Model-free asymptotically best forecasting of stationary economic time series

Bierens, H.J.

1986

SERIE RESEARCH MEMORANDA

MODEL-FREE ASYMPTOTICALLY

BEST FORECASTING OF STATIONARY

ECONOMIC TIME SERIES

H.J. Bierens

Researchmemorandum 1986-32

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
A M S T E R D A M

MODEL-FREE ASYMPTOTICALLY BEST FORECASTING
OF STATIONARY ECONOMIC TIME SERIES

by

Herman J. Bierens
Free University (VU)
Department of Economics.
P.O.Box 7161
1007 MC Amsterdam

Abstract. Given observations on a stationary economic vector time series process we show that the best ℓ periods ahead forecast (best in the sense of having minimal forecast error variance) of one of the variables can be consistently estimated by nonparametric regression on an ARMA memory index. Our approach is based on a combination of the ARMA memory index modeling approach of Bierens (1986a) with a modification to time series of the nonparametric kernel regression approach of Devroye and Wagner (1980). This approach is truly model-free, as no explicit specification of the distribution of the data generating process is needed.

October 1986

1. INTRODUCTION

Let (z_t) be a vector-valued economic time series process and let y_t be one of the components of z_t. Given that $z_1,z_2,\ldots,z_n$ are observable we wish to forecast $y_{n+\ell}$, $\ell \geq 1$. It is well-known that the best predictor for $y_{n+\ell}$ is the conditional expectation of $y_{n+\ell}$ relative to the entire past $z_n,z_{n-1},z_{n-2},\ldots$ of the this vector time series process, as then the forecast error variance is minimal. In this paper it will be shown that under mild conditions it is possible to estimate this conditional expectation consistently without specifying a model, i.e., we shall construct a random variable $\hat{y}^{(\ell)}_{t+\ell}$ depending on $z_1,z_2,\ldots,z_n$ such that

$$(1.1) \qquad \text{plim}_{n\to\infty}\{\hat{y}^{(\ell)}_{n+\ell} - E(y_{n+\ell}|z_n,z_{n-1},z_{n-2},\ldots)\} = 0,$$

where the construction of $\hat{y}^{(\ell)}_{n+\ell}$ does not involve any parametric specification of the conditional expectation involved.

The advantage of our approach over the traditional forcasting schemes like ARX models, ARMAX models, ARIMAX models and structural economic modeling is that we don't need to bother about possible model misspecification, as no model is specified. As is well-known, misspecification of parametric forecasting schemes will lead to inefficient forecasts, whereas our approach always leads to asymptotically best forecasts.

A crucial condition for the validity of (1.1) is that the process (z_t) is rational-valued:

$$(1.2) \qquad z_t \in Q^k,$$

where Q is the set of rational numbers. Thus each of the k components of z_t is assumed to be rational-valued. In practice this is hardly a condition, as economic time series are always reported in a finite number of decimal digits and are therefore always rational-valued.

Another condition is that the process is strictly stationary. This condition is more restrictive, but it is often possible to transform a nonstationary process into a stationary (or "almost" stationary) process by differencing. Thus, the components of z_t may be transformations of economic variables.

2. OUTLINE OF THE FORECASTING PROCEDURE

In this section we shall give the recipe for constructing the ℓ period ahead forecast $\hat{y}_{n+\ell}^{(\ell)}$. Our approach combines the ARMA memory index modeling approach of Bierens (1986a) with a modification to time series of the nonparametric kernel regression approach of Devroye and Wagner (1980).

An ARMA memory index x_t with constant term takes the form

$$(2.1)\qquad x_t = \mu + \textstyle\sum_{s=1}^{q}\gamma_s x_{t-s} + \sum_{s=0}^{p}\beta_s{}'z_{t-s}$$

$$= \mu/(1-\textstyle\sum_{s=1}^{q}\gamma_s) + \{(\sum_{s=0}^{p}\beta_s L^s)/(1-\sum_{s=1}^{q}\gamma_s L^s)\}'z_t,$$

where $p \geq 0$ and $q \geq 1$ are given integers, L is the backwards lag operator, μ is a constant, $\beta_0,\beta_1,\ldots,\beta_p$ are vectors in R^k and $\gamma = (\gamma_1,\ldots,\gamma_q)'$ ϵ R^q is such that the lag polynomial $1 - \sum_{s=1}^{q}\gamma_s L^s$ is invertible. A sufficient condition for the latter is that this lag polynomial has roots all outside the unit circle. Denoting these roots by $r_s(\gamma)$, we now define the set of admissible γ's by

$$(2.2)\qquad \Gamma = \{\gamma \in R^q : |r_s(\gamma)^{-1}| < 1 \text{ for } s = 1,2,..,q\}.$$

Thus the γ_s in (2.1) satisfy:

$$(2.3)\qquad (\gamma_1,\ldots,\gamma_q)' \in \Gamma.$$

For notational convenience, let

$$(2.4)\qquad \theta = (\mu,\beta_0{}',\ldots,\beta_p{}',\gamma_1,\ldots,\gamma_q)',$$

$$(2.5)\qquad \Theta = R \times R^{(p+1)k} \times \Gamma,$$

$$(2.6)\qquad x_t(\theta) = x_t.$$

In Bierens (1986a) we have shown that under mild conditions

$$(2.7)\qquad E(y_{t+\ell}|z_t,z_{t-1},z_{t-2},\ldots) = E\{y_{t+\ell}|x_t(\theta)\}$$

with probability 1, for all $\theta \in \Theta$ except in some subset S of Θ with Lebesgue measure zero. Moreover, the conditional expectation (2.7) can be written as

$$(2.8) \qquad E\{y_{t+\ell}|x_t(\theta)\} = g_\theta(x_t(\theta)),$$

where g_θ is a Borel measurable real function depending on θ. Cf. Chung (1974, Theorem 9.1.2). This function g_θ can for given θ be consistently estimated by nonparametric regression methods. However, there are two problems to deal with before we actually can apply nonparametric regression. First, the ARMA memory index $x_t(\theta)$ cannot be calculated in practice, as only $z_1,..,z_n$ are observable. Second, for arbitrary θ the function g_θ is likely be highly nonlinear and discontinuous (see Bierens (1986a, Sec. 5), which renders consistent nonparametric estimation of g_θ awkward (though not impossible). The first problem can be solved by replacing $x_t(\theta)$ by a truncated ARMA memory index $\tilde{x}_t(\theta)$, where for example

$$(2.9) \qquad \tilde{x}_t(\theta) = \mu + \textstyle\sum_{s=1}^{q}\gamma_s\tilde{x}_{t-s}(\theta) + \sum_{s=0}^{p}\beta_s{}'\tilde{z}_{t-s},$$
$$\text{with } \tilde{z}_t = z_t \text{ for } t \geq 1,$$
$$= 0 \text{ for } t < 1.$$

There are also other ways to define such a truncated ARMA memory index $\tilde{x}_t(\theta)$, for example let the recursive scheme (2.9) start from $\tilde{x}_t(\theta) = y_{t+\ell}$ for $1-\ell \leq t \leq p$ and $\tilde{x}_t(\theta) = y_1$ for $t \leq -\ell$. However, in the sequel we shall work with (2.9).

It is easy to verify that $\text{plim}_{n\to\infty}\{\tilde{x}_t(\theta)-x_t(\theta)\} = 0$, which suggests to use $\tilde{x}_t(\theta)$ instead of $x_t(\theta)$. Although the function g_θ will in general be nonlinear and discontinuous, we can force it towards a linear function by choosing θ such that $E(y_{t+\ell}-x_t(\theta))^2$ is minimal. Denote this optimal θ by θ_o, i.e., let

$$(2.10) \qquad \theta_o = \text{argmin } E(y_{t+\ell}-x_t(\theta))^2$$

and assume that

$$(2.11)\qquad \theta_o \notin S,$$

where S is the exeptional set with Lebesgue measure zero for which (2.7) does not hold with probability 1. Then for $\theta = \theta_o$, (2.7) goes through while g_θ will be 'close' to a linear function.

We now propose to estimate θ_o by

$$(2.12)\qquad \hat{\theta}_n = \operatorname{argmin}\{\textstyle\sum_{t=p+1}^{n-\ell}(y_{t+\ell}-\tilde{x}_t(\theta))^2;\ \theta \in \bar{\Theta}\},$$

where $\bar{\Theta}$ is a compact subset of Θ with $\theta_o \in \bar{\Theta}$. Note that n should satisfy

$$(2.13)\qquad n \geq \ell+p+q+(p+1)k+1,$$

as otherwise the number of parameters exceeds the number of observations. It will be shown that

$$(2.14)\qquad \operatorname{plim}_{n\to\infty}(\hat{\theta}_n - \theta_o) = 0.$$

Next, let

$$(2.15)\qquad x_{n,t} = \tilde{x}_t(\hat{\theta}_n).$$

We now propose to run a nonparametric regression of $y_{t+\ell}$ on $x_{n,t}$, using the kernel method. This kernel regression function estimator takes the form

$$(2.16)\quad \hat{g}_n^{(\ell)}(x|\gamma_n) = \{\textstyle\sum_{t=p+1}^{n-\ell}y_{t+\ell}K((x-x_{n,t})/\gamma_n)\}/\{\textstyle\sum_{t=p+1}^{n-\ell}K((x-x_{n,t})/\gamma_n)\},$$

where K(.) is a real function, called the <u>kernel</u>, and γ_n is a <u>window width</u> parameter. Cf. Bierens (1985) for a review of kernel regression estimation methods. Following Devroye and Wagner (1980) we choose for K a function satisfying $\inf_{|x|<M}K(x) > 0$ for every positive number M. A typical example of such a kernel is the density of the standard normal distribution:

$$(2.17)\qquad K(x) = e^{-\frac{1}{2}x^2}/\sqrt{(2\pi)}.$$

In the sequel of this paper we shall work further with this normal kernel. It will be shown that if we choose the window width such that

(2.18) $\quad \gamma_n = \tau n^{-\varepsilon}$ with $\tau > 0$, $0 < \varepsilon < 1/6$,

then

(2.19) $\quad \text{plim}_{n\to\infty}\{\hat{g}_n^{(\ell)}(x_{n,t}|\gamma_n) - g_{\theta_o}(x_n(\theta_o))\} = 0.$

Thus $\hat{y}_{n+\ell}^{(\ell)} = \hat{g}_n^{(\ell)}(x_{n,n}|\gamma_n)$ is a forecast of $y_{t+\ell}$ satisfying (1.1). However, although any γ_n satisfying (2.18) will do, the small sample performance of such a forecast is not invariant for the choice of γ_n. Therefore we propose to optimize the window width as follows. Choose constants a, b, c and d such that

(2.20) $\quad 0 < a < b < \infty$, $0 < c < d < 1/6$

and choose an integer t_o such that

(2.21) $\quad \ell+p+q+(p+1)k+1 < t_o < n-\ell.$

Let

(2.22) $\quad (\hat{\tau},\hat{\varepsilon}) = \text{argmin}\{\sum_{t=t_o}^{n-\ell}(y_{t+\ell}-\hat{g}_t^{(\ell)}(x_{t,t}|\tau n^{-\varepsilon}))^2;\ a\leq\tau\leq b;\ c\leq\varepsilon\leq d\}.$

Then the proposed forecast $\hat{y}_{n+\ell}^{(\ell)}$ is:

(2.23) $\quad \hat{y}_{n+\ell}^{(\ell)} = \hat{g}_n^{(\ell)}(x_{n,n}|\hat{\tau}n^{-\hat{\varepsilon}}).$

One could argue that the ARMA memory index may be considered as a model and that therefore our claim that our approach is nonparametric and model-free is not correct. The point is, however, that in principle the ARMA memory index can be chosen quite arbitrarily. For the asymptotic results the choice of p, q and θ in (2.9) does not matter (as long as $\theta \notin S$). The same applies to the window width, as long as (2.18) holds.

The reason for estimating θ and γ_n is not to estimate true parameters but to enhance the small sample performance of the predictor. Thus our approach is truly model-free. Consequently, the role of economic theory in our approach concerns only the selection of the variables in z_t and not the specification of a model.

3. ASSUMPTIONS AND MAIN RESULT

The conditions we need for the validity of the results in Section 2, apart from the rationality assumption, are conditions on the moments of z_t and conditions on the memory of the process (z_t). With respect to the latter we require that the memory of the process (z_t) vanishes, i.e., z_t and z_{t-m} should become independent as m goes to infinity. For this we shall use two concepts, namely the ν-stability concept and the α-mixing concept. These two concepts together impose very weak conditions on the memory of the process.

Definition 1. Let (x_t) be a stochastic process in a a subset X of a Euclidean space, with the structure:

$$(3.1) \qquad x_t = f_t(u_t, u_{t-1}, u_{t-2}, \ldots),$$

where (u_t) is a stochastic process in a Euclidean space U and (f_t) is a sequence of Borel measurable mappings from the space of one-sided infinite sequences in U into X. For $r > 0$ and $m \geq 1$, let

$$(3.2) \qquad \nu(m) = \sup_t E||x_t - E(x_t|u_t, u_{t-1}, \ldots, u_{t-m+1})||^r.$$

If $\lim_{m\to\infty}\nu(m) = 0$ then (x_t) is said to be ν-stable in L^r with respect to the base (u_t). Moreover, if ν is such that $\nu(m) = O(e^{-cm})$ for some $c > 0$ then the process (x_t) is said to be exponentially stable in L^r with respect to the base (u_t).

This concept restricts the memory of the process (x_t) in that the impact of the remote past of u_t on x_t vanishes. We note that if (x_t) and (u_t)

are strictly stationary and $E||x_t||^r < \infty$ then (x_t) is automatically ν-stable in L^r with respect to (u_t). See Bierens (1986b, Theorem A1). Thus, for example, stationary ARMA processes with Gaussian white noise errors are ν-stable in L^r with respect to the error process, for every $r > 0$. Moreover, these ARMA processes are also exponentially stable.

Since we assume stationarity, the ν-stability condition itself does not impose additional restrictions. However, stationary processes, though ν-stable, are not necessarily exponentially stable. It is the latter concept we need. For a further discussion of the ν-stability concept, see Bierens (1983,1986a,b). Also, see Bierens (1981) for a discussion of the related stochastic stability concept.

In order that for a ν-stable process (x_t) the impact of x_{t-m} on x_t vanishes as $m \to \infty$ we also need to restrict the memory of the base (u_t). The condition we employ for that is the α-mixing condition.

<u>Definition 2</u>. Let (u_t) be a stochastic process in a Euclidean space U. Let F_t^{∞} be the Borel field generated by $u_t, u_{t+1}, u_{t+2}, \ldots$ and let $F_{-\infty}^{t}$ be the Borel field generated by $u_t, u_{t-1}, u_{t-2}, \ldots$. Let for $m \geq 0$,

$$(3.3) \qquad \alpha(m) = \sup_t \sup_{E_1 \in F_t^{\infty}, E_2 \in F_{-\infty}^{t-m}} |P(E_1 \text{ and } E_2) - P(E_1)P(E_2)|.$$

Then the process (u_t) is said to be α-mixing if $\lim_{m\to\infty}\alpha(m) = 0$.

Examples of α-mixing processes are independent processes, finite-dependent processes, and Gaussian AR(1) processes. See Ibragimov and Linnik (1971) for the AR(1) case and White and Domowitz (1984) for a further discussion of the α-mixing concept.

We are now able to state our basic assumption.

<u>Assumption 1</u>. The process (z_t) is a strictly stationary process in Q^k, with

$$(3.4) \qquad E||z_t||^{4+\delta} < \infty \text{ for some } \delta > 0.$$

Moreover, (z_t) is exponentially stable in $L^{4+\delta}$ with respect to an α-mixing base, where

$$(3.5) \qquad \textstyle\sum_{m=0}^{\infty} \alpha(m)^{\delta/(4+\delta)} < \infty.$$

This assumption is sufficient for (2.7) to hold:

Theorem 1. Let Assumption 1 hold and let y_t be one of the components of z_t. There exists a subset S of Θ (cf. (2.5)) with Lebesgue measure zero such that for each integer $\ell \geq 1$, each t and each $\theta \in \Theta - S$,

$$P\{E(y_{t+\ell}|z_t, z_{t-1}, z_{t-2}, ..) = E(y_{t+\ell}|x_t(\theta)\} = 1, \tag{3.6}$$

where $x_t(\theta)$ is defined by (2.6).

Proof. Bierens (1986a).

Next, denote similarly to (2.2),

$$\bar{\Gamma} = \{\gamma \in R^q : |r_s(\gamma)^{-1}| \leq \rho \text{ for } s=1,2,..,q;\ 0 < \rho < 1\}, \tag{3.7}$$

let $\bar{C}$ be a closed hypercube in $R^{1+(p+1)k}$ and let

$$\bar{\Theta} = \bar{C} \times \bar{\Gamma}. \tag{3.8}$$

Then $\bar{\Theta}$ is a compact subset of Θ. Cf. Bierens (1986b, Section 2.1). We now assume:

Assumption 2.

(I) There exists a unique point θ_o in $\bar{\Theta}$ such that

$$E\{y_{t+\ell} - x_t(\theta_o)\}^2 = \inf_{\theta \in \bar{\Theta}} E\{y_{t+\ell} - x_t(\theta)\}^2.$$

Moreover,

(II) θ_o lies in an open subset of $\bar{\Theta}$,

(III) $\theta_o \notin S$ (cf. Theorem 1),

(IV) the matrix $(\partial/\partial\theta)(\partial/\partial\theta')E\{y_{t+\ell} - x_t(\theta)\}^2$ is non-singular at θ_o.

Since the set S has Lebesgue measure zero the condition (III) is not too restrictive. The other conditions are similar to those for consistency and asymptotic normality of nonlinear least squares estimators. Cf. Bierens (1981).

The assumptions 1 and 2 are all we need:

Theorem 2. Let $\hat{y}_{t+\ell}^{(\ell)}$ be defined as in Section 2, where the set $\bar{\Theta}$ in (2.12) is defined by (3.8). Under Assumptions 1 and 2 we have

$$\text{plim}_{n\to\infty}\{\hat{y}_{n+\ell}^{(\ell)} - E(y_{n+\ell}|z_n, z_{n-1}, z_{n-2}, \dots)\} = 0.$$

The rest of this paper is devoted to the proof of Theorem 2.

4. PROPERTIES OF THE ARMA MEMORY INDICES

In this section we derive the properties of the ARMA memory indices $x_t(\theta)$, $\tilde{x}_t(\theta)$ and $\tilde{x}_t(\hat{\theta}_n)$. We consider first the properties of the ARMA memory index (2.1) and its truncation (2.9). Observe from (2.1) that $x_t(\theta)$ can be written as

$$(4.1) \qquad x_t = x_t(\theta) = \phi(\theta) + \textstyle\sum_{s=0}^{\infty}\eta_s(\theta)'z_{t-s},$$

where

$$(4.2) \qquad \phi(\theta) = \mu/(1 - \textstyle\sum_{s=1}^{q}\gamma_s),$$

$$(4.3) \qquad \textstyle\sum_{s=0}^{\infty}\eta_s(\theta)L^s = (\sum_{s=0}^{p}\beta_s L^s)/(1 - \sum_{s=1}^{q}\gamma_s L^s).$$

For the compact set $\bar{\Theta}$ defined by (3.8) we have:

Lemma 1. For some constant $c_0 > 0$,

$$\sup_{\theta\in\bar{\Theta}}||\eta_s(\theta)|| \leq c_0(1+s)^q\rho^s,$$

where $\rho \in (0,1)$ is given by (3.7).

Proof: Appendix.

Lemma 2. For some constant $c_1 > 0$ and each component θ_i of θ,

$$\sup_{\theta\in\bar{\Theta}}||(\partial/\partial\theta_i)\eta_s(\theta)|| \leq c_1(1+s)^{2q+2}\rho^s.$$

Proof: Appendix.

Lemma 3. For some constant $c_2 > 0$ and each pair (θ_i,θ_j) of components of θ,

$$\sup_{\theta\in\bar{\Theta}}||(\partial/\partial\theta_i)(\partial/\partial\theta_j)\eta_s(\theta)|| \leq c_2(1+s)^{3q+4}\rho^s.$$

Proof: Appendix.

We shall need these lemmas in order to establish the rate of convergence to zero of $(x_{t,n} - x_t)$.

Next, consider the truncated ARMA memory index $\hat{x}_t(\theta)$ (cf.(2.9)). Clearly we have

$$(4.4) \qquad x_t(\theta) - \hat{x}_t(\theta) = \textstyle\sum_{s=t}^{\infty}\eta_s(\theta)'z_{t-s}.$$

It follows now from Lemma 1, Assumption 1(I) and Holder's inequality that for $r \leq 4+\delta$,

$$(4.5) \qquad E\sup_{\theta\in\bar{\Theta}}|x_t(\theta) - \hat{x}_t(\theta)|^r \leq E\{\textstyle\sum_{s=t}^{\infty}c_0(1+s)^q\rho^s||z_{t-s}||\}^r = O(t^{rq}\rho^{rt}).$$

This result proves part (I) of Lemma 4 below. Along the same lines we can prove parts (II) and (III), using Lemmas 2 and 3. Thus:

Lemma 4. There exists a constant c_* such that for $t \geq 1$ and $0 < r \leq 4+\delta$,

$$(I) \qquad E\sup_{\theta\in\bar{\Theta}}|x_t(\theta) - \hat{x}_t(\theta)|^r \leq c_*t^{qr}\rho^{rt}$$

and for all components θ_i and θ_j of θ,

(II) $\quad E\sup_{\theta\in\bar{\Theta}}|(\partial/\partial\theta_i)x_t(\theta) - (\partial/\partial\theta_i)\tilde{x}_t(\theta)|^r \leq c_* t^{(2q+2)r}\rho^{rt}$,

(III) $\quad E\sup_{\theta\in\bar{\Theta}}|(\partial/\partial\theta_i)(\partial/\partial\theta_j)x_t(\theta) - (\partial/\partial\theta_i)(\partial/\partial\theta_j)\tilde{x}_t(\theta)|^r$

$$\leq c_* t^{(3q+4)r}\rho^{rt}.$$

Next, let

(4.6) $\quad \hat{Q}(\theta) = 1/(n-\ell-p)\sum_{t=p+1}^{n-\ell}(y_{t+\ell} - x_t(\theta))^2$,

(4.7) $\quad \tilde{Q}(\theta) = 1/(n-\ell-p)\sum_{t=p+1}^{n-\ell}(y_{t+\ell} - \tilde{x}_t(\theta))^2$.

Then Lemma 4 implies:

<u>Lemma 5</u>.

(I) $\quad E\sup_{\theta\in\bar{\Theta}}|\hat{Q}(\theta) - \tilde{Q}(\theta)| = O(1/n)$.

For all components θ_i and θ_j of θ,

(II) $\quad E\sup_{\theta\in\bar{\Theta}}|(\partial/\partial\theta_i)\hat{Q}(\theta) - (\partial/\partial\theta_i)\tilde{Q}(\theta)| = O(1/n)$,

(III) $\quad E\sup_{\theta\in\bar{\Theta}}|(\partial/\partial\theta_i)(\partial/\partial\theta_j)\hat{Q}(\theta) - (\partial/\partial\theta_i)(\partial/\partial\theta_j)\tilde{Q}(\theta)| = O(1/n)$.

Moreover, from Bierens (1986b, Theorem A3) and Lemmas 1, 2 and 3 it follows:

<u>Lemma 6</u>. Under Assumption 1 we have:

(I) $\quad \text{plim}_{n\to\infty}\sup_{\theta\in\bar{\Theta}}|\hat{Q}(\theta) - \bar{Q}(\theta)| = 0$,

(II) $\quad \text{plim}_{n\to\infty}\sup_{\theta\in\bar{\Theta}}|(\partial/\partial\theta_i)\hat{Q}(\theta) - (\partial/\partial\theta_i)\bar{Q}(\theta)| = 0$,

(III) $\quad \text{plim}_{n\to\infty}\sup_{\theta\in\bar{\Theta}}|(\partial/\partial\theta_i)(\partial/\partial\theta_j)\hat{Q}(\theta) - (\partial/\partial\theta_i)(\partial/\partial\theta_j)\bar{Q}(\theta)| = 0$, where θ_i and θ_j are arbitrary components of θ and

(4.8) $\bar{Q}(\theta) = E\hat{Q}(\theta) = E(y_{t+\ell} - x_t(\theta))^2.$

From part (I) of the Lemmas 5 and 6 and Bierens (1981, Lemma 3.1.8) it follows now that under Assumptions 1 and 2,

(4.9) $\text{plim}_{n\to\infty}\hat{\theta}_n = \theta_o.$

The rate of convergence in probability of $\hat{\theta}_n$ to θ_o can be established as follows. Consider the Taylor expansion of $(\partial/\partial\theta_i)\tilde{Q}(\hat{\theta}_n)$ around θ_o:

(4.10) $(\partial/\partial\theta_i)\tilde{Q}(\hat{\theta}_n) = (\partial/\partial\theta_i)\tilde{Q}(\theta_o) + \{(\partial/\partial\theta)(\partial/\partial\theta_i)\tilde{Q}(\tilde{\theta}_i)\}(\hat{\theta}_n-\theta_o),$

where $\tilde{\theta}_i$ is a mean value. Since $\hat{\theta}_n$ converges in probability to an interior point θ_o of $\bar{\Theta}$, we have

(4.11) $P((\partial/\partial\theta_i)\tilde{Q}(\hat{\theta}_n) = 0) \to 1$ as $n \to \infty$.

Moreover, we have:

<u>Lemma 7</u>. Under Assumption 1, $\sqrt{(n/\ln(n))}(\partial/\partial\theta_i)\hat{Q}(\theta_o)$ is stochastically bounded.

<u>Proof</u>: Appendix.

This result, combined with Lemma 5(II), yields:

(4.12) $\sqrt{(n/\ln(n))}(\partial/\partial\theta_i)\tilde{Q}(\theta_o)$ is stochastically bounded.

Furthermore, (4.9) and Lemmas 5(III) and 6(III) imply

(4.13) $\text{plim}_{n\to\infty}(\partial/\partial\theta)(\partial/\partial\theta_i)\tilde{Q}(\tilde{\theta}_i) = (\partial/\partial\theta)(\partial/\partial\theta_i)\bar{Q}(\theta_o).$

From (4.10) through (4.13) we now conclude, similarly to the standard argument for deriving asymptotic normality results for nonlinear least squares, that:

Lemma 8. Under Assumptions 1 and 2, $\surd(n/\ln(n))(\hat{\theta}_n - \theta_o)$ is stochastically bounded.

Combining Lemmas 4(I) and 8 now yields:

Theorem 3. Under Assumptions 1 and 2,

(I) $\{\surd(n/\ln(n))\}(1/(n-p-\ell))\sum_{t=p+1}^{n-\ell}(\tilde{x}_t(\hat{\theta}_n) - x_t(\theta_o))$ is stochastically bounded,

(II) $\{\surd(n/\ln(n))\}(1/(n-p-\ell))\sum_{t=p+1}^{n-\ell}(\tilde{x}_t(\hat{\theta}_n) - x_t(\theta_o))^2$ is stochastically bounded,

(III) $\surd(n/\ln(n))(\tilde{x}_n(\hat{\theta}_n) - x_n(\theta_o))$ is stochastically bounded.

Finally we have:

Theorem 4. Under Assumptions 1 and 2,

(I) $E|x_t(\theta_o)|^{4+\delta} < \infty$.

(II) Moreover, the process $(x_t(\theta_o))$ is exponentially stable in $L^{4+\delta}$ with respect to an α-mixing base, where

(4.14) $\sum_{m=0}^{\infty}\alpha(m)^{\delta/(4+\delta)} < \infty$.

Proof: Appendix.

5. KERNEL REGRESSION FUNCTION ESTIMATORS

In this section we show that the kernel regression function estimator (2.16) with kernel (2.17) and window width (2.18) satisfies

$$(5.1)\qquad \hat{g}_n^{(\ell)}(x_{n,n}|\gamma_n) - E(y_{n+\ell}|x_n) \to 0 \text{ in prob.}$$

where $x_t = x_t(\theta_o)$ and $x_{n,t}$ is defined by (2.15). The proof involves two main steps. First we show

$$(5.2)\qquad \hat{g}_n^{(\ell)}(x_{n,n}|\gamma_n) - \tilde{g}_n^{(\ell)}(x_n|\gamma_n) \to 0 \text{ in prob.,}$$

where

$$(5.3)\qquad \tilde{g}_n^{(\ell)}(x|\gamma_n) = \{\textstyle\sum_{t=p+1}^{n-\ell} y_{t+\ell}K((x-x_t)/\gamma_n)\}/\{\textstyle\sum_{t=p+1}^{n-\ell}K((x-x_t)/\gamma_n)\}$$

and then we show

$$(5.4)\qquad \tilde{g}_n^{(\ell)}(x_n|\gamma_n) - g(x_n) \to 0 \text{ in prob.,}$$

where g is a Borel measurable real function such that

$$(5.5)\qquad g(x_n) = E(y_{n+\ell}|x_n) \text{ a.s.}$$

Denote

$$(5.6)\qquad \hat{a}_n = (1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell} y_{t+\ell}K((x_{n,n}-x_{n,t})/\gamma_n),$$

$$(5.7)\qquad \tilde{a}_n = (1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell} y_{t+\ell}K((x_n-x_t)/\gamma_n),$$

$$(5.8)\qquad \hat{b}_n = (1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell} K((x_{n,n}-x_{n,t})/\gamma_n),$$

$$(5.9)\qquad \tilde{b}_n = (1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell} K((x_n-x_t)/\gamma_n).$$

Since we choose for K the standard normal density, it follows from the well-known inversion formula for characteristic functions that

$$(5.10)\qquad K(x) = (1/(2\pi))\int_{-\infty}^{\infty}\exp(-i\xi x)\exp(-\tfrac{1}{2}\xi^2)d\xi,$$

hence

$$(5.11)\qquad K((x_{n,n}-x_{n,t})/\gamma_n)$$
$$= (1/(2\pi))\int_{-\infty}^{\infty}\exp(-i\xi(x_{n,n}-x_{n,t})/\gamma_n)\exp(-\tfrac{1}{2}\xi^2)d\xi$$
$$= (\gamma_n/(2\pi))\int_{-\infty}^{\infty}\exp(-i\xi x_{n,n})\exp(i\xi x_{n,t})\exp(-\tfrac{1}{2}\gamma_n^2\xi^2)d\xi,$$

and similarly

$$(5.12)\qquad K((x_n-x_t)/\gamma_n) = (\gamma_n/(2\pi))\int_{-\infty}^{\infty}\exp(-i\xi x_n)\exp(i\xi x_t)\exp(-\tfrac{1}{2}\gamma_n^2\xi^2)d\xi.$$

Thus we can write:

$$(5.13)\qquad \hat{a}_n = (\gamma_n/(2\pi))\int_{-\infty}^{\infty}\{(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}y_{t+\ell}\exp(i\xi x_{n,t})\}$$
$$\times\exp(-i\xi x_{n,n})\exp(-\tfrac{1}{2}\gamma_n^2\xi^2)d\xi,$$

$$(5.14)\qquad \tilde{a}_n = (\gamma_n/(2\pi))\int_{-\infty}^{\infty}\{(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}y_{t+\ell}\exp(i\xi x_t)\}$$
$$\times\exp(-i\xi x_n)\exp(-\tfrac{1}{2}\gamma_n^2\xi^2)d\xi$$

and consequently

$$(5.15)\qquad |\hat{a}_n - \tilde{a}_n| \le$$
$$(\gamma_n/(2\pi))\int_{-\infty}^{\infty}|(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}y_{t+\ell}\{\exp(i\xi x_{n,t})-\exp(i\xi x_t)\}|$$
$$\times\exp(-\tfrac{1}{2}\gamma_n^2\xi^2)d\xi + (\gamma_n/(2\pi))(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}|y_{t+\ell}|$$
$$\times\int_{-\infty}^{\infty}|\exp(-i\xi x_{n,n})-\exp(-i\xi x_n)|\exp(-\tfrac{1}{2}\gamma_n^2\xi^2)d\xi$$
$$\le (\gamma_n/(2\pi))\{(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}|y_{t+\ell}||x_{n,t}-x_t|$$
$$- (1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}|y_{t+\ell}||x_{n,n}-x_n|\}\int_{-\infty}^{\infty}|\xi|\exp(-\tfrac{1}{2}\gamma_n^2\xi^2)d\xi$$

$$\leq (1/(2\pi\gamma_n))\{(1/(n-\ell-p))\sum_{t=p+1}^{n-\ell} y_{t+\ell}^2\}^{\frac{1}{2}}$$

$$\times\{(1/(n-\ell-p))\sum_{t=p+1}^{n-\ell}(x_{n,t}-x_t)^2\}^{\frac{1}{2}}\int_{-\infty}^{\infty}|\xi|\exp(-\tfrac{1}{2}\xi^2)d\xi$$

$$+ (1/(2\pi\gamma_n))\{(1/(n-\ell-p))\sum_{t=p+1}^{n-\ell}|y_{t+\ell}|\}|x_{n,n}-x_n|$$

$$\times\int_{-\infty}^{\infty}|\xi|\exp(-\tfrac{1}{2}\xi^2)d\xi$$

It follows now from (5.15) and Theorem 3:

<u>Lemma 9</u>. Under Assumptions 1 and 2, $\gamma_n\sqrt{(n/\ln(n))}(\hat{a}_n - \tilde{a}_n)$ is stochastically bounded.

Similarly we have:

<u>Lemma 10</u>. Under Assumptions 1 and 2, $\gamma_n\sqrt{(n/\ln(n))}(\hat{b}_n - \tilde{b}_n)$ is stochastically bounded.

Next, observe that

$$(5.16)\qquad \hat{g}_n^{(\ell)}(x_{n,n}|\gamma_n) - \tilde{g}_n^{(\ell)}(x_n|\gamma_n) = \hat{a}_n/\hat{b}_n - \tilde{a}_n/\tilde{b}_n$$

$$= \{(\hat{a}_n-\tilde{a}_n)\tilde{b}_n - \tilde{a}_n(\hat{b}_n-\tilde{b}_n)\}/\{(\hat{b}_n-\tilde{b}_n)+\tilde{b}_n^2\}.$$

Since obviously $\tilde{a}_n$ and $\tilde{b}_n$ are stochastically bounded it follows from (5.16) and Lemmas 9 and 10 that (5.2) holds if

$$(5.17)\qquad \gamma_n\sqrt{(n/\ln(n))}\tilde{b}_n^2 \to \infty \text{ in prob.}$$

For proving (5.17) we need the following notation and lemmas. Let F be the distribution function of x_t and let

$$(5.18)\qquad \bar{b}_n = \int_{-\infty}^{\infty}K((x_n-x)/\gamma_n)dF(x).$$

<u>Lemma 11</u>. For any sequence (c_n) satisfying

$$(5.19) \qquad c_n > 0,\ c_n \to 0,\ c_n/\gamma_n \to 0$$

we have $P(\bar{b}_n \geq c_n) \to 1$ as $n \to \infty$.

<u>Proof</u>: This lemma follows easily from the proof of Devroye and Wagner (1980, Lemma 3).

<u>Lemma 12</u>. Under Assumptions 1 and 2,

$$E|\tilde{b}_n - \bar{b}_n| = O(\sqrt{(\ln(n)/n)}) + O(1/(n\gamma_n))$$

<u>Proof</u>: It follows from (5.9), (5.12) and (5.18) that

$$(5.20) \qquad |\tilde{b}_n - \bar{b}_n| \leq (\gamma_n/(2\pi))\int_{-\infty}^{\infty}|(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(\exp(i\xi x_t)-E(\exp(i\xi x_t)))| \times\exp(-\tfrac{1}{2}\gamma_n^2\xi^2)d\xi.$$

Now let

$$(5.21) \qquad x_t^{(m)} = E(x_t|u_t,u_{t-1},u_{t-2},\ldots,u_{t-m+1}),$$

where (u_t) is the α-mixing base of (x_t). Cf. Theorem 4. Then

$$(5.22) \qquad E|(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(\exp(i\xi x_t)-E(\exp(i\xi x_t)))|$$

$$\leq (2/(n-\ell-p))|\xi|\textstyle\sum_{t=p+1}^{n-\ell}E|x_t - x_t^{(m)}|$$

$$+ E|(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(\exp(i\xi x_t^{(m)})-E(\exp(i\xi x_t^{(m)})))|$$

$$\leq 2|\xi|\{E|x_t - x_t^{(m)}|^{4+\delta}\}^{1/(4+\delta)}$$

$$+ \{E|(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(\cos(\xi x_t^{(m)})-E(\cos(\xi x_t^{(m)})))|^2\}^{\frac{1}{2}}$$

$$+ \{E|(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(\sin(\xi x_t^{(m)})-E(\sin(\xi x_t^{(m)})))|^2\}^{\frac{1}{2}},$$

where the first term at the right hand side of (5.22) follows from Liapounov's inequality and the last two terms follow from the well-known formula $e^{ix} = \cos(x) + i.\sin(x)$. Since $\{\cos(\xi x_t^{(m)})\}$ and $\{\sin(\xi x_t^{(m)})\}$ are bounded α^*-mixing sequences, where

$$(5.23) \qquad \alpha^*(j) = 1 \quad \text{if } j \leq m+1, \qquad = \alpha(j-m-1) \text{ if } j > m+1$$

(cf. the proof of Lemma 7), it follows from Lemma A1 in the appendix (with $r = \infty$),

$$(5.24) \qquad E\{(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(\cos(\xi x_t^{(m)}) - E(\cos(\xi x_t^{(m)})))\}^2$$

$$- (1/(n-\ell-p)^2)\textstyle\sum_{t=p+1}^{n-\ell}E\{\cos(\xi x_t^{(m)}) - E(\cos(\xi x_t^{(m)}))\}^2$$

$$\leq 12\{m+1+\textstyle\sum_{j=0}^{\infty}\alpha(j)\}/(n-\ell-p)$$

uniformly in ξ. The same result holds for $\sin(\xi x_t^{(m)})$. Note that (4.16) implies $\sum_{j=0}^{\infty}\alpha(j) < \infty$. Consequently, for sufficiently large m and n the right hand side of (5.24) is bounded by $c.m/n$, where $c > 12$. Thus it follows from (5.22) and Theorem 4 that for some constants c_*, c_1 and c_2,

$$(5.25) \qquad E|(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(\exp(i\xi x_t) - E(\exp(i\xi x_t)))|$$

$$\leq c_1|\xi|\exp(-c_* m) + c_2\sqrt{(m/n)}.$$

Combining (5.20) and (5.25) we now see that there exist constants c_*, c_1^* and c_2^* such that for large m and n,

$$(5.26) \qquad E|\tilde{b}_n - \hat{b}_n| \leq c_1^*\exp(-c_* m)/\gamma_n + c_2^*\sqrt{(m/n)}.$$

Taking $m \sim \ln(n)/c_*$, Lemma 12 follows. Q.E.D.

Let us return to the proof of (5.17). From Lemma 12 and (5.9) and (5.18) it follows

$$(5.27)\qquad E|\gamma_n\sqrt{(n/\ln(n))}\ (\tilde{b}_n^2 - \bar{b}_n^2)| = E|\gamma_n\sqrt{(n/\ln(n))}\ (\tilde{b}_n-\bar{b}_n)(\tilde{b}_n+\bar{b}_n)|$$

$$\leq |\gamma_n\sqrt{(n/\ln(n))}(\tilde{b}_n-\bar{b}_n)|2\sup_x K(x) = O(\gamma_n) + O(1/\sqrt{(n.\ln(n))}) \to 0$$

as $\gamma_n \to 0$, and from Lemma 11 it follows

$$(5.28)\qquad P\{\gamma_n\sqrt{(n/\ln(n))}\bar{b}^2{}_n > c_n^2\gamma_n\sqrt{(n/\ln(n))}\} \to 1 \text{ as } n \to \infty.$$

Thus it follows from (5.27) and (5.28) that (5.17) holds if c_n can be chosen such that

$$(5.29)\qquad c_n > 0,\ c_n \to 0,\ c_n/\gamma_n \to 0,\ c_n^2\gamma_n\sqrt{(n/\ln(n))} \to \infty,$$

while $\gamma_n \to 0$. Thus, let

$$(5.30)\qquad \gamma_n = \tau n^{-\varepsilon},\ \tau > 0,\ \varepsilon > 0,$$

$$(5.31)\qquad c_n = \tau n^{-\varepsilon-\delta},\quad \delta > 0.$$

Then $c_n > 0$, $c_n \to 0$, $c_n/\gamma_n = n^{-\delta} \to 0$. Moreover, if $\varepsilon < 1/6$ we can choose δ such that $c_n^2\gamma_n\sqrt{(n/\ln(n))} \to \infty$. This proves:

<u>Theorem 5</u>. Let Assumptions 1 and 2 hold. If

$$(5.32)\qquad \gamma_n = \tau n^{-\varepsilon},\ \text{where } \tau > 0 \text{ and } 0 < \varepsilon < 1/6$$

then $\text{plim}_{n\to\infty}\{\hat{g}_n^{(\ell)}(x_{n,n}|\gamma_n) - \tilde{g}_n^{(\ell)}(x_n|\gamma_n)\} = 0$.

Next we turn to the proof of (5.4). Denote similarly to (5.18),

$$(5.33)\qquad \bar{a}_n = \int_{-\infty}^{\infty} g(x)K((x_n-x)/n)dF(x).$$

Then similarly to (5.20) we have

$$(5.34)\qquad |\tilde{a}_n - \bar{a}_n| \le (\gamma_n/(2\pi))\int_{-\infty}^{\infty}|(1/(n-\ell-p))\sum_{t=p+1}^{n-\ell}(y_{t+\ell}\exp(i\xi x_t) - E(y_{t+\ell}\exp(i\xi x_t)))|\exp(-\tfrac{1}{2}\gamma_n^2\xi^2)d\xi,$$

hence, similarly to Lemma 12 we have:

Lemma 13. Under Assumptions 1 and 2,

$$E|\tilde{a}_n - \bar{a}_n| = O(\surd(\ln(n)/n)) + O(1/(n\gamma_n)).$$

Thus

$$(5.35)\qquad \min(\surd(n/\ln(n)),\ n\gamma_n)(\tilde{a}_n - \bar{a}_n) \text{ is stochastically bounded.}$$

Since also Lemma 12 implies that

$$(5.36)\qquad \min(\surd(n/\ln(n)),\ n\gamma_n)(\tilde{b}_n - \bar{b}_n) \text{ is stochastically bounded,}$$

it follows

$$(5.37)\qquad \text{plim}_{n\to\infty}(\tilde{a}_n-\bar{a}_n)/\tilde{b}_n = \text{plim}_{n\to\infty}(\tilde{a}_n-\bar{a}_n)/((\tilde{b}_n-\bar{b}_n)+\bar{b}_n) = 0$$

if

$$(5.38)\qquad \text{plim}_{n\to\infty}\min(\surd(n/\ln(n)),\ n\gamma_n)\bar{b}_n = \infty.$$

According to Lemma 11 the latter is satisfied if there exists a sequence (c_n) satisfying (5.19) such that

$$(5.39)\qquad c_n\min(\surd(n/\ln(n)),\ n\gamma_n) \to \infty.$$

Taking c_n and γ_n the same as in (5.30) and (5.31), we see that (5.39) holds for γ_n satisfying (5.32). Thus:

Lemma 14. Under the conditions of Theorem 5 we have

$$\text{plim}_{n\to\infty}(\tilde{a}_n-\bar{a}_n)/\tilde{b}_n \to 0.$$

The next step is to show that

$$(5.40)\qquad \text{plim}_{n\to\infty}(\tilde{a}_n/\tilde{b}_n - \bar{a}_n/\bar{b}_n) = 0,$$

which is true if

$$(5.41)\qquad \text{plim}_{n\to\infty}(1/\tilde{b}_n - 1/\bar{b}_n) = \text{plim}_{n\to\infty}(\bar{b}_n-\tilde{b}_n)/((\tilde{b}_n-\bar{b}_n)\bar{b}_n+\bar{b}_n^2) = 0.$$

In view of (5.36) is suffices to show

$$(5.42)\qquad \text{plim}_{n\to\infty}\min(\sqrt{(n/\ln(n))},\ n\gamma_n)\bar{b}_n^2 = \infty$$

and in view of Lemma 11 it suffices for that to verify

$$(5.43)\qquad c_n^2\min(\sqrt{(n/\ln(n))},\ n\gamma_n) \to \infty$$

for c_n satisfying (5.19). Taking γ_n and c_n as in (5.30) and (5.31) we see that this is possible for γ_n satisfying (5.32). Thus:

Lemma 15. Under the conditions of Theorem 5,

$$\text{plim}_{n\to\infty}(\tilde{a}_n/\tilde{b}_n - \bar{a}_n/\bar{b}_n) = 0.$$

The last step is to show that $\text{plim}_{n\to\infty}(\bar{a}_n/\bar{b}_n - g(x_n)) = 0$. For that we need the following lemma.

Lemma 16. Let x be a random variable and let g be a Borel measurable real function on R such that $E|g(x)| < \infty$. For every $\varepsilon > 0$ there exists a uniformly continuous bounded function f such that $E|g(x)-f(x)| < \varepsilon$.

Proof: Dunford and Schwartz (1957, p.298), as quoted by Devroye and Wagner (1980) and Bierens (1986a, Lemma 4).

Now let $x = x_t$ and let g be defined by (5.5). Moreover, let

$$(5.44)\qquad \bar{a}_n^{(1)} = \int_{-\infty}^{\infty} f(x)K((x_n-x)/\gamma_n)dF(x),$$

where f is given in Lemma 16. Then

$$(5.45)\qquad E|(\bar{a}_n-\bar{a}_n^{(1)})/\bar{b}_n| \le E\{\int_{-\infty}^{\infty}|g(x)-f(x)|K((x_n-x)/\gamma_n)dF(x)\} / \{\int_{-\infty}^{\infty}K((x_n-x)/\gamma_n)dF(x)\}$$

$$= \int_{-\infty}^{\infty}|g(x)-f(x)|\int_{-\infty}^{\infty}\{K((w-x)/\gamma_n)/\int_{-\infty}^{\infty}K((w-y)/\gamma_n)dF(y)\} dF(w)dF(x)$$

$$\le C\int_{-\infty}^{\infty}|g(x)-f(x)|dF(x),$$

where C is a constant only depending on K(.). The last step follows from Devroye and Wagner (1980, Lemma 1). Thus:

$$(5.46)\qquad E|(\bar{a}_n-\bar{a}_n^{(1)})/\bar{b}_n| < \varepsilon C.$$

Next, let

$$(5.47)\qquad \bar{a}_n^{(2)} = f(x_n)\int_{-\infty}^{\infty}K((x_n-x)/\gamma_n)dF(x) = f(x_n)\bar{b}_n.$$

Then

$$(5.48)\qquad |\bar{a}_n^{(1)}-\bar{a}_n^{(2)}| \le \int_{-\infty}^{\infty}|f(x_n)-f(x)|K((x_n-x)/\gamma_n)dF(x).$$

Since f is uniformly continuous and bounded there exists a $\kappa_\varepsilon > 0$ and a constant M_ε such that

$$(5.49)\qquad |x_n-x| \le \kappa_\varepsilon \Rightarrow |f(x_n)-f(x)| < \varepsilon,$$

and

$$(5.50) \qquad \sup_x |f(x)| \le M_\varepsilon .$$

From (5.48), (5.49) and (5.50) it easily follows

$$(5.51) \qquad |\bar{a}_n^{(1)} - \bar{a}_n^{(2)}| \le \varepsilon \bar{b}_n + 2M_\varepsilon \exp(-\tfrac{1}{2}\kappa_\varepsilon^2/\gamma_n^2),$$

hence

$$(5.52) \qquad |(\bar{a}_n^{(1)} - \bar{a}_n^{(2)})/\bar{b}_n| \le \varepsilon + 2M_\varepsilon \exp(-\tfrac{1}{2}\kappa_\varepsilon^2/\gamma_n^2)/\bar{b}_n .$$

Since $P(\bar{b}_n/c_n > 1) \to 1$ and

$$(5.53) \qquad \lim_{n\to\infty} \exp(-\tfrac{1}{2}\kappa_\varepsilon^2/\gamma_n^2)/c_n = 0$$

for c_n and γ_n defined by (5.30) and (5.31), we have

$$(5.54) \qquad P(|(\bar{a}_n^{(1)} - \bar{a}_n^{(2)})/\bar{b}_n| \le 2\varepsilon) \to 1.$$

Now observe that

$$(5.55) \qquad \bar{a}_n/\bar{b}_n - g(x_n) = (\bar{a}_n - \bar{a}_n^{(1)})/\bar{b}_n + (\bar{a}_n^{(1)} - \bar{a}_n^{(2)})/\bar{b}_n + f(x_n) - g(x_n).$$

Since by Lemma 16,

$$(5.56) \qquad E|f(x_n) - g(x_n)| < \varepsilon$$

and ε is arbitrary, it is easy to verify from (5.46), (5.54), (5.55) and (5.56) that:

Lemma 17. Under the conditions of Theorem 5,

$$\text{plim}_{n\to\infty}\{\bar{a}_n/\bar{b}_n - g(x_n)\} = 0.$$

Combining Lemmas 14, 15 and 17 now yields:

Theorem 6. Under Assumptions 1 and 2 and condition (5.32),

$$\text{plim}_{n\to\infty}\{g_n^{(\ell)}(x_n|\gamma_n) - g(x_n)\} = 0.$$

Theorems 5 and 6 now prove Theorem 2 for the case that γ_n is chosen in advance according to (5.32). The proof for the case that γ_n is determined by (2.22) is not too hard and therefore left to the reader.

6. CONCLUSIONS AND DISCUSSION

In this paper we have shown that stationary economic time series can be asymptotically best forecasted without specifying the data generating process. The procedure we propose consists of three steps. First we fit a possibly misspecified ARMA model to the data and we use the estimated ARMA model to forecast the variable of interest in and out the sample. Then we regres the actual values of the forecasted variable on the corresponding ARMA forecasts, using a nonparametric kernel regression estimator. Finally, the out of sample forecast on the basis of the ARMA model is corrected by using the nonparametric kernel regression function as a 'filter', in order to correct for the impact of the misspecification of the ARMA model.

The practical significance of this approach depends on the length of the time series under review and the extent of the misspecification of the ARMA model used as a preliminary forecasting scheme. If this ARMA model is highly misspecified, then the function relating the best forecast to the ARMA forecast will be highly nonlinear and likely be highly discontinuous. From an asymptotic point of view this is no problem, as consistency of the nonparametric kernel regression estimator only requires that the regression function to be estimated is Borel measurable. However, nonparametric estimation of highly nonlinear and discontinuous functions will likely require much more data than nonparametric estimation of smooth functions in order to get the same degree of accuracy. Consequently, it makes sense to select the preliminary ARMA model as carefully as possible in order to get as close as possible to the true data generating process. Thus we advocate to choose an ARMA model already having a good forecasting performance. The nonparametric regression approach then provides the 'finishing touch' in that the impact of the (modest) misspecification of the ARMA model is filtered out by the kernel regression estimator. In practice one should therefore not expect dramatic improvements of the ARMA forecasts.

Our results may also contribute to the rational expectations theory. The rational expectations (RE) hypothesis states that expectations of economic agents are basically conditional expectations relative to all available information about the past of the economic process involved. Rational expectations theorists usually assume that also the true model is known to the public. Opponents of the RE hypothesis, in particular Shiller (1978), have argued that this assumption cannot be taken seriously, as economists are only now discovering these models. Indeed, this is a very strong point. However, our results show that in principle rational expectations can be learned without knowing the model, provided the data generating process is stationary. In this process of learning rational expectations the ARMA forecast plays the role of an adaptive expectation. The nonparametric regression approach then transforms the non-rational adaptive expectation into a rational expectation, and can therefore be considered as the actual 'learning by doing' procedure.

APPENDIX

Proof of Lemma 1:

Since $\theta \in \bar{\Theta}$ implies $\gamma = (\gamma_1,\ldots,\gamma_q)' \in \bar{\Gamma}$ (cf. (3.8), (3.9)), the roots $r_1(\gamma),\ldots,r_q(\gamma)$ of the lag polynomial $1-\sum_{s=1}^{q}\gamma_s L^s$ satisfy

$$\text{(A1)} \qquad |r_s(\gamma)^{-1}| \leq \rho.$$

Moreover

$$\text{(A2)} \qquad (1-\sum_{s=1}^{q}\gamma_s L^s)^{-1} = \{\Pi_{s=1}^{q}(1-r_s(\gamma)^{-1}L)\}^{-1}$$

$$= \sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty}\cdots\sum_{s_q=0}^{\infty}\Pi_{j=1}^{q} r_{s_j}(\gamma)^{-s_j} L^{\sum_{j=1}^{q}s_j}$$

$$= \sum_{s=0}^{\infty}\sum_{s_1+s_2+\ldots+s_q=s}\Pi_{j=1}^{q} r_{s_j}(\gamma)^{-s_j} L^s$$

$$= \sum_{s=0}^{\infty}\xi_s(\gamma)L^s,$$

say. Thus by (A1),

$$\text{(A3)} \qquad \sup_{\gamma\in\bar{\Gamma}} |\xi_s(\gamma)| \leq \sum_{s_1=0}^{s}\sum_{s_2=0}^{s}\cdots\sum_{s_q=0}^{s}\rho^s = (1+s)^q\rho^s.$$

Next, observe from (4.3) and (A2) that

$$\text{(A4)} \qquad \sum_{s=0}^{\infty}\eta_s(\theta)L^s = (\sum_{s=0}^{\infty}\xi_s(\gamma)L^s)(\sum_{s=0}^{p}\beta_s L^s)$$

$$= \sum_{s=0}^{\infty}\sum_{j=0}^{\min(s,p)}\beta_s\xi_{s-j}(\gamma)L^s.$$

Thus it follows from (A3) and (A4) that

$$\text{(A5)} \qquad \sup_{\theta\in\bar{\Theta}}||\eta_s(\theta)|| \leq \sum_{j=0}^{\min(s,p)}\sup_{\theta\in\bar{\Theta}}||\beta_j||(1+s-j)^q\rho^{s-j}$$

$$\leq \sup_{\theta\in\bar{\Theta}}||\theta||\sum_{j=0}^{p}(1+s)^q\rho^{s-p}$$

$$= (1+p)\rho^{-p}\sup_{\theta\in\bar{\Theta}}||\theta||(1+s)^q\rho^s. \qquad \text{Q.E.D.}$$

<u>Proof of Lemma 2</u>:

Observe from (2.1) that

(A6) $$(\partial x_t/\partial\gamma_j) = \sum_{s=1}^{q}\gamma_s(\partial x_{t-s}/\partial\gamma_j) + x_{t-j},\quad j=1,2,..,q,$$

(A7) $$(\partial x_t/\partial\beta_j') = \sum_{s=1}^{q}\gamma_s(\partial x_{t-s}/\partial\beta_j') + z_{t-j},\ j=0,1,..,p.$$

From (A6), (4.1) and (A2) it follows

(A8) $$(\partial x_t/\partial\gamma_j) = (1-\sum_{s=1}^{q}\gamma_s L^s)^{-1}\phi(\theta) + (\sum_{s=0}^{\infty}\xi_s(\gamma)L^s)(\sum_{s=j}^{\infty}\eta_{s-j}(\theta)L^s)'z_t$$

$$= (1-\sum_{s=1}^{q}\gamma_s L^s)^{-1}\phi(\theta) + (\sum_{s=0}^{\infty}(\sum_{i=0}^{\infty}\sum_{m=j}^{\infty}I(i+m=s)\xi_i(\gamma)\eta_{m-j}(\theta))L^s)'z_t$$

$$= (\partial/\partial\gamma_j)\phi(\theta) + (\sum_{s=0}^{\infty}(\partial/\partial\gamma_j)\eta_s(\theta)L^s)'z_t,$$

where I(.) is the indicator function. It follows now from Lemma 1, (A3) and (A8) that

(A9) $$||(\partial/\partial\gamma_j)\eta_s(\theta)|| \le \sum_{i=0}^{\infty}\sum_{m=j}^{\infty}I(i+m=s)(1+i)^q\rho^i c_0(1+m-j)^q\rho^{m-j}$$

$$= c_0(1+s)^{2q}\rho^{s-j}\sum_{i=0}^{\infty}\sum_{m=j}^{\infty}I(i+m=s) \le c_0\rho^{-q}(\sum_{i=0}^{s}\sum_{m=0}^{s}1)(1+s)^{2q}\rho^s$$

$$\le c_0\rho^{-q}(1+s)^{2q+2}\rho^s.$$

Furthermore, it follows from (A7), (A2) and (4.1),

(A10) $$(\partial x_t/\partial\beta_j') = \{(1-\sum_{s=1}^{q}\gamma_s L^s)^{-1}L^j\}z_t = (\sum_{s=0}^{\infty}\xi_s(\gamma)L^{s+j})z_t$$

$$= (\sum_{s=j}^{\infty}\xi_{s-j}(\gamma)L^s)z_t = (\sum_{s=0}^{\infty}(\partial/\partial\beta_j')\eta_s(\theta)L^s)z_t.$$

Thus for each component $\beta_{i,j}$ of β_j we have

(A11) $$||(\partial/\partial\beta_{i,j})\eta_s(\theta)|| = k|\xi_{s-j}(\gamma)| \le k(s+1-j)^q\rho^{s-j} \le k\rho^{-p}(1+s)^q\rho^s \le k\rho^{-p}(1+s)^{2q+2}\rho^s.$$

Finally, observe from (4.3) that

(A12) $$(\partial/\partial\mu)\eta_s(\theta) = 0.$$

The lemma follows now from (A10), A(12) and (A13). Q.E.D.

<u>Proof of Lemma 3</u>:

It follows from (A6) and (A7),

(A13) $$(\partial^2 x_t/\partial\gamma_i\partial\gamma_j) = \sum_{s=1}^{q}\gamma_s(\partial^2 x_{t-s}/\partial\gamma_i\partial\gamma_j)+(\partial x_{t-j}/\partial\gamma_i)+(\partial x_{t-i}/\partial\gamma_j),$$

(A14) $$(\partial^2 x_t/\partial\gamma_i\partial\beta_j') = \sum_{s=1}^{q}\gamma_s(\partial^2 x_{t-s}/\partial\gamma_i\partial\beta_j')+(\partial x_{t-i}/\partial\beta_j'),$$

(A15) $$(\partial^2 x_t/\partial\beta_i\partial\beta_j') = \sum_{s=1}^{q}\gamma_s(\partial^2 x_{t-s}/\partial\beta_i\partial\beta_j').$$

From (A15) it follows:

(A16) $$(\partial^2 x_t/\partial\beta_i\partial\beta_j') = 0.$$

From (A13), (A2) and (4.1) it follows

(A17) $$(\partial^2 x_t/\partial\gamma_i\partial\gamma_j) = (\sum_{s=0}^{\infty}\xi_s(\gamma)L^s)$$

$$\times\{\sum_{s=i}^{\infty}(\partial/\partial\gamma_i)\eta_{s-i}(\theta)+\sum_{s=j}^{\infty}(\partial/\partial\gamma_j)\eta_{s-j}(\theta)\}$$

$$= \sum_{s=0}^{\infty}\{\sum_{h=0}^{\infty}\sum_{m=i}^{\infty}I(h+m=s)\xi_s(\gamma)(\partial/\partial\gamma_i)\eta_{m-i}(\theta)'$$

$$+\sum_{h=0}^{\infty}\sum_{m=j}^{\infty}I(h+m=s)\xi_s(\gamma)(\partial/\partial\gamma_j)\eta_{m-j}(\theta)'\}L^s z_t$$

$$= (\sum_{s=0}^{\infty}(\partial/\partial\gamma_i)(\partial/\partial\gamma_j)\eta_s(\theta)L^s)'z_t.$$

Thus

$$(A18)\quad ||(\partial/\partial\gamma_i)(\partial/\partial\gamma_j)\eta_s(\theta)||$$

$$\leq \sum_{h=0}^{\infty}\sum_{m=i}^{\infty} I(h+m=s)(1+h)^q\rho^h c_1(1+m-i)^{2q+2}\rho^{m-i}$$

$$+ \sum_{h=0}^{\infty}\sum_{m=j}^{\infty} I(h+m=s)(1+h)^q\rho^h c_1(1+m-j)^{2q+2}\rho^{m-j}$$

$$\leq 2\rho^{-q}c_1(1+s)^{3q+2}\rho^s\sum_{h=0}^{\infty}\sum_{m=0}^{\infty} I(h+m=s) \leq 2\rho^{-q}c_1(1+s)^{3q+4}\rho^s.$$

By a similar argument we can show that (A14) implies

$$(A19)\quad ||(\partial/\partial\gamma_i)(\partial/\partial\beta_{m,j})\eta_s(\theta)|| \leq c_2(1+s)^{3q+4}\rho^s,$$

where $\beta_{m,j}$ is the m-th component of β_j. The lemma follows now from (A16), (A18) and (A19). Q.E.D.

<u>Proof of Lemma 7</u>:

We have

$$(A20)\quad (\partial/\partial\theta_i)\hat{Q}(\theta_o) = (-2/(n-\ell-p))\sum_{t=p+1}^{n-\ell}(y_{t-\ell}-x_t(\theta_o))(\partial/\partial\theta_i)x_t(\theta_o)$$

$$= (-2/(n-\ell-p))\sum_{t=p+1}^{n-\ell} v_t w_t,$$

say, where

$$(A21)\quad v_t = y_{t+\ell} - x_t(\theta_o) = y_{t+\ell} - \phi(\theta_o) - \sum_{s=0}^{\infty}\eta_s(\theta_o),$$

$$(A22)\quad w_t = (\partial/\partial\theta_i)x_t(\theta_o) = (\partial/\partial\theta_i)\phi(\theta_o) + \sum_{s=0}^{\infty}(\partial/\partial\theta_i)\eta_s(\theta_o).$$

Let (u_t) be the base of (z_t) and denote

$$(A23)\quad y_t^{(m)} = E(y_t|u_t,u_{t-1},u_{t-2},\dots,u_{t-m+1}),$$

$$(A24)\quad z_t^{(m)} = E(z_t|u_t,u_{t-1},u_{t-2},\dots,u_{t-m+1}),$$

$$(A25)\quad v_t^{(m)} = y_{t-\ell}^{(m)} - \phi(\theta_o) - \sum_{s=0}^{m}\eta_s(\theta_o)'z_{t-s}^{(m)},$$

(A26) $$w_t^{(m)} = (\partial/\partial\theta_i)\phi(\theta_o) + \sum_{s=0}^{m}(\partial/\partial\theta_i)\eta_s(\theta_o)'z_{t-s}^{(m)}.$$

Note that $v_t^{(m)}$ is a function of $u_{t+\ell}$, $u_{t+\ell-1}, \ldots, u_{t-2m+1}$ and $w_t^{(m)}$ is a function of $u_t, u_{t-1}, \ldots, u_{t-2m+1}$. From Assumption 1, Lemmas 1 and 2 and Hölder's inequality it follows

(A27) $$E|v_t - v_t^{(m)}|^{4+\delta} \leq E\{|y_{t+\ell} - y_{t+\ell}^{(m)}| + \sum_{s=0}^{m} c_0(1+s)^q\rho^s||z_{t-s} - z_{t-s}^{(m)}|| + \sum_{s=m+1}^{\infty} c_0(1+s)^q\rho^s||z_{t-s}||\}^{4+\delta}$$

$$\leq O\{E|y_{t+\ell} - y_{t+\ell}^{(m)}|^{4+\delta} + \sum_{s=0}^{m} c_0^{4+\delta}(1+s)^{q(4+\delta)}\rho^{s(4+\delta)}E||z_{t-s} - z_{t-s}^{(m)}||^{4+\delta} + \sum_{s=m+1}^{\infty} c_0^{4+\delta}(1+s)^{q(4+\delta)}\rho^{s(4+\delta)}E||z_{t-s}||^{4+\delta}\}$$

$$= O(\nu(m)) + O(m^{q(4+\delta)}\rho^{m(4+\delta)})$$

uniformly in t. Since by Assumption 1, $\nu(m) = O(e^{-cm})$ for some $c > 0$ and since

(A28) $$m^{q(4+\delta)}\rho^{m(4+\delta)} \leq 2\rho^{m(4+\frac{1}{2}\delta)} = 2e^{\ln(\rho)(4+\frac{1}{2}\delta)m}$$

for m sufficiently large, it follows now that

(A29) $$E|v_t - v_t^{(m)}|^{4+\delta} = O(e^{-cm}) \text{ for some } c > 0.$$

Moreover, similarly we have

(A30) $$E|w_t - w_t^{(m)}|^{4+\delta} = O(e^{-cm}) \text{ for some } c > 0.$$

Using (A29), (A30) and Schwartz' inequality it is now easy to show that

(A31) $$E|v_t w_t - v_t^{(m)} w_t^{(m)}|^{2+\frac{1}{2}\delta} = O(e^{-cm}) \text{ for some } c > 0.$$

Next, observe that $v_t^{(m)} w_t^{(m)}$ depends on $u_{t+\ell}, \ldots, u_{t-2m+1}$. Since the u_t's are α-mixing, the process $(v_t^{(m)} w_t^{(m)})$ is α^*-mixing, with

$$\text{(A32)}\qquad \alpha^*(j) = 1 \qquad \text{if } j \le \ell+2m+1,$$
$$= \alpha(j-\ell-2m-1) \text{ if } j > \ell+2m+1.$$

From Lemma A1 below and the fact that

$$\text{(A33)}\qquad E|v_t^{(m)}w_t^{(m)}|^{2+\frac{1}{2}\delta} \le (E|v_t^{(m)}|^{4+\delta})^{\frac{1}{2}}(E|w_t^{(m)}|^{4+\delta})^{\frac{1}{2}}$$
$$\le (\nu(m)+E|v_t|^{4+\delta})^{\frac{1}{2}}(\nu(m)+E|w_t|^{4+\delta})^{\frac{1}{2}} < \infty$$

it follows now that for some constants c^*, c^{**} and m and n sufficiently large,

$$\text{(A34)}\qquad E\{(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(v_t^{(m)}w_t^{(m)}-E(v_t^{(m)}w_t^{(m)}))\}^2$$
$$- (1/(n-\ell-p)^2)\textstyle\sum_{t=p+1}^{n-\ell}E(v_t^{(m)}w_t^{(m)}-E(v_t^{(m)}w_t^{(m)}))^2$$
$$\le c^*\{\textstyle\sum_{j=0}^{\infty}\alpha^*(j)^{\delta/(4+\delta)}\}/\{n-\ell-p\}$$
$$\le c^*\{\ell+1+2m+\textstyle\sum_{j=0}^{\infty}\alpha(j)^{\delta/(4+\delta)}\}/\{n-\ell-p\} \le c^{**}m/n$$

hence

$$\text{(A35)}\qquad E|(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(v_t^{(m)}w_t^{(m)}-Ev_t^{(m)}w_t^{(m)})| \le \sqrt{c^{**}}\sqrt{(m/n)}.$$

Combining (A31) and (A35), we see that for some positive constants c, $c^{(1)}$, $c^{(2)}$ and sufficiently large m and n,

$$\text{(A36)}\qquad E|(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(v_tw_t-Ev_tw_t)| \le c^{(1)}\sqrt{(m/n)} + c^{(2)}e^{-cm}.$$

Minimizing (A36) to m yields

$$\text{(A37)}\qquad E|(1/(n-\ell-p))\textstyle\sum_{t=p+1}^{n-\ell}(v_tw_t-Ev_tw_t)| = O(\sqrt{(\ln(n)/n)}).$$

By Chebishev's inequality, this proves the lemma. Q.E.D.

Lemma A1. Let (u_t) be an α-mixing process in R with $\sup_t\{E|u_t|^r\}^{1/r} = M < \infty$ for some $r > 2$. If $\sum_{m=0}^{\infty}\alpha(m)^{(r-2)/r} < \infty$ then

$$|E\{(1/n)\sum_{t=1}^{n}(u_t-E(u_t))\}^2 - (1/n^2)\sum_{t=1}^{n}E(u_t-E(u_t))^2|$$

$$\leq 12M^2(1/n)\sum_{m=0}^{\infty}\alpha(m)^{(r-2)/r}.$$

Proof: Let $a \in F_{t+m}^{\infty}$, $b \in F_{-\infty}^{t}$ (cf. Definition 2), with $\{E|a|^r\}^{1/r} < \infty$ and $\{E|b|^r\}^{1/r} < \infty$ for some $r > 2$. Then it follows from McLeish (1975, Lemma 2.1) that

(A38) $|E(ab)-E(a)E(b)| \leq 2(2^{(r-1)/r}+1)\alpha(m)^{(r-2)/r}\{E|a|^r\}^{1/r}\{E|b|^r\}^{1/r}.$

Taking $a = u_{t+m}-E(u_{t+m})$ and $b = u_t-E(u_t)$ we therefore have

(A39) $|Cov(u_{t+m},u_t)| \leq 6M^2\alpha(m)^{(r-2)/r}.$

Consequently,

(A40) $$|E\{(1/n)\sum_{t=1}^{n}(u_t-E(u_t))\}^2 -(1/n^2)\sum_{t=1}^{n}(u_t-E(u_t))^2|$$

$$\leq 2(1/n^2)\sum_{t=1}^{n-1}\sum_{j=t+1}^{n}|Cov(u_j,u_t)| \leq 2(1/n^2)\sum_{t=1}^{n}\sum_{j=0}^{\infty}|Cov(u_{t+j},u_t)|$$

$$\leq 12(1/n)M^2\sum_{j=0}^{\infty}\alpha(j)^{(r-2)/r}.$$ Q.E.D.

Proof of Theorem 4:

Let (u_t) be the base, let $z_t^{(m)}$ be defined by (A24) and let $x_t = x_t(\theta_o)$. Denote for even m:

(A41) $$x_t^*(m) = \phi(\theta) + \sum_{s=0}^{\frac{1}{2}m}\eta_s(\theta)'z_{t-s}^{(\frac{1}{2}m)}.$$

Then by Lemma 1 and Hölder's inequality,

(A42) $$E|x_t-x_t^*(m)|^{4+\delta} \leq E\{\sum_{s=0}^{\frac{1}{2}m}c_o(1+s)^{q}\rho^{s}||z_{t-s}-z_{t-s}^{(\frac{1}{2}m)}||$$

$$+ \sum_{s=\frac{1}{2}m+1}^{\infty}c_o(1+s)^{q}\rho^{s}||z_{t-s}||\}^{4+\delta}$$

$$\leq C\{\sum_{s=0}^{\frac{1}{2}m}(1+s)^{q(4+\delta)}\rho^{s(4+\delta)}E||z_{t-s}-z_{t-s}^{(\frac{1}{2}m)}||^{4+\delta}$$

$$+ \sum_{s=\frac{1}{2}m+1}^{\infty}(1+s)^{q(4+\delta)}\rho^{s(4+\delta)}E||z_{t-s}||^{4+\delta}\}$$

$$\leq C^{(1)}\nu(\tfrac{1}{2}m) + C^{(2)}\rho^{\frac{1}{2}m(4+\delta)}m^{q(4+\delta)}$$

for some constants C, $C^{(1)}$, $C^{(2)}$. It follows now easily from (A42) and Assumption 1 that there exist constants C_{**} and c such that for sufficiently large m,

(A43) $$E|x_t-x_t^*(m)|^{4+\delta} \leq C_{**}e^{-cm}.$$

Using Jensen's inequality for conditional expectations (cf. Bierens (1983, Lemma 3)), it follows from (A43) that part II of Theorem 4 holds. Proving part I is easy and therefore left to the reader. Q.E.D.

REFERENCES

1. Bierens, H.J. (1981). Pobust Methods and Asymptotic Theory in Nonlinear Econometrics. New York: Springer Verlag.
2. Bierens, H.J. (1983) .Uniform Consistency of Kernel Estimators of a Regression Function under Generalized Conditions. Journal of the American Statistical Association,77, 699-707
3. Bierens, H.J. (1985) .Kernel Estimators of Regression Functions. Invited paper presented at the Symposium on Nonparametric and Robust Estimation in the Fifth World Congress of the Econometric Society, August 1985 (forthcoming in :Advances in Econometrics 1985, Cambridge University Press)
4. Bierens, H.J. (1986a) .ARMA Memory Index Modeling of Economic Time Series (with discussion). Econometric Theory (forthcoming)
5. Bierens, H.J. (1986b) .ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands. Annals of Econometrics (forthcoming)
6. Chung, K.L. (1974) .A Course in Probability Theory. New York-London: Academic Press.
7. Devroye, L.P. and Wagner, T.J. (1980) .Distribution-Free Consistency Results in Non-parametric Discrimination and Function Estimation. Annals of Statistics,8, 231-239
8. Dunford, N. and Schwartz, J.T. (1957) .Linear Operators, Part I. General Theory. New York: Interscience.
9. Ibragimov, I.A. and Linnik, Yu.V. (1971) .Independent and Stationary Sequences of Random Variables. Groningen: Wolters-Noordhoff.
10. McLeish, D.L. (1975) .A Maximal Inequality and Dependent Strong Laws. Annals of Probability,3, 829-839
11. Shiller, R.J. (1978) .Rational Expectations and the Dynamic Structure of Macroeconometric Models: A Critical Review. Journal of Monetary Economics,4, 1-44
12. White, H. and Domowitz, I. (1984) .Nonlinear Regression with Dependent Observations. Econometrica,52, 143-161.