Biometrika (2004), **91**, 3, *pp.* 645–659

Printed in Great Britain

On multiple regression models with nonstationary correlated errors

By SUHASINI SUBBA RAO

Institut für Angewandte Mathematik, Universität Heidelberg, 69120 Heidelberg, Germany

suhasini@statlab.uni-heidelberg.de

Summary

We consider the estimation of parameters of a multiple regression model with nonstationary errors. We assume the nonstationary errors satisfy a time-dependent autoregressive process and describe a method for estimating the parameters of the regressors and the time-dependent autoregressive parameters. The parameters are rescaled as in nonparametric regression to obtain the asymptotic sampling properties of the estimators. The method is illustrated with an example taken from global temperature anomalies.

Some key words: Asymptotic normality; Consistency; Heteroscedastic errors; Local least squares; Local stationarity; Multiple regression; Nonstationary; Temperature anomaly; Time series.

1. Introduction

In many fields of research a time series $\{X_t\}$ is observed together with certain regressors which are believed to have a linear effect on the time series. The time series is then fitted with the multiple regression model

$$X_t = \alpha_1 f_{t,1} + \ldots + \alpha_q f_{t,q} + \varepsilon_t \quad (t = 1, \ldots, N), \tag{1}$$

where the regressors $\{f_{t,i} : t = 1, \ldots, N, i = 1, \ldots, q\}$ are observed. Often it is assumed that the errors $\{\varepsilon_t\}$ are independent and identically distributed, and ordinary least squares is used to estimate the parameters $\{\alpha_i\}$. However, it is quite plausible that the errors are dependent, in which case treating the errors as if they were independent and proceeding with the estimation could result in a severe loss of efficiency in the estimator. In this direction, several authors, see for example Durbin (1960) and Pierce (1971), have considered the case of dependent stationary errors, where the errors are assumed to satisfy a linear model.

A cause for concern is if the time series $\{X_t\}$ were observed over a long period of time, in which case it would seem quite likely that exogeneous factors could affect the stationarity of the errors. For example, it is believed that gas emissions over the past century may have caused ‘global warming’. Such an effect would almost certainly lead to a change in the structure of the temperature time series. In this case it would seem quite reasonable to change our working hypothesis and include the class of nonstationary models. Nonstationary time-varying autoregressive models have previously been studied by Subba Rao (1970), Ozaki & Tong (1975), Akaike & Kitagawa (1978), Priestley (1988) and Dahlhaus & Giraitis (1998); for a review on applications see Akaike & Kitagawa (1998). In this paper we model the errors by the regression model (1). We assume that the errors $\{\varepsilon_t\}$ satisfy the time-varying autoregressive model

$$\varepsilon_t = \beta_{t,1}\varepsilon_{t-1} + \ldots + \beta_{t,p}\varepsilon_{t-p} + \sigma_t\eta_t. \tag{2}$$

We describe a procedure for estimating both sets of parameters

$$\{\beta_{t,i} : t = 1, \ldots, N, i = 1, \ldots, p\}, \quad \{\alpha_i : i = 1, \ldots, q\}.$$

The case of ARMA errors will be considered in a future publication.

By using the rescaling device introduced by Dahlhaus (1997), we obtain the asymptotic sampling properties of the estimators given here.

The methods developed here are illustrated with a real example. We consider monthly global temperature anomalies from the northern and southern hemispheres observed during the period January 1865–November 2002. We will show that there is a significant upward linear trend in the global temperature and very interesting changes in the structure of the errors, most significantly between the years 1940 and 1960.

2. The estimation procedure

2·1. *The model*

We observe the time series $\{X_t\}$ and the regressors $\{f_{t,i} : t = 1, \ldots, N, i = 1, \ldots, q\}$ which satisfy the models (1) and (2). We assume the orders of the models p and q are known. To facilitate the study of the properties of the estimators we use the rescaling device introduced by Dahlhaus (1997), in which the parameters are rescaled as in nonparametric regression. We assume that $X_t = X_{t,N}, f_{t,i} = f_i(t/N)$, where

$$X_{t,N} = \alpha_1 f_1\left(\frac{t}{N}\right) + \ldots + \alpha_q f_q\left(\frac{t}{N}\right) + \varepsilon_{t,N} \quad (t = 1, \ldots, N) \tag{3}$$

and the errors $\{\varepsilon_{t,N}\}$ satisfy an autoregressive model of order p with time-dependent parameters:

$$\varepsilon_{t,N} = \beta_1\left(\frac{t}{N}\right)\varepsilon_{t-1,N} + \ldots + \beta_p\left(\frac{t}{N}\right)\varepsilon_{t-p,N} + \sigma\left(\frac{t}{N}\right)\eta_t, \tag{4}$$

where $\{\eta_t\}$ are independent, identically distributed random variables with $E(\eta_0) = 0$ and $E(\eta_0^2) = 1$. Let $\beta_u^{\mathrm{T}} = (\beta_1(u), \ldots, \beta_p(u))$ and $\alpha^{\mathrm{T}} = (\alpha_1, \ldots, \alpha_q)$. We call the autoregressive process of the type described in (4) a time-varying AR(p) process.

We obtain an asymptotically efficient estimator of $\beta_{t/N}$ and α through a two-stage procedure in which we estimate the time-varying autoregressive parameters $\beta_{t/N}$ and use these to estimate α. At both stages we minimise a local least squares criterion, using a kernel function. The rescaling technique is used to develop an asymptotic analysis of the parameters and it does not influence the actual estimation procedure.

From the models (3) and (4) we have

$$\begin{aligned}\sigma\left(\frac{t}{N}\right)\eta_t &= X_{t,N} - \sum_{i=1}^{q}\alpha_i f_i\left(\frac{t}{N}\right) - \sum_{j=1}^{p}\beta_j\left(\frac{t}{N}\right)\varepsilon_{t-j,N} \\ &= X_{t,N} - \sum_{i=1}^{q}\alpha_i f_i\left(\frac{t}{N}\right) - \sum_{j=1}^{p}\beta_j\left(\frac{t}{N}\right)\left\{X_{t-j,N} - \sum_{i=1}^{q}\alpha_i f_i\left(\frac{t-j}{N}\right)\right\} \\ &= -\sum_{j=0}^{p}\beta_j\left(\frac{t}{N}\right)X_{t-j,N} + \sum_{i=1}^{q}\sum_{j=0}^{p}\alpha_i\beta_j\left(\frac{t}{N}\right)f_i\left(\frac{t-j}{N}\right),\end{aligned} \tag{5}$$

where, for all $u \in [0, 1]$, $\beta_0(u) = -1$. In § 2·2 we use (5) to define a least squares criterion for estimating the time-varying autoregressive parameters and the regression parameters.

Let

$$g_i(b, k, N) = \sum_{j=0}^{p} b_j f_i\left(\frac{k-j}{N}\right), \quad \tilde{g}_i(b, u) = \sum_{j=0}^{p} b_j f_i(u),$$

$$g(b, k, N)^{\mathrm{T}} = (g_1(b, k, N), \ldots, g_q(b, k, N)), \quad \tilde{g}(b, u)^{\mathrm{T}} = (g_1(b, u), \ldots, g_q(b, u)),$$

$$f_{t,N}^{\mathrm{T}} = (f_1(t/N), \ldots, f_q(t/N)), \quad f_u^{\mathrm{T}} = (f_1(u), \ldots, f_q(u)),$$

$$\tilde{V}_1 = \int_0^1 \tilde{g}(\beta_u, u)\tilde{g}(\beta_u, u)^{\mathrm{T}} \sigma(u)^2 du, \quad \tilde{V}_2 = \int_0^1 \tilde{g}(\beta_u, u)\tilde{g}(\beta_u, u)^{\mathrm{T}} du. \tag{6}$$

Let

$$\Omega_1 = \{a = (a_1, \ldots, a_q) : \|a\|_2 \leqslant C_1\}, \quad \Omega_2 = \{b = (b_1, \ldots, b_p) : \|b\|_2 \leqslant C_2\}, \tag{7}$$

where C_1 and C_2 are finite constants and $\|.\|_2$ is the Euclidean norm.

2·2. *Estimation of the time-varying autoregressive parameters*

Though our aim is to estimate $\{\beta_{t_0/N} : t_0 = 1, \ldots, N\}$, we make our estimator more general. For each $u_0 \in (0, 1]$ if $|t_0/N - u_0| < 1/N$, or equivalently for each $u_0 \in (0, 1]$ if $\lceil u_0 N \rceil = t_0$, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x, we will use $\hat{\beta}_{t_0/N}$ as an estimator of β_{u_0}.

To construct the estimators we need to define a local least squares criterion. Let us consider (5). By relabelling $v_{t_0,N}(i, j) = \alpha_i \beta_j(t_0/N)$ we have

$$\sigma\left(\frac{t_0}{N}\right)\eta_{t_0} = \sum_{j=0}^{p} \beta_j\left(\frac{t_0}{N}\right) X_{t_0-j,N} - \sum_{i=1}^{q} \sum_{j=0}^{p} v_{t_0,N}(i, j) f_i\left(\frac{t_0-j}{N}\right).$$

Note that, if we treat the parameters $\{v_{t,N}(i, j)\}$ as a new set of parameters, the above model is linear in $\{\beta_j(t/N)\}$ and $\{v_{t,N}(i, j)\}$. Then least squares can be used to estimate the parameters $\{\beta_j(t/N)\}$ and $\{v_{t,N}(i, j)\}$. However, because of the nonstationarity of the errors we weight the least squares criterion using a kernel function, which leads to the following local least squares criterion:

$$\mathcal{L}_{t_0,N}(b, c) = \sum_{k=p}^{N} \frac{1}{bN} W\left(\frac{t_0/N - k/N}{b}\right)\left\{\sum_{j=0}^{p} b_j X_{k-j,N} - \sum_{i=1}^{q} \sum_{j=0}^{p} c_{ij} f_i\left(\frac{k-j}{N}\right)\right\}^2, \tag{8}$$

where $b_0 = -1$, $b^{\mathrm{T}} = (b_1, \ldots, b_p)$, $c^{\mathrm{T}} = (c_{10}, \ldots, c_{q0}, \ldots, c_{1p}, \ldots, c_{qp})$ and b is a bandwidth, which depends on N. The kernel function $W : [-\frac{1}{2}, \frac{1}{2}] \to \mathbb{R}$ is assumed to satisfy

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} W(x)dx = 1, \quad \int_{-\frac{1}{2}}^{\frac{1}{2}} W(x)xdx = 0.$$

Let $\|w\|_2^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} W(x)^2 dx$ and $w_2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} W(x)x^2 dx$. The kernel function naturally 'allows' us to derive the asymptotic properties: if we impose the restriction $|t_0/N - u_0| < 1/N$, and let $b \to 0$ and $bN \to \infty$ as $N \to \infty$, then the least squares criterion becomes increasingly localised about u_0. We will make this notion of localisation precise in later sections.

Let $v_{t,N}^{\mathrm{T}} = (v_{t,N}(1, 0), \ldots, v_{t,N}(q, 0), \ldots, v_{t,N}(1, p), \ldots, v_{t,N}(q, p))$. We define the parameter space

$$\Omega_3 = \{c = (c_{10}, \ldots, c_{q0}, \ldots, c_{1p}, \ldots, c_{qp}) : \|c\| \leqslant C_1(1 + C_2)\}$$

and use, as estimators of $\beta_{t_0/N}$ and $v_{t_0,N}$,

$$\hat{\beta}_{t_0,N}, \hat{c}_{t_0,N} = \arg \min_{b \in \Omega_2, c \in \Omega_3} \mathscr{L}_{t_0,N}(b, c). \tag{9}$$

We mention again that the minimisation is done by ignoring the fact that $v_{t_0,N}$ is a function of α_i and $\beta_j(t_0/N)$. As in Durbin's two-stage procedure we consider the minimisation with respect to b and c ignoring the dependence between $\beta_{t_0/N}$ and $v_{t_0,N}$ (Durbin, 1960). We will show in later sections that the estimators thus obtained are still asymptotically efficient.

2·3. *Estimation of the regression parameters*

We now describe the second stage of our estimation procedure. Given the time-varying autoregressive estimator $\hat{\beta}_{t,N}$, for $t = 1, \ldots, N$, we estimate α in (3). We partition the time series $\{X_{t,N}\}$ into blocks and estimate the parameters of the regressors α by choosing the $a \in C_1$ which minimises the average squared error over all the blocks.

Let $\tilde{b}$ denote the bandwidth, where $\tilde{b} \to 0$ and $\tilde{b}N \to \infty$ as $N \to \infty$. We note that $\tilde{b}$ need not be the same as the bandwidth b, defined in § 2·2; it must merely satisfy some rate constraints, which will be given in § 3·2. However, a rule of thumb is that the $\tilde{b}$ should depend on the smoothness properties of $\{\beta_j(.)\}_j$. We partition the same $\{X_{t,N}\}$ into blocks of length $\tilde{b}N$; we assume without loss of generality that $\tilde{b}N$ is an integer. It is clear there are $\lfloor 1/\tilde{b} \rfloor$ such blocks. Let $\lambda = \tilde{b}N$. Each block is disjoint and the rth block is of the form $(X_{(r-\frac{1}{2})\lambda,N}, \ldots, X_{(r+\frac{1}{2})\lambda-1,N})$, for $r = 1, \ldots, (\lfloor 1/\tilde{b} \rfloor - 1)$, where $\lfloor x \rfloor - 1$ is the largest integer less than or equal to x. Let $\hat{\beta}_{r\lambda,N} = (\hat{\beta}_{1,N}(r\tilde{b}), \ldots, \hat{\beta}_{p,N}(r\tilde{b}))^{\mathrm{T}}$. By substituting $\hat{\beta}_{r\tilde{b}N,N}$ and a into (5) we can estimate $\sigma(r\tilde{b})^2$ by

$$\frac{1}{\tilde{b}N} \sum_{k=(r-\frac{1}{2})\lambda}^{(r+\frac{1}{2})\lambda-1} \left\{ \sum_{j=0}^{p} \hat{\beta}_{j,N}(r\tilde{b}) X_{k-j,N} - \sum_{i=1}^{q} \sum_{j=0}^{p} a_i \hat{\beta}_{j,N}(r\tilde{b}) f_i\left(\frac{k-j}{N}\right) \right\}^2. \tag{10}$$

Now using (10) we define the global least squares criterion

$$T_N(a) = \tilde{b} \sum_{r=1}^{\lfloor 1/\tilde{b} \rfloor - 1} \frac{1}{\tilde{b}N} \sum_{k=(r-\frac{1}{2})\lambda}^{(r+\frac{1}{2})\lambda-1} \left\{ \sum_{j=0}^{p} \hat{\beta}_{j,N}(r\tilde{b}) X_{k-j,N} - \sum_{i=1}^{q} \sum_{j=0}^{p} a_i \hat{\beta}_{j,N}(r\tilde{b}) f_i\left(\frac{k-j}{N}\right) \right\}^2, \tag{11}$$

where $a^{\mathrm{T}} = (a_1, \ldots, a_q)$.

We estimate α by

$$\hat{\alpha}_N = \arg \min_{a \in \Omega_1} T_N(a). \tag{12}$$

It is clear we have ignored the observations both at the beginning and end of the sample, but this will not affect the asymptotic sampling properties of the estimator.

Remark 1. Note that the definition of $T_N(a)$ uses the rectangular window, $I(x) = 1$ if $x \in (0, 1]$ and $I(x) = 0$ otherwise. Therefore T_N is the same as

$$\tilde{b} \sum_{r=1}^{\lfloor 1/\tilde{b} \rfloor - 1} \sum_{k=p}^{N} \frac{1}{\tilde{b}N} I\left(\frac{k - rbN}{bN}\right) \left\{ \sum_{j=0}^{p} \hat{\beta}_{j,N}(r\tilde{b}) X_{k-j,N} - \sum_{i=1}^{q} \sum_{j=0}^{p} a_i \hat{\beta}_{j,N}(r\tilde{b}) f_i\left(\frac{k-j}{N}\right) \right\}^2.$$

Naturally other kernels can be used instead of the rectangular window. However, the inclusion of a general kernel makes the calculations in the Appendix more cumbersome with no additional benefit for the quality of the asymptotic properties of the estimators.

2·4. *Estimation of the time-varying variance*

To estimate $\sigma(u_0)^2$ we use the estimators of α and β_{u_0} obtained above. We estimate $\sigma(t_0/N)^2$ by

$$\hat{\sigma}^2_{t_0,N} = \sum_{k=p}^{N} \frac{1}{bN} W\left(\frac{u_0 - k/N}{b}\right) \left\{ \sum_{j=0}^{p} \hat{\beta}_{j,N}(u_0) X_{k-j,N} - \sum_{i=1}^{q} \sum_{j=0}^{p} \hat{\alpha}_i \hat{\beta}_{j,N}(u_0) f_i\left(\frac{k-j}{N}\right)\right\}^2, \tag{13}$$

where $|t_0/N - u_0| < 1/N$.

If $\sigma(u)^2$ is constant for all $u \in (0, 1]$, $\sigma(u)^2 = \sigma^2$, we estimate σ^2 by

$$\hat{\sigma}^2 = \tilde{b} \sum_{r=1}^{\lfloor 1/\tilde{b} \rfloor - 1} \frac{1}{\tilde{b}N} \sum_{k=(r-\frac{1}{2})\tilde{b}N}^{(r+\frac{1}{2})\tilde{b}N-1} \left\{ \sum_{j=0}^{p} \hat{\beta}_{j,N}(r\tilde{b}) X_{k-j,N} - \sum_{i=1}^{q} \sum_{j=0}^{p} \hat{\alpha}_i \hat{\beta}_{j,N}(r\tilde{b}) f_i\left(\frac{k-j}{N}\right)\right\}^2.$$

3. The sampling properties of the estimators

3·1. *Assumptions*

The proofs of consistency and asymptotic normality stated here are given in the Appendix. We also give an expression for the optimal bandwidth b according to a divergence criterion.

We make the following assumptions.

Assumption 1. The functions $\{\beta_j(.) : j = 1, \ldots, p\}$ and variance $\sigma(.)^2$ have continuous second derivatives and there exists a $\delta > 0$ such that, for each $u \in (0, 1]$, $1 - \sum_{j=1}^{p} \beta_j(u) z^j \neq 0$ when $|z| \leqslant 1 + \delta$.

Assumption 2. The regressors $\{f_i(u) : i = 1, \ldots, q\}$ are of bounded variation.

Assumption 3. Let Ω_1 and Ω_2 be as defined in (7). We assume $\alpha \in \text{int}(\Omega_1)$ and, for each $u \in [0, 1]$, $\beta_u \in \text{int}(\Omega_2)$, where $\text{int}(\Omega)$ denotes the interior of Ω.

Assumption 4. We require that $E(\eta_0^{4+\gamma}) < \infty$, for some $\gamma > 0$.

Assumption 5. For R_u as defined in (15), R_u is nonsingular for all $u \in (0, 1]$.

Assumption 6. Let $\mathscr{L}_{t_0,N}$, T_N, $\hat{\beta}_{t_0,N}$, $\hat{c}_{t_0,N}$ and $\hat{\alpha}_N$ be as defined in (8), (11), (9) and (12). Then for a large enough M we have for all $N \geqslant M$ that $\nabla \mathscr{L}_{t_0,N}(\hat{\beta}_{t_0,N}, \hat{c}_{t_0,N}) = 0$ and $\nabla T_N(\hat{\alpha}_N) = 0$.

Assumption 7. If $\tilde{V}_2$ is as defined in (6) then $\tilde{V}_2$ is nonsingular.

The above assumptions are fairly typical. Assumption 1 implies that the process $\{\varepsilon_{t,N}\}_t$ is in some sense ‘locally stationary’ and ensures that $\text{var}(\varepsilon_{t,N})$ is uniformly bounded in t and N. Assumption 2 implies that the regressors are bounded over the interval $(0, 1]$. Assumptions 3 and 4 are standard. We use Assumption 6 to prove asymptotic normality of estimators. This result can in fact be shown without using this assumption, but the proof is longer and not very instructive.

3·2. *Sampling distribution of the estimators*

All the results stated here use the idea of rescaled asymptotics. To illustrate the idea, consider the local average of a time-varying autoregressive process $\{Y_{t,N}\}$, about the time point t_0. Since the time-varying coefficients are only ‘slowly changing’, in some neighbourhood of the time point t_0, the observations $\{Y_{t_0+k,N}: k=-M,\ldots,M, \text{ with } M \ll N\}$ behave as if they came from a stationary process. Therefore, a local average or weighted local average about the time point t_0 could be considered as an estimator of $E(Y_{t_0,N})$. However, the limit as $M,N \to \infty$, with $M \ll N$, still makes no sense, since $E(Y_{t_0,N})$ may vary with t_0 and N. Thus we need a point of reference to define our limit. To this end, we impose the constraint $|t_0/N - u_0| < 1/N$, so that $t_0/N \doteq u_0$, and consider $\lim_{N\to\infty} E(Y_{t_0,N})$. A local average about the time point t_0 should converge to this limit. Roughly speaking, rescaled asymptotics involve the limit of a function $F_{t_0,N}$, either random or deterministic, as $N \to \infty$ under the restriction that $|t_0/N - u_0| < 1/N$.

We now consider the asymptotic properties of the estimator $\hat{\beta}_{t_0,N}$.

THEOREM 1. *Suppose that $\hat{\beta}_{t_0,N}$ is defined as in* (9). *Then under Assumptions* 1–4 *and if* $|u_0 - t_0/N| < 1/N$ *we have that*

$$\hat{\beta}_{t_0,N} \to \beta_{u_0},$$

in probability, where $b \to 0$ and $bN \to \infty$ as $N \to \infty$.

We now state a result concerning the asymptotic normality of $\hat{\beta}_{t_0,N}$. If the errors were stationary the asymptotic variance of $\hat{\beta}_{t_0,N}$ would involve the covariances between the errors. Here the errors are nonstationary so the covariances are dependent on location. This motivates us to use the rescaled limit discussed above, and in particular the rescaled covariance given by Dahlhaus & Giraitis (1998). For each $j \in \mathbb{Z}$ the rescaled covariance at location $u \in (0, 1]$ is

$$c(u,j) := \int_{-\pi}^{\pi} \frac{1}{|1 - \sum_{k=1}^{p} \beta_k(u) \exp(-ik\omega)|^2} \exp(ij\omega) d\omega. \tag{14}$$

The rescaled covariance $c(t/N, j)$ approximates the true covariance $\operatorname{cov}(X_{t,N}, X_{t-j,N})$, where the degree of approximation depends on N and the rescaled covariance is the true covariance of a stationary AR(p) process with the autoregressive parameters $\{\beta_j(u)\}_{j=1}^{p}$. We use the rescaled covariances to define the rescaled variance/covariance matrix. Let R_u denote a $(p \times p)$-dimensional matrix where the (i,j)th element of R_u is

$$(R_u)_{ij} = c(u, i-j). \tag{15}$$

THEOREM 2. *Suppose that $\hat{\beta}_{t_0,N}$ and R_u are defined as in* (9) *and* (15). *Then, under Assumptions* 1–6 *and for* $|u_0 - t_0/N| < 1/N$, *we have*

$$\surd(bN)(\hat{\beta}_{t_0,N} - \beta_{u_0}) - \surd(bN)\mu_b(u_0) \to \mathcal{N}\{0, \sigma(u_0)^2 \|w\|_2^2 R_{u_0}^{-1}\} \tag{16}$$

in distribution as $b \to \infty$ and $bN \to \infty$, $N \to \infty$, where

$$\mu_b(u) = b^2 w_2 R_u^{-1}(2R_u'\beta_u' + R_u\beta_u''), \tag{17}$$

$w_2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} W(x)^2 dx$. *Moreover R_u', R_u'' and β_u', β_u'' denote the first and second derivatives of R_u and β respectively; the derivatives are taken pointwise with respect to u over each element of the matrix and vector.*

The asymptotic bias $\mu_b(u)$ and variance $\sigma(u_0)^2\|w\|_2^2 R_{u_0}^{-1}$ are used to obtain the optimal bandwidth in § 3·3.

Now suppose that $x=(x_1,\ldots,x_p)$ is a p-dimensional vector, and let $\|x\|_\infty = \max_{1\leqslant i\leqslant p}(|x_i|)$.

COROLLARY 1. *Suppose that Assumptions* 1–7 *hold and* $\hat\beta_{t_0,N}$ *is as defined in* (9). *Then if* $|u_0 - t_0/N| < 1/N$ *we have the following*:

(i) *if* $b^4 = O\{(bN)^{-1}\}$, *then*

$$\|\hat\beta_{t_0,N} - \beta_{u_0}\|_\infty = O_p\left(\frac{1}{\sqrt{(bN)}} + b^2\right);$$

(ii) *if* $b^4 \ll (bN)^{-1}$, *then*

$$\sqrt{(bN)}(\hat\beta_{t_0,N} - \beta_{u_0}) \to \mathcal{N}\{0, \sigma(u_0)^2\|w\|_2^2 R_{u_0}^{-1}\},$$

in distribution, as $N\to\infty$.

We mention that the asymptotic variance of $\sqrt{(bN)}(\hat\beta_{t_0,N} - \beta_{u_0})$, which is $\sigma(u_0)^2\|w\|_2^2 R_{u_0}^{-1}$, is asymptotically equivalent to the variance of the estimator $\tilde\beta_{t_0,N}$, given by

$$\tilde\beta_{t_0,N} = \arg\min_{b\in\Omega_2}\sum_k \frac{1}{bN} W\left(\frac{t_0/N - k/N}{b}\right)(\varepsilon_{k,N} - b_1\varepsilon_{k-1,N} - \ldots - b_p\varepsilon_{k-p,N})^2,$$

where $\{\varepsilon_t\}$ satisfy the time-varying autoregressive process defined in (4); see Dahlhaus & Giraitis (1998).

We now consider the distributional properties of the estimator $\hat\alpha_N$.

THEOREM 3. *Let* $\hat\alpha_N$ *be as defined in* (12). *Then under Assumptions* 1–4 *we have*

$$\hat\alpha_N \to \alpha,$$

in probability, where $\tilde b\to 0$, $b\to 0$ *and* $\tilde b^2 bN\to\infty$ *as* $N\to\infty$.

THEOREM 4. *Let* $\hat\alpha_N$, $\tilde V_1$ *and* $\tilde V_2$ *be as defined in* (12) *and* (6) *respectively. Then under Assumptions* 1–7 *we have*

$$\sqrt{N}(\hat\alpha_N - \alpha) \to \mathcal{N}(0, 4\tilde V_2^{-1}\tilde V_1 \tilde V_2^{-1}), \tag{18}$$

in distribution, where $\tilde b\to 0$, $b\to 0$ *and* $\tilde b^3 bN\to\infty$ *as* $N\to\infty$.

Note that the asymptotic properties of $\hat\alpha_N$ are the same as the asymptotic properties of the estimators of the regression parameters under the assumption that $\{\beta_u\}$ are known.

Since $\{g(\hat\beta_{k,N}, k, N)\}$ are observed we can estimate both $\tilde V_1$ and $\tilde V_2$. By using the estimators $\hat\sigma_{t,N}^2$ of the time-varying variance, defined in (13), we can use

$$\hat V_1 = \frac{1}{N}\sum_{k=p}^{N} g(\hat\beta_{k,N}, k, N) g(\hat\beta_{k,N}, k, N)^{\mathrm{T}}\hat\sigma_{k,N}^2, \quad \hat V_2 = \frac{1}{N}\sum_{k=p}^{N} g(\hat\beta_{k,N}, k, N) g(\hat\beta_{k,N}, k, N)^{\mathrm{T}}$$

as estimators of $\hat V_1$ and $\hat V_2$ respectively. These estimators can be used to construct confidence intervals for α.

Remark 2. Note that, if $\sigma(u)^2 = \sigma^2$ for all $u \in (0, 1]$, then

$$\sqrt{N}(\hat{\alpha}_N - \alpha) \to \mathcal{N}(0, 4\sigma^2 \hat{V}_2^{-1})$$

in distribution.

3·3. *Optimal choice of bandwidth*

It is desirable to use the bandwidth which gives the 'best' estimator of β_u. By optimal bandwidth we mean the bandwidth which minimises a deviation criterion, which is usually the mean squared error.

From the results above we know that the bandwidth b appears both in the asymptotic bias and the variance of $\hat{\beta}_{t_0,N}$. Therefore, a bandwidth will need to be selected which is a compromise between the two terms and minimises a mean square criterion. It is worth noting we are using the term bias some what loosely: we are referring to $R_{u_0}^{-1}\mu_b(u_0)$, defined in (17), which strictly speaking is not the true bias $E(\hat{\beta}_{t_0,N} - \beta_{u_0})$.

In contrast, the bandwidth $\tilde{b}$ used in the second stage to estimate the parameters of the regressors behaves differently. Given that bandwidth $\tilde{b}$ satisfies sufficient rate constraints, then asymptotically $\hat{\alpha}_N$ is unbiased. Therefore in this section we will concentrate on evaluating the optimal bandwidth b, used to estimate the parameters of the time-varying autoregressive model.

For any $p + q(p+1)$-dimensional vector $v = (v_1, \ldots, v_{p+q(p+1)})$ let $v_\varepsilon^{\mathrm{T}} = (v_1, \ldots, v_p)$. By using (A2) for a large enough N we obtain

$$E\{(\hat{\beta}_{t_0,N} - \beta_{u_0})^{\mathrm{T}}(\hat{\beta}_{t_0,N} - \beta_{u_0})\} = E(J_{t_0,N}^{\mathrm{T}} J_{t_0,N}),$$

where

$$J_{t_0,N} = \{(\nabla^2 \mathscr{L}_{t_0,N}^*)^{-1} \nabla \mathscr{L}_{t_0,N}^*(\beta_{u_0}, v)\}_\varepsilon;$$

$\mathscr{L}_{t,N}^*$ is defined in (A1). However, it is clear that it is difficult to evaluate $E(J_{t,N}^{\mathrm{T}} J_{t,N})$ because of the random matrix $(\nabla^2 \mathscr{L}_{t,N}^*)^{-1}$. In such situations it is usual to replace the denominator with its limit, which is usually deterministic, and we do this here. It can be shown that $(\nabla^2 \mathscr{L}_{t_0,N}^*) \to V_{u_0}$ in probability, where $V_u = \mathrm{diag}(R_u, f_u f_u^{\mathrm{T}})$, and we replace $(\nabla^2 \mathscr{L}_{t_0,N}^*)^{-1}$ with the inverse of V_{u_0}. However, since V_{u_0} is singular, we use the Moore–Penrose generalised inverse $V_{u_0}^-$. Therefore we have that

$$\begin{aligned}(\hat{\beta}_{t_0,N} - \beta_{u_0}) &\simeq \{V_{u_0}^- \nabla \mathscr{L}_{t_0,N}^*(\beta_{u_0}, v_{t_0,N})\}_\varepsilon \\ &= \{\mathrm{diag}(R_{u_0}^{-1}, (f_{u_0} f_{u_0}^{\mathrm{T}})^-) \nabla \mathscr{L}_{t_0,N}^*(\beta_{u_0}, v_{t_0,N})\}_\varepsilon = R_{u_0}^{-1}\{\nabla \mathscr{L}_{t_0,N}^*(\beta_{u_0}, v_{t_0,N})\}_\varepsilon.\end{aligned}$$

Let $\tilde{J}_{t_0,N} = R_{u_0}^{-1}\{\nabla \mathscr{L}_{t_0,N}^*(\beta_{u_0}, v_{t_0,N})\}_\varepsilon$. Then by using (A4) we have

$$E(\tilde{J}_{t_0,N}^{\mathrm{T}} \tilde{J}_{t_0,N}) = \|w\|_2^2 \sigma(u_0)^2 \,\mathrm{tr}\,(R_{u_0}^{-1}) + \mathrm{tr}\,\{\mu_b(u_0)\mu_b(u_0)^{\mathrm{T}}\} + o\left(\frac{1}{bN} + b^4\right).$$

We use $E(\tilde{J}_{t_0,N}^{\mathrm{T}} \tilde{J}_{t_0,N})$ as the mean square criterion, instead of $E(J_{t_0,N}^{\mathrm{T}} J_{t_0,N})$, and use, as the optimal bandwidth,

$$b_{\mathrm{opt}} = \arg\min_b \int_0^1 [\|w\|_2^2 \sigma(u)^2 \,\mathrm{tr}\,(R_u^{-1}) + \mathrm{tr}\,\{\mu_b(u)\mu_b(u)^{\mathrm{T}}\}] du.$$

Therefore, the optimal bandwidth depends on the smoothness properties of $\{\beta_j(.)\}_j$. In practice R_u will have to be estimated from the residuals and the derivatives $\{\beta_j'(.)\}_j$ by taking first and second differences.

3·4. *Sampling properties of other estimation methods*

We describe two alternative estimation methods and state their sampling properties. The proofs in the later sections can be modified accordingly to prove the results stated here. One approach is to treat the errors as if they were independent and identically distributed, and to estimate the parameters of the regression model by least squares. These estimates are then used to estimate the residuals, which are then used to fit the autoregressive model. This method can be easily generalised to include nonstationary errors. In this case we use $\tilde{\alpha}_N$ as an estimator of the regression parameters, where

$$\tilde{\alpha}_N = \arg\min_{a \in \Omega_1} \frac{1}{N} \sum_{k=1}^{N} \left\{ X_{k,N} - \sum_{j=1}^{q} a_j f_j\left(\frac{k}{N}\right) \right\}^2 .$$

We let

$$\hat{\varepsilon}_{k,N} = X_{k,N} - \sum_{j=1}^{q} \tilde{\alpha}_j f_j\left(\frac{k}{N}\right),$$

where $\tilde{\alpha}_N = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_q)$. As an estimator of $\beta_{t_0/N}$ we use

$$\tilde{\beta}_{t_0,N} = \arg\min_{b \in \Omega_2} \sum_{k} \frac{1}{bN} W\left(\frac{t_0/N - k/N}{b}\right)\left(\hat{\varepsilon}_{k,N} - \sum_{j=1}^{p} b_j \hat{\varepsilon}_{k-j,N}\right)^2 .$$

It can be shown that

$$\sqrt{N}(\tilde{\alpha} - \alpha) \to \mathcal{N}(0, 4V_2^{-1} V_1 V_2^{-1}),$$

in distribution, where

$$V_1 = \int_0^1 f_u f_u^{\mathrm{T}} du, \quad V_2 = \int_0^1 f_u f_u^{\mathrm{T}} c(u, 0) du$$

and $\tilde{\beta}_{t_0,N}$ has the same asymptotic distribution as $\hat{\beta}_{t_0,N}$, given in (9).

A drawback of the above procedure is that V_2 contains $c(u, 0)$, which roughly speaking is the variance of the errors at rescaled time u. To remove this we can use a three-stage scheme and estimate α once again using $\{\tilde{\beta}_{t_0,N} : t_0 = 1, \ldots, N\}$. To do this we estimate the innovations $\sigma(k/N)\eta_k$ and use $\tilde{\alpha}_{2,N}$ as the actual estimator of α, where

$$\tilde{\alpha}_{2,N} = \arg\min_{a \in \Omega} \tilde{b} \sum_{r=1}^{\lfloor 1/\tilde{b} \rfloor - 1} \frac{1}{\tilde{b}N} \sum_{k=(r-\frac{1}{2})\tilde{b}N}^{(r+\frac{1}{2})\tilde{b}N - 1} \left\{ \sum_{j=0}^{p} \tilde{\beta}_{j,N}(r\tilde{b}) X_{k-j,N} - \sum_{i=1}^{q} \sum_{j=0}^{p} a_i \tilde{\beta}_{j,N}(r\tilde{b}) f_i\left(\frac{k-j}{N}\right) \right\}^2 .$$

Note the similarity between the definitions of $\tilde{\alpha}_{2,N}$ and $\hat{\alpha}_N$, given in (12). Asymptotically $\tilde{\alpha}_{2,N}$ has the same distribution as $\hat{\alpha}_N$, given in (18).

4. Data analysis: Global temperature anomalies

There has been intense speculation and research on global warming, and, if it exists, to find variables which cause this phenomenon. There is clear evidence (Subba Rao & Tsolaki, 2004) that the global temperatures are nonstationary and here our objective is to analyse the monthly temperature anomalies observed during the period January 1856 to November 2002. Our approach confirms the general belief that global warming exists and identifies various periods over which the structural changes may have occurred.

The data were obtained from the Climate Research Unit of the University of East Anglia, England (`http://www.cru.uea.ac.uk`). By anomalies we mean the difference of the temperatures from some reference value. Figures 1(a) and (b) display the temperature anomalies dataset from the two hemispheres. There is a clear seasonal component, which we model using a sine function with period 12 months. Moreover, the plots also suggest

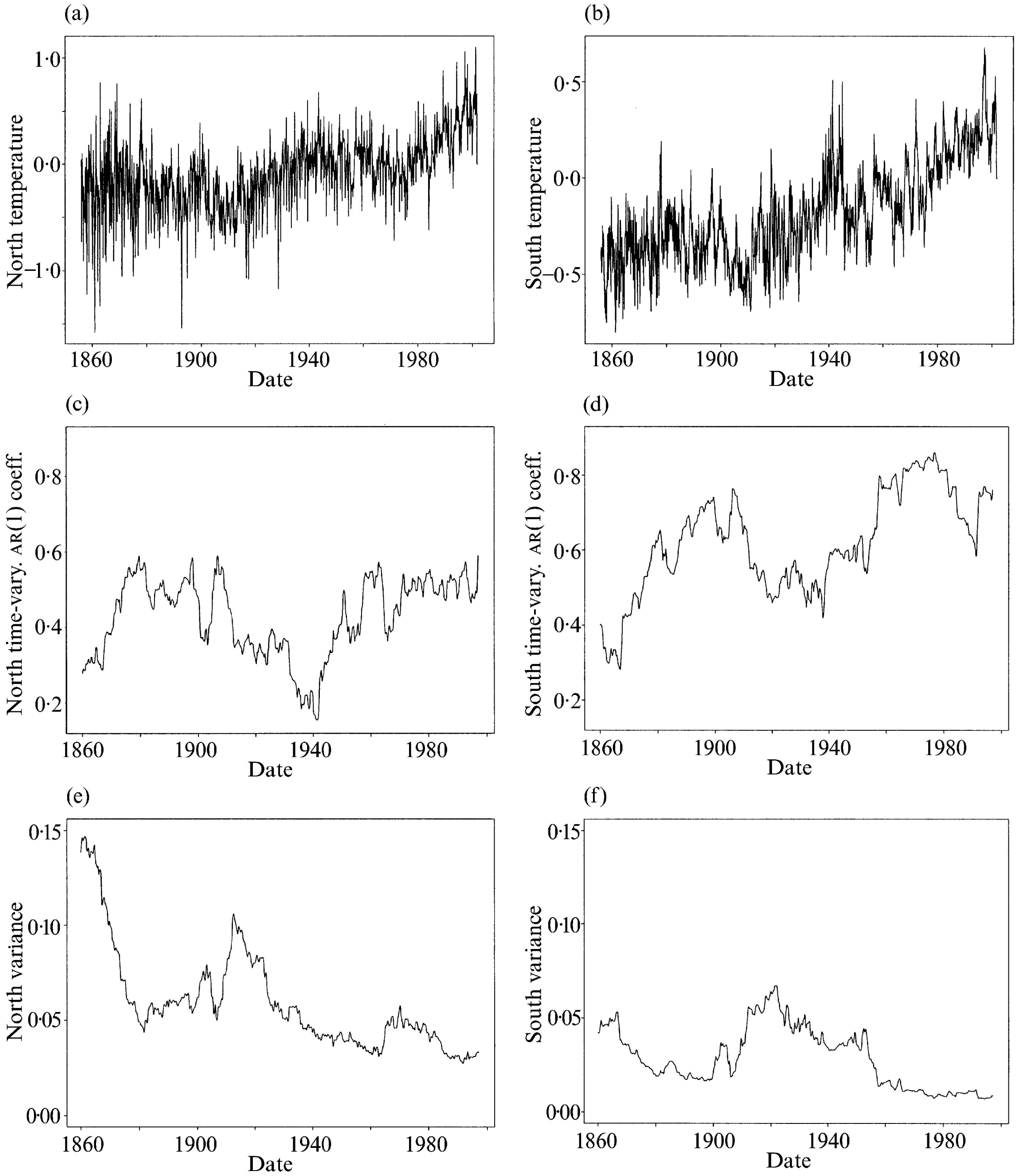

Fig. 1: Global temperature data. (a) northern hemisphere anomalies, (b) southern hemisphere anomalies, (c) northern hemisphere time-varying autoregressive parameters, (d) southern hemisphere time-varying autoregressive parameters, (e) northern hemisphere time-varying variance of the innovation, (f) southern hemisphere time-varying variance of the innovation.

that there exists a slight linear upwards trend which suggests a global warming effect. From the plot we also see a larger variance at the beginning and some regions of irregular fluctuations, which suggest that the data could be nonstationary. We use the two-stage method described above to fit models (3) and (4). More precisely we fit a model of the form

$$X_t = \alpha_0 + \alpha_1 t + \alpha_2 \sin(12t \times 2\pi) + \varepsilon_t, \quad (19)$$

where

$$\varepsilon_t = \beta_t \varepsilon_{t-1} + \sigma_t \eta_t \quad (t = 1, \ldots, 1764),$$

with X_t being the temperature anomalies. In the rescaled framework this is equivalent to

$$X_{t,1764} = \alpha_0 + \tilde{\alpha}_1 \left(\frac{t}{1764}\right) + \alpha_2 \sin\left(12 \times 1764 \times \frac{t}{1764} \times 2\pi\right) + \varepsilon_{t,1764},$$

where

$$\varepsilon_{t,1764} = \tilde{\beta}\left(\frac{t}{1764}\right)\varepsilon_{t-1,1764} + \sigma\left(\frac{t}{1764}\right)\eta_t \quad (t = 1, \ldots, 1764),$$

with $X_t = X_{t,1764}$, $\varepsilon_t = \varepsilon_{t,1764}$ $\tilde{\alpha}_1 = 1764\alpha_1$ and $\beta_t = \tilde{\beta}(t/1764)$; the regressors are $f_1(u) = u$ and $f_2(u) = \sin(12 \times 1764 \times u \times 2\pi)$ for $u \in (0, 1]$. Since the rescaling device is only used to derive the asymptotic sampling properties and does not affect the estimation method, we use model (19).

We use a rectangular kernel to estimate the time-varying AR(1) parameters with the bandwidth $b = 0{\cdot}1$. In view of endpoint considerations our estimates for the time-varying autoregressive parameters, for both hemispheres, only begin at time point $t = 88$ and end at time point $t = 1675$, which corresponds approximately to the years 1860–1998. The plots of the time-varying autoregressive parameters and variance have been smoothed using the S-Plus command `smooth`.

We fit the model in (19) for the monthly northern temperature anomalies. We obtain $\hat{\alpha}_0 = -0{\cdot}29$, $\hat{\alpha}_1 = 0{\cdot}00032$ and $\hat{\alpha}_2 = 0{\cdot}028$; the time-varying autoregressive estimates $\{\hat{\beta}_t\}_t$ are plotted in Fig. 1(c) and the time-varying variance estimates are plotted in Fig. 1(e). The sample variance-covariance matrix for the estimators of $(\alpha_0, \alpha_1, \alpha_2)$ is

$$\widehat{\text{var}}\begin{pmatrix}\hat{\alpha}_0\\ \hat{\alpha}_1\\ \hat{\alpha}_2\end{pmatrix} = \frac{1}{1586}\begin{pmatrix} 5{\cdot}95 & 0{\cdot}0030 & 0{\cdot}0030\\ 0{\cdot}0030 & 1{\cdot}5 \times 10^{-6} & 1{\cdot}5 \times 10^{-6}\\ 0{\cdot}0030 & 1{\cdot}5 \times 10^{-6} & 1{\cdot}5 \times 10^{-6}\end{pmatrix}.$$

Using the normal distribution and the sample variance of $\hat{\alpha}_1$, $\frac{1}{1586} 1{\cdot}5 \times 10^{-6}$, we see that the coefficient $\hat{\alpha}_1$ is significantly different from zero, indicating that there is evidence of an increase in temperature. For the southern hemisphere data, we obtain $\hat{\alpha}_0 = -0{\cdot}19$, $\hat{\alpha}_1 = 0{\cdot}00036$ and $\hat{\alpha}_2 = 0{\cdot}012$. The corresponding time-varying autoregressive estimates $\{\hat{\beta}_t\}$ are plotted in Fig. 1(d) and the time-varying variance estimates are plotted in Fig. 1(f). The sample variance-covariance matrix for the estimators of $(\alpha_0, \alpha_1, \alpha_2)$ is given by

$$\widehat{\text{var}}\begin{pmatrix}\hat{\alpha}_0\\ \hat{\alpha}_1\\ \hat{\alpha}_2\end{pmatrix} = \frac{1}{1586}\begin{pmatrix} 14 & 0{\cdot}0082 & 0{\cdot}0082\\ 0{\cdot}0082 & 4{\cdot}8 \times 10^{-6} & 4{\cdot}8 \times 10^{-6}\\ 0{\cdot}0082 & 4{\cdot}8 \times 10^{-6} & 4{\cdot}8 \times 10^{-6}\end{pmatrix}.$$

Again using the normal distribution we see that the coefficient $\hat{\alpha}_1$ is significantly different from zero, so that it is clear that the temperature in both hemispheres is increasing by approximately 0·004 degrees a year. The estimates of the time-varying autoregressive parameters in the two hemispheres identify common structural changes: the time-varying AR(1) parameter, $\hat{\beta}_t$, in both plots declines between 1900 and 1940, rises again between 1940 and 1960 and appears to level off after 1960. The analysis appears to confirm the belief (Parker et al., 1994) that there was a significant change in global weather. To some extent Wu et al. (2001) also support this view. They assume that the temperature anomalies satisfy the model

$$X_k = \mu_k + Z_k,$$

where $\{Z_k\}$ is a stationary process and μ_k are the means which are monotonically increasing. They develop a method for testing for changepoints, which they apply to the temperature anomaly data. They show that there are significant changes around the 1920's, and that from 1920 to 1945 there are a number of successive increasing changepoints.

Overall the autoregressive coefficient for the southern hemisphere seems to be larger than for the northern hemisphere, and the variation of the autoregressive coefficient in the southern hemisphere is larger. On the other hand, Figs 1(e) and (f) show that the variance seems to be greater in the northern hemisphere than in the southern hemisphere. It is well known that climatic changes differ between the hemispheres, partly because the southern hemisphere is covered by oceans, which can cause greater volatility in the weather.

ACKNOWLEDGEMENT

The author would like to thank Professor Rainer Dahlhaus for many useful and interesting suggestions, two anonymous referees for making many helpful suggestions and the editor for considerably improving the presentation of the paper. This research was supported by the Deutsche Forschungsgemeinshaft.

APPENDIX

Analysis of the sampling properties of the estimators

For many results in this section we will use some convergence results which are proved in a technical report available from the author.

First we prove weak consistency of $\hat{\beta}_{t_0,N}$. If the process were stationary we would normally show consistency of the estimator via convergence of the least squares criterion to its expectation. However, as the errors are nonstationary, we will use the rescaled limit

$$\tilde{\mathscr{L}}_u(b, c) = \sum_{j_1=0, j_2=0}^{p} b_{j_1} b_{j_2} \left[c(u, j_1 - j_2) + \left\{ \sum_{i=1}^{q} \alpha_i f_i(u) \right\}^2 \right]$$
$$- 2 \left\{ \sum_{j=0}^{p} \sum_{i=1}^{q} c_{ij} f_i(u) \right\} \left(\sum_{j=0}^{p} b_j \right) \left\{ \sum_{i=1}^{q} \alpha_i f_i(u) \right\} + \left\{ \sum_{i=1}^{q} \sum_{j=0}^{p} c_{ij} f_i(u) \right\}^2,$$

where $b^{\mathrm{T}} = (b_1, \ldots, b_p)$ and $c^{\mathrm{T}} = (c_{10}, \ldots, c_{q0}, \ldots, c_{1p}, \ldots, c_{qp})$.

Proof of Theorem 1. By using a result from the author's technical report we have $\mathscr{L}_{t_0,N}(\hat{\beta}_{t_0,N}, \hat{c}_{t_0,N}) \to \tilde{\mathscr{L}}_{u_0}(\beta_{u_0}, v_{t_0,N})$ in probability. To show that $\hat{\beta}_{t_0,N} \to \beta_{u_0}$ in probability, we must show that the minimum of $\tilde{\mathscr{L}}_{u_0}$ is unique for β_{u_0}. We see that $\tilde{\mathscr{L}}_{u_0}(b, c)$ attains its minimum, which is $\sigma(u_0)^2$, when

$$\sum_{j_1=0}^{p} b_{j_1} c(u_0, j_1 - j) = 0, \quad c_{ij} = \alpha_i^* b_j \quad (i = 1, \ldots, q, j = 1, \ldots, p),$$

where $\{\alpha_i^*\}$ satisfies $\sum_{i=1}^q \alpha_i^* f_i(u_0) = \sum_{i=1}^q \alpha_i f_i(u_0)$. It is clear that $\sum_{j_1=0}^p b_{j_1} c(u_0, j_1 - j) = 0$ are the Yule–Walker equations, which are satisfied only when $b_{j_1} = \beta_j(u_0)$. Thus we have that $\hat{\beta}_{t_0,N} \to \beta_{u_0}$ in probability, as asserted. □

We now investigate the distributional properties of $\hat{\beta}_{t_0,N}$, which are determined by the local least squares criterion $\mathscr{L}_{t_0,N}(b, c)$. Analysis of the random function $\mathscr{L}_{t_0,N}(b, c)$ is complicated. Instead we define an alternative criterion which will yield the same asymptotic properties as $\mathscr{L}_{t_0,N}(b, c)$. Our use of the alternative criterion was motivated by a similar device used by Durbin (1960). Let

$$\mathscr{L}^*_{t_0,N}(b, c^*) = \sum_{k=p}^{N} \frac{1}{bN} W\left(\frac{u_0 - k/N}{b}\right)\left\{X_{k,N} + \sum_{j=1}^{p} b_j \varepsilon_{k-j,N} - \sum_{i=1}^{q}\sum_{j=0}^{p} c^*_{ij} f_i\left(\frac{k-j}{N}\right)\right\}^2. \tag{A1}$$

We treat $\mathscr{L}^*_{t_0,N}(b, c^*)$ as if the time-varying AR errors $\{\varepsilon_{k,N}\}$ were observed. If

$$\hat{\beta}^*_{t_0,N}, \hat{c}^*_{t_0,N} = \arg \min_{b \in \Omega_2, c \in \Omega_3^*} \mathscr{L}^*_{t_0,N}(b, c),$$

then it is straightforward to show that $\hat{\beta}_{t_0,N} \simeq \hat{\beta}^*_{t_0,N}$. The transformation from $\mathscr{L}_{t_0,N}$ to $\mathscr{L}^*_{t_0,N}$ ensures that $\hat{\beta}_{t_0,N}$ and $\hat{c}^*_{t_0,N}$ are asymptotically uncorrelated, which simplifies the subsequent calculations.

In many likelihood methods it is common to establish asymptotic normality of the estimator through the gradient of the likelihood function. However, in the nonstationary case, the gradient of the localised least squares criterion will contain a bias which causes the estimators to be biased also. Nevertheless, by identifying the bias and treating the bias separately we will be able to show the asymptotic normality of the gradient of the localised least squares criterion. Under Assumption 6 we have

$$((\hat{\beta}_{t_0,N} - \beta_{u_0})^{\mathrm{T}}, (\hat{c}^*_{t_0,N} - v_{t_0,N})^{\mathrm{T}})^{\mathrm{T}} = \nabla^2 \mathscr{L}^*_{t_0,N}(\beta_{u_0}, v_{t_0,N})^{-1} \nabla \mathscr{L}^*_{t_0,N}(\beta_{u_0}, v_{t_0,N}). \tag{A2}$$

We decompose $\nabla \mathscr{L}^*_{t_0,N}(\beta_{u_0}, v_{t_0,N})$ into

$$\nabla \mathscr{L}^*_{t_0,N}(\beta_{u_0}, v_{t_0,N}) = \mathscr{S}_{t_0,N} + \mathscr{B}_{t_0,N},$$

where

$$\mathscr{S}_{t_0,N} = -2 \sum_{k=p}^{N} \frac{1}{bN} W\left(\frac{t_0/N - k/N}{b}\right) \sigma\left(\frac{k}{N}\right) \eta_k Y_{u-1,N},$$

$$\mathscr{B}_{t_0,N} = -2 \sum_{k=p}^{N} \frac{1}{bN} W\left(\frac{t_0/N - k/N}{b}\right)\left[\sum_{j=1}^{p}\left\{\beta_j(u_0) - \beta_j\left(\frac{k}{N}\right)\right\} \varepsilon_{k-j,N}\right] Y_{k-1,N}, \tag{A3}$$

and $Y^{\mathrm{T}}_{k-1,N} = (\varepsilon_{k-1,N}, \ldots, \varepsilon_{k-p,N}, f^{\mathrm{T}}_{k,N}, \ldots, f^{\mathrm{T}}_{k-p,N})$. It can be shown that

$$E(\mathscr{B}_{t_0,N})^{\mathrm{T}} = (2\mu_b(u_0)^{\mathrm{T}}, 0^{\mathrm{T}}_{q(p+1)}) + o(b^2), \quad E(\mathscr{S}_{t_0,N}) = 0_{p+q(p+1)},$$

$$\operatorname{var}(\mathscr{B}_{t_0,N}) = O\left(\frac{1}{N}\right), \quad \operatorname{var}(\mathscr{S}_{t_0,N}) = \frac{\|w\|_2^2}{bN} \sigma(u_0)^2 V_{u_0}, \quad \operatorname{cov}(\mathscr{B}_{t_0,N}, \mathscr{S}_{t_0,N}) = O\left(\frac{1}{N}\right), \tag{A4}$$

where $V_u = \operatorname{diag}(R_u, f_u f_u^{\mathrm{T}})$ and R_u is the variance-covariance matrix defined in (15), $\mu_b(u_0)$ is defined in (17) and 0_r is an r-dimensional zero vector. For any $p + q(p+1)$-dimensional vector $v = (v_1, \ldots, v_{p+q(p+1)})$ let $v_\varepsilon^{\mathrm{T}} = (v_1, \ldots, v_p)$. By using $\nabla^2 \mathscr{L}^*_{t_0,N}(\beta_{u_0}, v_{t_0,N}) \to V_{u_0}$ in probability, (A4) and standard methods it can be shown that

$$\surd(bN)(\hat{\beta}_{t_0,N} - \beta_{u_0}) = \tfrac{1}{2}\surd(bN) R^{-1}_{u_0} (\mathscr{S}_{t_0,N})_\varepsilon + \surd(bN)\mu_b(u_0) + o_p(1). \tag{A5}$$

Proof of Theorem 2. We have from (A5) that $\sqrt{(bN)}(\hat{\beta}_{t_0,N}-\beta_{u_0})-\frac{1}{2}\sqrt{(bN)}\mu_b(u_0)$ and $R_{u_0}^{-1}(\mathscr{S}_{t_0,N})$ have the same asymptotic distribution. Since $\mathscr{S}_{t_0,N}$ is the sum of martingale differences, by using the conditional Lindeberg condition (Hall & Heyde, 1980, Theorem 3.2) it is straightforward to show that $\sqrt{(bN)}\mathscr{S}_{t_0,N}\to\mathscr{N}(0,\sigma(u_0)^2\|w\|_2^2V_{u_0})$ in distribution. Therefore we have established the required result. □

Proof of Corollary 1. By using (A5), we have that

$$\|\hat{\beta}_{t_0,N}-\beta_{u_0}\|_\infty=O_p\{R_{u_0}^{-1}(\mathscr{S}_{t_0,N})_\varepsilon+\mu_b(u_0)\}=O_p\left(\frac{1}{\sqrt{(bN)}}+b^2\right), \tag{A6}$$

since $\mathrm{var}\{\sqrt{(bN)}\mathscr{S}_{t_0,N}\}=O(1)$ and $\mu_b(u_0)=O(b^2)$. This proves part (i).
Theorem 2 easily proves part (ii). □

We use the following bound in our investigation of $\hat{\alpha}_N$. By using Corollary 1 we have

$$\sup_{1\leqslant r\leqslant\lfloor 1/\tilde{b}\rfloor}\|\hat{\beta}_{r\tilde{b}N,N}-\beta_{r\tilde{b}}\|_\infty=O_p\left(\frac{1}{\tilde{b}\sqrt{(bN)}}+\frac{b^2}{\tilde{b}}\right). \tag{A7}$$

To derive the asymptotic sampling distribution of $\hat{\alpha}_N$ we need to consider the global least squares criterion $T_N(\hat{\alpha}_N)$. To simply the analysis we replace the estimators of the time-varying autoregressive parameters with the true parameters and use the corrected global least squares criterion

$$\widetilde{T}_N(a)=\sum_{r=1}^{\lfloor 1/\tilde{b}\rfloor}\tilde{b}\sum_{k=(r-1/2)\tilde{b}N}^{(r+1/2)\tilde{b}N-1}\frac{1}{\tilde{b}N}\left\{\sum_{j=0}^{p}\beta_j(r\tilde{b})X_{k-j,N}-\sum_{i=1}^{q}\sum_{j=0}^{p}a_i\beta_j(r\tilde{b})f_i\left(\frac{k-j}{N}\right)\right\}^2. \tag{A8}$$

In order to study the asymptotic properties of $\hat{\alpha}_N$ through $\widetilde{T}_N(a)$ we will need to show that $T_N(a)$ and $\widetilde{T}_N(a)$ are sufficiently close to each other. We shall assume in the analysis below that $\tilde{b}\sqrt{(bN)}\to\infty$, since from (A7) this implies that $\sup_{1\leqslant r\leqslant\lfloor 1/\tilde{b}\rfloor}\|\hat{\beta}_{r\tilde{b}N,N}-\beta_{r\tilde{b}}\|_\infty\to 0$, in probability. Under the assumption that $\tilde{b}\sqrt{(bN)}\to\infty$ and by using the results in our technical report we can show that

$$\sup_{a\in\Omega}\|T_N(a)-\widetilde{T}_N(a)\|=O_p\left(\frac{1}{\tilde{b}\sqrt{(bN)}}+b^2\right),\quad \|\nabla^2T_N(a)-\nabla^2\widetilde{T}_N(a)\|=O_p\left(\frac{1}{\tilde{b}\sqrt{(bN)}}+b^2\right).$$

Furthermore, we can show that $\nabla\widetilde{T}_N(a)$ at the true value $a=\alpha$ has the better rate

$$\sqrt{N}\nabla T_N(\alpha)=\sqrt{N}\nabla\widetilde{T}_N(\alpha)+O_p\left(\frac{1}{\tilde{b}^{3/2}\sqrt{(bN)}}+\frac{b^2}{\tilde{b}^{1/2}}\right).$$

Therefore, if $b^{3/2}\sqrt{(bN)}\to\infty$ and $b^2=o(\tilde{b}^{1/2})$ then $\sqrt{N}\nabla T_N(\alpha)=\sqrt{N}\nabla\widetilde{T}_N(\alpha)+o_p(1)$. We use these bounds in the results below.

Proof of Theorem 3. We first define the asymptotic limit function

$$\widetilde{T}(a)=\sum_{i_1,i_2=1}^{q}\{\alpha_{i_1}\alpha_{i_2}-\alpha_{i_1}a_{i_2}-\alpha_{i_2}a_{i_1}+a_{i_1}a_{i_2}\}\sum_{j_1,j_2=0}^{p}g(j_1,j_2,i_1,i_2),$$

where

$$g(j_1,j_2,i_1,i_2)=\int_0^1\beta_{j_1}(u)\beta_{j_2}(u)f_{i_1}(u)f_{i_2}(u)du.$$

Under the assumption $\tilde{b}^2bN\to\infty$, results from our technical report give that $T_N(a)\to\widetilde{T}(a)$ uniformly in α. Since $\widetilde{T}(a)$ has a unique minimum the arguments used in the proof of Theorem 1 prove Theorem 3. □

If $b \to 0$, $\tilde{b} \to 0$ and $\tilde{b}^2 N \to \infty$ as $N \to \infty$ it can be shown that

$$\sqrt{N}\tilde{T}_N(\alpha) = \sqrt{N} M_N + o_p(1), \tag{A9}$$

where

$$M_N = -2 \sum_{r=1}^{\lfloor 1/\tilde{b} \rfloor - 1} \tilde{b} \sum_{k=(r-1/2)\tilde{b}N}^{(r+1/2)\tilde{b}N-1} \frac{1}{\tilde{b}N} g(\beta_{k/N}, k, N)\eta_k.$$

Proof of Theorem 4. If $\tilde{b}^3 bN \to \infty$ it can be shown that $\nabla^2 T_N = \tilde{V}_2 + o_p(1)$. By using (A9) we can show that

$$\sqrt{N}(\hat{\alpha}_N - \alpha) = \{\tilde{V}_2^{-1} + o_p(1)\}\{\sqrt{N} M_N + o_p(1)\}. \tag{A10}$$

By using the Lindeberg condition we can show that $\sqrt{N} M_N \to \mathcal{N}(0, 4\tilde{V}_1)$ in distribution. Therefore, by using (A10) and the above we have the result. □

References

Akaike, H. & Kitagawa, G. (1978). A procedure for the modeling of non-stationary time-series. *Ann. Inst. Statist. Math.* **30**, 351–63.

Akaike, H. & Kitagawa, G. (1998). *The Practice of Time Series Analysis.* New York: Springer-Verlag.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **16**, 1–37.

Dahlhaus, R. & Giraitis, L. (1998). On the optimal length for the parameter estimates for locally stationary time series. *J. Time Ser. Anal.* **19**, 629–55.

Durbin, J. (1960). Estimation of parameters in time-series regression models. *J. R. Statist. Soc.* B **22**, 139–53.

Hall, P. & Heyde, C. (1980). *Martingale Limit Theory and its Application.* New York: Academic Press.

Ozaki, T. & Tong, H. (1975). On the fitting of non-stationary models in time series. In *Proceedings of the 8th Hawaii International Conference on System Science*, pp. 224–6. Western Periodical Company.

Parker, D., Jones, P., Bevan, A. & Follard, C. (1994). Interdecadal changes of surface temperature since the late 19th century. *Geophys. Res.* **99**, 14373–99.

Pierce, D. (1971). Least squares estimation in the regression model with autoregressive moving average errors. *Biometrika* **58**, 299–312.

Priestley, M. B. (1988). *Non-Linear and Non-Stationary Time Series Analysis.* London: Academic Press.

Subba Rao, T. (1970). The fitting of non-stationary time-series models with time-dependent parameters. *J. R. Statist. Soc.* B **32**, 312–22.

Subba Rao, T. & Tsolaki, E. (2004). Nonstationary time series analysis of monthly global temperature anomalies. In *Applications of Time Series Analysis in Geophysics*, Ed. D. Brillinger and E. Robinson, pp. 73–104. New York: Springer Verlag.

Wu, W., Woodroofe, M. & Mentz, G. (2001). Isotonic regression and the changepoint problem. *Biometrika*, **88**, 793–804.

[*Received February* 2003. *Revised October* 2003]