The Annals of Statistics
1982, Vol. 10, No. 3, 915–925

ANALYSIS OF TIME SERIES FROM MIXED DISTRIBUTIONS

By P. M. Robinson[1]

University of Surrey

Some stationary and non-stationary time series arise from mixed distributions, the probabilities attached to the occurrence of certain values being positive, while a continuum of possible values is also involved. Such series are modeled in terms of a stationary Gaussian process X_t, which is censored when it crosses certain thresholds. Procedures are proposed for estimating the autocorrelation function of X_t. Their strong consistency and asymptotic normality are established. We suggest tests of the hypothesis that X_t is white noise.

1. Introduction. Most of the methodology of time series analysis is best suited to data that are a realization of continuous random variables. A great deal is also known about certain stochastic processes for which the distributions are discrete. Some time series appear to arise from mixed distributions. For example, a rainfall series may contain a substantial proportion of zero values. External factors can censor an underlying continuous variable; examples that came to mind of data that may be so affected are riverflow data, sales data, and certain chemical processes. In econometrics and biostatistics, interest has sometimes focused on such models as the "Tobit" (Amemiya, 1973; Poirier, 1978; Robinson, 1982) where we observe

$$
\begin{aligned}
(1.1)\qquad Y_t &= \boldsymbol{\beta}\mathbf{z}_t + \sigma X_t, && \text{if } \boldsymbol{\beta}\mathbf{z}_t + \sigma X_t > 0,\\
&= 0, && \text{otherwise,}
\end{aligned}
$$

in which $\sigma > 0$, $\boldsymbol{\beta}$ is a row vector, and $\mathbf{z}_t$ is a column vector of explanatory variables. In all the literature except Robinson (1982), the unobservable stochastic process X_t has been assumed to be white noise.

The application of standard procedures to a time series from a mixed distribution could produce misleading results. For example, suppose that in (1.1) $\boldsymbol{\beta}$ and $\mathbf{z}_t$ are scalar, and $\mathbf{z}_t = 1$, all t. If X_t is stationary, so is Y_t, but one expects that the usual time series models fitted to Y_t would produce forecasts above the zero threshold in higher proportion than in the available data.

Let U_t be a real-valued process, observed at $t = 1, \ldots, T$. We model U_t in terms of the stationary Gaussian process X_t, for which

$$(1.2)\qquad EX_t = 0, \quad EX_t^2 = 1.$$

Denote the autocorrelation function of X_t by

$$\rho_u = EX_tX_{t-u} \quad u = 1, 2, \ldots .$$

For each t we observe

$$
\begin{aligned}
(1.3)\qquad U_t &= X_t \quad \text{if } X_t > b_t,\\
&= 0, \quad \text{if } X_t \le b_t,
\end{aligned}
$$

Received January 1979; revised December 1981.

[1] Research supported in part by Grant SOC78-05803 from the National Science Foundation. The constructive comments of the referees have led to several improvements. The final revision was written while the author was a Visiting Fellow in the Department of Statistics, Institute of Advanced Studies, Australian National University.

AMS 1980 *subject classifications.* Primary 62M10; secondary 60G15, 62M07.

Key words and phrases. Mixed distributions, Gaussian process, nonlinear least squares, moment estimator, strong consistency, central limit theorem, tests for white noise.

where b_t is known. A situation of upper rather than lower thresholds is transformed to (1.3) by a change of sign. An example of (1.3) is (1.1), with $U_t = (Y_t - \boldsymbol{\beta}\mathbf{z}_t)/\sigma$, for $Y_t > 0$, and $b_t = -\boldsymbol{\beta}\mathbf{z}/\sigma$. In practice $\boldsymbol{\beta}$ and σ are unknown, so U_t is unobservable and b_t is unknown. However maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$ of $\boldsymbol{\beta}$ and σ when the Y_t are independent are studied in Amemiya (1973) and in Robinson (1982) $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$ are shown to be strongly consistent and asymptotically normal (SCAN) even without independence for a wide class of correlated Gaussian processes X_t. One might form the $\hat{U}_t = (Y_t - \hat{\boldsymbol{\beta}}\mathbf{z}_t)/\hat{\sigma}$, for $Y_t > 0$, and $\hat{b}_t = -\hat{\boldsymbol{\beta}}\mathbf{z}_t/\hat{\sigma}$, and then apply the methods below for estimating the ρ_u with U_t, b_t replaced by $\hat{U}_t$, $\hat{b}_t$. It is possible to extend our proofs to show that these estimators of ρ_u are also SCAN; the arguments are fairly standard but lengthy and are omitted. Our estimators unfortunately are generally not robust to departures from Gaussianity but a possible extension relaxes the Gaussianity assumption by modeling Y_t as a nonlinear function of known form of X_t and possibly unknown parameters; in Poirier (1978) the model (1.1) is combined with Box-Cox transformations. The notion of an underlying continuous variable that can take negative values is of course entirely artificial in such cases as the rainfall example referred to above.

Our nonparametric estimates of ρ_u can be inserted in the usual formulas for weighted-covariance spectral estimates. They can also be used to identify and estimate a suitable finite-parameter model for X_t (for example an autoregressive moving average model) and as starting values in maximum likelihood estimation, although as shown in Robinson (1980) the likelihood may involve multidimensional integrals and therefore present computational difficulties. Our estimates of ρ_u can also be inserted in the expression given in Robinson (1982) to provide a consistent estimate of the limiting covariance matrix of $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}$ in the case of serial dependence. A further application of our estimates, one which we discuss in Section 5, is tests for serial independence of X_t.

We mention some other work in which a discrete-valued process arises from an underlying Gaussian X_t: only the sign of X_t is observed (Brillinger, 1968; Hinich, 1967); X_t is digitalized (McNeil, 1967); one observes an odd, bounded, nondecreasing function of X_t (Rodemich, 1966).

2. Nonlinear regression estimators. The regression of X_t on X_{t-u} is linear,

$$E(X_t \mid X_{t-u} = x) = \rho_u x.$$

When $b_t = -\infty$, all t, the ordinary least squares estimator of ρ_u is

$$\hat{\rho}_u = \left(\textstyle\sum_{t=u+1}^T X_t^2\right)^{-1} \textstyle\sum_{t=u+1}^T X_t X_{t-u}. \tag{2.1}$$

If X_t is ergodic, $\hat{\rho}_u$ is consistent for ρ_u.

The regression of X_t on X_{t-u}, conditional on $X_t > b_t$, is

$$E(X_t \mid X_t > b_t, X_{t-u} = x) = \rho_u x + \mu(b_t - \rho_u x; \rho_u) \tag{2.2}$$

for $x > b_{t-u}$, where

$$\mu(b; \rho) = \tau\phi_\tau(b)/\{1 - \Phi_\tau(b)\} \quad \tau = 1 - \rho^2,$$

and ϕ_τ and Φ_τ are the $N(0, \tau)$ probability density function and distribution function, respectively. The function $\mu(b; \rho)$ is the "hazard rate" for the $N(0, \tau)$ (Johnson and Kotz, 1970, page 278). In the least squares sense, (2.2) is the best predictor of X_t, conditional on knowledge that $X_{t-u} = x$ and that $X_t > b_t$. Define

$$q_t(\mathbf{b}; \rho) = \{X_t - \rho X_{t-u} - \mu(b - \rho X_{t-u}; \rho)\}^2 I_t(\mathbf{b}),$$

where $\mathbf{b} = (b, c)$ and $I_t(\mathbf{b}) = 1$ if $X_t > b$, and $X_{t-u} > c$; $= 0$, otherwise. Put

$$Q_u(\rho) = T^{-1} \textstyle\sum_{t=1}^T q_t(\mathbf{b}_t; \rho), \quad \mathbf{b}_t = (b_t, b_{t-u}),$$

and consider as an estimator of ρ_u a random variable $\hat{\rho}_{uA}$ for which

$$\min_{\rho \in \mathscr{R}} Q_u(\rho) = Q_u(\hat{\rho}_{uA})$$

where $\mathscr{R} = [\varepsilon - 1, 1 - \varepsilon]$, for some positive ε close to 0.

The censoring of X_t produces a regression function $\rho x + \mu(b - \rho x; \rho)$ that is nonlinear in x. It is monotone, having an asymptote ρx as $\operatorname{sgn}(\rho)x \to \infty$, but approaching b as $\operatorname{sgn}(\rho)x \to -\infty$ if $b > -\infty$. The nonlinearity in ρ of the regression necessitates use of numerical methods for the minimization of $Q_u(\rho)$. However Φ_τ, which is closely related to the error function, can be quickly computed by means of library functions on many computers, or alternatively by means of various approximations and expansions (Johnson and Kotz, 1970, pages 278–283).

The proposed nonlinear least squares (NLLS) method could also be used in the related problem of estimating the correlation coefficient from independent truncated or censored bivariate normal observations. For that problem, our method would be less efficient, asymptotically, than maximum likelihood. The latter approach can be adapted to our problem (although it would not be "maximum likelihood"), the objective function to be maximized being

$$L_u(\rho) = \sum \log \phi(X_t, X_{t-u}; \rho) + \sum \log \int_{-\infty}^{b_{t-u}} \phi(X_t, x; \rho)\, dx + \sum \log \int_{-\infty}^{b_t} \phi(x, X_{t-u}; \rho)\, dx + \sum \int\int_{-\infty}^{\mathbf{b}} \phi(x, y; \rho)\, dx\, dy,$$

where $\phi(\cdot, \cdot; \rho)$ is the standard bivariate normal density with correlation ρ and the four sums are respectively over $\{X_t > b_t, X_{t-u} > b_{t-u}\}$, $\{X_t > b_t, X_{t-u} \le b_{t-u}\}$, $\{X_t \le b_t, X_{t-u} > b_{t-u}\}$, $\{X_t \le b_t, X_{t-u} \le b_{t-u}\}$. Because they require computation of bivariate normal probabilities, these estimates are somewhat less easy to compute than those of this paper, but they will be more efficient, particularly when a large proportion of observations is censored.

3. Moment estimators. Because $EX_tX_{t-u} = \rho_u EX_t^2$, we can regard (2.1) as a moment estimator of ρ_u in the case $b_t = -\infty$, all t. The Gaussianity assumption in fact leads to a whole family of consistent moment estimators, but because the statistics $\sum_{t=u+1}^{T} X_tX_{t-u}$, $u = 0, \cdots, T-1$ are jointly sufficient for $\rho_1, \cdots, \rho_{T-1}$, estimators based on second moment statistics seem the only ones worth considering.

Turning to the censored case, define

$$m_{jk}(\mathbf{b}) = E\{(X_t - b)^j (X_{t-u} - c)^k I_t(\mathbf{b})\}, \quad j, k \ge 0.$$

It can be verified by integration by parts that

$$(1 - \rho_u^2) j m_{j-1,k} = m_{j+1,k} - \rho_u m_{j,k+1} + (b - \rho_u c) m_{jk}, \quad j \ge 1, k \ge 0, \tag{3.1}$$

$$(1 - \rho_u^2) k m_{j,k-1} = m_{j,k+1} - \rho_u m_{j+1,k} + (c - \rho_u b) m_{jk}, \quad j \ge 0, k \ge 1, \tag{3.2}$$

where reference to $\mathbf{b}$ is suppressed. Any number of consistent estimators can be formed from (3.1) and (3.2). In Rosenbaum (1961) moment estimators are proposed for the related problem of independent truncated bivariate normal variables (with unknown means and variances, and unknown truncation points that are constant over the observations). The relations (3.1) and (3.2) are derived by Rosenbaum for $j = 1$, $k = 0$ and $j = 0$, $k = 1$, respectively and a certain linear combination formed. A quadratic equation results, producing two possible estimates of ρ_u. It may be shown, analytically and by simulations, that generally both are between -1 and 1. This difficulty is not mentioned by Rosenbaum. It can be surmounted by eliminating the quadratic term. In deciding which two equations from the class (3.1), (3.2) to select, care must be exercised to avoid near-indeterminacy corresponding to very inefficient estimators. One possibility which involves moments no higher than order 3 is to choose $j = 1$, $k = 0$, and $j = k = 1$ in (3.1), which leads to

$$\hat{\rho}_{uB} = (c_u h_u - d_u e_u)/(f_u h_u - d_u g_u), \tag{3.3}$$

where

$$c_u = \frac{1}{T}\sum{}' X_t(X_t - b_t)(X_{t-u} - b_{t-u}), \quad d_u = \frac{1}{T}\sum{}' (X_{t-u} - b_{t-u}), \quad e_u = \frac{1}{T}\sum{}' X_t(X_t - b_t),$$

$$f_u = \frac{1}{T}\sum{}' (X_t - b_t)X_{t-u}(X_{t-u} - b_{t-u}), \quad g_u = \frac{1}{T}\sum{}' (X_t - b_t)X_{t-u}, \quad h_u = \frac{1}{T}\sum{}' 1,$$

the primed sums excluding terms for which $I_t(\mathbf{b}_t) = 0$.

4. Asymptotic properties. We establish that the estimators A and B are SCAN under mild weak-dependence conditions on X_t. We also show that $Q_u(\rho)$ has, almost surely (a.s.), a unique relative minimum for sufficiently large T. The latter property, which of course is known to hold in the uncensored case, $b_t \equiv -\infty$, is useful because it diminishes the need for a search of the parameter space prior to hill-climbing optimisation techniques.

For consistency of estimators, we make the following three assumptions.

ASSUMPTION A1. X_t is a stationary Gaussian process that satisfies (1.2) and has a spectral density function $S(\lambda)$ that is representable by

$$S(\lambda) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} \rho_u e^{-iu\lambda} = |P(e^{i\lambda})|^2 R(\lambda), \tag{4.1}$$

where $P(\cdot)$ is a polynomial and $R(\cdot)$ is a strictly positive function that satisfies a Hölder condition of order $\eta > 0$.

Condition A1 includes, for example, the case of an X_t generated by an autoregressive moving average process, where the moving average polynomial $P(z)$ can have roots on the unit circle.

ASSUMPTION A2. For all t, $b_t \le B < \infty$. For given u, the joint empirical distribution function, G_T, of $\mathbf{b}_1, \cdots, \mathbf{b}_T$ converges completely to a joint distribution function, G, as $T \to \infty$.

When $b_t = -\boldsymbol{\beta}\mathbf{z}_t/\sigma$, A2 implies a stability in the explanatory variables $\mathbf{z}_t$. In respect of $\hat{\rho}_{uB}$ only, we introduce also the following.

ASSUMPTION A3. The matrix

$$\int E\left\{\begin{bmatrix} (X_t - b)X_{t-u} \\ 1 \end{bmatrix}\begin{bmatrix} X_{t-u} - c \\ 1 \end{bmatrix}' I_t(\mathbf{b})\right\} dG(\mathbf{b}) \tag{4.2}$$

is non-singular, the prime denoting transposition.

The matrix expectation in the integrand of (4.2) becomes singular only as $b, c \to \infty$, so because b_t is bounded from above under Assumption A2, Assumption A3 seems reasonable.

THEOREM 1. *Under Assumptions* A1 *and* A2, *for every fixed u such that* $\rho_u \in \mathscr{R}$,

(i) $\lim_{T\to\infty}\hat{\rho}_{uA} = \rho_u$, a.s.

(ii) *For T sufficiently large,* $\hat{\rho}_{uA}$ *is* a.s. *the unique relative minimum of* $Q_u(\rho)$.

(iii) $\lim_{T\to\infty}\hat{\rho}_{uB} = \rho_u$, a.s., *if Assumption* A3 *also holds.*

The proof of this, and of Theorem 2 below, is contained in the Appendix. The central limit theorem is proved under the following three conditions.

ASSUMPTION B1. X_t satisfies Assumption A1, but $R(\lambda)$ is also differentiable, its derivative satisfying a Hölder condition of order $\eta > 0$.

ASSUMPTION B2. The b_t satisfy Assumption A2, for $u = 1, \cdots, \ell$, and also, for all u, v, $w = 1, 2, \cdots$, the joint empirical distribution function of $(b_1, b_{1-u}, b_{1-v}, b_{1-w}), \cdots (b_T, b_{T-u}, b_{T-v}, b_{T-w})$ converges completely to a joint distribution function as $T \to \infty$.

For the moment estimator, an additional condition is needed. Define $\mathbf{c}_u = (c_u, d_u, e_u, f_u, g_u, h_u)'$ and $\gamma_u = \lim_{T\to\infty} E\mathbf{c}_u$.

ASSUMPTION B3. $\lim_{T\to\infty} T^{1/2}(E\mathbf{c}_u - \gamma_u) = \mathbf{0}$.

A condition of this type, of rapid convergence of G_T to G, is not required for $\hat{\rho}_{uA}$, but it is satisfied if, for example, the explanatory variables $\mathbf{z}_t$ in (1.1), and thence b_t, are periodic functions of t.

THEOREM 2. *Under Assumptions* B1 *and* B2, *for any* $\ell > 0$ *such that* ρ_u *is an interior point of* $\mathscr{R}$, $u = 1, \cdots, \ell$,

$$T^{1/2}(\hat{\rho}_{1A} - \rho_1, \cdots, \hat{\rho}_{\ell A} - \rho_\ell) \text{ is asymptotically } N(\mathbf{0}, \mathbf{\Lambda}^{-1} \textstyle\sum \mathbf{\Lambda}^{-1}),$$

where $\mathbf{\Lambda}$ *is diagonal with* uth *diagonal element.*

$$\lambda_u = \lim_{T\to\infty} (\partial^2/\partial\rho^2) Q_u(\rho_u), \quad \text{a.s.},$$

and $\sum$ *has* uvth *element*

$$\sigma_{uv} = \lim_{T\to\infty} TE(\partial/\partial\rho) Q_u(\rho_u)(\partial/\partial\rho) Q_v(\rho_v). \tag{4.3}$$

Under Assumptions A3, B1, B2, B3, *for any* $\ell > 0$

$$T^{1/2}(\hat{\rho}_{1B} - \rho_1, \cdots, \hat{\rho}_{\ell B} - \rho_\ell) \text{ is asymptotically } N(\mathbf{0}, \mathbf{\Psi})$$

where $\mathbf{\Psi}$ *has* uv*th element* $\psi_{uv} = \boldsymbol{\alpha}_u' \mathbf{\Gamma}_{uv} \boldsymbol{\alpha}_v$, *with* $\mathbf{\Gamma}_{uv} = \lim_{T\to\infty} TE\{(\mathbf{c}_u - E\mathbf{c}_u)(\mathbf{c}_v - E\mathbf{c}_v)'\}$ *and* $\boldsymbol{\alpha}_u = p\lim_{T\to\infty} \mathbf{a}_u$, $\mathbf{a}_u$ *being the column vector of derivatives of* $\hat{\rho}_{uB}$ *with respect to* $\mathbf{c}_u$, *namely*

$$\mathbf{a}_u = (f_u h_u - d_u g_u)^{-1}(h_u, h_u j_u, -d_u, -h_u \hat{\rho}_{uB}, d_u \hat{\rho}_{uB} - d_u j_u)',$$

where $j_u = (c_u g_u - e_u f_u)/(f_u h_u - d_u g_u)$.

5. Tests for white noise. Expressions can be obtained for the asymptotic covariance matrices in Theorem 2 but generally these are rather complicated and difficult to estimate, compared to those for uncensored series (Robinson, 1977). The matrix $\mathbf{\Sigma}$ may be represented as an infinite series, after expansion of $\mu(b - \rho X_{t-u}; \rho)$ in Hermite polynomials. For the $\mathbf{\Gamma}_{uv}$ it is necessary to evaluate moments of the truncated quadrivariate normal distribution, for which the form for the characteristic function in Tallis (1961) may be useful. In any case $\mathbf{\Lambda}^{-1} \mathbf{\Sigma} \mathbf{\Lambda}^{-1}$ and $\mathbf{\Psi}$ are not generally the correct formulas when the b_t depend on estimated parameters.

Simplications result under certain hypotheses, notably white noise,

$$\rho_u = 0, \quad \text{all} \quad u > 0. \tag{5.1}$$

Write $\mu_t = \mu(b_t; 0)$, $F_t = 1 - \Phi_1(b_t)$, $\phi_t = \phi_1(b_t)$; then under (5.1) we have the following consistent estimators of σ_{uv}, λ_u and ψ_{uv}:

$$\hat{\sigma}_{uv} = 4T^{-1} \textstyle\sum \{1 - \mu_t(\mu_t - b_t)\}^3 F_t\{(1 + b_{t-u}\mu_{t-u})F_{t-u}\delta_{uv} + \phi_{t-u}\phi_{t-v}(1 - \delta_{uv})\},$$

$$\hat{\lambda}_u = 2T^{-1} \textstyle\sum \{1 - \mu_t(\mu_t - b_t)\}^3 (1 + b_{t-u}\mu_{t-u}) F_t F_{t-u},$$

$$\hat{\psi}_{uv} = T' \textstyle\sum \{2 - b_t(X_t - b_t)\}([\{1 - b_{t-u}(X_{t-u} - b_{t-u})\}\delta_{uv}$$

$$+ (X_{t-u} - b_{t-u})(X_{t-v} - b_{t-v})(1 - \delta_{uv})]h_u h_v - (X_{t-u} - b_{t-u})h_u\, d_v$$
$$- (X_{t-v} - b_{t-v})h_v\, d_u + d_u\, d_v)'/\{(f_u h_u - d_u g_u)(f_v h_v - d_v g_v)\},$$

where δ_{uv} is the Kronecker delta. The elementary proofs are omitted. Some rough comparisons of efficiency are possible under (5.1), particularly when $b_t \equiv b$. For all u,

$$\lim_{T\to\infty}\mathrm{Var}\; T^{1/2}\hat{\rho}_{uB}/\lim_{T\to\infty}\mathrm{Var}\; T^{1/2}\hat{\rho}_{uA} = \{2 - b(\mu - b)\}(1 + b\mu)(\mu - b)^{-2}.$$

This function always exceeds 1, and increases monotonically to ∞ as $b \to \infty$.

Many tests of (5.1) can be constructed. Statistics such as $T \sum \sum_{u,v=1}^{\ell} \hat{\rho}_u \kappa^{uv} \hat{\rho}_v$ are asymptotically χ^2_ℓ under (5.1), where $\hat{\rho}_u$ is either $\hat{\rho}_{uA}$ or $\hat{\rho}_{uB}$ and κ^{uv} is a consistent estimator of the (u, v)th element of $\Lambda\Sigma^{-1}\Lambda$ or Ψ^{-1}. A simpler statistic, which does not require actual estimation of the ρ_u, and which has the same asymptotic distribution under (5.1), is

$$T \sum_{u,v=1}^{\ell}\sum (\partial/\partial\rho)Q_u(0)\hat{\sigma}^{uv}(\partial/\partial\rho)Q_v(0)$$

where (see (A.4), (A.5) in the Appendix)

$$(\partial/\partial\rho)Q_u(0) = -2T^{-1} \sum (X_t - \mu_t)X_{t-u}\{1 - \mu_t(\mu_t - b_t)\}I_t(\mathbf{b}_t).$$

(When $b_t \equiv b$, Σ and Ψ are patterned matrices which are immediately invertible.) Tests that are likely to have more power arise from the function $L_u(\rho)$ of Section 2: another asymptotic χ^2_ℓ statistics is

$$T^{-1} \sum_{u=1}^{\ell} \{(\partial/\partial\rho)L_u(0)\}^2/\{1 - \mu_t(\mu_t - b_t) + \mu_t^2/\Phi_t\},$$

where $\Phi_t = 1 - F_t$ and

$$(\partial/\partial\rho)L_u(0) = \sum X_t X_{t-u} - \sum X_t \phi_{t-u}/\Phi_{t-u} - \sum X_{t-u}\phi_t/\Phi_t + \sum \phi_t\phi_{t-u}/\Phi_t\Phi_{t-u},$$

the sums in the last expression corresponding to those in the formula for $L_u(\rho)$.

Modified formulas that are appropriate when the b_t depend on estimated parameters, as in the case of model (1.1), are readily obtainable from a standard Taylor series argument.

6. Simulations. In order to evaluate and compare the practical performance of the estimators, a small simulation study was carried out. Three sequences of 1000 b_t were generated, such that $b_t = c + 0.25\cos(2\pi t/52)$, with c taken to be $-.5$, 0 and .5, respectively. In terms of model (1.1), $\beta = (-c, -0.25)$, $\mathbf{z}_t = (1, \cos(2\pi t/52))'$, $\sigma = 1$, and we have about twenty years of weekly data, containing a strong seasonal component. The conditions imposed on the b_t in our theorems are clearly satisfied. For each b_t sequence, 50 independent sequences of 200 and 1000 X_t were generated, with $\rho_u = (0.9)^u$, $u \geq 1$. The estimates $\hat{\rho}_{uA}$, $\hat{\rho}_{uB}$ were computed for $u = 1(1)16$; we report below only results for $u = 1(5)16$, but these are representative. In Table 1 the columns from left to right contain: u; true ρ_u; average (over 50 replications) estimated ρ_u; standard error; mean squared error; average value of $\sum I_t(\mathbf{b}_t)$ = effective degrees of freedom. In the 3rd through 5th columns, the left hand entries in each box refer to the NLLS estimator $\hat{\rho}_{uA}$ while the right hand ones refer to the moment estimator $\hat{\rho}_{uB}$. In the 3rd through 6th columns, the upper entries in the boxes are based on $T = 200$, the lower on $T = 1000$.

The performance of the estimators is generally very poor when $T = 200$, and in such samples, consideration should be given to more efficient estimators, such as those based on $L_u(\rho)$. For both sample sizes, a very strong tendency to underestimate is exhibited. When $T = 1000$ and $c = -.5$ or 0, both estimators perform quite well, with $\hat{\rho}_{uA}$ generally to be preferred. There is noticeable deterioration as u increases, which may largely be due to the associated decrease in $\sum I_t(\mathbf{b}_t)$. For $T = 1000$ and $c = .5$, $\hat{\rho}_{uA}$ still performs creditably but $\hat{\rho}_{uB}$ is very biased and unstable when u is large. The above results are based on putting $\hat{\rho}_{uB} = -1$ whenever (3.3) < -1; it never happened that (3.3) > 1. When $T = 1000$ we recorded (3.3) $- < 1$ on none of the $15 \times 50 = 750$ estimates computed for $c = -.5$; for

TABLE 1

$c = -.5$

u	ρ_u	AVE		SE		MSE		EDF
1	.900	.894	.857	.028	.082	.001	.008	126.60
		.898	.892	.015	.024	.000	.001	619.02
6	.314	.084	.071	.398	.313	.212	.157	99.56
		.299	.268	.161	.158	.026	.027	501.64
11	.109	−.108	−.059	.411	.362	.216	.159	87.36
		.107	.089	.206	.203	.042	.042	464.84
16	.038	−.138	−.112	.478	.456	.260	.230	79.58
		−.014	−.043	.216	.251	.049	.069	450.16

$c = 0$

u	ρ_u	AVE		SE		MSE		EDF
1	.900	.844	.846	.245	.094	.063	.012	94.26
		.899	.887	.018	.038	.000	.002	427.30
6	.314	.089	−.157	.369	.526	.187	.498	65.88
		.168	.209	.290	.258	.105	.078	293.60
11	.109	.019	−.184	.316	.648	.165	.506	55.98
		−.005	−.023	.235	.328	.068	.125	253.98
16	.038	−.050	−.253	.431	.545	.194	.381	50.50
		−.025	−.076	.236	.349	.060	.135	240.20

$c = .5$

u	ρ_u	AVE		SE		MSE		EDF
1	.900	.528	.769	.575	.287	.470	.099	48.02
		.902	.889	.019	.037	.000	.002	247.06
6	.314	−.030	−0.82	.496	.583	.364	.497	25.56
		.151	.061	.299	.499	.116	.313	137.60
11	.109	.055	.119	.597	.745	.359	.555	18.80
		.021	−.305	.329	.599	.116	.531	106.38
16	.038	−.058	−.159	.557	.732	.319	.574	15.42
		.053	−.190	.384	.644	.148	.467	97.38

$c = 0$ this event occurred twice for $c = .5$ 124 times, with nearly a third of the 50 replicates producing threshold values for high values of u. To compute $\hat{\rho}_{uA}$ we iterated from starting value 0, the $(j+1)$th iterate being

$$\hat{\rho}_u^{(j+1)} = \hat{\rho}_u^{(j)} - \frac{(\partial/\partial\rho)Q_u(\hat{\rho}_u^{(j)})}{2\sum'\{X_{t-u} + \mu'(b_t - \hat{\rho}_u^{(j)}X_{t-u};\, \hat{\rho}_u^{(j)})\}^2}$$

where the denominator is close to $(\partial^2/\partial\rho^2)Q_u(\hat{\rho}_u^{(j)})$ for large T. The iterations were halted as soon as $|\hat{\rho}_u^{(j+1)} - \hat{\rho}_u^{(j)}| \le .001$ and the number of iterative steps N was recorded for each estimate computed. When $T = 100$ and $c = -.5$ the average N increased from 3.38 for u

= 1 to 4.40 for $u = 16$; for $c = 0$ and $c = .5$ it was somewhat higher but never exceeded 7.

The same task was carried out for the case of independent uniformly distributed b_t's (when condition B3 is violated) and the overall message of the results was similar.

The computations were carried out partly on the University of British Columbia's Amdahl 470 V/6 and partly on the University of Surrey's PRIME network.

APPENDIX

Denote by $\mathscr{M}_t^u$, $t \le u$, the σ-field of events generated by $X_t, \cdots, X_u$, and the strong mixing coefficient

$$\alpha_r = \sup_{C \in \mathscr{M}_{-\infty}^t, D \in \mathscr{M}_{t+r}^\infty} |\Pr(C \cap D) - \Pr(C)\Pr(D)|,$$

for $r > 0$.

ASSUMPTION C1. Lex χ_t be a measurable function of $X_t, \cdots, X_{t-\ell}$, for fixed finite $\ell \ge 0$, such that $E\chi_t = 0$, $E|\chi_t|^\delta \le K < \infty$, some $\delta > 2$.

The following result will be useful in proving Theorem 1.

THEOREM A. *Let* $S_T = \chi_1 + \cdots + \chi_T$ *and let Assumptions* A1 *and* C1 *hold. Then* $\lim_{T\to\infty} T^{-1} S_T = 0$, a.s.

PROOF. Defining $S_{JT} = \chi_{J+1} + \cdots + \chi_{J+T}$,

$$ES_{JT}^2 = \sum_{J+1}^{J+T}\sum E\chi_s\chi_t \le KT(1 + \textstyle\sum_{r=1}^T \alpha_r^{1-2/\delta}), \tag{A.1}$$

by Ibragimov and Linnik (1971, Theorem 17.2.2). We can now apply Serfling (1970 page 1236), choosing for the functional $g(H_{JT})$ described there the right side of (A.1) because the conditions $g(H_{JT}) + g(H_{J+T,U}) \le g(H_{J,T+U})$, $1 \le T \le T + U$ and $g(H_{JT}) = O(T^2(\ln T)^{-2}(\ln\ln T)^{-2})$ are satisfied, the latter because $\alpha_r = O(r^{-\eta})$, under A1 (Ibragimov, 1970, Theorem 5). □

PROOF OF THEOREM 1. We first give the proofs for the NLLS estimator $\hat\rho_{uA}$. Abbreviate $Q_u(\rho)$ to $Q(\rho)$. Initially we shall show that

$$\lim_{T\to\infty} Q(\rho) = \bar{Q}(\rho), \quad \text{a.s.,} \quad \text{uniformly in } \rho \in \mathscr{R} \tag{A.2}$$

where

$$\bar{Q}(\rho) = \int q(\mathbf{b}; \rho)\, dG(\mathbf{b}), \qquad q(\mathbf{b}; \rho) = Eq_t(\mathbf{b}; \rho).$$

For any $\rho^* \in \mathscr{R}$

$$|Q(\rho) - \bar{Q}(\rho)| \le |Q(\rho) - Q(\rho^*)| + |Q(\rho^*) - EQ(\rho^*)| + |EQ(\rho^*) - \bar{Q}(\rho^*)| + |\bar{Q}(\rho^*) - \bar{Q}(\rho)|.$$

We put $Q(\rho^*) - EQ(\rho^*) = T^{-1}\sum_t \bar{q}_t$, $\bar{q}_t = q_t(b; \rho^*) - q(b; \rho^*)$, so that $E\bar{q}_t = 0$ and

$$E\bar{q}_t^4 \le Eq_t^4(b; \rho^*) \le 2^8 E[X_t^8 + X_{t-u}^8 + \mu(b - \rho^* X_{t-u}; \rho)^8].$$

From Johnson and Kotz (1970, page 279),

$$\mu(b; \rho) < \pi\sqrt{(b^2 + 2\pi)} + \pi(\pi - 1)b, \quad b > 0; \quad \mu(b; \rho) \downarrow 0, \quad \text{as} \quad b \to \infty, \tag{A.3}$$

so by Gaussianity and $b_t \le B < \infty$, it follows that $E\bar{q}_t^4 \le K$. Because $\bar{q}_t$ is measurable with respect to $\mathscr{M}_{t-u}^t$ we can apply Theorem A with $\chi_t = \bar{q}_t$, establishing $Q(\rho^*) - EQ(\rho^*) \to 0$, a.s. Next

$$|Q(\rho) - Q(\rho^*)| \leq |\rho - \rho^*| \, |T^{-1} \textstyle\sum_{t=1}^{T} (\partial/\partial\rho) q_t(b; \bar{\rho})|,$$

for $|\bar{\rho} - \rho| \leq |\rho^* - \rho|$, where

$$\text{(A.4)} \qquad \frac{\partial}{\partial\rho} q_t(\mathbf{b}; \rho) = -2\{X_t - \rho X_{t-u} - \mu(b - \rho X_{t-u}; \rho)\}\left\{X_{t-u} + \frac{\partial\mu(b - \rho X_{t-u}; \rho)}{\partial\rho}\right\} I_t(\mathbf{b}),$$

$$\text{(A.5)} \qquad \frac{\partial\mu(b - \rho x; \rho)}{\partial\rho} = \frac{\mu(b - \rho x; \rho)}{\tau}\left[\frac{(\rho b - x)}{\tau}\{\mu(b - \rho x; \rho) - (b - \rho x)\} - \rho\right].$$

It follows as before that $E\{(\partial/\partial\rho)q_t(b;\bar{\rho})\}^4 \leq K$, whence, by Theorem A, $T^{-1} \sum (\partial/\partial\rho)q_t(\mathbf{b}; \bar{\rho}) \to$ a.s. to a finite limit. Thus $Q(\rho)$ is equicontinuous. Because $q(\mathbf{b}; \rho)$ is continuous and bounded in $\mathbf{b}$, $EQ(\rho^*) - Q(\rho^*) \to 0$ from Jennrich (1969, Theorem 1). The continuity in ρ of $q(\mathbf{b}; \rho)$ implies that of $\bar{Q}(\rho)$, and this, and the compactness of $\mathscr{R}$, completes the proof of (A.2).

The convergence of $\hat{\rho}_{uA}$ follows from (A.2) by a standard type of argument (Jennrich, 1969, Theorem 6) once we prove

$$\text{(A.6)} \qquad Q(\rho) > Q(\rho_u), \quad \text{all} \quad \rho \in \mathscr{R}, \quad \rho \neq \rho_u.$$

Write

$$q_t(\mathbf{b}; \rho) = (y^2 - 2yz + z^2)I_t(\mathbf{b}), \quad y = X_t - \rho_u X_{t-u} - \mu(b - \rho_u X_{t-u}; \rho_u),$$

$$z = (\rho - \rho_u)X_{t-u} + \mu(b - \rho X_{t-u}; \rho) - \mu(b - \rho_u X_{t-u}; \rho_u).$$

Now $E(y \mid X_t > b, X_{t-u}) = 0$ implies $E\{yzI_t(\mathbf{b})\} = 0$. Thus

$$Eq_t(\mathbf{b}, \rho) = q(\mathbf{b}; \rho) = q(\mathbf{b}; \rho_u) + E\{z^2 I_t(\mathbf{b})\}.$$

Now for all $x > 0$ ($x < 0$), $\rho x + \mu(b - \rho x; \rho)$ is strictly monotone increasing (decreasing) in ρ. Thus, for all $b \leq B < \infty$, $c \leq B < \infty$, we have $E\{z^2 I(\mathbf{b})\} > 0$, $\rho \neq \rho_u$, that is $q(\mathbf{b}; \rho) > q(\mathbf{b}; \rho_u)$, $\rho \neq \rho_u$. Thus (A.6) is proved.

Part (ii) of the Theorem follows from the monotonicity of $\rho x + \mu(b - \rho x; \rho)$ mentioned above, which implies that the global minimum at ρ_u is the only relative minimum of z^2, for all X_{t-u}. Thus $q(\mathbf{b}; \rho)$, and thence $\bar{Q}(\rho)$, have a single relative minimum, at ρ_u. By uniform convergence, $Q(\rho)$ must therefore have a.s. a unique relative minimum for large enough T, and this must be $\hat{\rho}_{uA}$.

Part (iii) of the Theorem is a straightforward application of Theorem A and Jennrich (1969, Theorem 1), and indeed other members of the class of moment estimators discussed in Section 3 could be handled similarly. For $h, i, j, k \geq 0$,

$$T^{-1}\textstyle\sum' b_t^h b_{t-u}^i (X_t - b_t)^j (X_{t-u} - b_{t-u})^k$$

converges a.s. by Gaussianity and boundedness of b_t to

$$\int b^h c^i m_{jk}(\mathbf{b}) \, dG(\mathbf{b}).$$

The denominator of $\hat{\rho}_{uB}$ thus converges a.s. to the determinant of (4.2), which is non-zero under Assumption A3. On averaging the relations (3.1) over $\mathbf{b}$, it is seen that asymptotically they are satisfied by $\hat{\rho}_{uB}$. □

The proof of Theorem 2 uses the following.

THEOREM B. *Let χ_t satisfy Assumption* C1 *and let*

$$\text{(A.7)} \qquad \textstyle\sum_{j=1}^{\infty} \alpha_j^{1-2/\delta} < \infty.$$

Let the limits

$$\Omega_u = \lim_{T\to\infty} T^{-1} \textstyle\sum_{t=1}^{T-u} \chi_t \chi_{t-u}$$

exist for $u = 0, 1, \cdots$. *Then as* $T \to \infty$, $T^{1/2} S_T$ *is asymptotically normal with zero mean and variance*

$$\Omega_0 + 2 \textstyle\sum_{u=1}^{\infty} \Omega_u .$$

This extends Theorem 18.5.3 of Ibragimov and Linnik (1971) in two directions. First, our χ_t is a function of $X_t, \cdots, X_{t-\ell}$ instead of simply X_t; because ℓ is finite, this causes no problem. Second, whereas Ibragimov and Linnik's Theorem 18.5.3 is a central limit theorem for stationary processes, our χ_t are not assumed stationary, in order to deal with possible nonstationarity caused by unequal b_t. Inspection of the proof of the referenced Theorem 18.5.3 reveals that the uniform bound on $E|\chi_t|^\delta$, and existence of the limits Ω_u, can replace stationarity. We omit the full proof.

PROOF OF THEOREM 2. For the NLLS estimator, we have

$$O = (\partial/\partial\rho) Q_u(\hat\rho_{uA}) = (\partial/\partial\rho) Q_u(\rho_u) + (\partial^2/\partial\rho^2) Q_u(\bar\rho_u)(\hat\rho_{uA} - \rho_u)$$

for $|\bar\rho_u - \rho_u| \le |\hat\rho_{uA} - \rho_u|$, $u = 1, \cdots, \ell$. Let $\mathbf{d}$ be the $\ell \times 1$ vector with element $(\partial/\partial\rho) Q_u(\rho_u)$; then we show that

$$T^{1/2}\mathbf{d} \to N(\mathbf{O}, \mathbf{\Sigma}), \quad T \to \infty. \tag{A.8}$$

Writing $\chi_t = \sum_{u=1}^{\ell} \theta_u (\partial/\partial\rho) q_t(\mathbf{b}; \rho_u)$ we see that (A.8) is implied if for all sets of constants $\theta_1, \cdots, \theta_\ell$, $T^{-1/2} S_T$ is asymptotically normal. We have displayed $(\partial/\partial\rho) q_t(\mathbf{b}; \rho)$ in (A.4) and readily deduce that $E\chi_t = 0$ and

$$E|\chi_t|^\delta \le K\ell^{\delta-1} \textstyle\sum_{u=1}^{\ell} |\theta_u|^\delta \{E|X_{t-u} + (\partial/\partial\rho)\mu(b_t - \rho_u X_{t-u}; \rho_u)|^{2\delta}$$
$$\times E|X_t - \rho_u X_{t-u} - \mu(b_t - \rho_u X_{t-u}; \rho_u)|^{2\delta}\}.$$

From Gaussianity, (A.3), (A.5), $|\rho_u| < 1$ and $b_t \le B < \infty$ it follows that $E|\chi_t|^\delta \le K$ for any $\delta > 1$. By choosing $\delta > 2 + 2/\eta$ we will satisfy (A.7) because $\alpha_j \le Kj^{-1-\eta}$ under Assumption B1 (Ibragimov, 1970, Theorem 5). The limits (4.3) are seen to exist under B1 and B2, by applying the results obtained so far and Jennrich (1969, Theorem 1). The proof will be complete if $\lim(\partial^2/\partial\rho^2) Q_u(\bar\rho_u)$ exists and is a.s. non-zero, $u = 1, \cdots, \ell$. Since $\hat\rho_{uA} \to \rho_u$ a.s. from Theorem 1, it is sufficient for $(\partial^2/\partial\rho^2) Q_u(\rho)$ to converge uniformly in ρ within a neighbourhood of ρ_u, and for the limit to be positive at ρ_u. Now

$$(\partial^2/\partial\rho^2) Q_u(\rho) = -2T^{-1}\textstyle\sum[(\partial^2/\partial\rho^2)\mu(b_t - \rho X_{t-u}; \rho)\{X_t - \rho X_{t-u} - \mu(b_t - \rho X_{t-u}; \rho)\}$$
$$-\{X_{t-u} + (\partial/\partial\rho)\mu(b_t - \rho X_{t-u}; \rho)\}^2] I_t(\mathbf{b}_t),$$
$$(\partial^2/\partial\rho^2)\mu(b - \rho x; \rho) = \tau^{-2}\mu(b - \rho x; \rho)(\{\tau^{-1}(b - \rho x)[\mu(b - \rho x; \rho) - (b - \rho x)] - 1\}$$
$$+ \tau^{-1}(x - \rho b)^2[\tau^{-1}\{\mu(b - \rho x; \rho)$$
$$- (b - \rho x)\}\{2\mu(b - \rho x; \rho) - (b - \rho x)\} - 1]).$$

Uniform convergence, and the fact the limit is positive at ρ_u (by virtue of the strict monotonicity in ρ of $\rho x + \mu(b - \rho x; \rho)$), follow by arguments like those used in the proof of Theorem 1.

The proof for the moment estimators commences from

$$\textstyle\sum_{u=1}^{\ell} \theta_u T^{1/2}(\hat\rho_{uB} - \rho_u) = \sum_{u=1}^{\ell} \theta_u T^{1/2}(\mathbf{c}_u - E\mathbf{c}_u)'\bar{\mathbf{a}}_u + \sum_{u=1}^{\ell} \theta_u T^{1/2}(E\mathbf{c}_u - \gamma_u)'\bar{\mathbf{a}}_u \tag{A.9}$$

where $\bar{\mathbf{a}}_u$ is $\mathbf{a}_u$ evaluated at $\bar{\mathbf{c}}_u$, such that $\|\bar{\mathbf{c}}_u - \gamma_u\| \le \|\mathbf{c}_u - \gamma_u\|$. Because $\hat\rho_{uB} \to \rho_u$, it follows that $\bar{\mathbf{a}}_u \to$ a.s. to a finite limit, as in the proof of Theorem 1, and under Assumption B3, the second term on the right of (A.9) $\to 0$. It remains to show that $T^{1/2}(\mathbf{c}_u - E\mathbf{c}_u)$ is asymptotically normal with finite covariance matrix and this follows by the same sort of application of Theorem B as that used previously. □

REFERENCES

AMEMIYA, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica* **41** 997–1016.

BRILLINGER, D. R. (1968). Estimation of the cross-spectrum of a stationary bivariate Gaussian process from its zeros. *J. Roy. Statist. Soc. Ser. B.* **30** 145–159.

HINICH, M. (1967). Estimation of spectra after hard clipping of Gaussian processes. *Technometrics* **9** 391–400.

IBRAGIMOV, I. A. (1970). On the spectrum of stationary Gaussian sequences satisfying the strong mixing condition. II. Sufficient conditions. Mixing rate. *Theor. Probability Appl.* **15** 23–36.

IBRAGIMOV, I. A. and LINNIK, Y. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

JOHNSON, N. L. and KOTZ, S. (1970). *Distributions in Statistics. Continuous Univariate Distribution* **2.** Houghton-Mifflin, New York.

MCNEIL, D. R. (1967). Estimating the covariance and spectral density functions from a clipped stationary time series. *J. Roy. Statist. Soc. Ser. B.* **29** 180–195.

POIRIER, D. J. (1978). The use of the Box-Cox transform in limited dependent variable models. *J. Amer. Statist. Assoc.* **73** 284–287.

ROBINSON, P. M. (1977). Estimating variances and covariances of sample autocorrelations and autocovariances. *Aust. J. Statist.* **19** 236–240.

ROBINSON, P. M. (1980). Estimation and forecasting for time series containing censored and missing observations. In *Time Series Analysis* (O. D. Anderson, ed.) 167–182. North-Holland, Amsterdam.

ROBINSON, P. M. (1982). On the asymptotic properties of estimators of models containing limited dependent variables. *Econometrica* **50** 27–41.

RODEMICH, E. R. (1966). Spectral estimators using nonlinear functions. *Ann. Math. Statist.* **37** 1237–1256.

ROSENBAUM, S. (1961). Moments of a truncated bivariate normal distribution. *J. Roy. Statist. Soc. Ser. B.* **23** 405–408.

SERFLING, R. J. (1970). Convergence properties of S_n under moment restrictions. *Ann. Math. Statist.* **41** 1235–1248.

TALLIS, G. M. (1961). The moment generating function of the truncated multinormal distribution. *J. Roy. Statist. Soc. Ser. B.* **23** 233–239.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SURREY
GUILDFORD SURREY GU2 5XH
ENGLAND