ELSEVIER

Journal of Econometrics 70 (1996) 159–174

JOURNAL OF
Econometrics

Testing for structural change in a long-memory environment

Javier Hidalgo*, Peter M. Robinson

Department of Economics, London School of Economics, London WC2A 2AE, UK

Abstract

Long-memory time-series analysis is apt to be applied to economic time series which extend over many years, in which circumstances the possibility of structural breaks is likely to be entertained. Tests for a change in parameter values at a given time point are proposed in linear regression models with long-memory errors. Existing tests based on the assumption of serially independent or weakly dependent errors will typically be invalid in such an environment. The tests are derived in case of certain nonstochastic and stochastic regressors, and are given large-sample justification. A small Monte Carlo study of finite-sample behaviour is included.

Key words: Structural change; Long-memory
JEL classification: 210; 211

1. Introduction

Testing for economic structural change is currently a very active research area (see, e.g., the collections of articles in the recent volumes of Hackl, 1989, and Hackl and Westlund, 1991). Early tests for structural change involved simple statistical models, and independence of observations across time. In most time series, the serial independence assumption is unlikely to be reasonable, and more recent literature has allowed for forms of weak dependence, as well as for unit root nonstationarity. However, such dependence structures comprise only

* Corresponding author.

This article is based on research funded by the Economic and Social Research Council (ESRC) under reference number R000233609. We thank F. Diebold and referees for helpful comments on a previous version of the paper, and Ignacio Lobato for the use of a computer program.

0304-4076/96/$15.00 © 1996 Elsevier Science S.A. All rights reserved
SSDI 0304407694 01687 U

a small proportion of the possibilities, and in time-series analysis there has been increasing interest in processes which have longer memory than weakly dependent ones, while at the same time being stationary. In particular, consider a scalar, zero-mean covariance-stationary process u_t, $t = 0, 1, \ldots$, having autocovariance function $\gamma_j = Eu_0 u_j$ satisfying

$$\gamma_j \sim Cj^{2d-1} \quad \text{as} \quad j \to \infty, \tag{1}$$

where

$$0 < C < \infty, \qquad 0 < d < \tfrac{1}{2},$$

and '$\sim$' indicates that the ratio of left- and right-hand sides tends to one. Although the γ_j decay to zero, they do so slowly, such that the γ_j are not summable. In general, processes u_t satisfying (1) do not satisfy the sort of mixing and functions-of-mixing conditions employed in much recent econometric central limit theory, such theory forming the basis of the usual rules of approximate statistical inference. Two parametric models which possess property (1) are the 'fractional noise' process given by

$$\gamma_j = \tfrac{1}{2}\gamma_0 \{|j+1|^{2d+1} - 2|j|^{2d+1} + |j-1|^{2d+1}\}, \qquad j = 1, 2, \ldots, \tag{2}$$

and a number of versions of the autoregressive fractionally integrated moving average given by

$$(1-L)^d(1 - a_1 L - \cdots - a_p L^p)u_t = (1 + b_1 L + \cdots + b_q L^q)\varepsilon_t, \tag{3}$$

where L is the lag operator, the pth- and qth-order polynomials have no zeros in or on the unit circle, and ε_t is a sequence of zero-mean uncorrelated homoscedastic variates. In line with much recent econometric limit theory for weakly dependent sequences one could doubtless straightforwardly extend our results to allow for forms of stable heterogeneity in u_t, but while of some practical value, such an extension would entail little theoretical interest.

The present paper explores some implications of long memory dependence of form (1) for the problem of testing for a change in parameter values in simple models. The problem seems a natural one to study because long-memory time-series analysis is often applied to series which extend over a long period of time, and the longer the time period the greater the possibility of structural breaks. Recent econometric work has studied structural change problems in environments which, while not allowing for long-memory behaviour of form (1), are in other respects extremely general. We are content to begin here by looking at linear regression models with possibly few parameters and with regressors of a rather specific form. We also envisage only a single change point at a known time, testing the null of no change in parameter against this alternative. Again, recent work has allowed for more than one change point, which can occur at unknown times. Our simple focus is justified in part because many models still used by practitioners are simple, and in a number of problems it is the occurrence or nonoccurrence of a change at a specific time point which is of interest. It is also justified by the relative novelty of studying structural change in a long-memory environment. Indeed, the statistical methodology and theory of multivariate and nonlinear time series exhibiting long-memory behaviour are not at all well-developed even in case of constant-parameter models.

In this paper we consider models of form

$$y_t = \beta\left(\frac{t}{n}\right)' x_t + u_t, \tag{4}$$

where x_t is a $K \times 1$ vector of observable regressors, the prime indicates transposition, $\beta(s)$ is a $K \times 1$ vector such that

$$\begin{aligned} \beta(s) &= \beta_A, \qquad 0 \leqslant s \leqslant \tau, \\ &= \beta_B, \qquad \tau < s < 1, \end{aligned} \tag{5}$$

for a known number τ between zero and one, and where the unobservable error u_t has autocovariances γ_j satisfying (1). We are concerned with testing

$$\mathrm{H}_0\colon \quad \beta_A = \beta_B \tag{6}$$

against the alternative

$$\mathrm{H}_1\colon \quad \beta_A \neq \beta_B. \tag{7}$$

We shall consider both nonstochastic (specifically polynomial-in-time) and stochastic regressors. The case of stochastic x_t has hardly been explored in regression models with long-memory errors, even in case of no strucutral change; some preliminary results are in Robinson (1994a) and more comprehensive ones are in Robinson and Hidalgo (1994). We shall also adopt the simplifying yet classically important assumption that u_t is Gaussian. It would certainly be possible to extend our work to certain non-Gaussian u_t, such as nonlinear functionals of a Gaussian process (Taqqu, 1975) or linear processes (Ibragimov and Linnik, 1971; Yajima, 1991).

The following section proposes a test in case of certain nonstochastic x_t. Section 3 explores the stochastic case. Some Monte Carlo simulations are included in Section 4. Section 5 contains some brief final comments.

2. Nonstochastic regressors

We shall employ a type of Wald testing procedure, which involves assessing the significance of $\hat{\beta}_A - \hat{\beta}_B$, where $\hat{\beta}_A$ and $\hat{\beta}_B$ are estimates of β_A and β_B. Because the power of a test tends to vary with the efficiency of the parameter estimates employed, in general we would wish to employ a form of generalized least squares (GLS) estimate. The usual formula for GLS applies here, expressed in terms of the autocovariances γ_j. If the γ_j are known up to scale, this GLS estimate can be calculated. However, it is rather complicated to calculate and theoretically analyze, and there is evidence that for some simple nonstochastic regressors, ordinary least squares (OLS) has fairly good asymptotic relative efficiency when γ_j satisfies (1). (In case of weakly dependent u_t, OLS is actually asymptotically fully efficient for certain regressors, see Grenander, 1954.) Moreover, if the γ_j are not known up to scale, so that either we have a parametric model, or a semiparametric one such as (1), feasible GLS estimates are even more difficult to analyze, and apart from the recent paper of Dahlhaus (1992) little study of the asymptotic properties of feasible GLS estimates of regressions with long-memory errors has yet been made. Therefore we employ OLS estimates. Yajima (1991) has recently discussed the limiting distributional properties of OLS estimates of constant-parameter models with nonstochastic regressors and long-memory errors.

For given τ, define $h = [\tau n]$, where [.] indicates integer part. Correspondingly, introduce $X_1 = (x_1, \dots, x_h)'$, $X_2 = (x_{h+1}, \dots, x_n)'$, $Y_1 = (y_1, \dots, y_h)'$, $Y_2 = (y_{h+1}, \dots, y_n)'$, and then estimate β_A and β_B in (4) and (5) by

$$\hat{\beta}_A = (X_1' X_1')^{-1} X_1' Y_1, \qquad \hat{\beta}_B = (X_2' X_2)^{-1} X_2' Y_2. \tag{8}$$

Put

$$W = ((X_1' X_1)^{-1} X_1', -(X_2' X_2)^{-1} X_2')', \qquad u = (u_1, \dots, u_n)',$$

and let w_t be the tth column of W'. Under H_0 in (6),

$$\hat{\beta}_A - \hat{\beta}_B = W' u. \tag{9}$$

Assuming also that u_t is Gaussian with zero mean, (9) gives

$$\hat{\beta}_A - \hat{\beta}_B \sim \mathrm{N}(0, W' \Gamma W), \tag{10}$$

where Γ is the $n \times n$ Toeplitz matrix (γ_{s-t}).

The covariance matrix $W' \Gamma W$ depends on all the γ_j, $j = 0, 1, \dots, n-1$. In some important cases it will be possible to approximate $W' \Gamma W$ for large n in such a way that only the behaviour of γ_j for large j will be relevant, so that the semiparametric model (1) can be used. A leading case is that of time-trending polynomial regressors, that is

$$x_t = (1, t, \dots, t^{K-1})'.$$

Define

$$D = \operatorname{diag}\left\{n, \sum_1^n t^2, \dots, \sum_1^n t^{2(K-1)}\right\},$$

and note that as $n \to \infty$

$$D \sim \operatorname{diag}\left\{n, \frac{n^3}{3}, \dots, \frac{n^{2K-1}}{2K-1}\right\}.$$

Write

$$R_i = \lim_{n\to\infty} D^{-1/2} X_i' X_i D^{-1/2}, \qquad i = 1, 2,$$

so that $R_i = \Delta^{-1/2} Q_i \Delta^{-1/2}$, where

$$\Delta = \operatorname{diag}\{1, \tfrac{1}{3}, \dots, 1/(2K-1)\},$$

$$Q_1 = \left[\frac{\tau^{i+j-1}}{i+j-1}\right], \qquad Q_2 = \left[\frac{1-\tau^{i+j-1}}{i+j-1}\right],$$

where the (i,j)th elements of the matrices are indicated. Defining $\mu_t' = (1, t/n, \dots, (t/n)^{K-1})u_t$, we have, much as in, e.g., Lemma 2 of Robinson (1994c) (see also Yajima, 1988),

$$\lim_{n\to\infty} n^{-1-2d} \sum_{s,}^{h} \sum_{t=1}^{n} \mathrm{E}(\mu_s \mu_t') = \left[\int_0^\tau \int_0^\tau s^{i-1} t^{j-1} |s-t|^{2d-1}\, \mathrm{d}s\mathrm{d}t\right] = A_1,$$

$$\lim_{n\to\infty} n^{-1-2d} \sum_{s=1}^{h} \sum_{t=h+1}^{n} \mathrm{E}(\mu_s \mu_t') = \left[\int_0^\tau \int_\tau^1 s^{i-1} t^{j-1} |s-t|^{2d-1}\, \mathrm{d}s\mathrm{d}t\right] = A_2,$$

$$\lim_{n\to\infty} n^{-1-2d} \sum_{s=h+1}^{n} \sum_{t=1}^{h} \mathrm{E}(\mu_s \mu_t') = \left[\int_\tau^1 \int_0^\tau s^{i-1} t^{j-1} |s-t|^{2d-1}\, \mathrm{d}s\mathrm{d}t\right] = A_2',$$

$$\lim_{n\to\infty} n^{-1-2d} \sum\sum_{s,t=h+1}^{n} \mathrm{E}(\mu_s \mu_t') = \left[\int_\tau^1 \int_\tau^1 s^{i-1} t^{j-1} |s-t|^{2d-1}\, \mathrm{d}s\mathrm{d}t\right] = A_3,$$

where again the (i,j)th elements of the matrices are indicated. We deduce from the preceding arguments that, under H_0,

$$n^{-d} D^{1/2} (\hat\beta_A - \hat\beta_B) \to_d \mathrm{N}(0, \Omega), \tag{11}$$

where

$$\Omega = C\Delta^{1/2}(Q_1^{-1} A_1 Q_1^{-1} - Q_1^{-1} A_2 Q_2^{-1} - Q_2^{-1} A_2' Q_1^{-1} + Q_2^{-1} A_3 Q_2^{-1})\Delta^{1/2}. \tag{12}$$

The evaluation of A_1, A_2, and A_3 for given τ and d in general requires numerical integration. In the simplest case $K = 1$, corresponding to the simple classical change-point problem involving only an intercept parameter, a closed-form formula is available. We have

$$A_1 = \int_0^\tau \int_0^\tau |s-t|^{2d-1}\,\mathrm{d}s\mathrm{d}t = \frac{\tau^{2d+1}}{d(2d+1)},$$

$$A_2 = \int_0^\tau \int_\tau^1 |s-t|^{2d-1}\,\mathrm{d}s\mathrm{d}t = \frac{1-(1-\tau)^{2d+1}-\tau^{2d+1}}{2d(2d+1)},$$

$$A_3 = \int_\tau^1 \int_\tau^1 |s-t|^{2d-1}\,\mathrm{d}s\mathrm{d}t = \frac{(1-\tau)^{2d+1}}{d(2d+1)}.$$

(Notice that $A_2 \neq 0$, whereas for weakly dependent u_t, with summable γ_j, we would have $A_2 = 0$.) Because $\Delta = 1$ and $Q_1 = \tau$, $Q_2 = 1 - \tau$ in the pure intercept case,

$$\Omega = \frac{C}{d(2d+1)}\left[\frac{\tau^{2d+1}}{\tau^2} - \left\{\frac{1-(1-\tau)^{2d+1}-\tau^{2d+1}}{\tau(1-\tau)}\right\} + \frac{(1-\tau)^{2d+1}}{(1-\tau)^2}\right]$$
$$= \frac{C}{d(2d+1)}\left\{\frac{\tau^{2d}+(1-\tau)^{2d}-1}{\tau(1-\tau)}\right\} \quad \text{for} \quad 0<\tau<1. \tag{13}$$

Although τ is assumed known, in applications C and d will generally be unknown, so that the formula (12) for Ω, and its special case (13) when $K = 1$, will be infeasible, while d is also involved in the norming in (11). However, in the same way as was shown by Robinson (1994b) in the constant-parameter case (and for general $K \geqslant 1$),

$$n^{-2\hat{d}}(\hat{\beta}_A - \hat{\beta}_B)' D^{1/2} \hat{\Omega}^{-1} D^{1/2} (\hat{\beta}_A - \hat{\beta}_B) \to_d \chi^2_K, \tag{14}$$

if $\hat{\Omega}$ is computed from the above formulae (12) or (13) with C and d replaced by $\hat{C}$ and $\hat{d}$ such that, as $n \to \infty$, $\hat{C} \to_p C$ and $\hat{d} = d + o_p((\log n)^{-1})$. Notice that the usual 'autocorrelation-consistent covariance matrix' form of scaling which has been heavily stressed in recent econometric work, involving a spectral density estimate at zero frequency, is inappropriate here (because the spectral density of u_t is infinite at zero frequency) and would produce an invalid test.

The need for a d-estimate which differs by $o_p((\log n)^{-1})$ from d was pointed out by Robinson (1994b) in case of polynomial regression with constant parameters. Models (2) and (3) are special cases of parametric models, in which, for each j, γ_j is a uniquely defined function of j and of an unknown parameter vector which includes d as an element. For such correctly specified parametric models, Gaussian estimates have been shown to be root-n-consistent under various conditions, by Fox and Taqqu (1986) and Giraitis and Surgailis (1990). Such estimates have been applied to economic time series by Diebold et al. (1990) and Sowell (1990), and are clearly attractive if a parametric model can be specified with some certainty. In the present case they would be based on OLS residuals from regressions before and after the hypothesized structural break. However, if the parametric model is misspecified (for example if p or q is understated in (3)) the estimate of d will be inconsistent, and use of such an estimate in (14) will produce an asymptotically invalid test.

So far as (14) is concerned, estimates of parameters describing short- or medium-run behaviour (such as the ARMA coefficients in (3)) are useless, and moreover only a better-than-log n-consistent estimate of d, and a consistent estimate of C, are needed. It is possible to obtain such estimates in a much broader environment than that of a specific parametric model, under conditions on the autocovariances almost equivalent to (1), and when n is very large these have strong appeal. Two such types of estimates are based on the periodogram at low frequencies. Defining the OLS residuals

$$\begin{aligned} \hat{u}_t &= y_t - \hat{\beta}'_A x_t, \qquad 1 \leqslant t \leqslant h, \\ &= y_t - \hat{\beta}'_B x_t, \qquad h+1 \leqslant t \leqslant n, \end{aligned}$$

the periodogram is given by

$$I(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} \hat{u}_t \, \mathrm{e}^{it\lambda} \right|^2 .$$

Now define

$$\hat{F}(\lambda) = \frac{2\pi}{n} \sum_{j=1}^{[n\lambda/2\pi]} I(\lambda_j),$$

where [.] again denotes integer part and $\lambda_j = 2\pi j/n$. Following Robinson (1994b), we can then estimate d and C by

$$\hat{d}_1 = \tfrac{1}{2} - \frac{\log\{\hat{F}(q\lambda_m)/\hat{F}(\lambda_m)\}}{2\log q},$$

$$\hat{C}_1 = 2\Gamma(2-2\hat{d})\cos\left((\tfrac{1}{2}-\hat{d})\pi\right)\hat{F}(\lambda_m)\,\lambda_m^{2\hat{d}-1},$$

for any $q \in (0, 1)$, where m is a 'bandwidths' sequence that tends to infinity slower than n. Note that here we use the correspondence (under suitable conditions) between $\gamma_j \sim Cj^{2d-1}$ as $j \to \infty$, and $f(\lambda) \sim C\lambda^{-2d}/2\Gamma(1-2d)\cos((\frac{1}{2}-d)\pi)$ as $\lambda \to 0+$, where $f(\lambda)$ is the spectral density of u_t. It follows from Robinson (1994b) that under rather mild conditions (which do not include Gaussianity of u_t) $\hat{d}_1 = d + o_p((\log n)^{-1})$, $\hat{C}_1 \to_p C$. A much longer-established type of estimate is due to Geweke and Porter-Hudak (1983). Here the estimate of d is the slope coefficient in the OLS regression of $\log I(\lambda_j)$ on $-\log(4\sin^2\frac{1}{2}\lambda_j)$ and an intercept, for $j = 1, \ldots, m$. It has been applied to economic data by Diebold and Rudebusch (1989, 1991). However, Geweke and Porter-Hudak (1983) considered asymptotic properties only in case $-\frac{1}{2} < d < 0$ and even here their proof was incorrect, as shown by Robinson (1994c). Their proof requires *inter* *alia* that the $I(\lambda_j)/f(\lambda_j)$, where $f(\lambda_j)$ is the spectral density of u_t, are asymptotically independent and identically distributed as $n \to \infty$ for fixed j, and this is not true for $d > 0$ or $d < 0$. Following an earlier suggestion of Künsch (1986), Robinson (1994c) thus carried out the OLS regression of $\log I(\lambda_j)$ only over $j = l + 1, \dots, m$, where l is a trimming number which tends to infinity slower than m. Call this $\hat{d}_2$, and denote by $\hat{C}_2$ an estimate of C obtained from the intercept OLS estimate and the correspondence described above. (In fact Robinson, 1994c, obtained a more efficient estimate than this, and considered a multivariate time-series extension.) In case of Gaussian u_t, Robinson (1994c) established the asymptotic distribution properties of $\hat{C}_2$ and $\hat{d}_2$ and for suitable m these imply the desired properties $\hat{d}_2 = d + o_p(\log n)^{-1})$ and $\hat{C}_2 \to_p C$.

The form of the test statistic in (14) is strongly influenced by the property that the polynomial-in-time regressors have all their spectral mass concentrated at zero frequency. Other nonstochastic regressors, such as cosinusoids in time, will lead to formulae of a different type and will involve estimating other features of the autocovariance structure of u_t; see Yajima (1991) in the constant-parameter case. Different forms of test statistic are also liable to arise in case of stochastic regressors, to which we now turn.

3. Stochastic regressors

When $\{x_t\}$ is stochastic but independent of $\{u_t\}$ (which is not true when x_t includes lagged y_t), then (10) holds conditionally on $\{x_t\}$, under H_0, so that

$$(W'\Gamma W)^{-1/2}(\hat{\beta}_A - \hat{\beta}_B) \sim N(0, I_k)$$

unconditionally, where I_K is the K-rowed identity matrix. However, we are left with the problem of finding a feasible proxy $\hat{\Sigma}$ for $W'\Gamma W$ such that

$$(\hat{\beta}_A - \hat{\beta}_B)'\{\hat{\Sigma}^{-1} - (W'\Gamma W)^{-1}\}(\hat{\beta}_A - \hat{\beta}_B) \to_p 0 \quad \text{as} \quad n \to \infty,$$

in order to obtain that

$$(\hat{\beta}_A - \hat{\beta}_B)'\,\hat{\Sigma}^{-1}\,(\hat{\beta}_A - \hat{\beta}_B) \to_d \chi^2_K \quad \text{as} \quad n \to \infty.$$

In case of weakly dependent u_t and x_t, an autocorrelation-consistent covariance matrix estimate $\hat{\Sigma}$ can be employed which involves an estimate of the spectral density of $x_t u_t$ at zero frequency, much as in the estimation of the limiting covariance matrix of estimates of constant-parameter models, as indicated by Brillinger (1979) and on countless subsequent occasions in the literature. When u_t satisfies (1), matters are much more complicated. In case of constant-parameter regression, a discussion of this problem was initiated by Robinson (1994a) and developed by Robinson and Hidalgo (1994). The outcome depends on whether or not x_t has zero mean, whether or not there is an intercept, whether or not x_t is long memory, whether or not the elements of a vector x_t are cross-correlated, and on the combined extent of the memory of x_t and u_t. It is possible for the strong autocorrelation in u_t and x_t to be consistent with a finite spectral density matrix of $u_t x_t$ at zero frequency, when the standard method mentioned above is still applicable (though the usual sort of justification based on mixing processes is not). However, if $u_t x_t$ itself has an infinite spectral density matrix at zero frequency, then this standard method (whether employing a weighted autocovariance, autoregressive or other spectral density estimate) is inconsistent, and we can also end up with parameter estimates which have a singular limiting normal distribution, or which can be nonnormal. The same problems beset $\hat{\beta}_A - \hat{\beta}_B$, so that a test which may be valid for some x_t sequences will be asymptotically invalid for others.

We shall not discuss the various possibilities here but instead we propose a simple robust procedure. Let us consider only the case of a simple stochastic regression

$$y_t = \beta_1\left(\frac{t}{n}\right) + \beta_2\left(\frac{t}{n}\right) z_t + u_t, \tag{15}$$

so $K = 2$, $\beta_1(s)$, $\beta_2(s)$ are the components of $\beta(s)$, and z_t is a scalar stationary process, independent of u_t. The case of multiple stochastic regressors can readily be handled, although it complicates matters without giving any further insight regarding the proposed simple robust procedure. Again we test H_0 in (6) against H_1 in (7). Consider first $\hat{\beta}_A = (\hat{\beta}_{A1}, \hat{\beta}_{A2})'$ given by (8). Now

$$\hat{\beta}_{A1} = \bar{y}_A - \hat{\beta}_{A2}\bar{z}_A,$$

where $\bar{y}_A = h^{-1}\sum_{t=1}^{h} y_t$ and $\bar{z}_A = h^{-1}\sum_{t=1}^{h} z_t$. For any ζ we have

$$\begin{aligned}\hat{\beta}_{A1} + \zeta\hat{\beta}_{A2} &= \bar{y}_A - \hat{\beta}_{A2}(\bar{z}_A - \zeta)\\ &= \beta_{A1} + \zeta\beta_{A2} + \bar{u}_A - (\hat{\beta}_{A2} - \beta_{A2})(\bar{z}_A - \zeta).\end{aligned}$$

Similarly, in an obvious notation,

$$\begin{aligned}\hat{\beta}_{B1} + \zeta\hat{\beta}_{B2} &= \bar{y}_B - \hat{\beta}_{B2}(\bar{z}_B - \zeta)\\ &= \hat{\beta}_{B1} = \zeta\hat{\beta}_{B2} + \bar{u}_B - (\hat{\beta}_{B2} - \beta_{B2})(\bar{z}_B - \zeta).\end{aligned}$$

Now as in Section 2 we have, under (1) and Gaussianity,

$$n^{1/2-d}(\bar{u}_A - \bar{u}_B) \rightarrow_d \mathrm{N}\left[0, \frac{C}{d(2d+1)}\left\{\frac{\tau^{2d} + (1-\tau)^{2d} - 1}{\tau(1-\tau)}\right\}\right],$$

for $0 < \tau < 1$ (see (13)). Now consider

$$\hat{\beta}_{A2} - \beta_{A2} = \sum_{t=1}^{h}(z_t - \bar{z}_A)u_t \Big/ \sum_{t=1}^{h}(z_t - \bar{z}_A)^2.$$

If z_t is ergodic in the sense that $\delta_j = \operatorname{cov}(z_1, z_{1+j}) \to 0$, as $j \to \infty$, and $n^{-1}\sum_{t=1}^{n}(z_t - \mathrm{E}z_t)^2 \to_p \delta_0$ as $n \to \infty$, then also $h^{-1}\sum_{t=1}^{h}(z_t - \bar{z}_A)^2 \to_p \delta_0$. On the other hand,

$$\mathrm{E}\left\{\sum_{t=1}^{h} (z_t - \bar{z}_A)u_t\right\}^2 = \sum\sum_{s,t=1}^{h} \gamma_{s-t}\, \mathrm{E}\,\{(z_s - \bar{z}_A)(z_t - \bar{z}_A)\},$$

which is

$$\sum\sum_{s,t=1}^{h} \gamma_{s-t}\left\{\delta_{s-t} - \frac{1}{h}\sum_{j=1}^{h}(\delta_{s-j} + \delta_{t-j}) + \frac{1}{h^2}\sum\sum_{j,k=1}^{h}\delta_{j-k}\right\}.$$

For m sufficiently large and any $\varepsilon > 0$, there exists l such that $l \leqslant \varepsilon h$ and $|\delta_j| \leqslant \varepsilon$ for all $j > l$. Thus

$$\begin{aligned}\left|\sum\sum_{s,t=1}^{h}\gamma_{s-t}\delta_{s-t}\right| &\leqslant h\sum_{1-h}^{h-1}|\gamma_j\delta_j| \\ &\leqslant h\delta_0\sum_{j=1-l}^{l-1}|\gamma_j| + 2\varepsilon h\sum_{j=l}^{h-1}|\gamma_j| \\ &= \mathrm{O}(h(\varepsilon h)^{2d} + \varepsilon h^{2d+1}) \\ &= \mathrm{o}(h^{2d+1}) = \mathrm{o}(n^{2d+1}) \quad \text{as} \quad n \to \infty,\end{aligned}$$

because ε is arbitrary. We likewise deduce that, as $n \to \infty$, $h^{-1}\sum_{j=1}^{h}(\delta_{s-j} + \delta_{t-j}) \to 0$ uniformly in s, t and $h^{-2}\sum\sum_{j,k=1}^{h}\delta_{j-k} \to 0$. It follows that

$$\mathrm{E}\left\{\sum_{t=1}^{h}(z_t - \bar{z}_A)u_t\right\}^2 = \mathrm{o}\,(n^{2d+1}) \quad \text{as} \quad n \to \infty,$$

and thus, by Slutzky's theorem, for $\delta_0 > 0$,

$$\hat{\beta}_{A2} - \beta_{A2} = \mathrm{o}_p(n^{d-1/2}) \quad \text{as} \quad n \to \infty.$$

Likewise we deduce that

$$\hat{\beta}_{B2} - \beta_{B2} = \mathrm{o}_p(n^{d-1/2}) \quad \text{as} \quad n \to \infty.$$

Because $\bar{z}_A - \zeta$ and $\bar{z}_B - \zeta$ are both $\mathrm{O}_p(1)$ it follows that

$$(1, \zeta)\,(\hat{\beta}_A - \hat{\beta}_B) = \bar{u}_A - \bar{u}_B + \mathrm{o}_p(n^{d-1/2}),$$

and thus that

$$n^{1/2-d}(1,\zeta)\,(\hat{\beta}_A - \hat{\beta}_B) \to_d \mathrm{N}\left[0, \frac{C}{d(2d+1)}\left\{\frac{\tau^{2d} + (1-\tau)^{2d} - 1}{\tau(1-\tau)}\right\}\right]. \tag{16}$$

Now (16) holds not only under H_0 but whenever

$$\beta_{A1} - \beta_{B1} + \zeta(\beta_{A2} - \beta_{B2}) = 0,$$

that is for all departures from the null hypothesis such that

$$\frac{\beta_{A1} - \beta_{B1}}{\beta_{B2} - \beta_{A2}} = \zeta. \tag{17}$$

These are inconsistent directions of a test based on (16). But if (17) holds, then

$$\frac{\beta_{A1} - \beta_{B1}}{\beta_{B2} - \beta_{A2}} \neq \xi,$$

for $\xi \neq \zeta$. Now from the above argument, clearly

$$n^{1/2-d}(1, \xi)\,(\hat{\beta}_A - \hat{\beta}_B) \to_d \mathrm{N}\left[0, \frac{C}{d(2d+1)}\left\{\frac{\tau^{2d} + (1-\tau)^{2d} - 1}{\tau(1-\tau)}\right\}\right],$$

indeed

$$(1, \xi)\,(\hat{\beta}_A - \hat{\beta}_B) = \bar{u}_A - \bar{u}_B + \mathrm{o}_p(n^{d-1/2}).$$

Thus $n^{1/2-d}(1, \zeta)(\hat{\beta}_A - \hat{\beta}_B)$ and $n^{1/2-d}(\hat{\beta}_{A1} - \hat{\beta}_{B1})$ converge in distribution to the same random variable. It follows as in (16) that for any ζ, ξ, such that $\zeta \neq \xi$, and any $\pi \in (0, 1)$,

$$\frac{n^{1-2d}(\hat{\beta}_A - \hat{\beta}_B)'\left\{\pi\begin{bmatrix}1 & \zeta \\ \zeta & \zeta^2\end{bmatrix} + (1-\pi)\begin{bmatrix}1 & \xi \\ \xi & \xi^2\end{bmatrix}\right\}(\hat{\beta}_A - \hat{\beta}_B)}{\dfrac{C}{d(2d+1)}\left\{\dfrac{\tau^{2d} + (1-\tau)^{2d} - 1}{\tau(1-\tau)}\right\}} \to_d \chi_1^2, \tag{18}$$

under H_0, and that the corresponding test will be consistent against all departures (7). We could in fact consider a weighted average of additional terms of the same form though this does add to the ambiguity associated with the precise choice of statistic. Simple choices of ζ, ξ, and π in (18) are

$$\xi = 1, \quad \zeta = -1, \quad \pi = \tfrac{1}{2}. \tag{19}$$

Under (19), (18) reduces to

$$\frac{n^{1-2d}\{(\hat{\beta}_{A1} - \hat{\beta}_{B1})^2 + (\hat{\beta}_{A2} - \hat{\beta}_{B2})^2\}}{\dfrac{C}{d(2d+1)}\left\{\dfrac{\tau^{2d} + (1-\tau)^{2d} - 1}{\tau(1-\tau)}\right\}}. \tag{20}$$

Notice that under H_0 the first term in braces in the numerator determines the limiting distribution of (20), while under H_1 either or both terms in the braces in the numerator will cause (20) to tend to infinity in probability. Notice also the similarity of (20) to a statistic for the weakly dependent case in which $\hat{\beta}_{A1}$, $\hat{\beta}_{B1}$ are asymptotically independent of $\hat{\beta}_{A2}$, $\hat{\beta}_{B2}$. In any case, to obtain a feasible test statistic we have to insert consistent estimates of C and better-than-log n-consistent estimates of d, and these can be obtained from OLS residuals as indicated in the previous section.

4. Monte Carlo simulations

In the present section we study finite-sample performance of the test proposed in the previous section by means of a Monte Carlo study. We employ the model (15), where u_t and z_t are both Gaussian with zero mean and unit variance and the autocovariances γ_j and δ_j of u_t and z_t are given by

$$\gamma_j = Cj^{2d-1}, \qquad \delta_j = Bj^{2g-1}, \qquad j = 1, 2, \ldots, \tag{21}$$

for $B = C = 0.7$ and two different values of $d = g$, namely 0.2 and 0.3. We took $\tau = \frac{1}{2}$ throughout, and five different sets of parameter values β_{A1}, β_{B1}, β_{A2}, β_{B2} were employed. Throughout $\beta_{A1} = \beta_{A2} = 1$, while (β_{B1}, β_{B2}) was variously (1, 1), (1.5, 1.5), (2, 1.5), (0.5, 1.5), (0, 1.5). Thus in the first case we study size, while in the last four cases we study power, for various discrepancies in the intercept and slope values before and after the structural break.

All computations were carried out in double precision FORTRAN on the L.S.E.'s VAX computer, using a random number generator from the NAG library, the algorithm of Davies and Harte (1987) and the fast Fourier transform to generate the u_t, z_t. For each choice of parameter 5000 replications of series of lengths $T = 128$, 256, 512 were generated. We computed four different test statistics. Test 1 uses (20) with the true C and d values. Of course this test is almost certainly practically infeasible, but it was performed in order to help assess the impact of nuisance parameter estimation employed in other tests. Test 3 uses (20) with d and C replaced by $\hat{d}_1$ and $\hat{C}_1$. Test 2 uses (20) with $\hat{d}_1$ again, but in place of $\hat{C}_1$, $\hat{C}_3 = 0.1 \sum_{j=1}^{10} \hat{\gamma}_{N+j}$, where $\hat{\gamma}_1 = (n-\mathrm{j})^{-1} \sum_{t=j+1}^{n} \hat{u}_t \hat{u}_{t-j}$. $\hat{C}_3$ was suggested by Robinson (1994a), and is evidently algebraically far simpler than $\hat{C}_1$. Finally, test 4 follows conventional econometric practice by using a standard autocorrelation-consistent estimate $\hat{\Sigma}$, based on the incorrect assumption that the spectral density of u_t is finite at zero frequency, specifically the Bartlett estimate

$$\hat{\Sigma} = \begin{bmatrix} \hat{\gamma}_0 + 2\sum_{j=1}^{M}\left(1 - \frac{j}{M+1}\right)\hat{\gamma}_j & 0 \\ 0 & \dfrac{\hat{\eta}_{A0} + 2\sum_{j=1}^{M}\left(1 - \frac{j}{M+1}\right)\hat{\eta}_{Aj}}{\left[\sum_{1}^{h} z_j^2\right]\Big/ n} + \dfrac{\hat{\eta}_{B0} + 2\sum_{j=1}^{M}\left(1 - \frac{j}{M+1}\right)\hat{\eta}_{Bj}}{\left[\sum_{h+1}^{n} z_j^2\right]\Big/ n} \end{bmatrix},$$

where

$$\hat{\eta}_{Aj} = \frac{1}{n-j} \sum_{t=1}^{h-j} \hat{u}_t z_t \hat{u}_{t+j} z_{t+j}, \qquad \hat{\eta}_{Bj} = \frac{1}{n-j} \sum_{t=h+1}^{n-j} \hat{u}_t z_t \hat{u}_{t+j} z_{t+j}.$$

This estimate uses knowledge that the slope and intercept parameter estimates are asymptotically independent.

Table 1
Bandwidth values used in the tests

	$d = 0.2$			$d = 0.3$		
Test	$n = 128$	$n = 256$	$n = 512$	$n = 128$	$n = 256$	$n = 512$
2	24, 60	40, 125	90, 250	10, 60	20, 125	30, 250
3	90	124	135	70	150	300
4	13	13	19	13	13	19

For test 2 the value of (m, N) is indicated, and for tests 3 and 4 the cells indicate the values of m and M, respectively.

Table 2
$d = 0.2$

		$n = 128$			$n = 256$			$n = 512$		
(β_{B1}, β_{B2})	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
(1, 1)	1	9.92	4.74	0.92	10.10	4.70	0.75	9.60	4.10	0.82
size	2	11.66	6.55	2.40	9.20	4.92	1.84	9.46	5.48	1.76
	3	23.34	14.42	4.40	14.56	6.88	1.64	10.50	4.72	1.22
	4	38.00	30.10	17.80	36.10	27.14	15.00	34.12	26.00	13.30
(1.5, 1.5)	1	38.74	25.42	7.74	60.94	43.84	19.92	76.08	62.54	26.70
size-	2	33.28	21.16	5.44	43.53	29.08	7.90	62.52	43.36	14.58
corr.	3	42.22	28.02	9.60	37.44	23.22	5.90	74.56	59.64	25.00
power	4	40.06	27.46	8.94	62.68	46.72	21.90	84.84	73.20	43.22
(2.0, 1.5)	1	72.28	60.18	31.30	90.34	82.66	63.56	97.42	94.78	78.82
size-	2	57.10	42.96	17.22	73.24	59.68	26.06	90.90	80.50	46.58
corr.	3	74.32	63.12	34.06	78.52	65.80	32.04	96.92	93.30	73.88
power	4	67.86	53.96	25.46	87.98	79.38	56.90	98.00	95.20	81.48
(0.5, 1.5)	1	39.00	25.14	7.38	60.80	44.72	20.80	77.44	63.86	28.02
size-	2	33.56	21.32	5.60	42.48	29.04	8.14	63.44	44.12	15.44
corr.	3	43.20	28.82	9.30	37.42	23.02	5.78	75.60	61.12	25.60
power	4	40.84	27.54	8.00	62.26	47.18	22.24	85.72	74.44	44.30
(0.0, 1.5)	1	72.96	60.18	31.02	90.64	82.58	63.06	97.66	95.16	79.78
size-	2	57.02	42.16	17.50	12.44	58.34	25.96	91.52	80.62	47.08
corr.	3	74.02	62.50	34.92	78.22	64.74	32.18	97.20	94.04	74.36
power	4	68.66	55.22	25.20	88.20	79.36	56.66	98.40	95.96	82.20

Table 1 presents the values of m, N, and M used in the results displayed in Tables 2 and 3, which record for $d = 0.2$ and 0.3 respectively Monte Carlo sizes [for $(\beta_{B1}, \beta_{B2}) = (1.1)$] and size-corrected power [for $(\beta_{B1}, \beta_{B2}) = (1.5, 1.5)$, (2, 1.5), (0.5, 1.5), (0, 1.5)] for tests of level 10%, 5%, 1%. Other bandwidth values were also tried, but the ones for which results are reported on the whole gave the best results.

For the smaller values of n, the size of test 2 tends to be much closer to the asymptotic size than that of test 3, though the discrepancy is significantly smaller for $d = 0.2$ than for $d = 0.3$, and when $n = 512$, there is little difference, indeed it is hard to choose between these two tests an the infeasible test 1 in the latter case. Moreover, in terms of size test 2 is quite close to test 1 throughout. The invalidity of test 4 shows up through very large sizes.

Where size-corrected power is concerned, test 3 rivals the infeasible test 1, and is almost uniformly better than test 2, often much better. There is also an almost uniform superiority in power for $d = 0.2$ relative to $d = 0.3$. Moreover, power

Table 3
$d = 0.3$

		$n = 128$			$n = 256$			$n = 512$		
(β_{B1}, β_{B2})	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
(1, 1)	1	12.70	6.54	1.34	12.10	6.02	1.12	11.18	4.90	1.16
size	2	10.70	6.90	3.36	11.32	7.32	3.04	11.12	6.68	2.68
	3	30.78	20.66	8.18	22.86	13.70	4.36	14.94	7.26	2.04
	4	41.36	32.46	20.62	41.54	33.30	19.20	41.28	32.44	19.50
(1.5, 1.5)	1	37.14	24.26	7.16	47.72	30.70	11.86	61.88	46.26	17.30
size-	2	26.70	15.46	3.88	35.12	19.02	6.28	43.26	27.48	9.62
corr.	3	38.08	24.86	8.34	32.94	20.26	5.42	61.64	44.26	16.56
power	4	36.12	24.66	7.42	52.82	36.92	15.54	71.60	57.18	25.12
(2, 1.5)	1	69.90	57.66	28.14	82.08	71.10	44.86	92.08	85.88	61.12
size-	2	41.66	29.50	9.96	58.30	41.44	18.30	70.52	56.22	27.58
corr.	3	70.20	57.78	30.32	71.84	57.44	27.24	91.34	83.90	58.70
power	4	63.26	50.06	22.48	81.62	70.30	46.26	92.68	86.70	62.16
(0.5, 1.5)	1	38.08	23.88	6.88	48.14	32.28	11.26	63.12	47.34	17.06
size	2	25.96	16.18	4.14	34.00	19.32	6.52	43.86	28.54	10.16
corr.	3	39.42	25.38	7.96	32.50	19.60	5.28	62.96	45.66	16.70
power	4	36.68	24.48	6.68	53.74	37.86	16.32	72.10	58.30	25.94
(0, 1.5)	1	70.76	57.90	28.52	81.84	71.02	45.56	92.48	86.62	61.72
size-	2	41.44	28.90	10.92	57.06	40.84	18.10	71.08	56.68	28.34
corr.	3	69.58	57.20	31.28	70.70	55.96	27.10	92.06	85.00	59.22
power	4	64.62	50.76	22.44	81.38	69.98	46.34	93.40	87.14	62.70

seems to improve more slowly as n increases for $d = 0.3$ than for $d = 0.2$, as may possibly be explained by slower rates of convergence of constituent statistics in the former case. Test 4 does better in terms of size-corrected power than the other tests on the whole, but it is difficult to see what practical conclusions to draw from this experience in view of the very poor size properties we reported. We note in conclusion that even within the class of semiparametric estimates of C and d, the estimates $\hat{d}_1$, $\hat{C}_1$, and $\hat{C}_3$ which we have employed are unlikely to be the best. We stressed $\hat{d}_1$ partly due to its simplicity and its known consistency under mild conditions, but the modified version of the Geweke and Porter-Hudak (1983) estimates proposed by Robinson (1994c) are likely to be more efficient, while Robinson (1994d) has proposed an estimate which dominates the latter ones. It seems reasonable to suppose that a feasible version of (20) using such improved estimates of d and C would perform better than tests 2 and 3.

5. Final comments

While the paper has allowed for a degree of generality in its specification of the error structure, by employing a semiparametric rather than a parametric model for autocorrelations, it has focussed only on linear regression models (4), and simple and specific ones at that. Simple models are nevertheless of interest to practitioners, and though there is certainly scope for a more general treatment, it seems more complicated to treat models involving nonlinearity and simultaneity when there is a long-memory ingredient than in the weakly dependent situations stressed in the econometric literature. Another interesting direction for extending the research would be to allow for more than one, or an unknown, change point.

References

Brillinger, D.R., 1979, Confidence intervals for the crossvariance function, Selecta Statistica Canadiana 5, 3–16.

Dahlhaus, R., 1992, Efficient location and regression estimation for long range dependent regression models, Preprint.

Davies, R.B. and D.S. Harte, 1987, Tests for Hurst effect, Biometrika 74, 95–101.

Diebold, F.X. and G. Rudebusch, 1989, Long memory and persistence in aggregate output, Journal of Monetary Economics 24, 189–209.

Diebold, F.X. and G.R. Rudebusch, 1991, Is consumption too smooth? Long memory and the Deaton paradox, Review of Economics and Statistics 73, 1–9.

Diebold, F.X., S. Husted, and M. Rush, 1991, Real exchange rates under the gold standard, Journal of Political Economy 99, 1252–1271.

Fox, R. and M.S. Taqqu, 1986, Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, Annals of Statistics 14, 517–532.

Geweke, J. and S. Porter-Hudak, 1983, The estimation and application of long memory time series models, Journal of Time Series Analysis 4, 221–238.

Giraitis, L. and D. Surgailis, 1990, A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate, Probability Theory and Related Fields 86, 87–104.

Grenander, U., 1954, On the estimation of regression coefficients in the case of autocorrelated disturbance, Annals of Mathematical Statistics 25, 252–272.

Hackl, P., 1989, Statistical analysis and forecasting of economic structure change (Springer-Verlag, Berlin).

Hackl, P. and A. Westlund, 1991, Economic structural change, analysis and forecasting (Springer-Verlag, Berlin).

Ibragimov, I.A. and B. Linnik, 1971, Independent and stationary sequences of random variables (Wolters-Noordhoff, Groningen).

Künsch, H.R., 1986, Discrimination between monotonic trends and long-range dependence, Journal of Applied Probability 23, 1025–1030.

Robinson, P.M., 1994a, Time series with strong dependence, in: C.A. Sims, ed., Advances in econometrics: The sixth world congress, Vol. 1 (Cambridge University Press, Cambridge) 47–95.

Robinson, P.M., 1994b, Semiparametric analysis of long-memory time series, Annals of Statistics 22, 515–539.

Robinson, P.M., 1994c, Log-periodogram regression for time series with long range dependence, Annals of Statistics, forthcoming.

Robinson, P.M., 1994d, Gaussian semiparametric estimation of long range dependence, Annals of Statistics, forthcoming.

Robinson, P.M. and J.F. Hidalgo, 1994, Time series regression with long-range dependence, Preprint.

Sowell, F., 1990, Modeling long run behaviour with the fractional ARIMA model, Preprint.

Taqqu, M.S., 1975, Weak convergence to fractional Brownian motion and to the Rosenblatt process, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 31, 287–302.

Yajima, Y., 1988, On estimation of a regression model with long-memory stationary errors, Annals of Statistics 16, 791–807.

Yajima, Y., 1991, Asymptotic properties of the LSE in a regression model with long-memory stationary errors, Annals of Statistics 19, 158–177.