}_0 = n^{-1}\sum_{t=1}^{n} A_0(\hat{\theta}_0) \times \varphi(Z_t{*}, \hat{\theta}_0) + o_{P*}(1)$. Applying the same argument as in the proof of Theorem 1, it can be shown that $[D_{n2}{*}(\hat{\theta}_0) - \hat{\sigma}_1{*}n^{-1}h^{-(p+q)} - \sigma_2{*}n^{-1}h^{-q}] = o_{P*}((nh^{(p+q)/2})^{-1})$

$D_{n3}(\hat{\theta}_0) = o_{P*}((nh^{(p+q)/2})^{-1})$, and $D_{n4}(\hat{\theta}_0) = o_{P*}((nh^{(p+q)/2})^{-1})$. Now, by a similar proof to that of $nh^{(p+q)/2}D_{n1}(\theta_0) \to N[0, \sigma^2]$, we have under condition on X_n, $nh^{(p+q)/2}D_{n1}{*}(\hat{\theta}_0) \to N[0, \sigma^2]$. Also, one can show that $L_{n1}(\hat{\theta}_0, \hat{\theta}_n{*}), L_{n2}(\hat{\theta}_0, \hat{\theta}_n{*}),$

$L_{n3}(\hat{\theta}_0, \hat{\theta}_n{*}), L_{n4}(\hat{\theta}_0, \hat{\theta}_n{*}), \Gamma_{n1}{*}(\hat{\theta}_0, \hat{\theta}_n{*}), \Gamma_{n2}{*}(\hat{\theta}_0, \hat{\theta}_n{*}), \Gamma_{n3}{*}((\hat{\theta}_0, \hat{\theta}_n)^*)$, and $\Gamma_{n4}{*}(\hat{\theta}_0, \hat{\theta}_n{*})$

are $o_{p*}((nh^{(p+q)/2})^{-1})$. Taken together, conditional on X_n, the distribution of $J_n{*}$ converges to the standard normal in probability. Hence, the bootstrap critical value will approach the asymptotic critical value of J_n, and so the bootstrap test $J_n{*}$ is as consistent as the test J_n.

Appendix 3: Central Limit Theorem For Degenerate U-Statistics[3]

Let $\{Z_t\}$ be a strictly stationary, absolutely regular stochastic process with coefficient β_τ. Let $U_n = \sum\sum_{1 \le s < t \le n} H_n(Z_t, Z_s)$, where H_n depends on n and satisfies $\int H_n(x, y)dF(x) = 0$ for all y, and $F(\cdot)$ is the marginal distribution function of $\{Z_t\}$. Let $F_{i_1, \ldots, i_j}$ denote the joint distribution function of $\{Z_{i_1}, \ldots, Z_{i_j}\}$, $j = 2, 3, 4$; $dQ_{i_1, i_2}(z_{i_1}, z_{i_2})$ denote either $dF_{i_1, i_2}(z_{i_1}, z_{i_2})$ or $dF(z_{i_1})dF(z_{i_2})$; and $dQ_{i_1, i_2, i_3}(z_{i_1}, z_{i_2}, z_{i_3})$ be either $dF_{i_1, i_2, i_3}(z_{i_1}, z_{i_2}, z_{i_3})$ or $dF(z_{j_1})dF_{j_2, j_3}(z_{j_2}, z_{j_3})$, where $\{j_1, j_2, j_3\}$ is any possible permutation of $\{i_1, i_2, i_3\}$. Similarly, $dQ_{i_1, i_2, i_3, i_4}(z_{i_1}, z_{i_2}, z_{i_3}, z_{i_4})$ denote either $dF_{i_1, i_2, i_3, i_4}(z_{i_1}, z_{i_2}, z_{i_3}, z_{i_4})$, $dF_{i_1, i_3}(z_{i_1}, z_{i_3})dF_{i_2, i_4}(z_{i_2}, z_{i_4})$, or $dF(z_{j_1})dF_{j_2, j_3, j_4}(z_{i_2}, z_{i_3}, z_{i_4})$, where $\{j_1, j_2, j_3, j_4\}$ is any permutation of $\{i_1, i_2, i_3, i_4\}$. Also, let $\{\tilde{Z}_t\}_{t=1}^{n}$ be an i.i.d. sequence having the same marginal distribution as $\{Z_t\}$ and define $G(x, y) = E[H(Z_1, x)H(Z_1, y)]$; $\sigma_n^2 = \iint H^2(z_1, z_2)dF(z_1)dF(z_2) \equiv E[H^2(\tilde{Z}_1, \tilde{Z}_2)]$;

$$M_{n4} = max\left\{max_{s \ne t} E\left|H(Z_t, Z_s)\right|^{4(1+\delta)}, E\left|H(\tilde{Z}_2, \tilde{Z}_1)\right|^{4(1+\bar{\delta})}\right\};$$

$$M_{nij} = max\left\{max_{s \ne t, s' \ne t'} \int \left|H^i(z_t, z_s)H^j(z_{t'}, z_{s'})\right|^{1+\delta} dQ_{t, s, t', s'}(z_t, z_s, z_{t'}, z_{s'}),\right.$$
$$\left. max_{t \ne s, s'} \int \left|H^i(z_t, z_s)H^j(z_t, z_{s'})\right|^{1+\delta} dQ_{t, s, s'}(z_t, z_s, z_{s'}) \right\};$$

$(i, j) = (1, 0), (1, 1), (2, 2)$ or $(1, 3)$;

$$M_{nG11} = max\{max_{s \ne s'} \int \left|G(z_s, z_s)G(z_{s'}, z_{s'})\right|^{1+\delta} dQ_{s, s'}(z_s, z_{s'})$$
$$max_{s \ne s' \ne s''} \int \left|G(z_s, z_s)G(z_{s'}, z_{s''})\right|^{1+\delta} dQ_{s, s', s''}(z_s, z_{s'}, z_{s''})$$
$$max_{s \ne s' \ne s'' \ne s'''} \int \left|G(z_s, z_{s'})G(z_{s''}, z_{s'''})\right|^{1+\delta} dQ_{s, s', s'', s'''}(z_s, z_{s'}, z_{s''}, z_{s'''})\};$$

$$\mu_{n2} = max\{E[H^2(\tilde{Z}_1, \tilde{Z}_2)], max_{s \ne t} E[H^2(Z_t, Z_s)]\};\ \mu_{n4} = E[H^4(\tilde{Z}_1, \tilde{Z}_2)];$$

---

3. This appendix is adapted from Fan and Li (1999).

$$\tilde{\gamma}_{n14} = max_{s_1 \neq s_2} \int \{E[H(z, Z_{s_1})H(z, Z_{s_2})]\}^2 dF(z),$$

$$\tilde{\gamma}_{n22} = E[H^2(\tilde{Z}_1, \tilde{Z}_2)(H^2(\tilde{Z}_1, \tilde{Z}_3))],$$

$$\gamma_{nij} = max\{maxE[H^i(Z_t, Z_s)H^j(Z_{t'}, Z_{s'})], max\int E[H^i(Z_t, z_s)H^j(Z_{t'}, z_{s'})]dF_{s, s'}(z_s, z_{s'})\},$$

for $(i, j) = (1, 1), (2, 2)$ or $(1, 3)$, and the two maximums inside the curly brackets are taken for all $1 \le s \neq t \le n, 1 \le s' \neq t' \le n, s \neq s'$ or $t \neq t'$. If $t = t'$, $dF_{t, t}(z_t, z_t)$ means $dF(z_t)$,

$$\sigma_G^2 = E[G^2(Z_s, Z_s)]; \qquad \mu_{nG1} = max_{s \neq s'}|EG(Z_s, Z_{s'})|;$$

$$\mu_{nG2} = max\{max_{s \neq s'} EG^2(Z_s, Z_{s'}), EG^2(\tilde{Z}_1, \tilde{Z}_2)\};$$

$$M_{nG1} = max\{max_{s \neq s'} E(|G(Z_s, Z_{s'})|^{1+\delta}, E|G(\tilde{Z}_1, \tilde{Z}_{2'})|^{1+\delta})\};$$

$$\begin{aligned}\gamma_{nG11} = max\{&max_{s \neq s' \neq s''}|E[G(Z_s, Z_s)G(Z_{s'}, Z_{s''})]|, \\ &max_{s \neq s' \neq s''}|E[G(Z_s, Z_{s'})G(Z_s, Z_{s''})]|, \\ &max_{s \neq s' \neq s'' \neq s'''}|E[G(Z_s, Z_{s'})\dot{G}(Z_{s''}, Z_{s'''})]|\}.\end{aligned}$$

Theorem: Let $\bar{r} = r_n = [n^{1/4}], m = m_n = o(\bar{r}), k = k_n = [n/(r+m)]$. Consider the following assumptions:

A1. (i) $m^2\gamma_{n11}/\sigma_n^2 = o(1)$, (ii) $rm^3\gamma_{n13}/(n^2\sigma_n^4) = o(1)$, (iii) $rm^5\gamma_{n22}/(n^2\sigma_n^4) = o(1)$

(iv) $r\mu_{n2}/(n\sigma_n^2) = o(1)$, (v) $rm^2\mu_{n4}/(n^2\sigma_n^4) = o(1)$, (vi) $rm^3\tilde{r}_{n22}/(n\sigma_n^4) = o(1)$

(vii) $rm^4\tilde{\gamma}_{n14}/n\sigma_n^4 = o(1)$;

A2. (i) $m^2\mu_{nG2}/\sigma_n^4 = o(1)$, (ii) $m^4\gamma_{nG11}/\sigma_n^4 = o(1)$, (iii) $m^4\mu_{nG1}^2/\sigma_n^4 = o(1)$,

(iv) $m\sigma_G^2/(n\sigma_n^4) = o(1)$,

A3. (i) $nM_{n10}^{1/(1+\delta)}\beta_m^{\delta/(1+\delta)}/\sigma_n = o(1)$, (ii) $n^2M_{n11}^{1/(1+\delta)}\beta_m^{\delta/(1+\delta)}/\sigma_n^2 = o(1)$

(iii) $rm^2M_{n13}^{1/(1+\delta)}\beta_m^{\delta/(1+\delta)}/n\sigma_n^4 = o(1)$,

(iv) $r^3n(M_{n22} + M_{n4})^{1/(1+\delta)}\beta_m^{\delta/(1+\delta)}/\sigma_n^4 = o(1)$,

(v) $m^2n^2M_{nG11}{}^{1/(1+\delta)}\beta_m^{\delta/(1+\delta)}/\sigma_n^4 = o(1)$,

(vi) $m^2 n^2 {M_{nG1}}^{2/(1+\delta)} \beta_m^{2\delta/(1+\delta)} / \sigma_n^4 = o(1)$.

Let $\{Z_t\}$ be a strictly stationary, absolutely regular process, and assume that assumptions (A1) to (A3) hold. Then $\frac{\sqrt{2}U_n}{n\sigma_n} \to N(0, 1)$ in distribution.

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages. ***Les documents de travail sont publiés généralement dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue*.**

2001

2001-20	The Resolution of International Financial Crises: Private Finance and Public Funds	A. Haldane and M. Kruger
2001-19	Employment Effects of Restructuring in the Public Sector in North America	P. Fenton, I. Ip, and G. Wright
2001-18	Evaluating Factor Models: An Application to Forecasting Inflation in Canada	M.-A. Gosselin and G. Tkacz
2001-17	Why Do Central Banks Smooth Interest Rates?	G. Srour
2001-16	Implications of Uncertainty about Long-Run Inflation and the Price Level	G. Stuber
2001-15	Affine Term-Structure Models: Theory and Implementation	D.J. Bolder
2001-14	L'effet de la richesse sur la consommation aux États-Unis	Y. Desnoyers
2001-13	Predetermined Prices and the Persistent Effects of Money on Output	M.B. Devereux and J. Yetman
2001-12	Evaluating Linear and Non-Linear Time-Varying Forecast-Combination Methods	F. Li and G. Tkacz
2001-11	Gaining Credibility for Inflation Targets	J. Yetman
2001-10	The Future Prospects for National Financial Markets and Trading Centres	C. Gaa, S. Lumpkin, R. Ogrodnik, and P. Thurlow
2001-9	Testing for a Structural Break in the Volatility of Real GDP Growth in Canada	A. Debs
2001-8	How Rigid Are Nominal-Wage Rates?	A. Crawford
2001-7	Downward Nominal-Wage Rigidity: Micro Evidence from Tobit Models	A. Crawford and G. Wright
2001-6	The Zero Bound on Nominal Interest Rates: How Important Is It?	D. Amirault and B. O'Reilly
2001-5	Reactions of Canadian Interest Rates to Macroeconomic Announcements: Implications for Monetary Policy Transparency	T. Gravelle and R. Moessner
2001-4	On the Nature and the Stability of the Canadian Phillips Curve	M. Kichian

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à :

Publications Distribution, Bank of Canada
234 Wellington Street, Ottawa, Ontario K1A 0G9
E-mail: publications@bankofcanada.ca
Web site: http://www.bankofcanada.ca

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9
Adresse électronique : publications@banqueducanada.ca
Site Web : http://www.banqueducanada.ca