1984

Nonparametric Time Series Estimation of Joint DGP, Conditional DGP and Vector Autoregression

Radhey S. Singh

Aman Ullah

Follow this and additional works at: <https://ir.lib.uwo.ca/economicsresrpt>

Part of the [Economics Commons](#)

Citation of this paper:

Singh, Radhey S., Aman Ullah. "Nonparametric Time Series Estimation of Joint DGP, Conditional DGP and Vector Autoregression." Department of Economics Research Reports, 8423. London, ON: Department of Economics, University of Western Ontario (1984).

RESEARCH REPORT 8423
NONPARAMETRIC TIME SERIES ESTIMATION
OF JOINT DGP, CONDITIONAL DGP
AND VECTOR AUTOREGRESSION*

by

Radhey S. Singh and Aman Ullah

November, 1984

ABSTRACT

In this paper we develop nonparametric estimators of the joint time series data generating process (DGP) of (x_t, y_t) at different t -values, of conditional DGP, of the conditional mean of x_t given the past values of x and y , and, more generally, the conditional mean of (x_t, y_t) given their past values (vector autoregression). We establish, among other results, the central limit theorems for these estimators under far weaker mixing conditions than those used in Robinson [23], where only the x_t series is considered. Uniform consistency, and rate results for the consistencies of various estimators are also obtained. The results of the paper are useful in light of the fact that often the functional form of the dynamic regression is not known and also the assumption of the Gaussian process is not true.

*This paper was written when the first author visited the University of Western Ontario. The authors are thankful to NSERC for research support and to Virginia Ho for the research assistance. They are especially thankful to Peter Phillips and to a referee for giving them very constructive comments and suggestions which have led to a definite improvement over an earlier version of the paper.

1. INTRODUCTION

In time series econometrics and some other scientific disciplines one often analyzes the bivariate random processes of the type $\{x_t, t=0, \pm 1, \dots\}$ and $\{y_t, t=0, \pm 1, \dots\}$. A familiar study is that of the conditional mean (so-called dynamic regression function) of x_t given $\{x_{t-1}, \dots, x_{t-p}, y_t, \dots, y_{t-q}\}$ for some $p > 0$ and $q \geq 0$, or more generally the conditional mean of (x_t, y_t) given $\{x_{t-1}, \dots, x_{t-p}, y_{t-1}, \dots, y_{t-q}\}$ for some $p > 0$ and $q > 0$. The latter is usually referred to as a vector autoregression, when $p = q$.

For analyzing the behavior of x_t and y_t one almost invariably makes two assumptions. First, regarding the functional form of the autoregressive model and the other regarding the distribution of the disturbances (usually Gaussian). If $\{x_t\}$ and $\{y_t\}$ are in fact Gaussian processes, these assumptions may not be serious since a Gaussian process is uniquely characterized by its mean and autocovariances and such functionals can be asymptotically efficiently estimated, under some weak-dependence conditions on the process, by the corresponding functionals of sample mean and autocovariances. However, it is now well known that both the assumptions regarding the functional form and the disturbances are quite often questionable and their violations have varying effects on the econometric inference and policy implications.

The above problems can, however, be overcome if one can estimate the conditional mean (dynamic regression) nonparametrically. This can be achieved by estimating nonparametrically the joint data generating process (DGP) or density of all the variables involved.¹ For example, for any $p > p' \geq 0$ and $q \geq q' \geq 0$, we can estimate the regression of $(x_t, x_{t-1}, \dots, x_{t-p'}, y_t, y_{t-1}, \dots, y_{t-q'})$ on $(x_{t-p'-1}, \dots, x_{t-p}, y_{t-q'-1}, \dots, y_{t-q})$ nonparametrically

if we can estimate the joint conditional DGP of the former on the latter. This, of course, can be obtained if we can estimate the joint DGP of $(x_t, \dots, x_{t-p}, y_t, \dots, y_{t-q})$. Thus the whole problem reduces to the non-parametric estimation of the joint DGP of

$$\tilde{x}_t = (x_t, \dots, x_{t-p}) \text{ and } \tilde{y}_t = (y_t, \dots, y_{t-q})$$

for some $p \geq 0$ and $q \geq 0$. For some other applications of density estimates, see Singh [33] and Ullah and Singh [39].

With this motivation in view, our main thrust is to estimate non-parametrically the joint DGP of $(\tilde{x}_t, \tilde{y}_t)$ based on a single realization recorded (and available) of the pair $\{\tilde{x}_t, \tilde{y}_t\}$, $t=1, \dots, n$ for some $n > 0$.² Estimation of a multivariate DGP has been considered by Cacoullos [8], Epanechnikov [10], Singh [32, 35] and Singh and Ullah [36]; and of a regression function by Nadaraya [20], Watson [42], Schuster [30] and Singh and Ullah [36], among others, for i.i.d. processes. These works employ the Rosenblatt [24]-Parzen [21] type kernel density estimation technique. Roussas [28, 29], Rosenblatt [27] and Yakowitz [46] have considered estimation of a DGP or transition function for Markov processes. Some other dependent processes satisfying certain mixing conditions have been considered by Rosenblatt [25], and recently by Takahata [37, 38], Pham and Tran [22] and Robinson [23].

In this paper we consider estimation of the joint DGP of $(\tilde{x}_t, \tilde{y}_t)$, the conditional DGP (useful in testing for the independence) of a subset of components of $(\tilde{x}_t, \tilde{y}_t)$ given the rest and the conditional mean of a subset of the components of $(\tilde{x}_t, \tilde{y}_t)$ on the rest; and we establish, among other results, the central limit theorems (CLT) for these estimates under far weaker mixing conditions than those used in Robinson [23], where

only the $\{x_t\}$ process is considered. Uniform consistency, a feature not considered by Robinson, is also established for estimators of various functionals considered here. Rate results for pointwise and uniform consistencies of various estimators are also examined.

Thus our results are for the more general case of bivariate processes which can easily be extended for vector processes. Further we provide stronger results regarding the properties of estimators under weaker mixing conditions.

In Section 2 we state our basic assumptions on $\{(x_t, y_t)\}$ and f , the joint DGP of (x_t, y_t) . In Section 3 we present a class of Rosenblatt-Parzen type estimators for the multivariate DGP f and state various results on the bias, variance, pointwise and uniform consistency, and the distributions of the estimators of f . Similar results are established in Sections 4 and 5 respectively for estimators of the conditional DGP and autoregression functions considered in these sections. In Section 6 we present proofs of the theorems given in Section 3.

2. BASIC ASSUMPTIONS AND DEFINITIONS

Throughout the remainder of the paper we will assume that $\{x_t, t=0, \pm 1, \dots\}$ and $\{y_t, t=0, \pm 1, \dots\}$ are (jointly) strictly stationary real valued stochastic processes defined on a probability space.

We now define the strong mixing condition (Rosenblatt [25]) of a stationary stochastic process, say $\{v_t, t=0, \pm 1, \dots\}$ in a Euclidean space. Let $\mathcal{B}_t^j$ be the σ -field of events generated by stochastic vectors $v_t, v_{t+1}, \dots, v_{t+j}$. The process $\{v_t\}$ is said to be strongly mixing if α_j defined by

$$\alpha_j = \sup_{A \in \mathcal{B}_{-\infty}^t, B \in \mathcal{B}_{t+j}^\infty} |P(A \cap B) - P(A)P(B)|,$$

converges to zero as $j \rightarrow \infty$. This property is a joint property of $\{v_t\}$ and α_j . Notice that α_j is non-increasing in j with $1 \geq \alpha_1 \geq \alpha_2 \dots$. The

function α_j is called the mixing coefficient of $\{v_t\}$, and is a measure of the dependence of the processes $\{v_t\}_{-\infty < t < l}$ and $\{v_t\}_{l+j \leq t < \infty}$. In a strongly mixing process, the distant future is virtually independent of the past and the present (and vice versa). The m -dependent processes and Markov processes with finite state space (Billingsley [6], p. 167) are examples of strongly mixing processes. A Gaussian sequence with continuous spectral density bounded away from zero is also an example of strongly mixing processes (Ibragimov and Linnik [13], Th.17.3.1). A simple example of a strongly mixing process is an independent process with $\alpha_j \equiv 0$ for $j \geq 1$. Strongly mixing processes have been used in different contexts by various authors, including Rosenblatt [25], Hannan [11], Ibragimov [12] and Pham and Tran [22] to obtain asymptotic results for some statistics. Bierens [4] has considered kernel estimation of a regression function for a more restricted stationary process, namely, the ϕ -mixing process.³ Mixing conditions, mostly ϕ -mixing, have recently been used in econometrics, for example by Bierens [3, 5] and White and Domowitz [44], for proving consistency and asymptotic normality of least squares estimators of nonlinear time series regressions.

We assume that the joint process $\{(x_t, y_t)\}$ is strongly mixing with α_j being the mixing coefficient. Define

$$(2.1) \quad \theta_T = T^{-1} \sum_{j=1}^T (T-j) \alpha_j.$$

For some of our results (for example, for consistency of our estimators), we will make an assumption (specified in (2.2), see following (3.9)) weaker than

$$(2.2) \quad \theta_T = o(1) \quad \text{as } T \rightarrow \infty,$$

for some we assume (2.2) and for some others (for example, for asymptotic

normality of our estimators)

$$(2.2)^+ \quad \theta_T = o(1) \quad \text{as } T \rightarrow \infty$$

will be needed.⁴

Remark 2.1: Notice that (2.2) holds whenever $\sum_{j=1}^{\infty} \alpha_j < \infty$ since

$$\theta_T = \sum_{j=1}^T ((T-j)/T) \alpha_j \leq \sum_{j=1}^T \alpha_j \rightarrow \sum_{j=1}^{\infty} \alpha_j. \quad \text{But } \sum_{j=1}^{\infty} \alpha_j < \infty \text{ whenever } \sum_{j=1}^{\infty} \alpha_j^{1/2} < \infty, \text{ a}$$

condition invariably assumed for various CLT's involving mixing processes (see Sections 20, 21 of Billingsley [6], and Chapters 17, 18 of Ibragimov and Linnik [13]). Obviously (2.2) is also implied by (2.2)⁺. The assumption (2.2)⁺ (and hence also (2.2)) is weaker than " $\alpha_j = O(j^{-1-\epsilon})$ for some $\epsilon > 0$ " which, in turn, is far weaker than that made in (A3.1) of Robinson [23] for strongly mixing stationary processes. And, hence, as noted in Robinson's paper, it is satisfied by many Gaussian processes (Ibragimov [12]) and non-Gaussian autoregressive moving average processes (Pham and Tran [22]) and many linear and nonlinear functionals of such processes. We, however, use an assumption even weaker than (2.2) to prove pointwise as well as uniform consistency of our estimators of the joint DGP of (x_t, y_t) . See Remark 3.6 below.

Recall that the joint DGP of (x_t, y_t) is an $m = (p+q+2)$ -variate density function. Let $\tilde{w}$ be the point (of our interest) in $\mathbb{R}^m$ (the m -dimensional Euclidean space) at which we wish to estimate f . For some of our results we will use an assumption on f of the type:

(A)_r: For some integer $r \geq 1$, the r^{th} order partial derivatives of f are continuous and bounded in some neighborhood of $\tilde{w}$.

3. ESTIMATORS OF JOINT DGP, CONSISTENCY AND DISTRIBUTIONAL PROPERTIES

We give here a class of (Rosenblatt [24]-Parzen [21] type) estimators of the DGP f . Unless stated otherwise, the range of integration of functions on $\mathbb{R}^m$ will be on the whole space $\mathbb{R}^m$, and will not be indicated. As in Singh [35] let $\mathcal{K}_r$ be the class of all real-valued Borel-measurable bounded functions K on $\mathbb{R}^m$ such that

$$\int u_1^{i_1} \dots u_m^{i_m} K(u_1, \dots, u_m) = \begin{cases} 1 & \text{if } i_1 = \dots = i_m = 0 \\ 0 & \text{if } 0 < i_1 + \dots + i_m < r, \end{cases}$$

$$\int \|u\|^i |K(u)| < \infty \quad \text{for } i=0 \text{ and } r,$$

and $\|u\|^m |K(u)| \rightarrow 0$ as $\|u\| \rightarrow \infty$, where $\|u\|$ is the usual Euclidean norm on $\mathbb{R}^m$.

Remark 3.1: Notice that $\mathcal{K}_{r+1}$ is a subset of $\mathcal{K}_r$. For $r=0,1$ and 2 , the m -variate standard normal density is a member of $\mathcal{K}_r$, and so is the function $K(u_1, \dots, u_m) = 2^{-m} \prod_{j=1}^m I(-1 < u_j < 1)$ where $I(\cdot)$ stands for the indicator function. In fact, functions K of the type $K(u_1, \dots, u_m) = \prod_{j=1}^m K_j(u_j)$, where K_j are bounded symmetric (about zero) functions (not necessarily positive everywhere) on $\mathbb{R}^1$ with $\int K_j = 1$, $\int |K_j| < \infty$ and $u_t K_j(u_t) \rightarrow 0$ as $t \rightarrow \infty$ are members of $\mathcal{K}_r$ for $r=0,1$ and 2 . For $r=3$ and 4 , the function K given by $K(u_1, \dots, u_m) = \prod_{j=1}^m \gamma(u_j)$, where $\gamma(u) = (2\pi)^{-1/2} [2e^{-u^2/2} - 2^{-1/2} e^{-u^2/4}] I(-\infty < u < \infty)$ or $\gamma(u) = (2\pi)^{-1/2} \frac{1}{2}(3-u^2)e^{-u^2/2} I(-\infty < u < \infty)$ is a member of $\mathcal{K}_r$. For some other members of $\mathcal{K}_r$ for $r=0,1,\dots,6$, see Singh [35].

Let $0 < h = h_n$ be a function of n such that $h_n \rightarrow 0$ as $n \rightarrow \infty$. Let $w_t = (x_t, y_t)$. Suppose for some $r \geq 0$, (A_r) holds. For a fixed K in $\mathcal{K}_r$,

$$(3.1) \quad \delta_t(w) = h^{-m} K\left(\frac{w - w}{h}\right).$$

Then our proposed estimator of $f(\underline{w})$ based on $(\underline{x}_t, \underline{y}_t)$, $t=1, \dots, n$, is given by

$$(3.2) \quad f_n(\underline{w}) = n^{-1} \sum_{t=1}^n \delta_t(\underline{w}) .$$

Remark 3.2: Notice that for each fixed $r \geq 0$ and for each fixed K in $\mathcal{K}_r$ we have an estimate f_n of f of the type (3.2). For $r=0,1$ and 2 , if K in $\mathcal{K}_r$ is non-negative, then the estimate f_n of f will be non-negative. For $r > 2$, a K in $\mathcal{K}_r$ will be negative at some points in $\mathcal{R}^m$ and such K may give negative estimate f_n of f . However, as we will see, such estimates will give higher speed of consistency than those achieved with $r \leq 2$. In fact, we will see that the higher the value of r the better the speed of consistency of f_n .

We will now state our main results regarding f_n . All the convergence and rate results below are with respect to $n \rightarrow \infty$. In what follows, operators "E" and "var" stand for the expectation and variance respectively.

Theorem 3.1 (Asymptotic unbiasedness). At every continuity point $\underline{w}$ of f ,

$$(3.3) \quad B_n(\underline{w}) = (E f_n(\underline{w}) - f(\underline{w})) = o(1) .$$

And if f is uniformly continuous on $\mathcal{R}^m$, then

$$(3.3)' \quad \sup_{\underline{w}} |B_n(\underline{w})| = o(1),$$

where the supremum is taken over the whole space $\mathcal{R}^m$. Further, if for some $r \geq 1$, (A_r) holds and with this r , K in (3.1) belongs to $\mathcal{K}_r$, then

$$(3.4) \quad B_n(\underline{w}) = O(h^r) .$$

The result (3.4) holds uniformly on $\mathcal{R}^m$ if (A_r) holds uniformly in $\underline{w}$ on $\mathcal{R}^m$.

The proofs of all the theorems are given in Section 6 .

Remark 3.3: The results of Theorem 3.1 are valid for any joint stationary process $\{x_t, y_t\}$, regardless of the nature of their dependence.

Since $h \rightarrow 0$ as $n \rightarrow \infty$, (3.4) gives a speed of convergence for the asymptotic unbiasedness of f_n . This speed gets better and better as r gets large. There are, however, two points worth keeping in mind in this regard. First, as we have already seen, the construction of the kernel K belonging to $\mathcal{K}_r$ (and hence of the estimator f_n) becomes difficult as r gets large; second, as we will see, the choice of h converging to zero cannot be made arbitrarily small if we wish to control the $\text{var}(f_n)$.

We now introduce the notations

$$(3.5) \quad \eta_t(\tilde{w}) = \frac{h^{-m/2} \{K(\frac{\tilde{w}-w}{h}) - EK(\frac{\tilde{w}-w}{h})\}}{(f(w) \int K^2)^{1/2}}$$

and

$$(3.6) \quad \rho_s = E(\eta_t \eta_{t+s}) .$$

For the results in the remainder of this paper it is assumed that the joint process $\{(x_t, y_t)\}$ is strongly mixing.

Theorem 3.2 (Variance of f_n). For each fixed $n \geq 1$,

$$(3.7) \quad \text{var}(f_n) = ((nh^m)^{-1} f \int K^2) \{ \rho_0 + 2n^{-1} \sum_{i=1}^{n-1} (n-i) \rho_i \} .$$

Further, at every continuity point w of f ,

$$(3.8) \quad \text{var}(f_n(w)) = ((nh^m)^{-1} f(w) \int K^2) \{ 1 + o(\theta_n) + o(1) \} .$$

Remark 3.4: The result (3.7) holds even if the mixing assumption on the process $\{x_t, y_t\}$ is ignored.

The expression (3.7) suggests that for any finite n , the $\text{var}(f_n)$ is affected by serial dependence among $w_t = (x_t, y_t)$, $t=1, \dots, n$ (and hence by the serial dependence among the components of $\{x_t\}$ and $\{y_t\}$). Whereas it may be difficult to evaluate (or even to estimate) the exact value the contribution

of this serial dependence among w_t to $\text{var}(f_n)$, (3.8) shows that under the mixing condition it is only of the order $O((nh^m)^{-1}\theta_n)$.

If (2.2)⁺ holds, then (3.8) shows that

$$(3.8)' \quad \text{var}(f_n(w)) \sim (nh^m)^{-1} f(w) \int K^2.$$

Thus the asymptotic variance of f_n is the same as that obtained when w_t are assumed to be i.i.d. (Singh and Ullah [36]). On the other hand, assumption (2.2)⁺ for (3.8)' is far weaker than those made in the literature for mixing processes (e.g., see Ibragimov [12] for Gaussian processes, Pham and Tran [22] for non-Gaussian autoregressive moving average processes and Robinson [23] for strongly mixing processes).

Remark 3.5: Whereas (3.4) suggests that we can reduce the bias of the estimator f_n to any desired order by choosing $h = h_n$ converging to zero sufficiently fast, (3.7), (3.8) and (3.8)' suggest that we then, in fact, may inflate the $\text{var}(f_n)$. We give in (3.14) below a choice of h if our only concern is to control both the bias and the variance (and hence the mean square error). A further restriction, namely (3.14)' below, will be required if we are looking for the asymptotic normality of $(f_n - f)$, (see Remark 3.7 below).

Definition: For a positive sequence $\{a_n\}$, we say f_n is consistent for f with the rate $o(a_n)$ and write $|f_n - f| = o(a_n)$ in prob., if for every $\epsilon > 0$, $P[a_n^{-1} |f_n - f| > \epsilon] \rightarrow 0$ as $n \rightarrow \infty$. When $o(a_n)$ in this definition is replaced by $O(a_n)$, then we will mean that there exists a finite constant c such that $P[a_n^{-1} |f_n - f| \leq c] \rightarrow 1$ as $n \rightarrow \infty$.

In view of Chebyshev's inequality the results in the following corollary follow from those of Theorems 3.1 and 3.2.

Corollary 3.1 (Consistency of f_n): If

$$(3.9) \quad (nh^m)^{-1} = o(1)$$

and

$$(2.2)^- \quad (nh^m)^{-1} \theta_n = o(1),$$

then at every continuity point w of f ,

$$(3.10) \quad (f_n(w) - f(w)) = o(1) \text{ in prob.}$$

And if (3.4) and (3.8) hold, then

$$(3.11) \quad (f_n(w) - f(w)) = O(h^r + (nh^m)^{-1/2}(1 + \theta_n^{1/2})) \text{ in prob.}$$

Theorem 3.3 (Uniform Weak Consistency of f_n): Let the characteristic function of the function K involved in the definition of f_n be absolutely integrable. If (3.9) and (2.2)^- hold and f is uniformly continuous on $\mathcal{W}^m$, then f_n is uniformly (on $\mathcal{W}^m$) consistent for f , i.e.,

$$(3.12) \quad P[\sup_w |f_n(w) - f(w)| > \epsilon] = o(1) \text{ for every } \epsilon > 0.$$

And if (2.2) holds, and for some $r \geq 1$, (A_r) holds uniformly on $\mathcal{W}^m$ and with this r , K in (3.1) is in $\mathcal{X}_r$, then f_n is a uniformly consistent estimator of f with a rate $O(h^r + (nh^m)^{-1/2}(1 + \theta_n^{1/2}))$, that is

$$(3.13) \quad \sup_w |f_n(w) - f(w)| = O(h^r + (nh^m)^{-1/2}(1 + \theta_n^{1/2})) \text{ in prob.}$$

Remark 3.6: Notice that in view of (3.9), the assumption (2.2)^- on θ_n for (3.10) is weaker than (2.2), the boundedness of θ_n . It shows that f_n is consistent for f even if $\theta_n \rightarrow \infty$ so long as $(nh^m)^{-1} \theta_n \rightarrow 0$ as $n \rightarrow \infty$. In fact from (3.11) and (3.13); (2.2) is not necessary even if we wish to have some rate of convergence either for pointwise or for uniform consistency of f_n . However, if (2.2) does hold and h is chosen so that

$$(3.14) \quad h \sim n^{-1/(2r+m)}$$

then from (3.11)

$$(3.11)' \quad (f_n(\tilde{w}) - f(\tilde{w})) = O(n^{-r/(2r+m)}) \text{ in prob.},$$

and from (3.13),

$$(3.13)' \quad \sup_{\tilde{w}} |f_n(\tilde{w}) - f(\tilde{w})| = O(n^{-r/(2r+m)}) \text{ in prob.}$$

We will here not state any specific results for the mean square error of f_n . But, since $MSE(f_n) = (\text{Bias}(f_n))^2 + \text{var}(f_n)$, results with regard to $MSE(f_n)$ can be obtained from Theorems 1 and 2.

In what follows $N(a,b)$ will stand for the normal random variable with mean a and variance b . The symbol " $\xrightarrow{D}$ " will mean convergence in distribution.

Theorem 3.4 (Asymptotic Normality): If (2.2)⁺ holds, then at every continuity point $\tilde{w}$ of f ,

$$(3.15) \quad (nh^m)^{1/2} (f_n(\tilde{w}) - Ef_n(\tilde{w})) \xrightarrow{D} N(0, f(\tilde{w}) \int K^2),$$

In fact, if $\tilde{w}_1, \dots, \tilde{w}_\delta$ are δ distinct continuity points of f , then under (2.2)⁺, the asymptotic distribution of the vector

$$(nh^m)^{1/2} ((f_n(\tilde{w}_1) - Ef_n(\tilde{w}_1)), \dots, (f_n(\tilde{w}_\delta) - Ef_n(\tilde{w}_\delta)))$$

is a δ -variate normal distribution whose mean vector is 0 and the variance-covariance matrix is a diagonal matrix with $f(\tilde{w}_i) \int K^2$ as the i^{th} diagonal element.

Remark 3.7: If for some integer $r \geq 1$, (A_r) holds at $\tilde{w}$, and with this r , K in (3.1) is taken from $\mathcal{K}_r$, then choosing h so that

$$(3.14)' \quad nh^{2r+m} = o(1)$$

it follows from (3.4) and (3.15), that

$$(3.15)' \quad (nh^m)^{1/2} (f_n(\tilde{w}) - f(\tilde{w})) \xrightarrow{D} N(0, f(\tilde{w}) \int K^2).$$

Remark 3.8 (Estimation of the $\text{var}(f_n)$): It follows from Corollary 3.1 and Theorem 3.3 that the variance $f(\tilde{w}) \int K^2$ of $(nh^m)^{1/2} f_n(\tilde{w})$ appeared in (3.15) and (3.15)' can consistently be estimated by $\hat{f}_n(\tilde{w}) \int K^2$.

4. ESTIMATION OF THE CONDITIONAL DGP OF x_t GIVEN

$$x_{t-1}, \dots, x_{t-p}, y_t, \dots, y_{t-q}$$

Knowing that f_n introduced in Section 3 has some appealing properties as an estimator of f , the joint DGP of $\tilde{w}_t = (x_t, \dots, x_{t-p}, y_t, \dots, y_{t-q})$, we can use f_n to estimate the conditional density $g(w_1 | w_2, \dots, w_m)$ of x_t at w_1 given $x_{t-1} = w_2, \dots, x_{t-p} = w_{p+1}, y_t = w_{p+2}, \dots, y_{t-q} = w_{p+q+2}$ ($m = p+q+2$), since the latter is simply $f(w) / \int f(w) dw_1$. As a result, denoting the vector $(w_2, \dots, w_m)$ by $\tilde{w}^*$ so that $\tilde{w} = (w_1, \tilde{w}^*)$, our proposed estimator of $g(w_1 | \tilde{w}^*)$ is

$$(4.1) \quad g_n(w_1 | \tilde{w}^*) = \frac{f_n(\tilde{w})}{\int f_n(\tilde{w}) dw_1}$$

where f_n is as given in Section 3. Writing in (4.1) the exact value of f_n from Section 3, we get

$$(4.2) \quad g_n(w_1 | \tilde{w}^*) = \frac{h^{-1} \sum_{t=1}^n K((\tilde{w}_t - \tilde{w})/h)}{\sum_{t=1}^n K^*((\tilde{w}_t^* - \tilde{w}^*)/h)} = \frac{f_n(\tilde{w})}{f_n^*(\tilde{w}^*)}$$

where K is as in Section 3, $\tilde{w}_t^* = (x_{t-1}, \dots, x_{t-p}, y_t, \dots, y_{t-q})$, $K^*(\tilde{w}^*) = \int K(\tilde{w}) dw_1$ and

$$(4.3) \quad f_n^*(\tilde{w}^*) = (nh^{m-1})^{-1} \sum_{t=1}^n K^*((\tilde{w}_t^* - \tilde{w}^*)/h) .$$

Further we denote $\int f(\tilde{w}) dw_1$ by $f^*(\tilde{w}^*)$ so that $g(w_1 | \tilde{w}^*) = f(\tilde{w}) / f^*(\tilde{w}^*)$.

It is assumed throughout the remainder of this section that (2.2)⁺ holds. We also assume that $f^*(\tilde{w}^*) > 0$ (otherwise the conditional density $g(w_1 | \tilde{w}^*) = f(\tilde{w}) / f^*(\tilde{w}^*)$ is not defined, and the question of estimating g does not arise). Since the σ -field of events generated by $\{x_{t-1}, \dots, x_{t-p}, y_t, \dots, y_{t-q}\}$ is included in the σ -field of events generated by $\{x_t, \dots, x_{t-p}, y_t, \dots, y_{t-q}\}$, under the mixing condition on $\{(x_t, y_t)\}$ discussed earlier, it follows from the theory developed in Section 3, that f_n^* is a consistent

estimator of f^* . Hence pointwise (in w_1 and w^*) and uniform on $\{(w_1, w^*) | w_1 \in \mathcal{R}^1, w^* \in \mathcal{R}^{m-1}, f^*(w^*) \geq \epsilon > 0 \text{ for some } \epsilon\}$ consistency of $g_n^*(w_1 | w^*) = (f_n(w) / f_n^*(w^*))$, as an estimator of $g(w_1 | w^*) = (f(w) / f^*(w^*))$, follows from Section 3. It further follows from the theory developed in Section 3 that

$$\begin{aligned}
 (4.4) \quad & f_n(w) = Ef_n(w) + o(nh^m)^{-1/2} \quad \text{in prob.} \\
 & f_n^*(w^*) = Ef_n^*(w^*) + o(nh^{m-1})^{-1/2} \quad \text{in prob.} \\
 & \text{var}(f_n(w)) = (nh^m)^{-1} \{f(w) \int K^2 + o(1)\} \\
 & \text{var}(f_n^*(w^*)) = (nh^{m-1})^{-1} \{f^*(w^*) \int K^{*2} + o(1)\}.
 \end{aligned}$$

Now from (4.4) we have, in probability,

$$\begin{aligned}
 (4.5) \quad g_n(w_1 | w^*) &= \frac{f_n(w)}{f_n^*(w^*)} \\
 &= \frac{Ef_n(w)}{Ef_n^*(w^*)} \left[1 + \frac{f_n(w) - Ef_n(w)}{Ef_n(w)} - \frac{f_n^*(w^*) - Ef_n^*(w^*)}{Ef_n^*(w^*)} + o(nh^{m-1/2})^{-1} \right].
 \end{aligned}$$

Thus, the asymptotic unbiasedness of g_n (as an estimator of the conditional density g) and various rates of convergence follows from the general results in Theorem 3.1. Since by Hölder's inequality and (4.4), $nh^m \text{cov}(f_n(w), f_n^*(w^*)) \leq nh^m (\text{var } f_n(w) \cdot \text{var } f_n^*(w^*))^{1/2} = o(1)$, we have

$$\begin{aligned}
 \text{var}(g_n(w_1 | w^*)) &= \frac{(nh^m)^{-1} f(w) \int K^2 + \{(nh^{m-1})^{-1} f^*(w^*) \int K^{*2}\} (Ef_n(w) / Ef_n^*(w^*))^2}{(Ef_n^*(w^*))^2} \\
 &\quad + o(nh^{m-1/2})^{-1}.
 \end{aligned}$$

Hence from Theorem 3.1, we have

$$(4.6) \quad \text{var}(g_n(w_1 | w^*)) = (nh^m)^{-1} \left(\frac{g(w_1 | w^*)}{f^*(w^*)} \int K^2 \right) + o(nh^m)^{-1}.$$

Now we will investigate the asymptotic distribution of the estimator in (4.2). Since by Hölder's inequality $E[(f_n - Ef_n)(f_n^* - Ef_n^*)] \leq [\text{var}(f_n) \text{var}(f_n^*)]^{1/2}$ which is, in turn, $o(nh^{m-1/2})^{-1}$, (4.5) and the result in (3.15) applied to

f_n and f_n^* give

$$(4.7) \quad (nh^m)^{1/2} [g_n(w_1 | \tilde{w}^*) - E(g_n(w_1 | \tilde{w}^*))] \xrightarrow{D} N(0, \frac{g(w_1 | \tilde{w}^*)}{f^*(\tilde{w}^*)} \int K^2) ,$$

From (4.5) it follows that under the condition for (3.3), $Eg_n(w_1 | \tilde{w}^*) = g(w_1 | \tilde{w}^*) + o(1)$, and under the condition for (3.5) $Eg_n(w_1 | \tilde{w}^*) = g(w_1 | \tilde{w}^*) + O(h^r)$. Thus if h is chosen in accordance to (3.14)', we have

$$(4.8) \quad (nh^m)^{1/2} [g_n(w_1 | \tilde{w}^*) - g(w_1 | \tilde{w}^*)] \xrightarrow{D} N(0, \frac{g(w_1 | \tilde{w}^*)}{f^*(\tilde{w}^*)} \int K^2) .$$

It is shown in the latter part of the proof of Theorem 3.4, given in Section 6, that for any distinct continuity points $\tilde{w}, \tilde{\xi}$ in $\mathcal{R}^m$ of f , $nh^m \text{cov}(f_n(\tilde{w}), f_n(\tilde{\xi})) = o(1)$. Similarly $nh^{m-1} \text{cov}(f_n^*(\tilde{w}^*), f_n^*(\tilde{\xi}^*)) = o(1)$. We have already shown above that $nh^m \text{cov}(f_n(\tilde{w}), f_n^*(\tilde{w}^*)) = o(1)$. Hence from (4.5) and (4.8) it follows that $nh^m g_n(w_1 | \tilde{w}^*)$ and $nh^m g_n(\xi_1 | \tilde{\xi}^*)$ are asymptotically uncorrelated normal random variables whenever $w_1 \neq \xi_1$ or $\tilde{w}^* \neq \tilde{\xi}^*$ or both.

From the foregoing analysis it immediately follows that a consistent estimator of the $\text{var}(g_n(w_1 | \tilde{w}^*))$ is given by $(nh^m)^{-1} (g_n(w_1 | \tilde{w}^*) \int K^2 / f_n^*(\tilde{w}^*))$, and $100(1 - \beta)\%$ confidence interval for the conditional density $g(w_1 | \tilde{w}^*)$ is given by

$$(4.9) \quad g_n(w_1 | \tilde{w}^*) \pm z_{\beta/2} (nh^m)^{-\frac{1}{2}} \{g_n(w_1 | \tilde{w}^*) \int K^2 / f_n^*(\tilde{w}^*)\}^{\frac{1}{2}}$$

where $z_{\beta/2}$ is such that, if Z is the univariate standard normal r.v., then

$$P[-z_{\beta/2} < Z < z_{\beta/2}] = 1 - \beta.$$

Remark 4.1: Though we have here considered estimation of the conditional density of x_t given $x_{t-1}, \dots, x_{t-p}, y_t, \dots, y_{t-q}$, the estimator of the conditional density of any subset of $\{x_t, \dots, x_{t-p}, y_t, \dots, y_{t-q}\}$ given the rest can be constructed by the method discussed above.

5. ESTIMATION OF THE DYNAMIC REGRESSION OF

$$x_t \text{ on } x_{t-1}, \dots, x_{t-p}, y_t, \dots, y_{t-q}$$

The dynamic regression function of x_t on $\tilde{w}_t^* = (x_{t-1}, \dots, x_{t-p}, y_t, \dots, y_{t-q})$ is simply the conditional expectation of x_t given $\tilde{w}_t^*$, which can be computed if we know the conditional density of x_t given $\tilde{w}_t^*$. This latter function is estimated nonparametrically by g_n defined in (4.1). Hence a nonparametric estimator of the regression function of x_t on $\tilde{w}_t^*$, evaluated at a point $\tilde{w}^*$ of $\tilde{w}_t^*$ in $\mathbb{R}^{m-1}$ is given by

$$(5.1) \quad r_n(\tilde{w}^*) = \hat{E}(x_t | \tilde{w}_t^* = \tilde{w}^*) = \int w_1 g_n(w_1 | \tilde{w}^*) dw_1$$

$$= \frac{\sum_{t=1}^n \{x_t K^*((\tilde{w}_t^* - \tilde{w}^*)/h)\}}{\sum_{t=1}^n K^*((\tilde{w}_t^* - \tilde{w}^*)/h)}$$

where $K^*(\tilde{w}^*) = \int K(w_1, \dots, w_m) dw_1$, and K is as defined in Section 3, and is assumed to be symmetric about zero.

Since the σ -field (of events) generated by $\{x_t, \dots, x_{t-p}, y_t, \dots, y_{t-q}\}$ is included in the σ -field generated by $\{x_{t-1}, \dots, x_{t-p}, y_t, \dots, y_{t-q}\}$, the mixing condition of $\{(x_t, y_t)\}$ used for the various results in Section 3 applies for the results here.

We will denote

$$\int w_1 f(w_1, \dots, w_m) dw_1, \text{ by } \ell(\tilde{w}^*),$$

$$(nh^{m-1})^{-1} \sum_{t=1}^n x_t K\left(\frac{\tilde{w}_t^* - \tilde{w}^*}{h}\right) \text{ by } \ell_n(\tilde{w}^*),$$

and the true value of the regression at $\tilde{w}^*$ by $r(\tilde{w}^*)$. Then

$$(5.2) \quad r(\tilde{w}^*) = \frac{\ell(\tilde{w}^*)}{f^*(\tilde{w}^*)} \quad \text{and} \quad r_n(\tilde{w}^*) = \frac{\ell_n(\tilde{w}^*)}{f_n^*(\tilde{w}^*)}$$

where f^* and f_n^* are as introduced in Section 4.

The following results (5.3), (5.4) and (5.5) follow by arguments identical to those given (in Section 6) for the similar results in Section 3. It is assumed that $(2.2)^+$ holds, $E|x_t|^{2+} < \infty$ and f^* , l and $\int w_1^2 f(w_1, \tilde{w}^*) dw_1$ are continuous at $\tilde{w}^*$. Then

$$(5.3) \quad E(l_n(\tilde{w}^*)) = l(\tilde{w}^*) + o(1),$$

$$(5.4) \quad \left. \begin{aligned} nh^{m-1} \text{var}(l_n(\tilde{w}^*)) &= (\int w_1^2 f(w_1, \tilde{w}^*) dw_1) (\int K^{*2}) + o(1) \\ nh^{m-1} \text{var}(f_n^*(\tilde{w}^*)) &= f^*(\tilde{w}^*) \int K^{*2} + o(1) \\ \text{and } nh^{m-1} \text{cov}(l_n(\tilde{w}^*), f_n^*(\tilde{w}^*)) &= l(\tilde{w}^*) \int K^{*2} + o(1) \end{aligned} \right\}$$

Further, if (A_T) holds for l at $\tilde{w}^*$, then

$$(5.5) \quad E(l_n(\tilde{w}^*)) = l(\tilde{w}^*) + O(h^T),$$

Remark 5.1: Assumption $E|x_t|^{2+}$ is in fact needed only for the first equation of (5.4). For (5.3), (5.5) and the third equation of (5.4), only $E|x_t|^{1+}$ is needed.

By the use of Chebyshev's inequality, it follows from (5.5) that

$$(5.6) \quad \left. \begin{aligned} l_n(\tilde{w}^*) &= E l_n(\tilde{w}^*) + O(nh^{m-1})^{-1} \text{ in prob.} \\ f_n^*(\tilde{w}^*) &= E f_n^*(\tilde{w}^*) + O(nh^{m-1})^{-1} \text{ in prob.,} \end{aligned} \right\}$$

and hence we have, in probability

$$(5.7) \quad r_n(\tilde{w}^*) = \frac{E l_n(\tilde{w}^*)}{E f_n^*(\tilde{w}^*)} \left[1 + \frac{l(\tilde{w}^*) - E l_n(\tilde{w}^*)}{E l_n(\tilde{w}^*)} - \frac{f^*(\tilde{w}^*) - E f_n^*(\tilde{w}^*)}{E f_n^*(\tilde{w}^*)} + O(nh^{m-1})^{-1} \right].$$

Now asymptotic unbiasedness of $r_n(\tilde{w}^*)$, as an estimator of $r(\tilde{w}^*)$, the regression function of x_t on $\tilde{w}_t = (x_{t-1}, \dots, x_{t-p}, y_t, \dots, y_{t-q})$ evaluated at $\tilde{w}_t = \tilde{w}^*$, and its speed of convergence follows from (5.7) in conjunction with (5.3) and (5.5).

By the arguments identical to those used in Theorem 3.3 of Section 6 to prove that $(nh^m)^{1/2} \sup_{\tilde{w}} |f_n(\tilde{w}) - Ef_n(\tilde{w})| = o(1)$ in prob. whenever (2.2) holds, it can be shown that if $E|x_t|^2 < \infty$, then $(nh^{m-1})^{1/2} \sup_{\tilde{w}^*} |\ell_n^*(\tilde{w}^*) - E\ell_n^*(\tilde{w}^*)| = o(1)$ in prob. From these results and (5.7), it follows that

$$(5.8) \quad (nh^{m-1})^{1/2} \sup_{\tilde{w}^* \in B} |r_n(\tilde{w}^*) - Er_n(\tilde{w}^*)| = o(1) \quad \text{in prob.}$$

where B is a subset of $\mathcal{W}^{m-1}$ such that $\inf_{\tilde{w}^* \in B} f^*(\tilde{w}^*) \geq \epsilon$ for some $\epsilon > 0$. Hence, in view of (5.7), if $\ell(\tilde{w}^*)$ and $f^*(\tilde{w}^*)$ are uniformly continuous on B (implying $\sup_{\tilde{w}^* \in B} |E\ell_n(\tilde{w}^*) - \ell(\tilde{w}^*)| = o(1)$ and $\sup_{\tilde{w}^* \in B} |Ef_n^*(\tilde{w}^*) - f^*(\tilde{w}^*)| = o(1)$), then

$$\sup_{\tilde{w} \in B} |r_n(\tilde{w}^*) - r(\tilde{w}^*)| = o(1) \quad \text{in prob.}$$

Further, if (A_r) holds for $\ell(\tilde{w}^*)$ and $f^*(\tilde{w}^*)$ uniformly in $\tilde{w}^* \in B$, then the proof for the latter part of Theorem 3.1 implies $\sup_{\tilde{w}^* \in B} |E\ell_n(\tilde{w}^*) - \ell(\tilde{w}^*)| = o(h^r)$ and $\sup_{\tilde{w}^* \in B} |Ef_n^*(\tilde{w}^*) - f^*(\tilde{w}^*)| = o(h^r)$. Hence, (5.8) and (5.7) imply

$$\sup_{\tilde{w}^* \in B} |r_n(\tilde{w}^*) - r(\tilde{w}^*)| = o(\max\{(nh^{m-1})^{-1/2}, h^r\}) \quad \text{in prob.}$$

Now, (5.7) and (5.4) followed by (5.3) give

$$(5.9) \quad (nh^{m-1}) \text{var}(r_n(\tilde{w}^*)) = \left(\frac{\ell(\tilde{w}^*)}{f^*(\tilde{w}^*)} \right)^2 \left[\frac{\int \tilde{w}_1^2 f(\tilde{w}_1, \tilde{w}^*) d\tilde{w}_1}{\ell^2(\tilde{w}^*)} - \frac{1}{f^*(\tilde{w}^*)} \right] (\int K^2) + o(1) \\ = \Lambda(\tilde{w}^*) + o(1)$$

where

$$(5.10) \quad \Lambda(\tilde{w}^*) = \frac{\text{var}(x_t | \tilde{w}_t = \tilde{w}^*) \int K^2}{f^*(\tilde{w}^*)} + o(1).$$

It follows from the arguments (given in Section 6) to prove the asymptotic normality of f_n that

$$(nh^{m-1})^{1/2} (\ell_n(\tilde{w}^*) - E\ell_n(\tilde{w}^*)) \xrightarrow{D} N(0, (\int \tilde{w}_1^2 f(\tilde{w}_1, \tilde{w}^*) d\tilde{w}_1) \int K^2).$$

This result and (3.15) (applied to f_n^*) followed by (5.7) give

$$(5.11) \quad (nh^{m-1})^{1/2} (r_n(\tilde{w}^*) - Er_n(\tilde{w}^*)) \xrightarrow{D} N(0, \Lambda(\tilde{w}^*)) .$$

Finally, if (5.5) and the condition (A_r) for $\ell(\tilde{w}^*)$ and $f^*(\tilde{w}^*)$ hold at $\tilde{w}^*$, then we have from (5.7)

$$E(r_n(\tilde{w}^*)) = r(\tilde{w}^*) + O(h^r) + O(nh^{m-1})^{-1},$$

and hence, if h is chosen in accordance to

$$(3.14)'' \quad h^r(nh^{m-1})^{1/2} = o(1)$$

then from (5.9) we have

$$(5.12) \quad (nh^{m-1})^{1/2}(r_n(\tilde{w}^*) - r(\tilde{w}^*)) \xrightarrow{D} N(0, \Lambda(\tilde{w}^*)).$$

From our foregoing analysis, we consistently estimate $\Lambda(\tilde{w}^*)$, rewritten as,

$$\Lambda(\tilde{w}^*) = \frac{E(x_t^2 | \tilde{w}_t^* = \tilde{w}^*) - (E(x_t | \tilde{w}_t^* = \tilde{w}^*))^2}{f^*(\tilde{w}^*)} (\int K^2)$$

by

$$\Lambda_n(\tilde{w}^*) = \frac{\{a_n(\tilde{w}^*) - r_n^2(\tilde{w}^*)\}(\int K^2)}{(nh^{m-1})^{-1} \sum_{t=1}^n \{K^*((\tilde{w}_t^* - \tilde{w}^*)/h)\}}$$

where

$$a_n(\tilde{w}^*) = \frac{\sum_{t=1}^n \{x_t^2 K^*((\tilde{w}_t^* - \tilde{w}^*)/h)\}}{\sum_{t=1}^n \{K^*((\tilde{w}_t^* - \tilde{w}^*)/h)\}}.$$

Hence, a consistent estimator of $\text{var}(r_n(\tilde{w}^*))$ is $(nh^{m-1})^{-1} \Lambda_n(\tilde{w}^*)$, and from (5.12), $100(1-\beta)\%$ confidence interval for $r(\tilde{w}^*)$, the regression of x_t on $\tilde{w}_t^* = (x_{t-1}, \dots, x_{t-p}, y_t, \dots, y_{t-q})$, is given by

$$(5.13) \quad r_n(\tilde{w}^*) \pm z_{\beta/2} (nh^{m-1})^{-1/2} (\Lambda_n(\tilde{w}^*))^{1/2}$$

where $z_{\beta/2}$ is as defined at the end of Section 4.

In the proof (Section 6) of Theorem 3.4, it is shown that if

$\tilde{w}_1$ and $\tilde{w}_2$ are distinct points in $\mathcal{W}^m$, then $(nh^m) \text{cov}(f_n(\tilde{w}_1), f_n(\tilde{w}_2)) = o(1)$.

This result applies to f_n^* too; that is if $\tilde{w}^*$ and $\tilde{\xi}^*$ are distinct continuity points of f^* in $\mathcal{W}^{m-1}$, then

$$nh^{m-1} \text{cov}(f_n^*(\tilde{w}^*), f_n^*(\tilde{\xi}^*)) = o(1).$$

It also follows from the same arguments that for distinct $\tilde{w}^*$ and $\tilde{\xi}^*$ continuity points of ℓ and f^* in $\mathcal{R}^{m-1}$,

$$nh^{m-1} \text{cov}(\ell_n(\tilde{w}^*), \ell_n(\tilde{\xi}^*)) = o(1)$$

and

$$nh^{m-1} \text{cov}(\ell_n(\tilde{w}^*), f_n^*(\tilde{\xi}^*)) = o(1) .$$

These results in conjunction with (5.7) prove that

$$nh^{m-1} \text{cov}(r_n(\tilde{w}^*), r_n(\tilde{\xi}^*)) = o(1)$$

whenever $\tilde{w}^*$ and $\tilde{\xi}^*$ are distinct. Hence, from (5.12) and the Cramer-Wold technique (Theorem 7.7, Billingsley [6]), we conclude that if $\tilde{w}_1^*, \dots, \tilde{w}_\delta^*$ are distinct points in $\mathcal{R}^{m-1}$ at which $\int \tilde{w}_1^2 f(\tilde{w}_1, \tilde{w}^*) d\tilde{w}_1$ is continuous and $f^*(\tilde{w}^*)$ and $\ell(\tilde{w}^*)$ satisfy condition (A_r) , then taking h satisfying (3.14)'', the asymptotic distribution of the vector $nh^{m-1}((r_n(\tilde{w}_1^*) - r(\tilde{w}_1^*)), \dots, (r_n(\tilde{w}_\delta^*) - r(\tilde{w}_\delta^*)))$ is a δ -variate normal whose mean vector is 0 and variance-covariance matrix is the $\delta \times \delta$ diagonal matrix whose i^{th} element is $\Lambda(\tilde{w}_i^*) \int K^2$.

5.1 Illustrative Example

As an illustration of the econometric model specification by non-parametric technique, consider the model of testing the efficiency of the foreign exchange market. In the extensive empirical literature on this subject (see Levich [16] for a survey), the regression model usually formulated is

$$(5.14) \quad S_t = a_0 + a_1 F_{t-1} + u_t, \text{ or } ES_t | F_{t-1} = a_0 + a_1 F_{t-1}$$

where S_t is the logarithm of the spot rate, F_t is the logarithm of the forward rate and u_t is the random error. The market efficiency is then examined by testing whether or not the forward rate is an unbiased predictor of the future spot rate, that is, by testing the null hypothesis that $a_0 = 0$ and $a_1 = 1$; see, for example, Longworth [17], and a recent study by Baillie

et al [2] on the extent of bias in the foreign exchange market.

Our main interest here is to analyze the relationship between S_t and F_{t-1} by the nonparametric technique. For this purpose we considered the Canadian-U.S. exchange market with 90-day forward rates (see OECD historical statistics 1960-1979). The sample period was 1970.7 to 1978.12. We estimated the conditional mean $ES_t | F_{t-3}$ by the nonparametric technique given in (5.1). This is

$$(5.15) \quad \hat{E}(S_t | F_{t-3}) = \sum_t S_t \ell_t(F_{t-3})$$

where

$$(5.16) \quad \ell_t(F_{t-3}) = \frac{K(\frac{F_{t-3} - F}{h})}{\sum_t K(\frac{F_{t-3} - F}{h})}.$$

The conditional mean in (5.15) was estimated by using the following kernel functions:

$$(i) \quad K_1\left(\frac{F_{t-3} - F}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{F_{t-3} - F}{h}\right)^2} \quad \text{and } h = n^{-\frac{1}{5}} s$$

$$(ii) \quad K_2\left(\frac{F_{t-3} - F}{h}\right) = \frac{1}{2}\left[3 - \left(\frac{F_{t-3} - F}{h}\right)^2\right] K_1\left(\frac{F_{t-3} - F}{h}\right) \quad \text{and } h = n^{-\frac{1}{7}} s$$

where s^2 is the sample variance of F . The explanation for the choices of h is given below, and the kernels are chosen in accordance to Remark 3.1.

Regarding the choice of h , we note from (3.4) and (3.8) that as h decreases the bias decreases but variance increases. Therefore we must choose h in such a way that the bias and variance both remain under control. In fact, we would like to choose h so that the mean-squared error = $\{(\text{bias})^2 + \text{variance}\}$ be minimum. A choice of such h for the independent case is given in Singh [34]. Here, if in view of (3.4) and (3.8) we choose $h \sim n^{-1/(1+2r)}$ (for $m=1$) then the $(\text{bias})^2$ and variance both become of the order $O(n^{-2r/(1+2r)})$.

From (3.1) and (3.2) the variance of $f_n(w)$ will get inflated whenever $\text{var} K(\frac{w_1 - w}{h})$ is large, which in turn will be large whenever $\text{var}(w_1) = \sigma_1^2$ (say) is large. This effect of σ_1^2 on the $\text{var} K(\frac{w_1 - w}{h})$ will be eliminated to some extent if h is also proportional to σ_1 . Therefore a better choice of h is the one which is proportional to $\sigma_1 n^{-1/(1+2r)}$. Since $\sigma_1 = (\text{var } F_{t-3})^{1/2}$ is normally unknown, we recommend replacing it by $s = (\Sigma(F_{t-3} - \bar{F})^2/n)^{1/2}$ which is treated as a constant.⁵

The nonparametric estimates of the conditional means, $E(S_t | F_{t-3})$, under kernels K_1 in (i) and K_2 in (ii) are plotted in Figures 1 and 2, respectively. In these figures, + represents the observed data; the linear regression is based on the least squares estimate of $E(S_t | F_{t-3}) = a_0 + a_1 F_{t-3}$ in (5.14), and the nonlinear regression is based on the nonparametric estimate of $E(S_t | F_{t-3})$ from (5.15). These figures show that the true relationship between S_t and F_{t-3} may be nonlinear.⁶ We also note that the least squares residual sum of squares based on (5.14) was .0348, and the nonparametric residual $(S_t - ES_t | F_{t-3})$ sum of squares corresponding to kernels K_1 and K_2 were .0326 and .0317, respectively. Thus the nonparametric specification performed better than the linear regression specification in (5.14).⁷

Figure 1 here

Figure 2 here

6. PROOFS OF THEOREMS

In this section we prove the theorems stated in Section 3.

Proof of Theorem 3.1: From (3.2) and (3.1), we have

$$(6.1) \quad Ef_n(\underline{w}) = \int K(\underline{u}) f(\underline{w} + h\underline{u}) d\underline{u}.$$

Since $\int K = 1$, from (3.3) we write

$$B_n(\underline{w}) = \int K(\underline{u}) (f(\underline{w} + h\underline{u}) - f(\underline{w})) d\underline{u} = I_1 + I_2$$

where, for some $\epsilon > 0$, I_1 and I_2 are the integrals over the subsets $\|\underline{h}\underline{u}\| < \epsilon$ and $\|\underline{h}\underline{u}\| \geq \epsilon$ of $\mathbb{R}^m$, respectively. Since f is continuous at $\underline{w}$ and K is integrable, choosing ϵ sufficiently small, I_1 can be made arbitrarily close to zero.

Having so chosen ϵ , I_2 can be bounded by

$$\int_{\|\underline{u}\| \geq h^{-1}\epsilon} |K(\underline{u})| f(\underline{w} + h\underline{u}) d\underline{u} + f(\underline{w}) \int_{\|\underline{u}\| \geq h^{-1}\epsilon} |K(\underline{u})| d\underline{u}.$$

Since $h^{-1} \rightarrow \infty$ as $n \rightarrow \infty$, the second integral is $o(1)$. Also, the first integral, written as $\int_{\|\underline{u}\| \geq h^{-1}\epsilon} \|\underline{u}\|^{-m} (\|\underline{u}\|^m |K(\underline{u})|) f(\underline{w} + h\underline{u}) d\underline{u}$, is bounded by

$$\epsilon^{-m} \sup_{\|\underline{u}\| \geq h^{-1}\epsilon} (\|\underline{u}\|^m |K(\underline{u})|) \cdot h^m \int f(\underline{w} + h\underline{u}) d\underline{u} \text{ which, in turn, is } o(1), \text{ since}$$

$h^m \int f(\underline{w} + h\underline{u}) d\underline{u} = 1$ and $\|\underline{u}\|^m |K(\underline{u})| \rightarrow 0$ as $\|\underline{u}\| \rightarrow \infty$. This proves (3.3). Proof of (3.3)' follows on the same lines with applications of uniform continuity on $\mathbb{R}^m$ (and hence also of boundedness) of f at appropriate places. Proofs for (3.4) and for the last result of the theorem follow from the arguments similar to those given for the proof of Theorem 1 ($p=0$ case) of Singh [35].

□

Proof of Theorem 3.2: Since K is a Borel-measurable function on $\mathcal{R}^m$, the stationarity of $\{(x_t, y_t)\}$ implies the stationarity of $\{\eta_t\}$ defined in (3.5) (Ibragimov and Linnik [13], Section 18.6). Write $S_n = \sum_{t=1}^n \eta_t$. Then by stationarity,

$$(6.2) \quad \begin{aligned} ES_n^2 &= n\rho_0 + 2\sum_{i=1}^{n-1} \sum_{j=1}^{n-i} E(\eta_j \eta_{j+i}) \\ &= n\rho_0 + 2\sum_{i=1}^{n-1} (n-i)\rho_i, \end{aligned}$$

where ρ_i for $i=0,1,\dots$, are as defined in (3.6). Hence, since $n^{-1}h^{-m/2}(f \int K^2)^{1/2} S_n = (f_n - Ef_n)$, (3.7) follows.

To obtain (3.8), we will estimate ρ_0 and ρ_i in (3.7). It follows from (3.5) that

$$(6.3) \quad (f(\tilde{w}) \int K^2) E \eta_{\tilde{t}}^2(\tilde{w}) = [(\int K^2(u) f(\tilde{w} + hu) du) - h^m (\int K(u) f(\tilde{w} + hu) du)^2] .$$

Since f is continuous at $\tilde{w}$, by arguments used to prove (3.3), it follows that the right-hand side of (6.3) is $f(\tilde{w}) \int K^2 + o(1) + O(h^m)$. Hence by stationarity $\rho_0 = E \eta_{\tilde{t}}^2 = 1 + o(1)$. Now consider ρ_i for $i \neq 0$. Again, since K is a Borel-measurable function on $\mathcal{R}^m$ and $\{(x_t, y_t)\}$ is strongly mixing, the process $\{\gamma_t\}$, where $\gamma_t = K(\frac{\tilde{w}_t - \tilde{w}}{h})$, is strongly mixing (Ibragimov and Linnik [13], Section 18.6, or Billingsley [6], Section 21). Let $\{\alpha'_j\}$ be the sequence of mixing coefficients of this process. Since $(h^m(f(\tilde{w}) \int K^2)\rho_i)$ can be written as $\{E(\gamma_j \gamma_{j+i}) - E\gamma_j E\gamma_{j+i}\} = E\gamma_j \{E(\gamma_{j+i} | \mathcal{B}_{-\infty}^j) - E(\gamma_{j+i})\}$ and K (and hence γ_j) is bounded, say by c , employing the proof of Lemma 5.2 of Dvoretzky [9] (or the proof of Lemma 2.1 of McLeish [18] via Dvoretzky's lemma), it follows by stationarity that $|\rho_i(\tilde{w})| \leq 4c h^{-m}(f(\tilde{w}) \int K^2)^{-1} \alpha'_j E|K(\frac{\tilde{w}_1 - \tilde{w}}{h})|$. But $E|K((\tilde{w}_1 - \tilde{w})/h)| = h^m f(\tilde{w}) \int |K| + o(h^m)$ by the arguments used to prove (3.3). Hence for a constant c_0 , $|\rho_i| \leq \alpha'_i(c_0 + o(1))$. Now using these estimates of ρ_0 and ρ_i for $i \neq 0$ in (3.7), we get

$$(6.4) \quad nh^m(f(\tilde{w}) \int K^2)^{-1} \text{var}(f_n(\tilde{w})) = 1 + O(\theta'_n) + o(1)$$

where $\theta'_n = n^{-1} \sum_{i=1}^n (n-i) \alpha'_i$. But, with $p' = \max\{p, q\}$, since y_t is $\{x_{t-p'}, \dots, x_{t+p'}, y_{t-p'}, \dots, y_{t+p'}\}$ -measurable, by the arguments used to prove 18.6.2 in Ibragimov and Linnik [13], it follows that $\theta'_n = o(\theta_n)$. Hence (6.4) gives (3.8). □

Proof of Theorem 3.3: Let ψ be the characteristic function of K .

Then the inversion formula for the characteristic function gives

$$K(\underline{u}) = (2\pi)^{-m} \int \exp(-i\underline{v}' \underline{u}) \psi(\underline{v}) d\underline{v}$$

Choose an arbitrary but fixed point $\underline{w}_0$ in the domain of f . Since f is uniformly continuous, it is also continuous at $\underline{w}_0$. From the above expression for K , we can write

$$(f_n(\underline{w}) - E f_n(\underline{w})) = (2\pi)^{-m} \int \exp(i\underline{v}' (\underline{w} - \underline{w}_0)/h) \lambda_n(\underline{v}/h) \psi(\underline{v}) d\underline{v}$$

where $\lambda_n(\underline{v}) = n^{-1} \sum_{t=1}^n \zeta_t(\underline{v})$ and $\zeta_t(\underline{v}) = h^{-m} \{ \exp(-i\underline{v}' (\underline{w}_t - \underline{w}_0)) - E \exp(-i\underline{v}' (\underline{w}_t - \underline{w}_0)) \}$.

It follows from Lemma 1 of Jennrich [14] that $\sup_{\underline{w}} |f_n(\underline{w}) - E f_n(\underline{w})|$ is a random variable, and hence we can talk of its expectation. Hence

$$(6.5) \quad E \sup_{\underline{w}} |f_n(\underline{w}) - E f_n(\underline{w})| \leq (2\pi)^{-m} \int E |\lambda_n(\underline{v}/h)| |\psi(\underline{v})| d\underline{v}.$$

We now evaluate $E |\lambda_n(\underline{v}/h)|$. By stationarity it follows that

$$E |\lambda_n(\underline{v}/h)|^2 = n^{-1} E \zeta_1^2(\underline{v}/h) + 2n^{-2} \sum_{i=1}^{n-1} \{ (n-i) E (\zeta_1(\underline{v}/h) \zeta_{1+i}(\underline{v}/h)) \}.$$

But

$$\begin{aligned} E \zeta_1^2(\underline{v}/h) &\leq h^{-2m} E \{ \exp(-i\underline{v}' (\underline{w}_1 - \underline{w}_0)/h) \}^2 \\ &\leq h^{-m} \int f(\underline{w}_0 + h\underline{\xi}) d\underline{\xi} = h^{-m} (f(\underline{w}_0) + o(1)) \end{aligned}$$

since f is continuous at $\underline{w}_0$ and hence arguments used to prove Theorem 3.3 apply here. By these arguments and those used in the latter part of the proof of Theorem 3.2 it also follows that

$$2n^{-2} \sum_{i=1}^{n-1} \{ (n-i) E (\zeta_1(\underline{v}/h) \zeta_{1+i}(\underline{v}/h)) \} = (nh^m)^{-1} (o(\theta_n) + o(1)).$$

Thus, since $E |\lambda_n| \leq (E |\lambda_n|^2)^{1/2}$, we conclude that

$$E|\lambda_n(v/h)| = (nh^m)^{-1/2}(O(1) + O(\theta_n^{1/2})),$$

and, since ψ is absolutely integrable, it follows from (6.5) that

$$(6.6) \quad (nh^m)^{1/2} E \sup_{\tilde{w}} |f_n(\tilde{w}) - Ef_n(\tilde{w})| = O(1 + \theta_n^{1/2}).$$

Now (6.6) and Chebyshev's inequality establish

$$(6.7) \quad (nh^m)^{1/2} \sup_{\tilde{w}} |f_n(\tilde{w}) - Ef_n(\tilde{w})| = O(1 + \theta_n^{1/2}) \text{ in prob.}$$

Hence, if (2.2) holds, then

$$(nh^m)^{1/2} \sup_{\tilde{w}} |f_n(\tilde{w}) - Ef_n(\tilde{w})| = O(1) \text{ in prob.}$$

The conclusions (3.12) and (3.13) of Theorem 3.3 now follow from (6.7) and Theorem 3.1. □

Proof of Theorem 3.4: We will prove this theorem by invoking

Theorem 3 of Bradley [7]. First we will show that the conditions of his theorem are satisfied in our case. Let $\eta_t = \eta_t^n$ be defined as in (3.5). Let $S_n = \sum_{t=1}^n \eta_t$. Then from (3.8) it follows that

$$(6.8) \quad \begin{aligned} \text{var}(S_n) &= n^2 h^m (f \int K^2)^{-1} \text{var}(f_n) \\ &= n + O(n\theta_n) + o(n). \end{aligned}$$

In fact, by stationarity, it can be proved by the arguments leading to (6.8) that for any J and N such that $0 \leq J < J+N \leq n$

$$\text{var}(\eta_{J+1}^n + \dots + \eta_{J+N}^n) = N + O(N\theta_N) + o(N).$$

Hence for $N = 1, 2, \dots$,

$$(6.9) \quad \begin{aligned} s_N^2 &= \inf_{n, J, L} \{ \text{var}(\eta_{J+1}^n + \dots + \eta_L^n) \mid J \geq 0, J+N \leq L \leq n \} \\ &= (N + O(N\theta_N) + o(N)) \rightarrow \infty \text{ as } N \rightarrow \infty. \end{aligned}$$

Further, for any $0 < \epsilon \leq 1$, by Minkowski inequality we have

$$\begin{aligned} E |\eta_t|^{2+\epsilon} &\leq 4h^{-m(2+\epsilon)/2} (E \int K^2)^{-(2+\epsilon)/2} E \left| K \left(\frac{w_t - w}{h} \right) \right|^{2+\epsilon} \\ &= 4h^{-m(2+\epsilon)/2} (E \int K^2)^{-(2+\epsilon)/2} h^m (E \int |K|^{2+\epsilon} + o(1)) \end{aligned}$$

by the arguments applied to prove (3.3). Hence for any $0 < \epsilon \leq 1$,

$$(6.10) \quad E |\eta_t|^{2+\epsilon} = O(h^{-m\epsilon/2}).$$

It is proved in the proof of Theorem 3.2 that the sequence of the mixing coefficients of the process $\{\gamma_t\}$, where $\gamma_t = K\left(\frac{w_t - w}{h}\right)$, (and hence of the process $\{\eta_t\}$) converges to zero (first letting $n \rightarrow \infty$) whenever $\{\alpha_j\}$ does. This conclusion and the estimates obtained for ρ_i , $i=0,1,\dots,n-1$, in the proof of Theorem 3.2 prove that α_N^* and ρ_N^* both converge to zero as $N \rightarrow \infty$, where α_N^* and ρ_N^* are respectively α_N and ρ_N of Bradley [7] defined for our process $\{\eta_t^n\}$. Hence all the conditions of Theorem 3 of Bradley [7] with $k(n) = n$ for the process $\{\eta_t^n\}$ are satisfied, and we conclude that

$$(6.11) \quad \frac{S_n}{(\text{var}(S_n))^{1/2}} \xrightarrow{D} N(0,1).$$

But by (6.8), $\text{var}(S_n) = n(1 + O(\theta_n) + o(1))$. Hence $n^{-1/2} S_n \xrightarrow{D} N(0,1)$ under (2.2)⁺. This concludes (3.15) since $n^{-1/2} S_n = (nh^m)^{1/2} (f_n - Ef_n) / (E \int K^2)^{1/2}$.

Now we will prove the second part of the theorem. We have proved that at every continuity point w_i of f ,

$$(nh^m)^{1/2} (f_n(w_i) - Ef_n(w_i)) \xrightarrow{D} N(0, f(w_i) \int K^2).$$

Hence, in view of the Cramer-Wold technique (Theorem 7.7, Billingsley [6]), the second part of the theorem follows if we can prove that for any two distinct continuity points w and ξ of f ,

$$(6.12) \quad nh^m \operatorname{cov}(f_n(\tilde{w}), f_n(\tilde{\xi})) = o(1) .$$

From the definition of η_t , to prove (6.12), it suffices to show that

$$(6.13) \quad E \left[\sum_{j=1}^n \eta_j(\tilde{w}) \left(\sum_{j=1}^n \eta_j(\tilde{\xi}) \right) \right] = o(n) .$$

Denote $\eta_j(\tilde{w})$ and $\eta_j(\tilde{\xi})$ by η_j and η'_j respectively. Then, the left-hand side of (6.13) is

$$(6.14) \quad \sum_{j=1}^n \eta_j \eta'_j + \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} \eta_j \eta'_{j+i} + \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} \eta'_j \eta_{j+i} .$$

The first term in (6.14) is bounded by $2n(f(\tilde{w})f(\tilde{\xi}))^{-1/2} \int K^2$ times

$$h^{-m} E \left| K\left(\frac{w - \tilde{w}}{h}\right) K\left(\frac{w - \tilde{\xi}}{h}\right) \right|$$

which is $o(1)$ by the arguments given in the latter part of the proof of Lemma 2.2 of Cacoullos [8].

By the arguments identical to those used in the proof of Theorem 3.2 to show that $\sum_{i=1}^{n-1} \sum_{j=1}^{n-i} E(\eta_j \eta_{j+i})$ in (6.2) is $O(n\theta_n)$, it follows that both the second and third terms in (6.14) are of the order $O(n\theta_n)$, which is $o(n)$ by (2.2)⁺. □

FIGURE 1 NONPARAMETRIC REGRESSION KERNEL K1

FIGURE 2 NONPARAMETRIC REGRESSION KERNEL K2

REFERENCES

- [1] Ahmad, I. A. Strong consistency of density estimation by orthogonal series methods for dependent variables with applications. Annals of the Institute of Statistical Mathematics 31 (1979): 279-288.
- [2] Baillie, R. T., R. E. Lippens & P. C. McMahon. Testing rational expectations and efficiency in the foreign exchange market. Econometrica 51 (1983): 513-535.
- [3] Bierens, H. J. A uniform weak law of large numbers under ϕ -mixing, with application to nonlinear least squares estimation. Statistica Neerlandica 36 (1982): 81-86.
- [4] Bierens, H. J. Uniform consistency of kernel estimators of a regression function under generalised conditions. Journal of the American Statistical Association 78 (1983): 699-707.
- [5] Bierens, H. J. Model specification testing of time series regressions. Journal of Econometrics 27 (1984): forthcoming.
- [6] Billingsley, P. Convergence of Probability Measures. New York-London: John Wiley, 1968.
- [7] Bradley, R. C. Central limit theorems under weak dependence. Journal of Multivariate Analysis 10 (1968): 1-16.
- [8] Cacoullos, T. Estimation of a multivariate density. Annals of the Institute of Statistical Mathematics 18 (1966): 179-189.
- [9] Dvoretzky, A. Asymptotic normality for sums of dependent random variables. In Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, 1972, 513-535, University of California Press.

- [10] Epanechnikov, V. A. Nonparametric estimation of a multivariate probability density. Theory of Probability and Applications (1969): 153-158.
- [11] Hannan, E. J. Multiple Time Series. New York: Wiley, 1970.
- [12] Ibragimov, I. A. On the spectrum of stationary Gaussian processes satisfying the strong mixing condition II. Sufficient conditions. Mixing rate. Theory of Probability and Applications 15 (1970): 23-36.
- [13] Ibragimov, I. A. & Yu. V. Linnik. Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff: Groningen, 1971.
- [14] Jennrich, R. I. Asymptotic properties of nonlinear least squares estimates. Annals of Mathematical Statistics 40 (1969): 633-643.
- [15] Kronmal, R. & M. Tarter. The estimation of probability densities and cumulatives by Fourier series methods. Journal of the American Statistical Association 38 (1968): 482-493.
- [16] Levich, R. M. On the efficiency of markets for foreign exchange. In R. Dornbusch and J. A. Frenkel (eds.), International Economic Policy, Theory and Evidence. John Hopkins University Press: 1979.
- [17] Longworth, D. Testing the efficiency of the Canadian-U.S. exchange market under the assumption of no risk premium. Journal of Finance 36 (1981): 43-49.
- [18] McLeish, D. L. A maximal inequality and dependent strong laws. The Annals of Probability 3 (1975): 829-839.
- [19] Moore, D. S. & J. W. Yackel. Consistency properties of nearest neighbor density estimates. Annals of Statistics 5 (1977): 143-154.
- [20] Nadaraya, E. A. On estimating regression. Theory of Probability and Applications 9 (1964): 141-142.

- [21] Parzen, E. J. On estimation of a probability density and mode.
Annals of Mathematical Statistics 35 (1964): 1065-1076.
- [22] Pham, T. D. & L. T. Tran. The strong mixing property of the autoregressive moving average time series model. Seminaire de Statistique Grenoble (1980): 59-76.
- [23] Robinson, P. M. Nonparametric estimators for time series. Journal of Time Series Analysis (1983): 185-208.
- [24] Rosenblatt, M. Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics 27 (1956): 832-837.
- [25] Rosenblatt, M. A central limit theorem and a strong mixing condition. Proceedings of the National Academy of Science, USA 42 (1956): 43-47.
- [26] Rosenblatt, M. Conditional probability density and regression estimators.
In P. R. Krishnaiah (ed.), Multivariate Analysis II, 25-31. New York: Academic Press, 1969.
- [27] Rosenblatt, M. Density estimates and Markov processes. In M. C. Puri (ed.), Nonparametric Techniques in Statistical Inference, 199-210. Cambridge: Cambridge University Press, 1970.
- [28] Roussas, G. Nonparametric estimation in Markov processes. Annals of the Institute of Statistical Mathematics 21 (1967): 73-87.
- [29] Roussas, G. Nonparametric estimation of the transition distribution function of a Markov process. Annals of Mathematical Statistics 40 (1969): 1386-1400.
- [30] Schuster, E. F. Joint asymptotic distribution of the estimated regression function at a number of distinct points. Annals of Mathematical Statistics 43 (1972): 84-88.

- [31] Schwartz, S. C. Estimation of a probability density by an orthogonal series. Annals of Mathematical Statistics 38 (1967): 1261-2165.
- [32] Singh, R. S. Nonparametric estimation of mixed partial derivatives of a multivariate density. Journal of Multivariate Analysis 6 (1976): 555-567.
- [33] Singh, R. S. Applications of estimators of a density and its derivatives to certain statistical problems. Journal of the Royal Statistical Society, Series B 39 (1977): 357-363.
- [34] Singh, R. S. Mean squared errors of estimates of a density and its derivatives. Biometrika 66 (1979): 177-180.
- [35] Singh, R. S. Speed of convergence in nonparametric estimation of a multivariate μ -density and its mixed partial derivatives. Journal of Statistical Planning and Inference 5 (1981): 287-298.
- [36] Singh, R. S. & A. Ullah. Estimation of multivariate DGP. University of Western Ontario Manuscript (1984).
- [37] Takahata, H. Limiting behaviour of density estimates for stationary asymptotically uncorrelated processes. Bulletin of Tokyo Gakugei University, Series IV 29 (1977): 1-9.
- [38] Takahata, H. Almost sure convergence of density estimators for weakly dependent stationary processes. Bulletin of Tokyo Gakugei University, Series IV 32 (1980): 11-32.
- [39] Ullah, A. & R. S. Singh. Estimation of density and its applications in econometrics. University of Western Ontario Manuscript (1984).
- [40] Van Ryzin, J. On histogram method of density estimation. Communications in Statistics 2 (1973): 493-506.

- [41] Watson, G. S. Smooth regression analysis. Sankhya, Series A 26 (1964): 359-372.
- [42] Watson, G. S. Density estimation by orthogonal series. Annals of Mathematical Statistics 40 (1969): 1496-1498.
- [43] Wegman, E. J. Maximum likelihood estimation of a unimodal density function. Annals of Mathematical Statistics 41 (1970): 457-471.
- [44] White, H. & I. Domowitz. Nonlinear regression with dependent observations. Econometrica 52 (1984): 143-161.
- [45] Wolverton, C. T. & T. J. Wagner. Asymptotically optimal discriminant functions for pattern classification. IEEE Transaction of Information Theory, IT-15 (1969): 258-265.
- [46] Yakowitz, S. Nonparametric estimation of Markov transition functions. Annals of Statistics 7 (1979): 671-679.
- [47] Yamato, H. Sequential estimation of a continuous probability density function and mode. Bulletin of Mathematical Statistics 14 (1971): 1-12.

FOOTNOTES

¹DGP here is referred as what is called the probability density function in the statistical literature.

²We are assuming that for some $\eta' > 0$ and $\eta'' > 0$, a single realization of $\{x_t, t=1-p, \dots, -1, 0, 1, \dots, \eta'\}$ and of $\{y_t, t=1-q, \dots, -1, 0, 1, \dots, \eta''\}$ is recorded, and $n = \min\{\eta', \eta''\}$. If x_t only for $t=1, \dots, \eta'$ ($>p \geq 0$) and y_t only for $t=1, \dots, \eta''$ ($>q \geq 0$) are realized, then our $\tilde{x}_t$ and $\tilde{y}_t$ will represent $(x_{t+p}, \dots, x_t)$ for $t=1, \dots, \eta' - p$ and $(y_{t+q}, \dots, y_t)$ for $t=1, \dots, \eta'' - q$ respectively.

³If there is a nonnegative function ϕ_j converging to zero as $j \rightarrow \infty$ such that for every $A \in \mathcal{B}_{-\infty}^t$ and $B \in \mathcal{B}_{t+j}^\infty$, $|P(A \cap B) - P(A)P(B)| \leq \phi_j P(A)$, then $\{\tilde{y}_t\}$ is said to be a ϕ -mixing process. Notice that every ϕ -mixing process is a strong mixing.

⁴The condition (2.2) is used to prove consistency of our estimates, whereas $(2.2)^+$ is used to obtain an asymptotic expression for the variances. It is shown that under $(2.2)^+$, the asymptotic variance of the various estimators are the same as those obtained for i.i.d. process (Singh and Ullah [36]). The conditions $(2.2)^+$ is also used to prove the central limit theorems for our estimators.

⁵Notice that s is a consistent estimator of $\sigma_1 = (\text{var} F_{t-3})^{\frac{1}{2}}$ and hence for large n it can be treated as a constant. In general, however, s is random which makes h random. The analysis for random h is beyond the scope of this paper and will be dealt with in a separate study.

⁶We also observed on the basis of histogram that the errors were not symmetrically distributed.

⁷ Although not reported here, the variability in our nonparametric regression can be calculated by using the formula for the confidence interval in (5.13).