Journal of Econometrics 26 (1984) 323–353. North-Holland

# MODEL SPECIFICATION TESTING OF TIME SERIES REGRESSIONS

Herman J. BIERENS*

*University of Amsterdam, 1011 NH Amsterdam, The Netherlands*

Received December 1982, final version received January 1984

Given the specification of the lag length and functional form of a (non)linear time series regression we shall propose a test of the null hypothesis that the expectation of the error conditional on the exogenous variables, all lagged exogenous variables and all lagged dependent variables equals zero with probability 1. In the case that the data-generating process is strictly stationary this test is consistent with respect to the alternative hypothesis that the null is false. The test is also applicable for a particular class of non-stationary time series regressions, although in that case consistency with respect to all possible alternatives is no longer guaranteed. The test involved is a generalization of a test proposed in Bierens (1982b). Moreover, we also present a similar but simpler test of the hypothesis that the errors are martingale differences.

## 1. Introduction

In a previous paper [Bierens (1982b)] we have introduced two model specification tests of nonlinear regression models. These tests test the null hypothesis that the expectation of the errors conditional on the regressors equals zero with probability 1 against the alternative hypothesis that the null is false. Both tests are consistent under the maintained hypothesis that the data-generating process is i.i.d.; i.e., if the maintained hypothesis is true then the type II error vanishes as the number of observations increases to infinity. This implies that asymptotically *any* misspecification of the model will be detected. In the present paper we shall extend one of these tests (i.e., test 1) to (non)linear time series regressions, using the results in Bierens (1981, 1982a). Roughly speaking, we shall test the null hypothesis that the regression function equals the conditional expectation function, where the conditioning is on the explanatory variables and all lagged dependent and explanatory variables. In the case that the data-generating process is strictly stationary we employ the

*I wish to thank an anonymous referee for his very helpful comments. An earlier version of this paper was written while I was spending a three-month term at the University of Minnesota, whose kind hospitality is gratefully acknowledged. Research in this paper was financed, in part, by a grant from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

general alternative hypothesis that the null is false. In the non-stationary case the alternative is slightly less general, but still it will cover a wide range of misspecifications. In fact, by testing this null hypothesis we do not only test whether the functional form is correctly specified, but we then also test the absence of autocorrelation of the errors, as the null hypothesis involved implies that the errors are martingale differences. The test involved appears to be laborious for large models. Therefore we also propose a similar but much simpler test of the hypothesis that the errors are martingale differences, as in practice misspecification of the model will usually imply that the martingale difference hypothesis fails to hold.

The plan of the paper is as follows. In section 2 we discuss the model and the null hypothesis. In section 3 we consider the asymptotic properties of the least squares estimators of the parameters of non-linear time series regressions. Section 4 shows how the null hypothesis can be identified. Section 5 outlines the test procedure and the asymptotic theory of the test. In section 6 we demonstrate the performance of the test by some numerical examples. In section 7 we show that by sacrificing some small sample power the test can substantially be simplified. In section 8 we discuss the test of the martingale difference hypothesis. The results in the sections 4 through 8 have been derived under the assumption that the data-generating process is strictly stationary. In section 9 we show what happens if we drop the stationarity assumption.

Most of the proofs of our results will be given in an extended version of the present paper [Bierens (1983)] which is available from the author on request.

## 2. The model and the null hypothesis

Consider a vector time series process $(z_j)$ in a Euclidean space $Z$. It is well-known that if we wish to forecast $z_j$ on basis of the history up to period $j-1$ of the process involved, the best forecasting scheme (i.e., the one which minimizes the variance of the forecast error) is just the expectation of $z_j$ conditional on *all* lagged $z_j$. Consequently the best time series regression model is the one which represents this conditional expectation, as such a model employs the maximum amount of explanatory power contained in the past of the process.

The expectation of $z_j$ conditional on all lagged $z_j$'s is essentially a $Z$-valued (Borel measurable) function $G_j$, say, of the one-sided infinite sequence of lagged $z_j$'s [compare Chung (1974, theorem 9.1.2)], i.e.,

$$\mathrm{E}(z_j|z_{j-1}, z_{j-2}, \dots) = G_j(z_{j-1}, z_{j-2}, \dots) \quad \text{a.s.} \tag{1}$$

Defining

$$e_j = z_j - \mathrm{E}(z_j|z_{j-1}, z_{j-2}, \dots), \tag{2}$$

the model then takes the tautological form

$$z_j = G_j(z_{j-1}, z_{j-2}, \ldots) + e_j. \tag{3}$$

If the process $(z_j)$ is strictly stationary then the regression functions $G_j$ are time invariant, i.e.,

$$G_j = G \quad \text{for all } j, \tag{4}$$

but if not there is no guarantee that such a common regression function $G$ exists.

In specifying this model the usual practice is to assume that:

(a) there is a common regression function, so that (4) holds;

(b) this common regression function is a function of only a finite number of lagged $z_j$'s, say,

$$G(z_{j-1}, z_{j-2}, \ldots) = G(z_{j-1}, z_{j-2}, \ldots, z_{j-m}); \tag{5}$$

(c) the regression function $G$ equals a known $Z$-valued function $F$ of the $m$ lagged $z_j$'s and an unknown parameter vector $\theta_0$,

$$G(z_{j-1}, z_{j-2}, \ldots, z_{j-m}) = F(z_{j-1}, z_{j-2}, \ldots, z_{j-m}, \theta_0) \quad \text{a.s.} \tag{6}$$

Thus the regression model is specified as

$$z_j = F(z_{j-1}, z_{j-2}, \ldots, z_{j-m}, \theta_0) + e_j. \tag{7}$$

If this specification is correct, i.e., if indeed for this particular lag length $m$ and functional form $F$ we have

$$\mathrm{E}(z_j | z_{j-1}, z_{j-2}, \ldots) = F(z_{j-1}, z_{j-2}, \ldots, z_{j-m}, \theta_0) \quad \text{a.s.,} \tag{8}$$

for all $j$ and some $\theta_0$, then the error process $(e_j)$ satisfies

$$\mathrm{E}(e_j | z_{j-1}, z_{j-2}, \ldots) = 0 \quad \text{a.s.} \tag{9}$$

Models of the type (7) satisfying (9), play, for example, a role in the alternative approach to macroeconomic modelling advocated by Sims (1980, 1981). Moreover, least squares estimation of such models under the assumptions of strict stationarity and ergodicity has been considered by Sims (1976) and Klimko and Nelson (1978).

Specification of a time series model does not only involve the choice of the lag length and the functional form but also the choice of the variables to be taken into account. In fact the choice of the variables is crucial, as generally the functional form and the lag length of the model will change if we add or delete components of $z_j$. For example, consider the bivariate time series process

$$z_{1,j} = \rho z_{1,j-1} + \zeta_{j-1} \sum_{l=1}^{\infty} \lambda_j^l z_{1,j-l} + \varepsilon_j, \qquad 0 < |\lambda_j| < 1,$$

$$z_{2,j} = \mu z_{2,j-1} + \zeta_j, \tag{10}$$

$$z_j' = (z_{1,j}, z_{2,j}),$$

where the error processes $(\varepsilon_j)$ and $(\zeta_j)$ are independent with zero mean. Then

$$\mathrm{E}(z_{1,j}|z_{j-1}, z_{j-2}, \dots) = \rho z_{1,j-1} + (z_{2,j-1} - \mu z_{2,j-2}) \sum_{l=1}^{\infty} \lambda_j^l z_{1,j-l},$$

$$\mathrm{E}(z_{2,j}|z_{j-1}, z_{j-2}, \dots) = \mu z_{2,j-1}, \tag{11}$$

whereas

$$\mathrm{E}(z_{1,j}|z_{1,j-1}, z_{1,j-2}, \dots) = \rho z_{1,j-1}. \tag{12}$$

Thus in the former case the regression function corresponding to $z_{1,j}$ is a function of all lagged $z_{1,j}$'s and some lagged $z_{2,j}$'s depending on the time index $j$, whereas in the latter case the regression function is a time invariant function of only one lagged $z_{1,j}$.

The problem of how to select the variables to be included in an economic time series model is in the first instance an economic problem rather than a statistical problem. The initial choice of the variables should be made on basis of economic theory. After then statistical inference may help us to determine whether the specified model is appropriate for modelling the data under review, and on basis of the results of causality tests [see Granger (1969) and Sims (1977)] one may even reduce the data set by deleting redundant variables. In the following, however, we shall limit our attention to the problem of how to test the correctness of the model specification, given the data-generating process. The problem of selecting the variables will therefore not be further considered.

In this paper we shall propose a test of the hypothesis that for $j \geq 1$ the model (7) represents the mathematical expectation of $z_j$ conditional on the *whole* past of the process, on basis of a data set

$$\{z_{-m+1}, \ldots, z_0, z_1, \ldots, z_n\} \quad \text{of size} \quad m+n, \tag{13}$$

where $m$ is fixed. Asymptotic results will be derived for $n \to \infty$. In particular we shall focus on the first equation of the system (7). Partitioning $z_j$ and $e_j$ in

$$z_j' = (y_j, x_j') \in \boldsymbol{R} \times \boldsymbol{R}^s, \qquad e_j' = (u_j, v_j') \in \boldsymbol{R} \times \boldsymbol{R}^s, \tag{14}$$

this first equation can be written as

$$y_j = f(y_{j-1}, \ldots, y_{j-p}, x_{j-1}, \ldots, x_{j-q}, \theta_0) + u_j, \qquad j \geq 1, \tag{15}$$

where $f$ is a known Borel measurable real function on $\boldsymbol{R}^{p+q \cdot s} \times \Theta$ with $\Theta$ a parameter space, and

$$\max(p, q) \leq m. \tag{16}$$

In order that the regression model (15) represents the conditional expectation of $y_j$ given all lagged $z_j$'s the following two conditions should be satisfied:

*Assumption 1.* There exist positive integers $p$ and $q$ such that for all non-negative integers $l$ and all $j \geq 1$,

$$\mathrm{E}(y_j | y_{j-1}, \ldots, y_{j-p}, x_{j-1}, \ldots, x_{j-q})$$
$$= \mathrm{E}(y_j | y_{j-1}, \ldots, y_{j-p-l}, x_{j-1}, \ldots, x_{j-q-l}) \quad \text{a.s.}$$

*Assumption 2.* Given the lag lengths $p$ and $q$ the specified regression function $f$ is such that for some $\theta_0 \in \Theta$ and all $j \geq 1$,

$$\mathrm{E}(y_j | y_{j-1}, \ldots, y_{j-p}, x_{j-1}, \ldots, x_{j-q})$$
$$= f(y_{j-1}, \ldots, y_{j-p}, x_{j-1}, \ldots, x_{j-q}, \theta_0) \quad \text{a.s.}$$

Now the null hypothesis to be tested is:

$$\mathrm{H}_0: \quad \text{Assumptions 1 and 2 hold.} \tag{17}$$

In the case that the data-generating process is strictly stationary, this null hypothesis will be tested against the alternative hypothesis that the null is false. In the non-stationary case, however, this general alternative hypothesis is no longer manageable, due to the fact that under the alternative the true regression model may now depend on the time index $j$. For example it will be impossible to distinguish the case that Assumptions 1 and 2 fail to hold for a finite number of $j \geq 1$, as asymptotically this case is 'almost' the same as the null hypothesis (17). Nevertheless, also in the case of non-stationary data it will be possible to test the null against an alternative hypothesis which covers a wide range of misspecifications, as will be shown in section 9.

Before we discuss the test we first pay attention to the conditions under which the least squares estimators of the parameters of non-linear time series regression models are consistent and asymptotically normally distributed, as these conditions and results will also be needed for deriving the asymptotic properties of our model specification test.

## 3. Stochastic stability, $\phi$-mixing and least squares estimation

In Bierens (1982a) we have proved weak consistency and asymptotic normality of least squares estimators of the parameters of a non-linear time series regression model, assuming that the errors are independent and the data are generated by a *stochastically stable* process with respect to a $\phi$-mixing *base*. In this section we show that the independence assumption on the errors can be relaxed; i.e., we show that Assumptions 1 and 2, together with the stochastic stability and $\phi$-mixing conditions and some regularity conditions, are sufficient for weak consistency and asymptotic normality of least squares estimators. Moreover, we shall also weaken the moment conditions and the continuity condition on the regression function which have been employed in Bierens (1982a).

If the data-generating process is strictly stationary consistency and asymptotic normality of least squares estimators of non-linear time series regressions can also be derived by assuming ergodicity [see Hannan (1971), Robinson (1972), Sims (1976), Klimko and Nelson (1978) and Hansen (1982)]. However, the condition that the data-generating process is stochastically stable with respect to a $\phi$-mixing base is a sufficient condition for a weak form of ergodicity; hence this condition is of lower level than ergodicity. Moreover, the stochastic stability concept is closely related to the usual stability condition for linear autoregressive processes, and is therefore easier to verify than ergodicity. Furthermore, our approach allows for a unified treatment of stationary and non-stationary time series.

The stochastic stability concept has been introduced in Bierens (1981, ch. 5). In order to make this concept clear we consider a non-linear vector autoregressive stochastic process of the form

$$z_j = \Gamma_j(z_{j-1}, z_{j-2}, \ldots) + v_j, \qquad -\infty < j < \infty, \tag{18}$$

where the $z_j$ and $v_j$ are (possibly non-stationary) vector time series processes in $\boldsymbol{R}^r$, and the $\Gamma_j$ are Borel measurable mappings from the space of one-sided infinite sequences in $\boldsymbol{R}^r$ into $\boldsymbol{R}^r$.

By $m$ times backwards substitution of (18) we can write

$$z_j = \Gamma^*_{m,j}(v_j, v_{j-1}, \ldots, v_{j-m}, z_{j-m-1}, z_{j-m-2}, \ldots), \qquad -\infty < j < \infty. \tag{19}$$

Roughly speaking this stochastic process is a stable process (in a similar sense as for deterministic difference equations) if the impact of the initial values $z_{j_0-1}, z_{j_0-2}, \ldots, j_0 < j$, on the distribution of $z_j$ vanishes if $j - j_0 \to \infty$. In that case we can 'approximate' $z_j$ by a function of a finite number of $v_j$'s, say, $v_j, v_{j-1}, \ldots, v_{j-m}$, for example by replacing the random vectors $z_{j-m-l}$, $l \geq 1$, in (19) by appropriate constant vectors $c_{j-m-l}$, $l \geq 1$, say. Thus we then approximate $z_j$ by

$$z_j^{(m)} = \Gamma^*_{m,j}(v_j, v_{j-1}, \ldots, v_{j-m}, c_{j-m-1}, c_{j-m-2}, \ldots).$$

Another way of approximating $z_j$ is to take the expectation of $z_j$ conditional on $v_j, \ldots, v_{j-m}$, i.e.,

$$z_j^{(m)} = \mathrm{E}(z_j | v_j, v_{j-1}, \ldots, v_{j-m}).$$

Now if for such an approximation, only depending on $v_j, \ldots, v_{j-m}$, there exists a sequence $(m_n)$ of positive integers such that together

$$m_n = \mathrm{o}(n) \quad \text{for} \quad n \to \infty,$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mathrm{P}\left(\|z_j - z_j^{(m_n)}\| > \delta\right) = 0 \quad \text{for every} \quad \delta > 0,$$

then the process $(z_j)$ is said to be *stochastically stable* with respect to the *base* $(v_j)$ [see Bierens (1981, definition 5.4.1)].

As is shown in Bierens (1981, section 5.1), stochastic stability of linear autoregressions is closely related to the usual stability condition that the lag operator on $z_j$ has roots all outside the unit circle. If, for example, the $z_j$ and $v_j$ are real valued and if the $\Gamma_j$ are linear time invariant functions of $p$ lagged

$z_j$'s, then (18) can be written as

$$z_j = \alpha_1 z_{j-1} + \cdots + \alpha_p z_{j-p} + v_j + \gamma,$$

or equivalently

$$A(L) z_j = v_j + \gamma,$$

where

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p$$

is a polynomial lag-operator. If all the roots of $A(L)$ lie outside the unit circle, then $(z_j)$ is stochastically stable with respect to the base $(v_j)$ [see Bierens (1981, theorem 5.1.2)].

Note that stationarity is not required for stochastic stability. For example, it is not hard to show [using the approach in Bierens (1981, p. 158)] that the univariate first-order autoregressive process $(z_j)$, defined by

$$z_j = \theta_j z_{j-1} + v_j, \qquad \sup_{-\infty < j < \infty} |\theta_j| < 1,$$

is stochastically stable with respect to $(v_j)$.

It is not hard to verify that if $(z_j)$ is stochastically stable with respect to a base, then for any positive integer $k$ $\{(z_j, z_{j-1}, \ldots, z_{j-k})\}$ is stochastically stable with respect to that base. More generally we have:

*Lemma 1.* *Let $(z_j)$ be a stochastically stable process with respect to a base $(v_j)$. Let $k$ and $l_1, \ldots, l_k$ be arbitrary positive integers and let $z_j^{*\prime} = (z_j', z_{j-l_1}', \ldots, z_{j-l_k}')$. Then $(z_j^*)$ is stochastically stable with respect to $(v_j)$.*

The stochastic stability concept is only meaningful if the base satisfies certain conditions, for it is clear from the definition that every stochastic process is stochastically stable with respect to itself. The condition on the base we shall impose is that of $\phi$-mixing. $\phi$-mixing processes are well known in the mathematical–statistical literature [see, for example, Billingsley (1968)]. For convenience we shall recall its definition. Let $(v_j)$ be a stationary process in $\boldsymbol{R}^q$, let $\mathscr{F}_{-\infty}^{j}$ be the Borel field generated by $v_j, v_{j-1}, v_{j-2}, \ldots,$ and let $\mathscr{F}_j^{\infty}$ be the Borel field generated by $v_j, v_{j+1}, v_{j+2}, \ldots$ . The process $(v_j)$ is said to be $\phi$-mixing if, for each non-negative integer $i$, $E_1 \in \mathscr{F}_{-\infty}^{j}$ and $E_2 \in \mathscr{F}_{j+i}^{\infty}$ together imply

$$|\mathrm{P}(E_1 \cap E_2) - \mathrm{P}(E_1)\mathrm{P}(E_2)| \leq \phi(i)\mathrm{P}(E_1),$$

where $\phi$ is a real function on the non-negative integers such that

$$\lim_{i\to\infty} \phi(i) = 0.$$

If the process $(v_j)$ is non-stationary, then $\phi$ will also depend on $j$; thus

$$|\mathrm{P}(E_1 \cap E_2) - \mathrm{P}(E_1)\mathrm{P}(E_2)| \leq \phi(i, j)\mathrm{P}(E_1),$$

for $E_1 \in \mathscr{F}^{j}_{-\infty}$, $E_2 \in \mathscr{F}^{\infty}_{j+i}$. If for each $j$,

$$\lim_{i\to\infty} \phi(i, j) = 0,$$

then $(v_j)$ will be called a *non-stationary* $\phi$-mixing process [see Bierens (1982a)].

Examples of $\phi$-mixing processes are independent processes, $m$-dependent processes and stationary Markov processes with finite state space [see Billingsley (1968, pp. 167–168)]. On the other hand, Gaussian autoregressive processes are not $\phi$-mixing. Thus, although the $\phi$-mixing concept allows for an asymptotically vanishing memory, it is still rather restrictive. In combination with the stochastic stability concept, however, we get a very weak condition which is satisfied for most stochastic processes of interest in econometrics. For example, ARMA processes are stochastically stable with respect to a $\phi$-mixing base (though not generally $\phi$-mixing themselves) under the usual conditions on the error process (Gaussian white noise) and the lag-polynomial (all roots outside the unit circle).

The usefulness of the stochastic stability concept combined with the $\phi$-mixing concept is demonstrated by the following uniform weak law of large numbers:

*Lemma 2. Let $Z$ be a Euclidean space. Let $(z_j)$ be a $Z$-valued stationary stochastically stable process with respect to a $\phi$-mixing base. Let $\psi(z,\theta)$ for each $z \in Z$ be a continuous real function on $\Theta$ and for each $\theta \in \Theta$ a Borel measurable real function on $Z$, where $\Theta$ is a compact subset of a Euclidean space. Moreover, assume*

$$\mathrm{E} \sup_{\theta\in\Theta} |\psi(z_j,\theta)| < \infty.$$

*Then*

$$\operatorname*{plim}_{n\to\infty} \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{j=1}^{n} \{\psi(z_j,\theta) - \mathrm{E}\,\psi(z_j,\theta)\} \right| = 0.$$

*Proof*. Bierens (1983).

For extending this result to the non-stationary case we need an additional condition which ensures that $(1/n)\Sigma_{j=1}^{n}\mathrm{E}\,\psi(z_j,\theta)$ converges uniformly on $\Theta$ to a limit function. The condition we shall employ involves the concept of *proper convergence* of distribution functions [see Feller (1966, p. 243)]. A sequence $(H_n)$ of distribution functions is said to converge properly if there exists a distribution function $H$, say, such that $\lim_{n\to\infty}H_n = H$ pointwise in the continuity points of $H$. The generalization of Lemma 2 [and of Theorem 2 of Bierens (1982a)] can now be stated as follows:

*Lemma 3. Let $Z$, $\Theta$ and $\psi$ be as in Lemma 2, and let $(z_j)$ be a $Z$-valued, stochastically stable process with respect to a non-stationary $\phi$-mixing base, where $\phi$ is such that*

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi(i,j)=0.$$

*Let $H_j$ be the distribution function of $z_j$ and suppose*

$$\frac{1}{n}\sum_{j=1}^{n}H_j\to H \quad \textit{properly}.$$

*Moreover, assume*

$$\sup_{n\geq 1}\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\sup_{\theta\in\Theta}|\psi(z_j,\theta)|^{1+\delta}<\infty \quad \textit{for some} \quad \delta>0.$$

*Then*

$$\operatorname*{plim}_{n\to\infty}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{j=1}^{n}\psi(z_j,\theta)-\int_Z\psi(z,\theta)\,\mathrm{d}H(z)\right|=0.$$

*Proof*. Bierens (1983).

Note that Lemmas 2 and 3 carry over if the process $(z_j)$ is merely $\phi$-mixing, for a stochastic process is always stochastically stable with respect to itself.

Next we turn to least squares estimation. For convenience we put

$$w_j'=\left(y_{j-1},\ldots,y_{j-p},x_{j-1}',\ldots,x_{j-q}'\right), \tag{20}$$

so that model (15) now becomes

$$y_j = f(w_j, \theta_0) + u_j. \tag{21}$$

*Assumption 3.* The data-generating process $\{(y_j, x_j)\}$ is stochastically stable with respect to a (non-stationary) $\phi$-mixing base, where in the non-stationary case $\phi$ is such that

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \phi(i, j) = 0.$$

From Lemma 1 and (14) and (20) it then follows that also $\{(w_j, z_j)\}$ is stochastically stable with respect to a $\phi$-mixing base, and the same applies to $\{(w_j, z_j, z_{j-1}, \dots, z_{j-l})\}$ for any positive integer $l$. Now put

$$\hat{Q}(\theta) = \frac{1}{n} \sum_{j=1}^{n} \{ y_j - f(w_j, \theta)\}^2, \tag{22}$$

and suppose that, in addition to Assumption 3, the following assumptions hold:

*Assumption 4.* The real function $f(w, \theta)$ is for each $w \in \boldsymbol{R}^{p+q\cdot s}$ continuous on $\Theta$ and for each $\theta \in \Theta$ Borel measurable on $\boldsymbol{R}^{p+q\cdot s}$, where $\Theta$ is a compact subset of $\boldsymbol{R}^r$.

*Assumption 5.* Let

$$\sup_{n\geq 1} \frac{1}{n} \sum_{j=1}^{n} \mathrm{E} \sup_{\theta\in\Theta} |f(w_j, \theta)|^{2+\delta} < \infty,$$

$$\sup_{n\geq 1} \frac{1}{n} \sum_{j=1}^{n} \mathrm{E}|y_j|^{2+\delta} < \infty \quad \text{for some} \quad \delta > 0.$$

*Assumption 6.* The distribution functions $H_j(y, w)$ of the $(y_j, w_j)$ satisfy

$$\frac{1}{n} \sum_{j=1}^{n} H_j \to \overline{H} \quad \text{properly.}$$

Then $\hat{Q}(\theta)$ satisfies the conditions of Lemma 3, hence

$$\operatorname*{plim}_{n\to\infty} \sup_{\theta\in\Theta} |\hat{Q}(\theta) - Q(\theta)| = 0, \tag{23}$$

where

$$Q(\theta)=\int\{y-f(w,\theta)\}^2\,\mathrm{d}\bar{H}(y,w). \tag{24}$$

*Assumption 7.* There exists a unique point $\theta_*\in\Theta$ such that

$$Q(\theta_*)=\inf_{\theta\in\Theta}Q(\theta).$$

Under this assumption and from (23) and Lemma 3.1.8 of Bierens (1981) it now follows that the least squares estimator $\hat{\theta}$, defined by

$$\hat{\theta}\in\Theta,\qquad \hat{Q}(\hat{\theta})=\inf_{\theta\in\Theta}\hat{Q}(\theta), \tag{25}$$

is a consistent estimator of $\theta_*$. Thus we have:

*Theorem 1. If Assumptions 3 through 7 are satisfied, then*

$$\operatorname*{plim}_{n\to\infty}\hat{\theta}=\theta_*.$$

Note that we do not require that Assumptions 1 and 2 hold. Thus Theorem 1 holds regardless whether or not the null hypothesis is true. However, if Assumptions 1 and 2 are satisfied then obviously the $\theta_0$ in Assumption 2 is the same as the $\theta_*$ in Assumption 7. Thus:

*Theorem 2. If Assumptions 1 through 7 are satisfied, then*

$$\operatorname*{plim}_{n\to\infty}\hat{\theta}=\theta_0(=\theta_*).$$

For proving asymptotic normality of $\hat{\theta}$ we need the following additional assumptions:

*Assumption 8.* The parameter space $\Theta$ is convex. If Assumptions 1 and 2 hold, then $\theta_0$ is an interior point of $\Theta$.

*Assumption 9.* The first and second derivatives of $f(w,\theta)$ to $\theta$ are, for each $w\in\boldsymbol{R}^{p+q\cdot s}$, continuous functions on $\Theta$ and, for each $\theta\in\Theta$, Borel measurable functions on $\boldsymbol{R}^{p+q\cdot s}$. Moreover, for some $\delta > 0$,

$$\sup_{n\geq 1}\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\sup_{\theta\in\Theta}|(\partial/\partial\theta_i)f(w_j,\theta)|^{2+\delta}<\infty,\qquad i=1,\ldots,r,$$

$$\sup_{n\geq 1}\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\sup_{\theta\in\Theta}\Big[|y_j-f(w_j,\theta)|^{2+\delta}\cdot|(\partial/\partial\theta_{i_1})f(w_j,\theta)|^{1+\delta}$$

$$\cdot|(\partial/\partial\theta_{i_2})f(w_j,\theta)|^{1+\delta}\Big]<\infty,$$

$$\sup_{n\geq 1}\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\sup_{\theta\in\Theta}|(\partial/\partial\theta_{i_1})(\partial/\partial\theta_{i_2})f(w_j,\theta)|^{1+\delta}<\infty,$$

$$i_1,i_2=1,\ldots,r.$$

*Assumption 10.* The matrix

$$A_1=\int\{(\partial/\partial\theta')f(w_j,\theta_*)\}\{(\partial/\partial\theta)f(w_j,\theta_*)\}\,\mathrm{d}\overline{H}(y,w)$$

is non-singular.

Moreover, since by Assumptions 1 and 2 the errors $u_j$ of model (15) are martingale differences, we need a martingale central limit theorem. The following convenient central limit theorem is a corollary of the martingale central limit theorem of Brown (1971).

*Lemma 4. Let $(v_j)_{j=1}^{\infty}$ be a sequence of martingale differences, i.e.,*

$$\mathrm{E}(v_j|v_{j-1},\ldots,v_1)=0\quad\text{a.s.}\quad\textit{for}\quad j=2,3,\ldots$$

*If*

$$\sup_{n\geq 1}\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}|v_j|^{2+\delta}<\infty\quad\textit{for some}\quad\delta>0,$$

*and*

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}v_j^2=\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}v_j^2=\sigma^2\in(0,\infty),$$

*then*

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}v_j\to\mathrm{N}(0,\sigma^2)\quad\textit{in distr.}\tag{26}$$

*If the $v_j$'s are identically distributed, then a sufficient condition for* (26) *is*

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} v_j^2 = \mathrm{E} v_1^2 = \sigma^2 \in (0, \infty).$$

This central limit theorem is reminicent of the well-known central limit theorem of Liapounov for independent random variables; it can be derived from Brown's (1971) result in a similar way as Liapounov's proof.

Denoting

$$A_2 = \int \{y - f(w, \theta_*)\}^2 \{(\partial/\partial\theta') f(w, \theta_*)\}$$
$$\times \{(\partial/\partial\theta) f(w, \theta_*)\} \, \mathrm{d}\bar{H}(y, w), \tag{27}$$

$$\hat{A}_1 = \frac{1}{n} \sum_{j=1}^{n} \{(\partial/\partial\theta') f(w_j, \hat{\theta})\} \{(\partial/\partial\theta) f(w_j, \hat{\theta})\}, \tag{28}$$

$$\hat{A}_2 = \frac{1}{n} \sum_{j=1}^{n} \{y_j - f(w_j, \hat{\theta})\}^2 \{(\partial/\partial\theta') f(w_j, \hat{\theta})\} \{(\partial/\partial\theta) f(w_j, \hat{\theta})\}, \tag{29}$$

and using Lemmas 2, 3 and 4, it is not too hard to show along the same lines as the argument in Bierens (1981, sec. 3.1) that the following theorem holds:

*Theorem 3. Under Assumptions 1 through 10, we have*

$$\sqrt{n}\,(\hat{\theta} - \theta_0) \to \mathrm{N}\left[0, A_1^{-1} A_2 A_1^{-1}\right] \quad \textit{in distr.},$$

*and*

$$\operatorname*{plim}_{n\to\infty} \hat{A}_1^{-1} \hat{A}_2 \hat{A}_1^{-1} = A_1^{-1} A_2 A_1^{-1}. \tag{30}$$

This result is similar to Theorem 3.2 of Klimko and Nelson (1978), which has been derived under the assumptions of strict stationarity and ergodicity, and to Corollary 3.3 of Domowitz and White (1982). However, the assumptions of Domowitz and White are different from ours in that they assume that the data-generating process is mixing itself, whereas we assume that the data are generated by functions of a one-sided infinite mixing sequence.

Note that in the case of a strictly stationary data-generating process we may set the $\delta$ in Assumptions 5 and 9 equal to zero, as we then may refer to Lemma 2 instead of Lemma 3. Of course, in that case we also do not need Assumption 6, as then $H_j = \bar{H}$ for all $j$.

In the definition of the matrices $A_1$ and $A_2$ we have used $\theta_*$ instead of $\theta_0$. In view of Theorem 3 there is no loss of generality in doing so, as under the conditions of Theorem 3 we have $\theta_0 = \theta_*$. The reason for using $\theta_*$ is, however, that then $\hat{A}_1$ and $\hat{A}_2$ remain consistent estimators of $A_1$ and $A_2$, respectively, if Assumptions 1 and 2 fail to hold. These results will be needed in the asymptotic theory of our model specification test under the alternative hypothesis. Thus part (30) of Theorem 3 goes through if Assumptions 1 and 2 do not hold, but then the matrix $A_1^{-1}A_2A_1^{-1}$ need no longer be the variance of the limiting distribution of $\hat{\theta}$.

In the sequel of this paper we shall discuss our model specification testing approach. As our approach is rather complicated it is convenient to derive our results first under the stationarity assumption. Thus throughout sections 4 through 8 we assume:

*Assumption 11.* The data-generating process $(z_j)$, where $z_j' = (y_j, x_j') \in \boldsymbol{R} \times \boldsymbol{R}^s$, is strictly stationary.

The non-stationary case will be considered in section 9.

## 4. Identification of the null hypothesis

In this section we show that there exists a one-to-one correspondence between the truth of the null hypothesis and the stochastic behaviour of particular sample moments, so that these sample moments determine whether the null is true. In other words, the sample moments involved identify the null hypothesis.

It will be convenient to restate the null hypothesis as follows:

$$\text{H}_0\text{: There exists a } \theta_0 \in \Theta \text{ such that for all positive integers } l,\ \text{P}[\text{E}(y_j - f(w_j, \theta_0) | z_{j-1}, \ldots, z_{j-l}) = 0] = 1. \tag{31}$$

The equivalence of (17) and (31) is easy to verify, using the notations (14) and (20). Obviously this null hypothesis is false if and only if:

$$\text{H}_1\text{: For all } \theta \in \Theta \text{ there exists at least one positive integer } l \text{ such that } \text{P}[\text{E}(y_j - f(w_j, \theta) | z_{j-1}, \ldots, z_{j-l}) = 0] < 1. \tag{32}$$

The following theorem shows that there exists a one-to-one correspondence between the above hypotheses and the properties of particular mathematical expectations:

*Theorem 4.* *Let $v$ be a random variable satisfying* $\mathrm{E}|v| < \infty$ *and let $z$ be a random vector in a Euclidean space Z. Then*

(I) $\mathrm{P}[\mathrm{E}(v|z)=0] < 1$ *if and only if* $\mathrm{E}v\mathrm{e}^{\mathrm{i}t'z} \neq 0$ *for some non-random vector* $t \in Z$.

*If in addition z is bounded, then*

(II) $\mathrm{P}[\mathrm{E}(v|z)=0] < 1$ *if and only if* $\mathrm{E}v\mathrm{e}^{\mathrm{i}t'z} \neq 0$ *for some non-random vector t in an arbitrarily small neighborhood of the origin of Z.*

*Proof.* Bierens (1982b, theorem 1).

Denoting

$$\zeta_l(\theta, t_1, \ldots, t_l) = \mathrm{E}\big(y_j - f(w_j, \theta)\big)\mathrm{e}^{\mathrm{i}(t_1' z_{j-1} + \cdots + t_l' z_{j-l})}, \tag{33}$$

it thus follows from part I of Theorem 4 that $\mathrm{H}_0$ is true if for some $\theta_0 \in \Theta$ and every $l \geq 1$, $\zeta_l(\theta_0, t_1, \ldots, t_l) = 0$ for all $t_1, \ldots, t_l \in \boldsymbol{R}^{1+s}$, and that $\mathrm{H}_1$ is true if for all $\theta \in \Theta$ there exists an $l \geq 1$ and vectors $t_1, \ldots, t_l$ in $\boldsymbol{R}^{1+s}$ such that $\zeta_l(\theta, t_1, \ldots, t_l) \neq 0$. Consequently we have:

*Corollary 1.* *Let H be an arbitrary integrable positive real function on* $\boldsymbol{R}^{1+s}$ *and let*

$$\eta_l(\theta) = \int_{\boldsymbol{R}^{1+s}} \cdots \int_{\boldsymbol{R}^{1+s}} |\zeta_l(\theta, t_1, \ldots, t_l)|^2 H(t_1) \cdots H(t_l)\, \mathrm{d}t_1 \cdots \mathrm{d}t_l.$$

*Then $H_0$ is true if for some $\theta_0 \in \Theta$, $\eta_l(\theta_0) = 0$ for all $l \geq 1$, and $H_1$ is true if* $\inf_{\theta \in \Theta} \eta_l(\theta) > 0$ *for some $l \geq 1$.*

Also part II of Theorem 4 can be used for identifying the hypotheses $\mathrm{H}_0$ and $\mathrm{H}_1$. Let $\Phi$ be a bounded continuous 1-1 mapping from $\boldsymbol{R}^{1+s}$ into $\boldsymbol{R}^{1+s}$; for example, let $\Phi$ be a vector of functions atan$(\cdot)$. Then $\{\Phi(z_{j-1}), \ldots, \Phi(z_{j-l})\}$ generate the same Borel field as $z_{j-1}, \ldots, z_{j-l}$; hence

$$\begin{aligned} &\mathrm{E}\big(y_j - f(x_j, w_j)|z_{j-1}, \ldots, z_{j-l}\big) \\ &= \mathrm{E}\big(y_j - f(w_j, \theta)|\Phi(z_{j-1}), \ldots, \Phi(z_{j-l})\big) \quad \text{a.s.} \end{aligned} \tag{34}$$

Thus if we replace the $z_j$'s in (33) by $\Phi(z_j)$, i.e.,

$$\zeta_l^*(\theta, t_1, \ldots, t_l) = \mathrm{E}\big(y_j - f(w_j, \theta)\big)\mathrm{e}^{\mathrm{i}(t_1'\Phi(z_{j-1}) + \cdots + t_l'\Phi(z_{j-1}))}, \tag{35}$$

then it follows from (34) and part II of Theorem 4 that $\mathrm{H}_0$ is true if for some $\theta_0 \in \Theta$ and all $l \geq 1$, $\zeta_l^*(\theta_0, t_1, \ldots, t_l) = 0$ for all $t_1, \ldots, t_l$ in an arbitrarily small neighborhood of the origin of $\boldsymbol{R}^{1+s}$, and $\mathrm{H}_1$ is true if for all $\theta \in \Theta$ there exists an $l \geq 1$ and vectors $t_1, \ldots, t_l$ in an arbitrarily small neighborhood of the origin of $\boldsymbol{R}^{1+s}$ such that $\zeta_l^*(\theta, t_1, \ldots, t_l) \neq 0$. Similarly to Corollary 1 we thus have:

*Corollary 2. Let $N_0$ be an arbitrarily small neighborhood of the origin of $\boldsymbol{R}^{1+s}$ and let*

$$\eta_l^*(\theta) = \int_{N_0} \cdots \int_{N_0} |\zeta_l^*(\theta, t_1, \ldots, t_l)|^2 \,\mathrm{d}t_1, \ldots, \mathrm{d}t_l.$$

*Then $H_0$ is true if for some $\theta_0 \in \Theta$, $\eta_l^*(\theta_0) = 0$ for all $l \geq 1$, and $H_1$ is true if $\inf_{\theta \in \Theta} \eta_l^*(\theta) > 0$ for some $l \geq 1$.*

Both corollaries can be used as a basis of a consistent test of $\mathrm{H}_0$ versus $\mathrm{H}_1$. However, since the analysis in our previous paper [Bierens (1982b)] has been based on similar results as in Corollary 2, it is convenient to work further on basis of this corollary, as we then can refer to these previous results.

## 5. The test

We consider a data set of $n+m$ observations, $(y_{-m+1}, x_{-m+1}), \ldots, (y_0, x_0), (y_1, x_1), \ldots, (y_n, x_n)$, on the strictly stationary vector time series process $\{(y_j, x_j)\}_{-\infty}^{+\infty}$ in $\boldsymbol{R} \times \boldsymbol{R}^s$ considered before, where $m \geq \max(p, q)$, with $p$ and $q$ the specified lag lengths of $y_j$ and $x_j$, respectively. Moreover, we shall use the following notations:

$$\begin{aligned} \bar{z}_j &= \Phi(z_j) \quad \text{if} \quad j \geq -m+1, \\ &= 0 \qquad\quad \text{if} \quad j < -m+1, \end{aligned} \tag{36}$$

where $z_j' = (y_j, x_j')$ and $\Phi$ is the bounded continuous 1-1 mapping from $\boldsymbol{R}^{1+s}$

into $\boldsymbol{R}^{1+s}$ considered in Corollary 2;

$$\hat{\xi}_l(\theta, t_1, \ldots, t_l) = \frac{1}{n} \sum_{j=1}^{n} \left(y_j - f(w_j, \theta)\right) \mathrm{e}^{\mathrm{i}(t_1' \bar{z}_{j-1} + \cdots + t_l' \bar{z}_{j-l})}; \tag{37}$$

$$\hat{\eta}_l(\theta) = \int_{N_0} \cdots \int_{N_0} |\hat{\xi}_l(\theta, t_1, \ldots, t_l)|^2 \, \mathrm{d}t_1, \ldots, \mathrm{d}t_l; \tag{38}$$

$$\bar{\eta}_l(\theta) = \int_{N_0} \cdots \int_{N_0} |\mathrm{E}\hat{\xi}_l(\theta, t_1, \ldots, t_l)|^2 \, \mathrm{d}t_1, \ldots, \mathrm{d}t_l, \tag{39}$$

where $N_0$ is an arbitrarily small neighborhood of the origin of $\boldsymbol{R}^{1+s}$, for example,

$$N_0 = \mathop{\times}_{k=1}^{s+1} [-\varepsilon_k, \varepsilon_k] \quad \text{with} \quad \varepsilon_k > 0. \tag{40}$$

Furthermore, let

$$\hat{\eta}(\theta) = \sum_{l=1}^{L_n} \gamma_l \hat{\eta}_l(\theta), \tag{41}$$

$$\bar{\eta}(\theta) = \sum_{l=1}^{\infty} \gamma_l \bar{\eta}_l(\theta), \tag{42}$$

where $(\gamma_l)$ is an arbitrary sequence of real numbers satisfying

$$\gamma_l > 0, \qquad \sum_{l=1}^{\infty} \gamma_l \left( \int_{N_0} \mathrm{d}t \right)^l < \infty, \tag{43}$$

and $(L_n)$ is a sequence of positive integers satisfying

$$L_n \to \infty \quad \text{as} \quad n \to \infty. \tag{44}$$

From Corollary 2 we now may conclude without further discussion that:

*Theorem 5.* *$H_0$ is true if* $\inf_{\theta \in \Theta} \bar{\eta}(\theta) = 0$ *and $H_1$ is true if* $\inf_{\theta \in \Theta} \bar{\eta}(\theta) > 0$.

It can be shown [see Bierens (1983)] that under fairly weak conditions

$$\operatorname*{plim}_{n\to\infty} \sup_{\theta\in\Theta} |\hat{\eta}(\theta) - \bar{\eta}(\theta)| = 0, \tag{45}$$

under $H_0$ as well as under $H_1$. From (45) and Theorems 1, 2 and 5 it now follows that $\operatorname{plim}_{n\to\infty}\hat{\eta}(\hat{\theta}) = \bar{\eta}(\theta_0) = 0$ if $H_0$ is true and $\operatorname{plim}_{n\to\infty}\hat{\eta}(\hat{\theta}) = \eta(\theta_*) > 0$ if $H_1$ is true, which suggests to use $\hat{\eta}(\hat{\theta})$ as a test statistic. However, similarly to Bierens (1982, sec. 3), it follows that a much stronger result applies if the null hypothesis is true, i.e., it appears that $n\hat{\eta}(\hat{\theta})$ converges in distribution to a non-negative random variable. Although the limiting distribution involved is of an unknown type, we can estimate its first moment $\mu$, say, consistently. Denoting this consistent estimate by $\hat{\mu}$, we then have by Chebishev's inequality for first absolute moments

$$\lim_{n\to\infty} \sup \mathrm{P}\left[n\hat{\eta}(\hat{\theta}) > \frac{1}{\alpha}\hat{\mu}\right] \le \alpha \quad \text{for every} \quad \alpha > 0, \tag{46}$$

if $H_0$ is true. Under $H_1$, however, we have $\operatorname{plim}\hat{\eta}(\hat{\theta}) > 0$, hence $\operatorname{plim} n\hat{\eta}(\hat{\theta}) = \infty$. So if $\hat{\mu}$ remains stochastically bounded under $H_1$, which will appear to be the case, then obviously

$$\lim_{n\to\infty} \mathrm{P}\left[n\hat{\eta}(\hat{\theta}) > \frac{1}{\alpha}\hat{\mu}\right] = 1 \quad \text{for every} \quad \alpha > 0, \tag{47}$$

if $H_1$ is true. Thus conducting the test at the $\alpha \cdot 100\%$ significance level (where $0 < \alpha < 1$) we accept $H_0$ if $\hat{\mu}/n\hat{\eta}(\hat{\theta}) \ge \alpha$ and we reject $H_0$ if not. Since by (47) the type II error vanishes for $n \to \infty$, this test is consistent.

Observe from (37), (38) and (41) that $\hat{\eta}(\theta)$ can be written

$$\hat{\eta}(\theta) = \frac{1}{n^2} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \left(y_{j_1} - f\left(w_{j_1}, \theta\right)\right)\left(y_{j_2} - f\left(w_{j_2}, \theta\right)\right)$$

$$\times \sum_{l=1}^{L_n} \gamma_l \prod_{k=1}^{l} \int_{N_0} \mathrm{e}^{\mathrm{i}t'(\bar{z}_{j_1-k} - \bar{z}_{j_2-k})}\,\mathrm{d}t. \tag{48}$$

In particular, if we choose $N_0$ as in (40) and if we denote

$$\left(\bar{z}_{1,j}, \ldots, \bar{z}_{1+s,j}\right) = \bar{z}_j', \tag{49}$$

then

$$\int_{N_0} \mathrm{e}^{\mathrm{i}t'(\bar{z}_{j_1-k}-\bar{z}_{j_2-k})}\,\mathrm{d}t = \prod_{r=1}^{s+1} 2\cdot\frac{\sin\left[\varepsilon_r\left(\bar{z}_{r,j_1-k}-\bar{z}_{r,j_2-k}\right)\right]}{\bar{z}_{r,j_1-k}-\bar{z}_{r,j_2-k}}, \tag{50}$$

and consequently

$$\begin{aligned}\hat{\eta}(\theta) = {} & \frac{1}{n^2}\sum_{j=1}^{n}\left(y_j - f(w_j,\theta)\right)^2 \sum_{l=1}^{L_n}\gamma_l\left(\prod_{r=1}^{s+1}(2\varepsilon_r)\right)^l \\ & + 2\cdot\frac{1}{n^2}\sum_{j_1=1}^{n-1}\sum_{j_2=j_1+1}^{n}\left(y_{j_1}-f(w_{j_1},\theta)\right)\left(y_{j_2}-f(w_{j_2},\theta)\right) \\ & \times \sum_{l=1}^{L_n}\gamma_l\prod_{k=1}^{l}\sum_{r=1}^{s+1} 2\cdot\frac{\sin\left[\varepsilon_r\left(\bar{z}_{r,j_1-k}-\bar{z}_{r,j_2-k}\right)\right]}{\bar{z}_{r,j_1-k}-\bar{z}_{r,j_2-k}}. \end{aligned} \tag{51}$$

Moreover, it appears [see Bierens (1983)] that $\hat{\mu}$ can be written as

$$\begin{aligned}\hat{\mu} = {} & \hat{\sigma}^2\sum_{l=1}^{L_n}\gamma_l\left(\int_{N_0}\mathrm{d}t\right)^l - \frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\left(\hat{u}_{j_1}^2+\hat{u}_{j_2}^2\right)\left\{(\partial/\partial\theta)f\left(w_{j_1},\hat{\theta}\right)\right\} \\ & \times \hat{A}_1^{-1}\left\{(\partial/\partial\theta')f\left(w_{j_2},\hat{\theta}\right)\right\}\sum_{l=1}^{L_n}\gamma_l\prod_{k=1}^{l}\int_{N_0}\mathrm{e}^{\mathrm{i}t'(\bar{z}_{j_1-k}-\bar{z}_{j_2-k})}\,\mathrm{d}t \\ & + \frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\left\{(\partial/\partial\theta)f\left(w_{j_1},\hat{\theta}\right)\right\}\hat{A}_1^{-1}\hat{A}_2\hat{A}_1^{-1}\left\{(\partial/\partial\theta')f\left(w_{j_2},\hat{\theta}\right)\right\} \\ & \times \sum_{l=1}^{L_n}\gamma_l\prod_{k=1}^{l}\int_{N_0}\mathrm{e}^{\mathrm{i}t'(\bar{z}_{j_1-k}-\bar{z}_{j_2-k})}\,\mathrm{d}t, \end{aligned} \tag{52}$$

where

$$\hat{\mu}_j = y_j - f\left(w_j,\hat{\theta}\right), \tag{53}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}\hat{u}_j^2, \tag{54}$$

and $\hat{A}_1$ and $\hat{A}_2$ are defined by (28) and (29), respectively.

Again choosing $N_0$ as in (40) it follows from (50) and (52) that

$$\hat{\mu} = \left\{\hat{\sigma}^2 - \frac{1}{n}\mathrm{tr}\left(\hat{A}_1^{-1}\hat{A}_2\right)\right\} \sum_{l=1}^{L_n} \gamma_l \left(\prod_{r=1}^{s+1} 2\varepsilon_r\right)^l$$

$$+2 \cdot \frac{1}{n^2} \sum_{j_1=1}^{n-1} \sum_{j_2=j_1+1}^{n} \left[\left\{(\partial/\partial\theta) f\left(w_{j_1}, \hat{\theta}\right)\right\}\right.$$

$$\times \hat{A}_1^{-1}\hat{A}_2\hat{A}_1^{-1}\left\{(\partial/\partial\theta') f\left(w_{j_2}, \hat{\theta}\right)\right\}$$

$$\left. -\left(\hat{u}_{j_1}^2 + \hat{u}_{j_2}^2\right)\left\{(\partial/\partial\theta) f\left(w_{j_1}, \hat{\theta}\right)\right\} \hat{A}_1^{-1}\left\{(\partial/\partial\theta') f\left(w_{j_2}, \hat{\theta}\right)\right\}\right]$$

$$\times \sum_{l=1}^{L_n} \gamma_l \prod_{k=1}^{l} \prod_{r=1}^{s+1} 2 \cdot \frac{\sin\left[\varepsilon_r\left(\bar{z}_{r,j_1-k} - \bar{z}_{r,j_2-k}\right)\right]}{\bar{z}_{r,j_1-k} - \bar{z}_{r,j_2-k}}. \tag{55}$$

For fixed $l$ the asymptotic behaviour of $\hat{\eta}_l(\hat{\theta})$ is quite similar to that of the statistics $\hat{\eta}$ in Bierens (1982b), due to the fact that instead of referring to Lemma 1 and the Theorems 3 and 4 in that paper; we now may refer to the present Lemma 2 and Theorems 1, 2 and 3. Similarly to Bierens (1982, theorems 6 and 7) we therefore have:

*Theorem 6. Under Assumptions 1 through 11,*

$$\lim_{n\to\infty} \sup \mathrm{P}\left[n\hat{\eta}(\hat{\theta}) > \hat{\mu}/\alpha\right] \le \alpha \quad \textit{for every} \quad \alpha > 0.$$

*Proof.* Bierens (1983).

*Theorem 7. If the null hypothesis is false, then under Assumptions 3 through 7 and 9 through 11, we have*

$$\lim_{n\to\infty} \mathrm{P}\left[n\hat{\eta}(\hat{\theta}) > \hat{\mu}/\alpha\right] = 1 \quad \textit{for every} \quad \alpha > 0.$$

*Proof.* Bierens (1983).

## 6. Some numerical experiments

In this section we shall demonstrate the performance of our test by applying it to artificial data sets generated by a univariate autoregressive stochastic process of the form

$$y_j = \alpha y_{j-1} + \beta y_{j-2} + u_j, \qquad u_j \sim \mathrm{NID}(0,1), \quad -\infty < j < +\infty. \tag{56}$$

We consider two cases, namely

Case 1: $\alpha = 0.8, \quad \beta = 0,$
Case 2: $\alpha = 0, \quad \beta = 0.8.$

From Bierens (1981, theorem 5.1.2) it follows that in both cases the process $(y_j)$ is stochastically stable with respect to the independent, hence $\phi$-mixing, base $(u_j)$.

The null hypothesis to be tested is

$$\mathrm{H}_0: \quad \mathrm{E}\left[y_j | y_{j-1}, \ldots, y_{j-l}\right] = \theta_1 y_{j-1} + \theta_2 \quad \text{a.s.}$$
$$\text{for all} \quad l \geq 1 \quad \text{and some} \quad \theta_1, \theta_2.$$

Clearly only in Case 1 this null hypothesis is true.

We have chosen

$$\begin{aligned} \bar{z}_j &= \operatorname{atan}(y_j) \quad \text{if} \quad j \geq 0, \\ &= 0 \qquad\qquad \text{if} \quad j < 0, \end{aligned} \qquad [\text{cf. } (36)] \qquad (57)$$

$$N_0 = [-\varepsilon, \varepsilon], \qquad [\text{cf. } (40)] \qquad (58)$$

$$\begin{aligned} \gamma_l &= 1 \qquad \text{for} \quad l = 1, \ldots, 10m, \\ &= \frac{1}{(l-10m)!} \quad \text{for} \quad l > 10m, \end{aligned} \qquad [\text{cf. } (41), (43)] \qquad (59)$$

$$L_n = [\sqrt{n}\,]. \qquad [\text{cf. } (41), (43)] \qquad (60)$$

For each of the two cases we generated an artificial data set of size 101 ($n = 100$ and $m = 1$) and we carried out the test for various values of $\varepsilon$. The results are given in table 1.

From these results we see that the test statistic $\hat{\mu}/n\hat{\eta}(\hat{\theta})$ is very sensitive to the choice of $\varepsilon$, especially if the null is false. The best choice of $\varepsilon$ seems to be

Table 1

Test results for various $\varepsilon$.

| | $\hat{\mu}/n\hat{\eta}(\hat{\theta})$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\varepsilon = 0.25$ | $\varepsilon = 0.5$ | $\varepsilon = 0.75$ | $\varepsilon = 1.0$ | $\varepsilon = 1.25$ | $\varepsilon = 1.5$ | $\varepsilon = 1.75$ | $\varepsilon = 2.0$ |
| Case 1 | 1.27 | 1.19 | 1.21 | 1.18 | 1.13 | 1.09 | 1.05 | 1.03 |
| Case 2 | 0.04 | 0.06 | 0.08 | 0.12 | 0.18 | 0.27 | 0.40 | 0.55 |

Table 2

Test results for $\varepsilon = 0.25$.

| | $\hat{\mu}/n\hat{\eta}(\hat{\theta})$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Case 1 | 1.04 | 1.13 | 3.16 | 3.18 | 0.27 | 0.43 | 1.02 | 0.43 | 0.88 | 1.58 |
| Case 2 | 0.05 | 0.05 | 0.05 | 0.06 | 0.03 | 0.05 | 0.05 | 0.06 | 0.08 | 0.05 |

0.25 (at least for the two examples under review), so for each case we did 10 additional simulations with $\varepsilon = 0.25$. The results involved are given in table 2.

From table 2 we see that the test works quite well; i.e., if we choose a significance level of 10%, then for each pair of simulations the true null (Case 1) is accepted and the false null (Case 2) is rejected. However, the results in table 1 suggest that the small sample power of the test crucially depends on the appropriate choice of $\varepsilon$, or in the general case on the choice of $N_0$. Some further experiments with other models indicate that the optimal $N_0$ depends on the extent of the misspecification, so that there may be no general rule for choosing $N_0$.

In our previous paper Bierens (1982b) we have proposed to standardize the variables $z_j$ before transforming them by the bounded 1-1 mapping $\Phi$ [see (36)]. It was argued that this would preserve sufficient variation in the $\bar{z}_j = \Phi(z_j)$ and that this would not affect the asymptotic properties. In the present case, however, it is not clear whether standardization is allowed.

A sufficient condition for this is that $\Phi$ has bounded continuous first derivatives and that $(1/\sqrt{n})\Sigma_{j=1}^{n}(z_j - \mathrm{E}z_j)$ and $(1/\sqrt{n})\Sigma_{j=1}^{n}(z_j'z_j - \mathrm{E}z_j'z_j)$ remain stochastically bounded.

In the independent case considered in Bierens (1982b) the latter conditions follow easily, but not in the dependent case under review. Therefore we have not standardized $y_j$ in (57).

## 7. A simplification

A serious drawback of our approach is that it is so laborious for models with a lot of explanatory variables and parameters. Especially the calculation of $\hat{\mu}$ will then cost a large amount of computer time. This is due to the fact that the calculation of (51) requires about $\frac{1}{2}n^2 \cdot L_n(s+1)$ times the calculation of functions of the type $\sin(\cdot)/(\cdot)$, whereas the calculation of (55) requires in addition about $\frac{1}{2}n^2$ times the calculation of the two quadratic forms

$$(\partial/\partial\theta)f(w_{j_1}, \hat{\theta})\hat{A}_1^{-1}\hat{A}_2\hat{A}_1^{-1}(\partial/\partial\theta')f(w_{j_2}, \hat{\theta}),$$

and

$$(\partial/\partial\theta)f(w_{j_1}, \hat{\theta})\hat{A}_1^{-1}(\partial/\partial\theta')f(w_{j_2}, \hat{\theta}).$$

However, if we make the additional assumption:

*Assumption 12.* Under the null hypothesis,

$$\mathrm{E}\left[u_j^2 | y_{j-1},\ldots,y_{j-l},x_{j-1},\ldots,x_{j-l}\right] = \mathrm{E}u_j^2 = \sigma^2 \quad \text{a.s.} \quad \text{for} \quad l \geq 1,$$

then

$$\hat{\hat{\mu}} = \hat{\sigma}^2 \sum_{l=1}^{L_n} \gamma_l \left( \int_{N_0} \mathrm{d}t \right)^l \geq \hat{\mu} \quad \text{a.s.} \quad \text{for large } n. \tag{61}$$

From Theorems 6 and 7 and (61) it thus follows:

*Theorem 8. Under Assumptions 1 through 12,*

$$\lim_{n\to\infty} \sup \mathrm{P}(n\hat{\eta}(\hat{\theta}) \geq \hat{\hat{\mu}}/\alpha) \leq \alpha \quad \textit{for every} \quad \alpha > 0,$$

*whereas if $H_0$ is false, then under Assumptions 3 through 7 and 9 through 11,*

$$\lim_{n\to\infty} \mathrm{P}(n\hat{\eta}(\hat{\theta}) \geq \hat{\hat{\mu}}/\alpha) = 1 \quad \text{for every} \quad \alpha > 0.$$

By using $\hat{\hat{\mu}}/n\hat{\eta}(\hat{\theta})$ as a test statistic instead of $\hat{\mu}/n\hat{\eta}(\hat{\theta})$ we sacrifice some power in favour of computational convenience, but the test will still be consistent, even if Assumption 12 is not satisfied.

## 8. Testing the martingale difference hypothesis

It is clear from Assumptions 1 and 2 that, under the null hypothesis (17), the expectation of the errors $u_j$ of model (15) conditional on all lagged $u_j$ equals zero a.s. In other words: the null hypothesis implies that the errors $u_j$ are martingale differences. If the null is false it is theoretically possible that this conditional expectation remains zero a.s., for Assumptions 1 and 2 are stronger conditions than the martingale difference condition, but in practice this will hardly occur. Thus testing the null hypothesis that the errors are martingale differences, i.e.,

$$\mathrm{H}_0^*: \quad \mathrm{E}\left[u_j | u_{j-1},\ldots,u_{j-l}\right] = 0 \quad \text{a.s.} \quad \text{for every} \quad l \geq 1, \tag{62}$$

where $u_j = y_j - f(w_j, \theta_*)$ with $\theta_*$ the probability limit of the least squares estimator $\hat{\theta}$, will in practice be almost sufficient for detecting misspecification.

We can test $H_0^*$ in a similar way as before. Thus let

$$\hat{z}_j = \phi(\hat{u}_j) \quad \text{if} \quad j \geq -m+1, \qquad = 0 \quad \text{if} \quad j < -m+1, \tag{63}$$

$$\tilde{\eta}_l = \int_{-\varepsilon}^{\varepsilon} \cdots \int_{-\varepsilon}^{\varepsilon} \left| \frac{1}{n} \sum_{j=1}^{n} \hat{u}_j e^{i\sum_{k=1}^{l} t_k \hat{z}_{j-k}} \right|^2 dt_1 \ldots dt_l, \tag{64}$$

$$\tilde{\eta} = \sum_{l}^{L_n} \gamma_l \tilde{\eta}_l, \tag{65}$$

$$\tilde{\bar{\mu}} = \hat{\sigma}^2 \sum_{l=1}^{L_n} \gamma_l (2\varepsilon)^l, \tag{66}$$

where the $\hat{u}_j$ are the least squares residuals, $\phi$ is a bounded 1-1 mapping from $\boldsymbol{R}$ into $\boldsymbol{R}$ with *continuous and bounded first derivative* $\phi'$, $(\gamma_l)$ is a sequence of positive numbers satisfying

$$\sum_{l=1}^{\infty} \gamma_l l^2 (2\varepsilon)^l < \infty, \tag{67}$$

and $L_n$ is as before. Then similarly to Theorem 8 we have:

*Theorem 9. Under Assumptions 1 through 12,*

$$\lim_{n\to\infty} \sup \mathrm{P}(n\tilde{\eta} \geq \tilde{\bar{\mu}}/\alpha) \leq \alpha \quad \textit{for every} \quad \alpha > 0.$$

*If $H_0^*$ is false then, under Assumptions 3 through 7 and 9 through 11,*

$$\lim_{n\to\infty} \mathrm{P}(n\tilde{\eta} \geq \tilde{\bar{\mu}}/\alpha) = 1 \quad \textit{for every} \quad \alpha > 0.$$

*Proof.* Bierens (1983).

In order to see how this test works out we repeated the experiments in section 6. Thus for each of the two cases involved we generated an artificial date set of size 101 and we calculated the values of the test statistic $\tilde{\bar{\mu}}/n\tilde{\eta}$ for various $\varepsilon$'s. The results are reported in table 3.

Table 3

Test results for various $\varepsilon$.

| | $\tilde{\tilde{\mu}}/n\tilde{\eta}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\varepsilon=0.25$ | $\varepsilon=0.5$ | $\varepsilon=0.75$ | $\varepsilon=1.0$ | $\varepsilon=1.25$ | $\varepsilon=1.5$ | $\varepsilon=1.75$ | $\varepsilon=2.0$ |
| Case 1 | 361.70 | 72.94 | 23.73 | 9.83 | 4.98 | 2.99 | 2.06 | 1.58 |
| Case 2 | 4.72 | 0.69 | 0.26 | 0.15 | 0.11 | 0.11 | 0.12 | 0.14 |

Table 4

Test results for $\varepsilon = 1.5$.

| | $\tilde{\tilde{\mu}}/n\tilde{\eta}$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Case 1 | 4.61 | 4.11 | 2.51 | 2.44 | 2.32 | 2.67 | 2.72 | 2.88 | 3.93 | 4.26 |
| Case 2 | 0.18 | 0.12 | 0.17 | 0.12 | 0.16 | 0.32 | 0.09 | 0.14 | 0.10 | 0.15 |

These results indicate that $\varepsilon = 1.5$ is a good choice, at least for Case 2, so we repeated this experiment 10 times for $\varepsilon = 1.5$. The results involved can be found in table 4.

We see from table 4 that this test is somewhat more conservative than the previous test, which is probably due to the fact that $\tilde{\tilde{\mu}}$ is only an upperbound of the consistent estimate of the first moment of the limiting distribution of $n\tilde{\eta}$. Therefore in applying this test one should not choose the significance level too small, which is the price we have to pay for computational convenience.

## 9. Nonstationarity

We shall now deal with the question what will happen if we apply our model specification tests to nonstationary data. The kind of non-stationarity we consider is similarly to that in section 3. Thus we retain Assumptions 3 and 6. However, in order that Lemma 3 is applicable to random functions of the type (37) we need the following extension of Assumption 6:

*Assumption 13.* For any positive integer $l$ the distribution functions $H_{j,l}$ of $(y_j, z'_{j-1}, \ldots, z'_{j-l})$ satisfy

$$\frac{1}{n}\sum_{j=1}^{n} H_{j,l} \to \overline{H}_l \quad \text{properly.}$$

Redefining the regression function $f(w_j, \theta)$ as

$$f^*(z_{j-1}, \ldots, z_{j-m}, \theta) = f(w_j, \theta), \tag{68}$$

which is no loss of generality, and denoting

$$\bar{\xi}_l(\theta, t_1, \ldots, t_l) = \int \{ y - f^*(z^{(1)}, \ldots, z^{(m)}, \theta) \} e^{i\{t_1'\Phi(z^{(1)}) + \cdots + t_l'\Phi(z^{(l)})\}}$$

$$\times \mathrm{d}\bar{H}_{l_*}(y, z^{(1)}, \ldots, z^{(l_*)}), \tag{69}$$

where

$$l_* = \max(l, m), \tag{70}$$

it now follows from Lemma 3:

*Lemma 5. Under Assumptions 3, 4, 5 and 13,*

$$\operatorname*{plim}_{n\to\infty} \sup_{\theta\in\Theta, t_1\in N_0, \ldots, t_l\in N_0} |\hat{\xi}_l(\theta, t_1, \ldots, t_l) - \bar{\xi}_l(\theta, t_1, \ldots, t_l)| = 0,$$

*for all $l \geq 1$ and any compact subset $N_0$ of $\boldsymbol{R}^{1+s}$, where $\hat{\xi}_l$ is defined by (37).*

If the null hypothesis (17) is true, then it follows from Theorem 4 that $\mathrm{E}\hat{\xi}_l(\theta_0, t_1, \ldots, t_l) \equiv 0$ for every $l \geq 1$. Hence, since

$$\bar{\xi}_l(\theta, t_1, \ldots, t_l) = \lim_{n\to\infty} \mathrm{E}\hat{\xi}_l(\theta, t_1, \ldots, t_l), \tag{71}$$

we then also have $\bar{\xi}_l(\theta_0, t_1, \ldots, t_l) \equiv 0$ for $l \geq 1$.

Next, let $\bar{g}_l(z^{(1)}, \ldots, z^{(l)})$ be the conditional expectation function implied by $\bar{H}_l(y, z^{(1)}, \ldots, z^{(l)})$. Thus $\bar{g}_l$ is such that for any random drawing $(y_*, z^{(1)}_*, \ldots, z^{(l)}_*)$ from the distribution $\bar{H}_l$,

$$\mathrm{E}[y_* | z^{(1)}_*, \ldots, z^{(l)}_*] = \bar{g}_l(z^{(1)}_*, \ldots, z^{(l)}_*) \quad \text{a.s.} \tag{72}$$

Then it follows from part II of Theorem 4 that, for $l \geq m$,

$$\bar{\xi}_l(\theta, t_1, \ldots, t_l) = 0 \quad \text{for all} \quad t_1, \ldots, t_l \in N_0, \tag{73}$$

where $N_0$ is an arbitrary neighborhood of the origin of $\boldsymbol{R}^{1+s}$, if and only if

$$f^*(z^{(1)}_*, \ldots, z^{(m)}_*, \theta) = \bar{g}_l(z^{(1)}_*, \ldots, z^{(l)}_*) \quad \text{a.s.} \tag{74}$$

Thus redefining $\bar{\eta}_l(\theta)$ as

$$\bar{\eta}_l(\theta) = \int_{N_0} \cdots \int_{N_0} |\bar{\xi}_l(\theta, t_1, \ldots, t_l)|^2 \,\mathrm{d}t_1 \cdots \mathrm{d}t_l, \quad \text{[cf. (39)]} \tag{75}$$

retaining the definitions of $\hat{\eta}(\theta)$ and $\bar{\eta}(\theta)$ [see (41) and (42)], and restating the alternative hypothesis as follows:

$\mathrm{H}_1^*$: For each $\theta \in \Theta$ there is an $l \geq m$ such that, for a random drawing $(y_*, z_*^{(1)}, \ldots, z_*^{(l)})$ from $\bar{H}_l$,

$$\mathrm{P}\{f^*(z_*^{(1)}, \ldots, z_*^{(m)}, \theta) = \mathrm{E}[y_* | z_*^{(1)}, \ldots, z_*^{(l)}]\} < 1,$$

we now have the following modification of Theorem 5:

*Theorem 10. Let Assumptions 3, 4, 5 and 13 be satisfied. Then*

$$H_0 \text{ implies } \inf_{\theta \in \Theta} \bar{\eta}(\theta) = 0,$$

*and*

$$H_1^* \text{ implies } \inf_{\theta \in \Theta} \bar{\eta}(\theta) > 0.$$

However, the reverse of Theorem 10 is not generally true. Although

$$\inf_{\theta \in \Theta} \bar{\eta}(\theta) > 0 \quad \text{implies that } \mathrm{H}_1^* \text{ is true,} \tag{76}$$

it is possible that

$$\inf_{\theta \in \Theta} \bar{\eta}(\theta) = 0 \quad \text{while } \mathrm{H}_0 \text{ is false,}$$

namely in the case that the conditional expectations $\mathrm{E}[y_j | z_{j-1}, \ldots, z_{j-l}]$ are not time invariant. For example, consider the following univariate non-stationary time series process

$$y_j = \theta_j y_{j-1} + u_j, \tag{77}$$

where the $u_j$'s are NID(0,1) and the sequence $(\theta_j)$ of parameters satisfies

$$\sup_{-\infty < j < \infty} |\theta_j| < 1, \quad \lim_{|j| \to \infty} \theta_j = \theta_* \quad (\text{hence } |\theta_*| < 1), \tag{78}$$

but

$$\theta_j \neq \theta_* \quad \text{for all } j. \tag{79}$$

By backwards substitution of (77) we get

$$y_j = u_j + \theta_j u_{j-1} + \theta_j \theta_{j-1} u_{j-1} + \theta_j \theta_{j-1} \theta_{j-2} u_{j-2} + \cdots$$
$$= u_j + \theta_j \sum_{s=1}^{\infty} \left( \prod_{r=1}^{s} \theta_{j-s} \right) u_{j-s}, \tag{80}$$

hence $y_j$ is normally distributed with zero mean and variance

$$\sigma_j^2 = 1 + \theta_j^2 \sum_{s=1}^{\infty} \prod_{r=1}^{s} \theta_{j-s}^2. \tag{81}$$

Moreover, it is easy to verify that $(y_j, y_{j-1}, \ldots, y_{j-l})$ is $(l+1)$-variate normally distributed with zero mean vector and variance matrix

$$\Omega_j^{(l)} = \left( \mathrm{E} y_{j-j_1} y_{j-j_2} \right), \qquad j_1, j_2 = 0, 1, \ldots, l, \tag{82}$$

where

$$\begin{aligned} \mathrm{E} y_{j-j_1} y_{j-j_2} &= \sigma_{j-j_1}^2 = \sigma_{j-j_2}^2 && \text{if} \quad j_1 = j_2, \\ &= \left( \prod_{s=j_1}^{j_2} \theta_{j-s} \right) \sigma_{j-j_1}^2 && \text{if} \quad j_1 < j_2, \\ &= \left( \prod_{s=j_2}^{j_1} \theta_{j-s} \right) \sigma_{j-j_2}^2 && \text{if} \quad j_1 > j_2. \end{aligned} \tag{83}$$

Since (78) implies

$$\lim_{j \to \infty} \sigma_j^2 = \sum_{s=0}^{\infty} \theta_*^{2s}, \tag{84}$$

and

$$\lim_{j \to \infty} \mathrm{E} y_{j-j_1} y_{j-j_2} = \theta_*^{|j_1 - j_2|} \lim_{j \to \infty} \sigma_j^2 = \theta_*^{|j_1 - j_2|} \sum_{s=0}^{\infty} \theta_*^{2s}, \tag{85}$$

we have

$$\lim_{j \to \infty} \Omega_j^{(l)} = \Omega^{(l)} = \left( \theta_*^{|j_1 - j_2|} \sum_{s=0}^{\infty} \theta_*^{2s} \right), \qquad j_1, j_2 = 0, 1, \ldots, l. \tag{86}$$

In its turn (86) implies that for the example under review the distribution function $\overline{H}_l$ in Assumption 13 equals the distribution function of the $(l+1)$-

variate normal distribution with zero mean vector and variance matrix $\Omega^{(l)}$; hence the conditional expectation function $\bar{g}_l$ implied by $\overline{H}_l(y, y^{(1)}, \ldots, y^{(l)})$ is

$$\bar{g}_l(y^{(1)}, \ldots, y^{(l)}) = \theta_* y^{(1)}. \tag{87}$$

Thus, if we would specify the regression model (77) as

$$y_j = \theta_0 y_{j-1} + u_j, \tag{88}$$

then we would find

$$\inf_{\theta \in \Theta} \bar{\eta}(\theta) = \bar{\eta}(\theta_*) = 0, \tag{89}$$

while clearly the null hypothesis is false.

Furthermore, it is not hard to prove that in this example the 'limit' parameter $\theta_*$ is just the point $\theta_*$ considered in Assumption 7, so that, in view of Theorem 1, the least squares estimator $\hat{\theta}$ of the parameter $\theta_0$ in the misspecified model (88) converges in probability to $\theta_*$,

$$\operatorname*{plim}_{n\to\infty} \hat{\theta} = \theta_*.$$

However, $\sqrt{n}(\hat{\theta} - \theta_*)$ is only asymptotically normal distributed if

$$\operatorname*{plim}_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\theta_j - \theta_*) y_{j-1}^2 = 0,$$

which requires a further condition on the rate of convergence of $\theta_j$ to $\theta_*$, for example the condition

$$\frac{1}{n} \sum_{j=1}^{n} |\theta_j - \theta_*| = \mathrm{o}(1/\sqrt{n}).$$

This example shows that in the non-stationary case there is a class of misspecifications, contained in the gap between the null hypothesis (17) and the alternative hypothesis $\mathrm{H}_1^*$, which may not be detectable by our model specification test. On the other hand, the alternative hypothesis $\mathrm{H}_1^*$ contains a very wide class of severe misspecifications of which it is important to know whether they occur. Therefore it seems useful to apply our test even when the data-generating process is non-stationary.

The modification of the results in sections 4 through 8 to the non-stationary case does not involve particular difficulties. By referring to Lemma 3 instead of

Lemma 2 and replacing Assumption 11 by Assumption 13 the results simply go through or only need minor adjustments. In particular, statements such as 'if $H_0$ is false' (Theorems 7 and 8) should now be replaced by 'if $H_1^*$ is true'.

## References

Bierens, H.J., 1981, Robust methods and asymptotic theory in nonlinear econometrics, Lecture notes in economics and mathematical systems, Vol. 192 (Springer-Verlag, New York).

Bierens, H.J., 1982a, A uniform weak law of large numbers under $\phi$-mixing with application to nonlinear least squares estimation, Statistica Neerlandica 36, 81–86.

Bierens, H.J., 1982b, Consistent model specification tests, Journal of Econometrics 20, 105–134.

Bierens, H.J., 1983, Model specification testing of time series regressions, Research memorandum no. 8321 (Department of Economics, University of Amsterdam, Amsterdam).

Billingsley, P., 1968, Convergence of probability measures (Wiley, New York).

Brown, B.M., 1971, Martingale central limit theorems, Annals of Mathematical Statistics 42, 59–66.

Chung, K.L., 1974, A course in probability theory (Academic Press, New York).

Domowitz, I. and H. White, 1982, Misspecified models with dependent observations, Journal of Econometrics 20, 35–58.

Feller, W., 1966, An introduction to probability theory and its applications, Vol. II (Wiley, New York).

Granger, C.W.J., 1969, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37, 424–438.

Hannan, E.J., 1971, Nonlinear time-series regression, Journal of Applied Probability 8, 767–780.

Hansen, L.P., 1982, Large sample properties of generalized method of moment estimators, Econometrica 50, 1029–1054.

Klimko, L.A. and P.I. Nelson, 1978, On conditional least squares estimation for stochastic processes, Annals of Statistics 6, 629–642.

Robinson, P.M., 1972, Nonlinear regression for multiple time-series, Journal of Applied Probability 9, 758–768.

Sims, C.A., 1976, Asymptotic distribution theory for a class of nonlinear estimation methods, Discussion paper 76-69 (Center for Economic Research, University of Minnesota, Minneapolis, MN).

Sims, C.A., 1977, Exogeneity and causal ordening in macroeconomic models, in: C.A. Sims, ed., New methods in business cycle research (Federal Reserve Bank, Minneapolis, MN) 23–43.

Sims, C.A., 1980, Macroeconomics and reality, Econometrica 48, 1–48.

Sims, C.A., 1981, An autoregressive index model for the U.S. 1948–1975, in: J. Kmenta and J.B. Ramsey, eds., Large-scale macro-econometric models (North-Holland, Amsterdam) 283–327.