

# Establishing Conditions for the Functional Central Limit Theorem in Nonlinear and Semiparametric Time Series Processes

James Davidson\*  
Cardiff University

February 2001

JEL Classification: C10, C22.

Keywords: Near-epoch dependence, FCLT, nonlinear, bilinear, GARCH, TAR.

## Abstract

This paper considers methods of deriving sufficient conditions for the central limit theorem and functional central limit theorem to hold in a broad class of time series processes, including nonlinear processes and semiparametric linear processes. The common thread linking these results is the concept of near-epoch dependence on a mixing process, since powerful limit results are available under this limited-dependence property. The particular case of near-epoch dependence on an independent process provides a convenient framework for dealing with a range of nonlinear cases, including the bilinear, GARCH, and threshold autoregressive models. It is shown in particular that even SETAR processes with a unit root regime have short memory, under the right conditions. A simulation approach is also demonstrated, applicable to cases that are analytically intractable. A new FCLT is given for semiparametric linear processes, where the forcing processes are of the NED-on-mixing type, under conditions that are evidently close to necessary.

---

\*Email: davidsonje@cf.ac.uk. Research supported by the ESRC under award L138251025. This paper is based on parts of a working paper circulated under the title "When is a Time Series I(0)?" I am grateful to Michael Jansson, Ron Gallant, an Associate Editor, and two anonymous referees for comments which have materially improved the paper. I retain all responsibility for errors.

# 1 Introduction

For statistical inference in time series, it is usually necessary to rely on asymptotic convergence results such as the central limit theorem (CLT) and functional central limit theorem (FCLT). The latter result, in particular, is central to the recent research on integrated processes and cointegration models. To cite one seminal example from a large and growing literature, Phillips (1987) derives distributional results for unit-root autoregressive processes. His results rely on a FCLT due to Herrndorf (1984) which assumes a mixing condition on the increments of the process.

As a benchmark, the basic result of this type is as follows. Let the stochastic process  $X_n : [0, 1] \mapsto \mathbb{R}$  be defined by

$$X_n(\xi) = \sigma_n^{-1} \sum_{t=1}^{[n\xi]} (x_t - Ex_t) \quad 0 < \xi \leq 1 \quad (1.1)$$

where  $\sigma_n^2 = \text{Var}(\sum_{t=1}^n x_t)$ , and  $[a]$  denotes the largest integer not exceeding  $a$ . Introduce the following assumptions.

1.  $x_t$  is either  $\alpha$ -mixing of size  $-r/(r-2)$  for  $r > 2$  or  $\phi$ -mixing of size  $-r/(2r-2)$ , for  $r \geq 2$ .<sup>1</sup>
2.  $\sup_t E|x_t - Ex_t|^r < \infty$ , and if  $r = 2$  then  $\{(x_t - Ex_t)^2\}$  is uniformly integrable.
3.  $\frac{\sigma_n^2}{n} \rightarrow \sigma^2 > 0$ , as  $n \rightarrow \infty$ .

Let  $\xrightarrow{d}$  denote convergence in distribution, and let  $B$  denote standard Brownian motion on  $[0, 1]$ .

**Theorem 1.1** If Assumptions 1, 2 and 3 hold, then  $X_n \xrightarrow{d} B$ .

This is a special case of De Jong and Davidson's (2000) Theorem 3.1, but the strong-mixing ( $\alpha$ -mixing) case is essentially the same as the one cited by Phillips (1987). Considering the case  $\xi = 1$  yields a CLT, and for this result, Assumption 3 can be relaxed to permit forms of global heteroscedasticity; see De Jong (1997) and Davidson (1992, 1993) for details.<sup>2</sup> Henceforth we will refer only to the FCLT to avoid repetition. There is effectively no loss of generality, since while the corresponding CLT may permit more heterogeneity, the short-memory requirements are always common to both results.

In applications of Theorem 1.1, to determine the limiting distribution of the Dickey-Fuller (1979) or Phillips-Perron (1988) statistics, for example, note that the increments of the observed time series must satisfy Assumptions 1–3. Corrections for autocorrelation may be needed to estimate the variance consistently, as in the Phillips-Perron nonparametric correction, and this is also, in effect, the role of the lagged increments in the augmented Dickey-Fuller regression. However, these corrections are irrelevant to the weak convergence of the partial sum process. To apply results such as Theorem 1.1, the prime consideration must be whether the economic time series having increments  $x_t$  does possess the properties specified for it.

While the flexibility and model-independence of mixing conditions are attractive features, there are well-known practical limitations with this approach to limiting the memory of a process. Since the strong and uniform mixing conditions bind on the supremum over all possible random

---

<sup>1</sup>Among other references on mixing conditions see Davidson (1994), Chapter 14, for definitions and examples.

<sup>2</sup>Without Assumption 3 it may still be possible to show convergence to a Gaussian a.s. continuous limit process, but this process is not Brownian motion.

events involving sequence coordinates having a given time separation, the relevant distributions must be fully specified to establish that they hold. They are therefore difficult or impossible to verify in practical applications. Moreover, they are known to fail in some seemingly innocent cases. Stable first-order linear autoregressive models with Bernoulli-distributed i.i.d. shocks are not strong mixing (Andrews 1984). Known sufficient conditions for strong mixing specify not merely a continuous distribution, but a smoothness condition on the density (see Davidson 1994, Theorem 14.9). Establishing that uniform mixing holds for linear models is even harder, and the best known sufficient condition (e.g. Davidson 1994, Theorem 14.14) requires the shock process to be bounded with probability 1, ruling out normality for example.

This paper considers alternatives to Assumption 1 that are either weaker than mixing, or more easily verifiable in applications, or both. The main vehicle for the analysis is the concept of near-epoch dependence (NED) on a mixing process, first introduced to econometricians by Gallant and White (1988). This approach to specifying conditions for the FCLT was pioneered by McLeish (1975), although the concept originates with Billingsley (1968) and Ibragimov (1962). It is a nonparametric restriction on the memory of a process that nonetheless constrains only a sequence of low-order moments. It has the benefits of both holding in cases where mixing fails, and, as we show below, of being potentially verifiable, as the property of a range of popular nonlinear time series models.

Let  $x_t(\dots, v_{t-1}, v_t, v_{t+1}, \dots)$  denote a random sequence whose coordinates are measurable functions of another random process  $\{v_s, -\infty < s < \infty\}$ , which is in general vector-valued. Define  $\mathcal{F}_s^t = \sigma(v_s, \dots, v_t)$  for any  $s \leq t$ , and also let  $E_{t-m}^{t+m}(\cdot)$  denote the expectation conditional on  $\mathcal{F}_{t-m}^{t+m}$ . For future reference, note that we shall also use the standard notation  $\mathcal{F}_t$  to denote  $\mathcal{F}_{-\infty}^t$ . The following definition is adapted from Davidson (1994), Definition 17.1. We reproduce it here for convenience, since features will need to be specifically cited in what follows.

**Definition 1**  $x_t$  is said to be near-epoch dependent on  $\{v_s\}$  in  $L_p$ -norm (or  $L_p$ -NED) for  $p > 0$  if<sup>3</sup>

$$\|x_t - E_{t-m}^{t+m}x_t\|_p \leq d_t\nu(m) \quad (1.2)$$

where  $d_t$  is a sequence of positive constants, and  $\nu(m) \rightarrow 0$  as  $m \rightarrow \infty$ . It is said to be  $L_p$ -NED of size  $-\mu$  if  $\nu(m) = O(m^{-\mu-\varepsilon})$  for  $\varepsilon > 0$ .

In the case where  $\nu(m) = O(e^{-\delta m})$  for  $\delta > 0$ , such that the size is nominally equal to  $-\infty$ , we also say the process is geometrically  $L_p$ -NED. The scale constants  $d_t$  allow for possible nonstationarity and in many cases, including all those examined in this paper, can be set equal to 1. More generally, it is desirable in applications to require that  $d_t = O(\|x_t\|_p)$ , since otherwise the concept can become vacuous. Note that NED is a condition on the mapping from  $\{v_s, -\infty < s < \infty\}$  to  $x_t$ , and so says nothing about the amount of dependence in the  $x_t$  series itself. It becomes useful when combined with a mixing condition on  $v_s$ , and in particular, independence of this series.

The NED-on-mixing property, subject to suitable size and moment restrictions, proves sufficient for the FCLT to hold. Consider the following assumption.

- 1'.  $x_t$  is  $L_2$ -NED of size  $-\frac{1}{2}$  on a process  $\{v_s\}$  with respect to constants  $d_t \leq \|x_t\|_r$ , where  $v_s$  is either  $\alpha$ -mixing of size  $-r/(r-2)$  for  $r > 2$  or  $\phi$ -mixing of size  $-r/(2r-2)$ , for  $r \geq 2$ .

The following result is another case of De Jong and Davidson's (2000b) Theorem 3.1.

**Theorem 1.2** If Assumptions 1', 2 and 3 hold, then  $X_n \xrightarrow{d} B$ .

---

<sup>3</sup> $\|\cdot\|_p$  denotes the  $L_p$ -norm,  $(E|\cdot|^p)^{1/p}$ .

The noteworthy feature of this result is that the mixing and moment conditions are identical to Theorem 1.1. The extension to NED functions of mixing processes represents a kind of free lunch, with no penalties on the permitted mixing sizes or other conditions. The CLT obtained by considering the case  $\xi = 1$  is similarly the most general currently available in respect of memory properties, although for this case Assumption 3 can be relaxed, as before.

The linear process is a standard example. Consider the general MA( $\infty$ ) model

$$x_t = \sum_{k=0}^{\infty} \theta_k u_{t-k} \quad (1.3)$$

where  $u_t$  (to be thought of as an element of  $\mathbf{v}_t$ , or possibly coincident with it) is bounded in  $L_2$ -norm with mean 0, but otherwise is unspecified. If the MA coefficients are absolutely summable, the process is  $L_2$ -NED on  $\{u_t\}$  with respect to NED numbers

$$\nu(m) = \sum_{j=m+1}^{\infty} |\theta_j|$$

and constants

$$d_t = 2 \sup_{s \leq t} \|u_s\|_2$$

(see Davidson 1994, Example 17.3.). Thus, if  $|\theta_j| = O(j^{-1-\mu-\varepsilon})$  for  $\varepsilon > 0$  the NED size is  $-\mu$ . To satisfy the  $L_2$ -NED conditions of Theorem 1.2 we would need  $|\theta_j| = O(j^{-3/2-\varepsilon})$ . However, since the shocks are themselves permitted to be dependent, note that this result is not comparable with the standard linear process setup, with i.i.d. shocks.

So much is well known. The present paper extends the theory in two directions, distinct but having in common the application of the NED concept. First, in Section 2, the property of  $L_2$ -NED on an independent process is used to unify the treatment of a range of popular nonlinear time series models. Parametric conditions are derived for these models to satisfy the assumptions of Theorem 1.2. The familiar case of ARMA processes is treated first, in Section 2.1, to establish methods and notation. We then consider in turn, bilinear models (Section 2.2), GARCH models (Section 2.3), and threshold autoregressive models (Section 2.4), looking in particular at the mixed stable/unit root case (Section 2.5). Section 2.6 considers an analogous, analytically intractable case, the ESTAR model, in which simulation is used as an alternative means of checking the conditions. Section 2.7 then offers some generalizations of the approach for smooth nonlinear autoregressive forms.

Second, in Section 3, it is shown that under linear process assumptions, the previously established sufficient conditions for the FCLT can be weakened quite dramatically, well beyond what the discussion of the case (1.3) above might suggest. In particular the result would apply, under mild summability assumptions, to a linear process driven by any of the sequences discussed in Section 2. This result is of particular interest, because it appears to get close to defining necessary memory conditions for the FCLT.

Section 4 concludes the paper, and the main proofs are gathered in the Appendix.

## 2 Applications of the NED Approach

The NED property offers an approach to proving the FCLT for a range of time series models, since there is hope that the  $L_p$  norm in (1.2), for the case  $p = 2$ , can be evaluated and bounded in specific cases. In many cases of practical interest,  $u_t$  is specified as an independent shock process, and then Theorem 1.2 offers more generality than necessary. The NED numbers fully determine

the restrictions on the memory of the observed process. In the examples that follow we always assume independent shocks, although this is principally for tractability of the derivations.

As in most dynamic econometric models, the observed process  $x_t$  is made to depend only on current and lagged values of  $u_t$ , and hence is  $\mathcal{F}_t$ -measurable. In particular, the operator  $E_{t-m}^{t+m}$  is the same as  $E_{t-m}^t$ . Another feature of the results that follow is that, because it specifically bounds a second moment, Assumption 1' yields Assumption 2 as a by-product. We do not find the former without the latter. This generalizes the familiar fact in linear processes, that a memory limitation is implicit in the property of wide-sense stationarity. These two properties are often confused, and we should emphasize that the coincidence need not obtain in more general classes of model than the ones considered here.

## 2.1 ARMA Models

This case is well known but we describe it for completeness, and because the techniques illuminate the treatment of the nonlinear cases. Let the ARMA( $p, q$ ) model be

$$x_t = \sum_{j=1}^p \lambda_j x_{t-j} + u_t + \sum_{j=1}^q \theta_j u_{t-j}. \quad (2.1)$$

where  $u_t \sim iid(0, \sigma^2)$ . Set  $q = p$ , for simplicity, and without loss of generality since the excess terms can be set to zero as required, and write the model in companion form. Letting  $\mathbf{x}_t = (x_t, x_{t-1}, \dots, x_{t-p+1})'$  and  $\mathbf{u}_t = (u_t, u_{t-1}, \dots, u_{t-p+1})'$ , define  $p \times p$  matrices

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & \lambda_{p-1} & \lambda_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \mathbf{\Theta} = \begin{bmatrix} \theta_1 & \cdots & \theta_{p-1} & \theta_p \\ -1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & -1 & 0 \end{bmatrix} \quad (2.2)$$

so that

$$\begin{aligned} \mathbf{x}_t &= \mathbf{\Lambda} \mathbf{x}_{t-1} + \mathbf{u}_t + \mathbf{\Theta} \mathbf{u}_{t-1} \\ &= \mathbf{\Lambda}^m \mathbf{x}_{t-m} + \mathbf{u}_t + \sum_{j=0}^{m-2} \mathbf{\Lambda}^j (\mathbf{\Lambda} + \mathbf{\Theta}) \mathbf{u}_{t-j-1} + \mathbf{\Lambda}^{m-1} \mathbf{\Theta} \mathbf{u}_{t-m} \end{aligned} \quad (2.3)$$

for any  $m \geq 2$ . The process is covariance stationary subject to the usual stability condition that  $\max_i |\mu_i| < 1$ , where the  $\mu_i$  are the eigenvalues of  $\mathbf{\Lambda}$ . Under the assumptions, we have

$$\mathbf{x}_t - E_{t-m}^{t+m} \mathbf{x}_t = \mathbf{\Lambda}^{m-p} (\mathbf{x}_{t-m+p} - E_{t-m}^{t+m} \mathbf{x}_{t-m+p}). \quad (2.4)$$

The object of interest is the  $L_2$ -norm of the first element of (2.4). Let  $\|\cdot\|_2$  with a vector argument represent the vector of the  $L_2$ -norms of the elements, and similarly let  $|\cdot|$  represent the vector of absolute values. Then, letting

$$\mathbf{c} = (1, 0, \dots, 0)' \quad (p \times 1) \quad (2.5)$$

note that

$$\begin{aligned} \|x_t - E_{t-m}^{t+m} x_t\|_2 &= \|\mathbf{c}' \mathbf{\Lambda}^{m-p} (\mathbf{x}_{t-m+p} - E_{t-m}^{t+m} \mathbf{x}_{t-m+p})\|_2 \\ &\leq 2 \|\mathbf{c}' \mathbf{\Lambda}^{m-p}\| \|\mathbf{x}_{t-m+p}\|_2 \\ &= O(\max_i |\mu_i|^m) \end{aligned} \quad (2.6)$$

where the inequality is obtained from the Minkowski and Jensen inequalities and the assumption of stationarity. The conclusion is summarised as follows.

Proposition 2.1 The covariance stationary ARMA( $p, q$ ) process is geometrically  $L_2$ -NED on the shock process with respect to constants  $d_t = 1$ .

A point to note is that the stated condition suffices to satisfy Assumptions 1' and 2. To ensure condition 3 holds requires the additional stipulation of invertibility, excluding unit roots in the moving average component.

A feature of the NED approach is the ease of generalization to allow for heterogeneous marginal distributions. The above argument is straightforwardly modified for the case where  $u_t \sim ihd(0, \sigma_t^2)$ , denoting that the random variables are independent but heterogeneously distributed, with time dependent variances in particular. If for example  $\sigma_t^2 = O(t^\alpha)$  for  $\alpha > 0$ , it is easily verified that the foregoing proposition holds with  $d_t = Ct^{\alpha/2}$ , for some  $C > 0$ . While these conditions can be shown to be compatible with a central limit theorem, they do not lead to a FCLT with regular Brownian motion as the limit process.

For the sake of simplicity, the analysis following focuses on stationary processes by assuming the driving processes to be i.i.d. Following the current literature, the models to be considered are mainly varieties of nonlinear finite-order difference equation. Since their solution involves iterated products of random variables, it is not surprising to find that they mostly share the geometric  $L_2$ -NED property with the linear ARMA form, subject to conditions sufficient for them to possess bounded  $L_2$  norms. These are constant under stationarity of the driving process, but in each case that follows, note how the assumption can be relaxed to the i.h.d. case, leading, given the existence of the requisite marginal moments, to modified results allowing  $d_t$  to be a function of time.

## 2.2 Bilinear Models

The books of Priestley (1988) and Tong (1990) are leading references on bilinear models. For a recent econometric application see Davidson and Peel (1998). Chapter 4.1 of Priestley (1988) analyses the so-called BL( $p, 0, p, 1$ ) model, which is the case with  $m = p$  of

$$x_t = \sum_{j=1}^p \lambda_j x_{t-j} + \sum_{j=1}^m \beta_j x_{t-j} u_{t-1} + u_t. \quad (2.7)$$

This in turn represents a sub-class of the class of BL( $p, r, m, k$ ) models defined by Subba Rao (1981), which involve terms in  $x_{t-j} u_{t-i}$  for  $j = 1, \dots, m$  and  $i = 2, \dots, k$ , and  $u_{t-j}$  for  $j = 0, \dots, r$ . The setting of  $m = p$  and  $r = 0$  sacrifices effectively no generality, since the MA coefficients are not involved in the NED analysis, similarly to the ARMA case.

Following Priestley, let  $\mathbf{A}$  be defined as in (2.2) and let

$$\mathbf{B} = \begin{bmatrix} \beta_1 & \cdots & \beta_{p-1} & \beta_p \\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix} \quad (p \times p) \quad (2.8)$$

and  $\mathbf{C}$  be defined by (2.5). Defining  $\mathbf{x}_t$  ( $p \times 1$ ) as before, the companion form of the model is

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{x}_{t-1}u_{t-1} + \mathbf{C}u_t. \quad (2.9)$$

Define the  $p \times p$  matrices

$$\mathbf{\Sigma} = E(\mathbf{x}_t \mathbf{x}_t') \quad \mathbf{\Xi} = E(u_t \mathbf{x}_t \mathbf{x}_t') \quad \mathbf{\Phi} = E(u_t^2 \mathbf{x}_t \mathbf{x}_t')$$

and let  $E(u_t^4) = \mu_4$ . While not necessarily assuming Gaussianity, let  $E u_t^3 = 0$ . This simplifies the expressions following, specifically the form of the matrix  $\mathbf{P}$ , but does not affect the nature of the solution and the conditions for its existence. Then, straightforward though still tedious manipulation shows that  $\Xi = \mathbf{K} + \mathbf{K}'$  where

$$\mathbf{K} = \sigma^4(\mathbf{I} - \mathbf{\Lambda})^{-1}\mathbf{B}\mathbf{c}\mathbf{c}' \quad (2.10)$$

and

$$\text{Vec } \mathbf{\Sigma} = (\mathbf{I} - \mathbf{\Lambda} \otimes \mathbf{\Lambda} - \sigma^2 \mathbf{B} \otimes \mathbf{B})^{-1}(\text{Vec } \mathbf{P} + (\mu_4 - \sigma^4)\mathbf{B} \otimes \mathbf{B} \text{Vec } \mathbf{c}\mathbf{c}') \quad (2.11)$$

$$\text{Vec } \mathbf{\Phi} = (\mathbf{I} - \mathbf{\Lambda} \otimes \mathbf{\Lambda} - \sigma^2 \mathbf{B} \otimes \mathbf{B})^{-1}(\sigma^2 \text{Vec } \mathbf{P} + (\mu_4 - \sigma^4)(\mathbf{I} - \mathbf{\Lambda} \otimes \mathbf{\Lambda}) \text{Vec } \mathbf{c}\mathbf{c}') \quad (2.12)$$

where

$$\mathbf{P} = \mathbf{\Lambda}\mathbf{\Xi}\mathbf{B}' + \mathbf{B}\mathbf{\Xi}\mathbf{\Lambda}' + \sigma^2\mathbf{c}\mathbf{c}' \quad (2.13)$$

subject to the stability condition represented by  $\max_i |\mu_i| < 1$ , where  $\mu_i$  are the eigenvalues of  $\mathbf{\Lambda} \otimes \mathbf{\Lambda} + \sigma^2 \mathbf{B} \otimes \mathbf{B}$ .

Letting  $\Psi_t(j) = \prod_{i=1}^j (\mathbf{\Lambda} + \mathbf{B}u_{t-i})$  ( $p \times p$ ) and  $\mathbf{w}_t = (\mathbf{\Lambda} + \mathbf{B}u_t)\mathbf{x}_t$  ( $p \times 1$ ), note that

$$\mathbf{x}_t - E_{t-m}^{t+m}\mathbf{x}_t = \Psi_t(m)(\mathbf{w}_{t-m-1} - E_{t-m}^{t+m}\mathbf{w}_{t-m-1}) \quad (2.14)$$

where

$$E(\mathbf{w}_t\mathbf{w}_t') = \mathbf{\Lambda}\mathbf{\Sigma}\mathbf{\Lambda}' + \mathbf{\Lambda}\mathbf{\Xi}\mathbf{B}' + \mathbf{B}\mathbf{\Xi}\mathbf{\Lambda}' + \mathbf{B}\mathbf{\Phi}\mathbf{B}' < \infty \quad (2.15)$$

and by the independence of the terms,

$$\text{Vec } E[\Psi_t(j)\mathbf{c}\mathbf{c}'\Psi_t(j)'] = (\mathbf{\Lambda} \otimes \mathbf{\Lambda} + \sigma^2 \mathbf{B} \otimes \mathbf{B})^j \text{Vec } \mathbf{c}\mathbf{c}'. \quad (2.16)$$

Since  $\Psi_t(m)$  and  $\mathbf{w}_{t-m-1}$  are also independent,

$$\begin{aligned} \|x_t - E_{t-m}^{t+m}x_t\|_2 &= \|\mathbf{c}'\Psi_t(m)(\mathbf{w}_{t-m-1} - E_{t-m}^{t+m}\mathbf{w}_{t-m-1})\|_2 \\ &\leq 2 \text{tr}[E(\Psi_t(m)\mathbf{c}\mathbf{c}'\Psi_t(m)')E(\mathbf{w}_{t-m-1}\mathbf{w}_{t-m-1}')^{1/2}] \\ &= O(\max_i |\mu_i|^{m/2}) \end{aligned} \quad (2.17)$$

according to (2.16), using the Jensen inequality similarly to (2.6). The conclusion may be stated as

**Proposition 2.2** The covariance stationary BL( $p, 0, p, 1$ ) process is geometrically  $L_2$ -NED on the shock process with respect to constants  $d_t = 1$ .

Extending this result to BL( $p, r, m, 1$ ) is trivial, for reasons already remarked. Moving to the general BL( $p, r, m, k$ ) is not trivial, however, both because of the complexity of the relevant moment expressions, and because these involve moments of the  $u_t$  of higher order, depending on  $k$ . Thus, consider the expression corresponding to (2.16) when  $\Psi_t(j) = \prod_{i=1}^j (\mathbf{\Lambda} + \sum_{l=0}^{k-1} \mathbf{B}_i u_{t-i-l})$ . However, subject to the existence of the required moments, parameter restrictions for the geometric  $L_2$ -NED property could in principle be derived by elaboration of the above arguments.

### 2.3 GARCH Models

On GARCH models of persistent volatility see, for example, Bollerslev (1986), and Engle (1995). In this model

$$u_t = h_t^{1/2} \varepsilon_t \quad (2.18)$$

where  $\varepsilon_t \sim iid(0, 1)$  and  $h_t$  is an  $\mathcal{F}_{t-1}$ -measurable process. Hansen (1991) has derived conditions for NED in the GARCH(1,1) case. In the GARCH( $p, p$ ) model that we consider here, without loss of generality,

$$\begin{aligned} h_t &= \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \\ &= \alpha_0 + \sum_{i=1}^p [\alpha_i (\varepsilon_{t-i}^2 - 1) + \delta_i] h_{t-i} \end{aligned} \quad (2.19)$$

where  $\delta_i = \alpha_i + \beta_i$ . The following lemma holds for any case of (2.18) in which  $h_t$  is bounded below by a constant  $\alpha_0 > 0$ . (All proofs are given in the Appendix.)

Lemma 2.1  $\|u_t - E_{t-m}^{t+m} u_t\|_2 \leq \alpha_0^{-1/2} \|h_t - E_{t-m}^{t+m} h_t\|_2$ .

It therefore suffices to show that  $h_t$  is  $L_2$ -NED on  $\{\varepsilon_t\}$ , and as with the ARMA and BL models the key step is to express the process as the sum of a  $\mathcal{F}_{t-m}^{t+m}$ -measurable term and a remainder. The companion form of the model is

$$\begin{aligned} \mathbf{h}_t &= \mathbf{c} \alpha_0 + (\mathbf{\Delta} + \mathbf{A} \mathbf{S}_{t-1}) \mathbf{h}_{t-1} \\ &= \mathbf{c} \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} \prod_{k=1}^j (\mathbf{A} \mathbf{S}_{t-k} + \mathbf{\Delta}) \mathbf{c} \end{aligned} \quad (2.20)$$

where  $\mathbf{h}_t = (h_t, \dots, h_{t-p+1})'$  and  $\mathbf{S}_t = \text{diag}\{\varepsilon_t^2 - 1, \dots, \varepsilon_{t-p+1}^2 - 1\}$ . Here,  $\mathbf{c}$  is as in (2.5) and  $\mathbf{\Delta}$  and  $\mathbf{A}$  have the same structure as  $\mathbf{\Lambda}$  in (2.2) and  $\mathbf{B}$  in (2.8) respectively, with the  $\delta_i$  and  $\alpha_i$ , respectively, in their top rows. Note the condition for covariance stationarity, that  $\mathbf{\Delta}$  has all its eigenvalues within the unit circle such that

$$E(\mathbf{h}_t) = \alpha_0 (\mathbf{I} - \mathbf{\Delta})^{-1} \mathbf{c}. \quad (2.21)$$

Let  $\mathbf{\Upsilon}_t(m) = \alpha_0 \prod_{k=1}^{m-p+1} (\mathbf{A} \mathbf{S}_{t-k} + \mathbf{\Delta})$ . Noting that  $\mathbf{A} \mathbf{S}_s + \mathbf{\Delta}$  is  $\mathcal{F}_{t-m}^{t+m}$ -measurable for  $s = t-1, \dots, t-m+p$ ,

$$\mathbf{h}_t - E_{t-m}^{t+m} \mathbf{h}_t = \mathbf{\Upsilon}_t(m) (\mathbf{h}_{t-m+p-1} - E_{t-m}^{t+m} \mathbf{h}_{t-m+p-1}). \quad (2.22)$$

Suppose that  $h_t$  is bounded in  $L_2$ -norm, or equivalently, that the process  $x_t$  has bounded fourth moments. Then

$$\begin{aligned} \|h_t - E_{t-m}^{t+m} h_t\|_2 &= \|\mathbf{c}' \mathbf{\Upsilon}_t(m) (\mathbf{h}_{t-m+p-1} - E_{t-m}^{t+m} \mathbf{h}_{t-m+p-1})\|_2 \\ &\leq \sum_{j=1}^p \|\{\mathbf{\Upsilon}_t(m)\}_{1j} (h_{t-m+p-j} - E_{t-m}^{t+m} h_{t-m+p-j})\|_2 \\ &\leq 2 \sum_{j=1}^p \|\{\mathbf{\Upsilon}_t(m)\}_{1j}\|_2 \|h_{t-m+p-j}\|_2 \end{aligned} \quad (2.23)$$



where the first inequality is Minkowski's. The second uses the fact, easily verified, that  $h_{t-m+p-j}$  is independent of  $\{\Upsilon_t(m-p+1)\}_{1j}$  by assumption, and then Jensen's inequality, as before.

To obtain the conditions for  $\|h_t\|_2 < \infty$ , not depending on  $t$ , let

$$\Sigma = E(\mathbf{h}_t \mathbf{h}_t') \quad \Xi = E(\mathbf{h}_t \mathbf{h}_t' S_t)$$

and note that

$$E(S_t \mathbf{h}_t \mathbf{h}_t' S_t) = \mu_4 \text{dg } \Sigma$$

where  $\mu_4 = E\varepsilon_t^4 - 1$ , the variance of  $\varepsilon_t^2 - 1$ , and  $\text{dg } \Sigma$  represents the  $p \times p$  diagonal matrix having the same diagonal as  $\Sigma$ . Using the first equality in (2.20) to substitute for  $\mathbf{h}_t$  and taking expectations yields

$$\begin{aligned} \Sigma &= \Delta \Sigma \Delta' + \mu_4 \mathbf{A} \text{dg } \Sigma \mathbf{A}' + \Delta \Xi \mathbf{A}' + \mathbf{A} \Xi' \Delta' + \alpha_0^2 \mathbf{F} \\ \Xi &= \Delta \Xi \mathbf{J}' + \mu_4 \mathbf{A} \text{dg } \Sigma \mathbf{J}' \end{aligned} \quad (2.24)$$

where

$$\mathbf{J} = \begin{bmatrix} \mathbf{0}' & 0 \\ \mathbf{I}_{p-1} & \mathbf{0} \end{bmatrix} \quad (p \times p).$$

and

$$\mathbf{F} = \mathbf{c} \mathbf{c}' + \Delta (\mathbf{I} - \Delta)^{-1} \mathbf{c} \mathbf{c}' + \mathbf{c} \mathbf{c}' (\mathbf{I} - \Delta')^{-1} \Delta'.$$

Define the  $p^2 \times p^2$  permutation matrix  $\mathbf{P}$ , such that  $\text{Vec}(\Xi') = \mathbf{P} \text{Vec } \Xi$ , and the  $p^2 \times p^2$  deletion matrix  $\mathbf{D}$ , such that  $\text{Vec}(\text{dg } \Sigma) = \mathbf{D} \text{Vec } \Sigma$ . Then the stable solution of (2.24) for  $\Sigma$  is given by

$$\text{Vec } \Sigma = \alpha_0^2 (\mathbf{I} - \mathbf{M})^{-1} \text{Vec } \mathbf{F}$$

where

$$\mathbf{M} = \Delta \otimes \Delta + \mu_4 (\mathbf{A} \otimes \mathbf{A}) \mathbf{D} + ((\mathbf{A} \otimes \Delta) + (\Delta \otimes \mathbf{A}) \mathbf{P}) [\mathbf{I} - (\mathbf{J} \otimes \Delta)]^{-1} \mu_4 (\mathbf{J} \otimes \mathbf{A}) \mathbf{D}. \quad (2.25)$$

Accordingly, the condition for existence of the fourth order moments of the GARCH( $p, p$ ) (and more generally the GARCH( $p, q$ ) by setting the excess terms to 0 in formulae) is that the eigenvalues of both  $\Delta$  and  $\mathbf{M}$  all lie strictly inside the unit circle. Subject to this condition it is clear, in view of the geometric rates of convergence of the sequences of cross moments, easily deduced from (2.20), that the process is fourth-order stationary in the sense of, e.g., Hannan (1970, page 209). Since the eigenvalue condition is necessary as well as sufficient, we may refer to it as the fourth-order stationarity condition. Note that to obtain this condition, we have invoked parameter restrictions plus the finiteness of the fourth moment of the i.i.d. driving process  $\varepsilon_t$ , but not Gaussianity of the latter process.

Moreover, it is clear from consideration of the second equality of (2.20) that  $E(\Upsilon_t(m) \Upsilon_t(m)')$  is the  $m$ th term in the expansion of  $\Sigma$  in geometric series, and hence that  $\|\{\Upsilon_t(m)\}_{1j}\|_2 = O(\max_i |\mu_i|^{m/2})$ , where the  $\mu_i$  are the eigenvalues of  $\mathbf{M}$  in (2.25). The conclusions of this section are therefore summarised as follows.

**Proposition 2.3** A fourth-order stationary GARCH( $p, q$ ) process is geometrically  $L_2$ -NED on the underlying i.i.d. process, with respect to constants  $d_t = 1$ .

A further point to notice here is that the argument can be adapted to prove that  $x_t$  is geometrically  $L_1$ -NED, subject only to the covariance stationarity condition. This is sufficient to show that, for example, the sample mean of the process converges in probability. However, it does not suffice for the FCLT.

The pure GARCH process is of course a fourth-order stationary martingale difference on the assumptions, and Assumption 1' is therefore not directly required for the proof of the FCLT, although it suffices. However, FCLTs for martingale differences call for a weak law of large numbers to hold in the squares of the process (Davidson 1994, Theorem 24.3). For this purpose, the  $L_1$ -NED property of the squares can be invoked (Davidson 1994, Theorem 17.9), and Proposition 2.3 establishes fourth-order stationarity as a sufficient condition.

The result may be also extended by, for example, letting the process drive an ARMA. Let  $x_t$  in (2.1) have the MA( $\infty$ ) representation

$$x_t = \sum_{j=0}^{\infty} \rho_j u_{t-j} \quad |\rho_j| = O(|\mu_A|^j) \quad (2.26)$$

where  $\mu_A$  is the absolutely largest eigenvalue of the companion form in (2.3). Also let  $u_t = h_t^{1/2} \varepsilon_t$  where  $\varepsilon_t \sim iid(0,1)$  and  $h_t$  is generated by (2.19) and let  $\mu_G$  denote the absolutely largest eigenvalue of  $M$  in (2.25). Then, combining previous results with Minkowski's inequality yields

$$\begin{aligned} \|x_t - E_{t-m}^{t+m} x_t\|_2 &= \|c'(x_t - E_{t-m}^{t+m} x_t)\|_2 \\ &\leq \|c'(u_t - E_{t-m}^{t+m} u_t)\|_2 \\ &\quad + \sum_{j=0}^{m-p} \|c' \Lambda^j (\Lambda + \Theta)(u_{t-j-1} - E_{t-m}^{t+m} u_{t-j-1})\|_2 \\ &\quad + \sum_{j=m-p-1}^{\infty} \|c' \Lambda^j (\Lambda + \Theta) u_{t-j-1}\|_2 \\ &= O\left(\sum_{j=0}^{m-p} |\mu_A|^j |\mu_G|^{m-p-j} + \sum_{j=m-p-1}^{\infty} |\mu_A|^j\right) \\ &= O((m-p+1) \max\{|\mu_A|^{m-p}, |\mu_G|^{m-p}\}). \end{aligned} \quad (2.27)$$

Since the exponential term in the final member of (2.27) dominates, the conclusion is as follows.

**Proposition 2.4** A stable ARMA driven by a fourth-order stationary GARCH process is geometrically  $L_2$ -NED on the underlying i.i.d. process, with respect to constants  $d_t = 1$ .

However, it may be noted that a more general result is obtained by combining the GARCH result with Theorem 3.1 to be introduced below, since under the assumptions, the driving GARCH process satisfies the conditions imposed on  $x_t$  in (3.1).

## 2.4 Switching and Threshold Autoregressions

A wide class of nonlinear autoregressive models take the general form

$$x_t = \sum_{j=1}^p \phi_{j,t-1} x_{t-j} + u_t \quad (2.28)$$

where  $u_t \sim iid(0, \sigma^2)$  and  $\phi_{jt}$  is a  $\mathcal{F}_t$ -measurable random variable that in general depends on  $x_s$  for  $s \leq t$ . The bilinear models already considered are members of this class. However, note that

$\mathcal{F}_t$  can be larger than the natural filtration  $\sigma(x_s, s \leq t)$ . As a referee points out, it is one of the strengths of the NED approach that such models can be easily treated.

An important set of cases is defined by

$$\phi_{jt} = \sum_{i=1}^N \lambda_{ji} I_t^i \quad (2.29)$$

where the  $\lambda_{ji}$  for  $i = 1, \dots, N$  and  $j = 1, \dots, p$  are coefficients, and the  $I_t^i$  are  $\mathcal{F}_t$ -measurable indicator functions of which one equals unity, and the rest zero. Models of this type include the self-exciting threshold autoregression (SETAR) models, in which  $I_t^i$  is the indicator of  $x_t$  itself falling in a particular interval (see e.g. Priestley 1988 Chapter 4.2, Tong 1990 Chapter 3.3, Granger and Teräsvirta 1993 Chapter 4.1) and the Markov-switching autoregression (Hamilton 1994, Chapter 22.4) in which the  $I_t^i$  constitute an independently generated Markov chain. A third group are the smooth transition autoregressive (STAR) models, in which the indicator functions are replaced by smooth functions of the data having the unit interval as range, such as distribution functions or logistic forms.

The companion form for these models is

$$\mathbf{x}_t = \mathbf{\Phi}_{t-1} \mathbf{x}_{t-1} + \mathbf{C} u_t \quad (2.30)$$

where  $\mathbf{C}$  was defined in (2.5) and

$$\mathbf{\Phi}_{t-1} = \begin{bmatrix} \phi_{1,t-1} & \cdots & \phi_{p-1,t-1} & \phi_{p,t-1} \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}.$$

The usual expansion yields, for  $m > p$ ,

$$\mathbf{x}_t = \sum_{j=1}^{m-p} \prod_{k=1}^j \mathbf{\Phi}_{t-k} \mathbf{C} u_{t-j} + \prod_{k=1}^{m-p+1} \mathbf{\Phi}_{t-k} \mathbf{x}_{t-m+p-1} + \mathbf{C} u_t. \quad (2.31)$$

Using the Minkowski, Jensen and Hölder inequalities, we have for some  $r \geq 1$ ,<sup>4</sup>

$$\begin{aligned} \|x_t - E_{t-m}^{t+m} x_t\|_2 &= \left\| \mathbf{C}' \prod_{k=1}^{m-p+1} \mathbf{\Phi}_{t-k} \mathbf{x}_{t-m+p-1} - \mathbf{C}' \prod_{k=1}^{m-p+1} \mathbf{\Phi}_{t-k} E_{t-m}^{t+m}(\mathbf{x}_{t-m+p-1}) \right\|_2 \\ &\leq 2 \left\| \mathbf{C}' \prod_{k=1}^{m-p+1} \mathbf{\Phi}_{t-k} \right\|_{2r/(r-1)} \|\mathbf{x}_{t-m+p-1}\|_{2r}. \end{aligned} \quad (2.32)$$

In the case  $r = 1$ , the first factor of the majorant of (2.32) is interpreted as the sup-norm of the argument.

For the case of (2.29),

$$\mathbf{\Phi}_t = \sum_{i=1}^N I_t^i \mathbf{\Lambda}_i$$

---

<sup>4</sup>As previously, the  $L_2$ -norm with vector argument is interpreted as the vector of  $L_2$ -norms.

where the  $\mathbf{\Lambda}_i$  are defined as in (2.3), and

$$\left\| \mathbf{c}' \prod_{k=1}^{m-p+1} \mathbf{\Phi}_{t-k} \right\|_{\infty} = O(|\mu^*|^{m-p+1})$$

where  $|\mu^*|$  denotes the largest modulus of the eigenvalues of the set  $\mathbf{\Lambda}_1, \dots, \mathbf{\Lambda}_N$ . If  $|\mu^*| < 1$  then the eigenvalues of  $\mathbf{\Phi}_t$  are stable with probability 1, and the TAR is accordingly covariance stationary. Letting  $\mathbf{\Lambda}^*$  denote the member of the set to which  $\mu^*$  corresponds, we have

$$\begin{aligned} E(x_t^2) &= \sigma^2 + \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \mathbf{c}' E \prod_{k=1}^j \prod_{i=1}^l \mathbf{\Phi}_{t-k} \mathbf{c} \mathbf{c}' \mathbf{\Phi}'_{t-i} \mathbf{c} u_{t-j} u_{t-i} \\ &\leq \sigma^2 + \sigma^2 \left( \sum_{j=1}^{\infty} \mathbf{c}' \mathbf{\Lambda}^{*j} \mathbf{c} \right)^2 < \infty. \end{aligned}$$

Chan and Tong (1985) and Tong (1990) develop the concept of geometric ergodicity as a restriction on the behaviour of nonlinear stochastic difference equations like the above. In stationary processes, this restriction implies strong mixing at the geometric rate so that the conditions of Theorem 1.1 are satisfied. It is therefore worth emphasising the advantages conveyed by the NED approach in these cases, in addition to the relative ease of establishing the property already demonstrated, and the powerful asymptotic results available. The extension to nonstationary cases, featuring trending moments of the shock processes for example, has already been indicated. However, another important gain in generality is that NED depends in no way on the distribution of the shocks, beyond the existence of the requisite moments. By contrast, establishing geometric ergodicity in a typical application of the above type requires the shock distribution to be continuous.

## 2.5 Unit Root SETAR Models

A more interesting case is where maximum eigenvalues of unity are admissible with positive probability, so that the process behaves under certain conditions as a unit root process. Where stability with probability 1 is not available, the case  $r > 1$  in (2.32) may be considered, requiring the process to have finite  $L_{2r}$  norm. Then we must consider the behaviour of the  $m$ -fold product in the majorant as  $m$  increases.

A simple first-order example of the self-exciting type will be analysed. Let

$$x_t = \begin{cases} x_{t-1} + u_t, & |x_{t-1}| < a \\ \lambda x_{t-1} + u_t, & \text{otherwise} \end{cases} \quad (2.33)$$

for  $0 \leq \lambda < 1$  and  $a > 0$ , where  $u_t \sim iid(0, \sigma^2)$ , and assume further that  $u_t$  is continuous, the p.d.f. having finite second derivative at 0, and  $E|u_t|^{2r} < \infty$  for  $r > 1$ . It could for example be Gaussian. Models of this type can arise in the theory of exchange rate bands, and Ss-type inventory models. The following result is proved in the Appendix.

**Proposition 2.5** Under the stated assumptions,  $x_t$  in (2.33) is geometrically  $L_2$ -NED on  $u_t$  with

$$\nu(m) = O \left( \left( g(a) + (1 - g(a)) \lambda^{2r/(r-1)} \right)^{m(r-1)/2r} \right) \quad (2.34)$$

where, letting  $f$  denote the p.d.f. of  $u_t$ ,

$$g(a) = (2a)^{-1} \int_{-a}^a \int_{-a}^a f(x-y) dx dy \quad (2.35)$$

and  $g(a) < 1$ .

Figure 1 plots the function  $g(a)$  for the case where  $f$  is the standard Gaussian. Application of

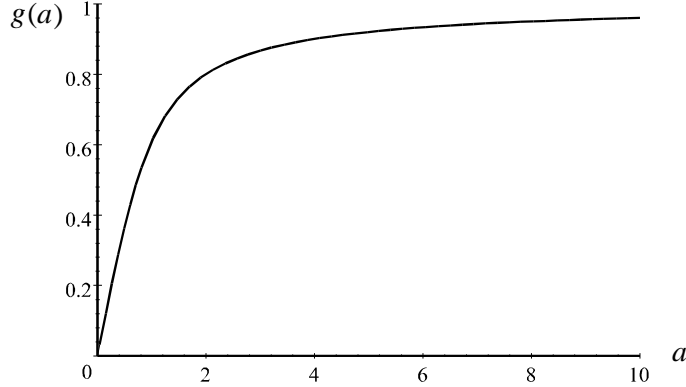


Figure 1: The standard Gaussian case.

(2.34) shows how the memory of the process is bounded depending on the width of the band, measured in units of the standard deviation of the shock process. A large value of  $r$  is evidently necessary if the rate of decline of  $\nu(m)$  is not to be very slow even for relatively small values of  $a$ , and hence formula (2.34) is most useful for cases where all moments exist. Otherwise, the inequality in (2.32) could suffer a lack of sharpness, such that with  $r$  close to 1, the dependence might in practice decline faster than the derived bound indicates.

## 2.6 The ESTAR Model

A close relative of the last example is the exponential smooth transition autoregression (ESTAR) taking the form

$$x_t = e^{-\beta x_{t-1}^2} x_{t-1} + u_t. \quad (2.36)$$

where  $\beta > 0$  is a parameter and, as usual,  $u_t \sim iid(0, \sigma^2)$  and  $E|u_t|^{2r} < \infty$  for  $r > 1$ . This is another way to implement the idea of a process which has random walk characteristics in the mid-range, and mean reversion properties at extreme values, but in this case the transition is smooth. The autoregressive coefficient  $\phi_{t-1} = e^{-\beta x_{t-1}^2}$ , corresponding to  $\Phi_{t-1}$  in (2.30) with  $p = 1$ , is equal to 1 only with probability 0.

Analytic solution of the difference equation is by no means straightforward in this type of model, and no simple way to evaluate the factors of (2.32) suggests itself. However, it is easy to compute the relevant moments by simulation. The results of this exercise, using standard Gaussian disturbances and a sample size of 100,000, are reported in Table 1. In the case of  $\left\| \prod_{k=1}^{m+1} \phi_{t-k} \right\|_{2r/(r-1)}$ , the norms have been estimated for  $m = 1, \dots, 100$  and the logarithm of this series has been regressed on  $m$ .<sup>5</sup> The  $R^2$ s in these regressions are all greater than 0.97 and exceed 0.999 for the smaller values of  $\beta$ , so the exponential approximation is good. The antilogarithms of the slope coefficients appear as the left-hand factors in each cell of the table, raised to power  $m$ , while the right-hand factors are the estimated  $L_{2r}$ -norms of the process. In contrast to the threshold model, there is no serious penalty here in setting  $r$  close to unity. The difference is

---

<sup>5</sup>The norm may be too small for the logarithm to be computed for the larger values of  $m$ , and in such cases the number of lags is truncated.

	$\beta = 1$	$\beta = 0.1$	$\beta = 0.01$	$\beta = 0.001$
$r = 1.1$	$0.565^m \times 1.08$	$0.904^m \times 1.40$	$0.981^m \times 2.30$	$0.995^m \times 4.01$
$r = 50$	$0.499^m \times 4.09$	$0.876^m \times 4.81$	$0.963^m \times 6.85$	$0.988^m \times 10.87$

Table 1: Numerical evaluation of bound (2.32) in the ESTAR Model. The factor 2 has been omitted.

that in the former model  $\phi_t = 1$  with positive probability, so that taking a large power does not shrink it, but that does not occur in this case.

## 2.7 Smooth Nonlinear Autoregressions

Finally, consider methods for tackling a general class of models which encompasses some of those discussed above. These have the form

$$x_t = f(u_t, x_{t-1}) \quad (2.37)$$

where  $f$  is differentiable at least with respect to its second argument. Assume, as previously, that  $u_t \sim iid(0, \sigma^2)$ , with  $\|u_t\|_{2r} = B < \infty$  for  $r > 1$ . The approach here is to consider the recursive solution

$$\begin{aligned} x_t &= f(u_t, f(u_{t-1}, f(u_{t-2}, \dots))) \\ &= g(u_t, u_{t-1}, u_{t-2}, \dots) \end{aligned} \quad (2.38)$$

Denoting the latter function by  $g_t$ , also define  $g_t^m = g(u_t, \dots, u_{t-m}, 0, 0, \dots)$ . Then note that

$$g_t - g_t^m = \sum_{j=m+1}^{\infty} G_{t-j} u_{t-j} \quad (2.39)$$

by the mean value theorem, where using the chain rule of differentiation and (2.38), we can write

$$G_{t-j} = \left( \frac{\partial g_t}{\partial u_{t-j}} \right)^* = \prod_{k=1}^m f_{2,t-k}.$$

Here the ‘\*’ denotes that the derivative is evaluated at points  $u_{t-j}^* \in [0, u_{t-j}]$ , and

$$f_{2,t-k} = f_2(u_{t-k}, g(u_{t-k-1}, \dots, u_{t-m}, u_{t-m-1}^*, \dots)) \quad (2.40)$$

denotes the derivative of  $f$  with respect to its second argument, evaluated at the points  $u_{t-m-k}^*$ , for  $k \geq 0$ . Now use the fact that  $E_{t-m}^{t+m}(\cdot)$  is the minimum mean-square  $\mathcal{F}_{t-m}^{t+m}$ -measurable approximation of its argument, and then the Hölder inequality, to obtain, much as before,

$$\begin{aligned} \|g_t - E_{t-m}^{t+m} g_t\|_2 &\leq \|g_t - g_t^m\|_2 \\ &\leq B \sum_{j=m+1}^{\infty} \left\| \prod_{k=1}^m f_{2,t-k} \right\|_{2r/(r-1)} \end{aligned} \quad (2.41)$$

This case was considered by Gallant and White (1988) (also see Davidson 1994 Chapter 17.1) although subject to the stringent stability condition,  $|f_{2,t-k}| \leq b < 1$  with probability 1. In this case it is obviously easy to bound (2.41), but the restriction is stronger than strictly necessary, in the light of the earlier discussion. As an illustration of the calculations involved, one might take the ESTAR example considered above, although our approach there is preferable since it exploited the additivity of the error term. The cost of the extra generality is to introduce the infinite series into the NED bound, but since in general the terms of the sum decline exponentially (if they do decline) there is no extra penalty here. The sum from  $m$  to  $\infty$  in (2.41) declines at the same rate as the  $m$ th term.

### 3 Linear Forms

Let

$$z_t = \sum_{k=0}^{\infty} \theta_k x_{t-k} \quad (3.1)$$

and define for  $z_t$  the standardised stochastic process  $Z_n$  on  $[0, 1]$ , having the form of (1.1). The following theorem is proved in the Appendix.

**Theorem 3.1**  $Z_n \xrightarrow{d} B$  if the process  $\{x_s, -\infty < s < \infty\}$  in (3.1) satisfies Assumptions 1', 2 and 3, and the sequence  $\{\theta_j\}$  satisfies the following conditions:  $|\theta_j|$  is regularly varying at infinity,<sup>6</sup>

$$0 < \left| \sum_{j=0}^{\infty} \theta_j \right| < \infty \quad (3.2)$$

and

$$\sum_{k=0}^{\infty} \left( \sum_{j=1+k}^{n+k} \theta_j \right)^2 = o(n). \quad (3.3)$$

In other words, this result says that any process satisfying the assumptions of Theorem 1.2 can be replaced by a moving average of itself, under the specified conditions, and the weak convergence is preserved. This is a corollary to Theorem 3.1 of Davidson and de Jong (2000), which establishes the FCLT for fractionally integrated processes. It operates by the trick due to Davydov (1970), of re-ordering the partial sums of  $Z_n$  to collect terms with the same time index, so converting a dependence problem effectively into one of heteroscedasticity. In the fully general version of Theorem 1.2 given by De Jong and Davidson (2000b), the variances of the process are allowed to trend like  $t^\alpha$  for any  $\alpha > -1$ , and exploiting this fact yields the result. However, it cannot be generalized beyond the linear case. There is no way yet known of further improving general (model-independent) dependence conditions over those given in Theorem 1.2.

The novel condition here is (3.3). Considering the case  $n = 1$  shows that it implies square-summability of the coefficients. However, consideration of the following numerical lemma, proved in the Appendix, shows that it is weaker than absolute summability.

**Lemma 3.1** For  $\varepsilon > 0$ ,

$$\sum_{k=1}^{\infty} (k^{-\varepsilon} - (n+k)^{-\varepsilon})^2 = o(n).$$

If the coefficients all take the same sign, the finiteness specified in (3.2) is equivalent to absolute summability of the sequence, and condition (3.3) holds in consequence. If  $|\theta_j| = O(j^{-1-\varepsilon})$  for  $\varepsilon > 0$ , then

$$\begin{aligned} \left| \sum_{j=1+k}^{n+k} \theta_j \right| &\leq \sum_{j=1+k}^{n+k} |\theta_j| \\ &\leq C \sum_{j=1+k}^{n+k} j^{-1-\varepsilon} \end{aligned}$$

---

<sup>6</sup>That is,  $|\theta_j| = j^\rho L(j)$  for real  $\rho$  where  $L(xj)/L(j) \rightarrow 1$  as  $j \rightarrow \infty$  for all  $x > 0$ .  $L$  is called a slowly varying function, and can be arbitrary for finite  $j$ . This assumption can certainly be relaxed, by specifying suitable regularly varying bounding sequences. However, such an extension would complicate the statement of the result while adding little useful generality; in particular, note that the finite-order MA case is covered directly by Theorem 1.2.

$$\leq \frac{C}{\varepsilon} (k^{-\varepsilon} - (n+k)^{-\varepsilon}) \quad (3.4)$$

for  $C > 0$ .<sup>7</sup>

If the coefficients may differ in sign, however, condition (3.3) becomes the binding condition. Consider an example with alternating signs,  $\theta_j = (-1)^j j^{-1/2-\varepsilon}$ ,  $\varepsilon > 0$ .<sup>8</sup> This sequence is square-summable but not absolutely summable. However, defining

$$\theta_j^* = \begin{cases} \frac{1}{2}(\theta_j + \theta_{j+1}), & 1+k \leq j \leq n+k \text{ and } j-k \text{ odd} \\ \theta_{j-1}^*, & 1+k \leq j \leq n+k \text{ and } j-k \text{ even} \end{cases}$$

note that if  $n$  is even,

$$\sum_{j=1+k}^{n+k} \theta_j = \sum_{j=1+k}^{n+k} \theta_j^* \quad (3.5)$$

and it can be verified that  $\varepsilon(j+1)^{-3/2-\varepsilon} < \theta_j^* < \varepsilon j^{-3/2-\varepsilon}$ . If  $n$  is odd, the same equality holds with  $\theta_j^*$  so defined for  $j < n+k$  and  $\theta_{n+k}^* = \theta_{n+k}$ . Then, similarly to (3.4),

$$\left| \sum_{j=1+k}^{n+k} \theta_j \right| \leq C_1 (k^{-1/2-\varepsilon} - (n+k)^{-1/2-\varepsilon}) + C_2 (n+k)^{-1/2-\varepsilon} \quad (3.6)$$

for  $C_1 > 0$  and  $C_2 \geq 0$ , with  $C_2 = 0$  if  $n$  is even. Both terms on the majorant side of (3.6) satisfy the condition that their sum of squares over  $k = 1, 2, \dots$  is  $o(n)$ .

This example can be elaborated with more general patterns of sign change, such that the partial signed sums grow appropriately. Consider

$$\theta_j = j^{-1/2-\varepsilon} \cos(2\pi j/N)$$

for positive finite integer  $N$ . A construction on the lines of (3.5) is clearly possible, off-setting pairs of positive and negative terms. A sinusoidal lag distribution involving changes of sign will in general allow longer memory than a monotone one. These examples give an indication of the generality of condition (3.3), noting that the  $-\frac{1}{2}$ 's in the exponents of the first majorant term of (3.6) are actually surplus to the requirements of square-summability. However, a counter-example is provided by the case  $\theta_j = j^{-1/2-\varepsilon} \cos(2\pi j^\gamma/N)$  for  $\gamma < 1$ . In this case the period of the fluctuations is increasing with the lag, and the 'off-setting' device fails in the tail, as  $k$  increases.

An example of (3.1) in which the assumptions of Theorem 3.1 are violated is the fractionally integrated, or  $I(d)$ , model. In this case,

$$\theta_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} = O(j^{d-1}). \quad (3.7)$$

When  $d = 0$ ,  $\theta_j = 0$  for  $j > 0$  and the assumptions of Theorem 3.1 reduce to those of Theorem 1.2. The long memory case, in which  $0 < d < \frac{1}{2}$ , violates the upper bound specified in condition (3.2) because the lag coefficients are positive and non-summable. As the above-cited theorem of Davidson and de Jong (2000) shows, the limit of the partial sums is not  $B$  in this case but

---

<sup>7</sup>For simplicity, we ignore the possible role of slowly varying components here. A version of the argument can also be derived for the summable case  $|\theta_j| = O(j^{-1}(\log j)^{-1-\delta})$ ,  $\delta > 0$ .

<sup>8</sup>There is no suggestion that this is a realistic model of any observed process. It is taken simply as a tractable case, for purposes of illustration.



fractional Brownian motion, a Gaussian process having positively correlated increments. This strongly indicates that the summability condition is necessary.

The negative fractional model in which  $-\frac{1}{2} < d < 0$  also violates (3.2), for although its coefficients are summable, they sum identically to 0. The process is constructed as the simple difference of a nonstationary fractional process (with  $\frac{1}{2} < d < 1$ ). The limit of the partial sums is in this case a Gaussian process with negatively correlated increments. However, note that without the sum-to-zero property a process with  $|\theta_k| = O(k^{d-1})$  for  $d < 0$  is one of the cases covered by Theorem 3.1, and the limit process is ordinary Brownian motion.

## 4 Conclusion

The objectives of this paper have been twofold. First, it has considered the operationalization of the near-epoch dependence assumption, for establishing the FCLT conditions in nonlinear models. The condition is easily checked for a range of models, and indeed, most of the classes of model dealt with in the monographs of Priestley (1988), Tong (1990) and Granger and Teräsvirta (1993) are covered by our results. Second, it has presented a new FCLT for linear processes driven by dependent sequences of random variables, whose conditions appear close to necessity for this class of models. It is known that different limits obtain when the summability conditions are violated.

The one limiting feature of the former analysis has been the need to assume that the shock processes driving these models are independent. It has been pointed out that relaxing the stationarity assumption for the shock processes is generally rather trivial, but the analysis has used the independence in a crucial way to simplify the derivations. This is a limitation, in the sense that the theory itself allows the underlying process to exhibit unspecified local dependence, subject to a mixing condition. The difficulty in exploiting this generalization lies simply in the greater difficulty of establishing the NED property. Of course, it can be argued that modelling nonlinear dependence reduces the need to allow for unspecified dependence, and Proposition 2.4 illustrates this point. Extending the class of switching and threshold models to allow for conditional heteroscedasticity represents one of several avenues for future research, and the conjecture that geometric  $L_2$ -NED properties hold under comparable assumptions is a plausible one.

## A Appendix

### Proof of Lemma 2.1

First, noting that  $u_t$  and  $h_t$  are independent and  $u_t$  is measurable with respect to  $\mathcal{F}_{t-m}^{t+m}$  for  $m \geq 0$ ,

$$\|x_t - E_{t-m}^{t+m} x_t\|_2 = \|h_t^{1/2} - E_{t-m}^{t+m} h_t^{1/2}\|_2. \quad (\text{A-1})$$

Next, since  $h_t$  and  $E_{t-m}^{t+m} h_t$  are both bounded below by  $\alpha_0$  we have

$$\left| h_t^{1/2} - (E_{t-m}^{t+m} h_t)^{1/2} \right| \leq \alpha_0^{-1/2} |h_t - E_{t-m}^{t+m} h_t| \quad \text{a.s.}$$

and hence

$$\|h_t^{1/2} - (E_{t-m}^{t+m} h_t)^{1/2}\|_2 \leq \alpha_0^{-1/2} \|h_t - E_{t-m}^{t+m} h_t\|_2. \quad (\text{A-2})$$

Also note that

$$\|h_t^{1/2} - E_{t-m}^{t+m} h_t^{1/2}\|_2 \leq \|h_t^{1/2} - (E_{t-m}^{t+m} h_t)^{1/2}\|_2 \quad (\text{A-3})$$

in view of the fact that  $E_{t-m}^{t+m}(\cdot)$  is the minimum MSE,  $\mathcal{F}_{t-m}^{t+m}$ -measurable approximation of its argument. Combining (A-2), (A-3) and (A-1) yields the result. ■

## Proof of Proposition 2.5

We evaluate the bound in (2.32). First verify that  $\|x_t\|_{2r} < \infty$  for  $r > 1$ . Let

$$x_{1t} = x_t I(|x_{t-1}| < a)$$

so that  $\|x_{1t}\|_{2r} \leq a$ . Then note that  $\|x_t\|_{2r} \leq \|x_{1t}\|_{2r} + \|x_t - x_{1t}\|_{2r}$ , using Minkowski's inequality, where

$$\begin{aligned} \|x_t - x_{1t}\|_{2r} &\leq \lambda \|x_{t-1}\|_{2r} + \|u_t\|_{2r} \\ &\leq \sum_{j=0}^{m-1} \lambda^j \|u_{t-j}\|_{2r} + a\lambda^m \\ &< \infty \end{aligned}$$

where time  $t - m$  is the most recent occasion prior to  $t$  in which the process is contained in the band, and the last inequality holds for any  $m$ , finite or infinite.

Next, note that  $\prod_{k=1}^m \phi_{t-k}$  is a random variable whose range consists of the values  $\lambda^j$  for integers  $j = 0, \dots, m$ , depending on the number of occasions that the band is breached in time periods  $t - m$  to  $t - 1$ . If  $f_n(\cdot)$  denotes the p.d.f. of  $x_t$  conditional on the process having stayed within the band for the preceding  $n$  periods, there exists the recurrence relation

$$f_n(x) = \int_{-a}^a f_{n-1}(y) f(x - y) dy, \quad n = 1, 2, 3, \dots$$

where  $f(\cdot)$  denotes the p.d.f. of  $u_t$  and  $f_0(x) = f(x - x_0)$ , and  $x_0$  denotes the point at which the process is reset within the band following a breach. (See Cox and Miller 1965, Chapter 2.3). Note that  $|x_0| \geq \lambda a$ . The probability of still being inside the band at step  $n$  is therefore

$$\begin{aligned} \int_{-a}^a f_n(x) dx &= \int_{-a}^a f_{n-1}(y) \int_{-a}^a f(x - y) dx dy \\ &= \prod_{j=0}^n g_j \end{aligned}$$

where

$$\begin{aligned} g_0(a, \lambda) &= \int_{-a}^a f(x - x_0) dx \\ &\leq \int_{a(\lambda-1)}^{a(\lambda+1)} f(u) du \leq 1 \end{aligned}$$

and

$$g_n(a) = \frac{\int_{-a}^a f_{n-1}(y) \int_{-a}^a f(x - y) dx dy}{\int_{-a}^a f_{n-1}(y) dy} < 1 \quad (\text{A-4})$$

for  $n \geq 1$ . The inequality in (A-4) is strict because  $\int_{-a}^a f(x - y) dx \leq 1$  for  $-a \leq y \leq a$  and the latter inequality must be strict for  $|y|$  near enough to  $a$  because the support of the distribution contains an open interval around 0 by assumption. Note that  $g_0$  is increasing in  $a$  and decreasing in  $\lambda$ , whereas  $g_n$ , which is the conditional probability that the band is not breached at step  $n$ , is increasing in  $a$  and decreasing in  $n$ , with  $\inf_a g_n = 0$ ,  $\sup_a g_n = 1$ , and  $g(a) = \inf_n g_n$  is defined in (2.35).

To calculate the  $L_{2r/(r-1)}$  norm of  $\prod_{k=1}^m \phi_{t-k}$  exactly from these formulae is feasible given a form for  $f$ , but obviously very complicated. Instead, consider the parallel problem in which  $g_n = g(a)$  for every  $n$ . Also assume that the band is breached at exactly time  $t - m$ . In this case, we would have, exactly,

$$\begin{aligned} \left\| \prod_{k=1}^m \phi_{t-k} \right\|_{2r/(r-1)} &= \left( \sum_{j=0}^m \lambda^{2rj/(r-1)} \binom{m}{j} g(a)^{m-j} (1-g(a))^j \right)^{(r-1)/2r} \\ &= \left( g(a) + (1-g(a)) \lambda^{2r/(r-1)} \right)^{m(r-1)/2r}. \end{aligned} \quad (\text{A-5})$$

The second assumption lowers the probability of the first breach occurring subsequent to time  $t - m$ , and so increases  $\left\| \prod_{k=1}^m \phi_{t-k} \right\|_{2r/(r-1)}$ , and is innocuous from the point of view of bounding this norm. Moreover, since  $g_n \rightarrow g(a)$ , the approximation in (A-5) becomes arbitrarily close as  $m$  increases. ■

### Proof of Theorem 3.1

This proceeds by adapting the proof of Theorem 3.1 of Davidson and de Jong (2000), (henceforth, DdJ). Consider the quantities

$$a_{nt}(\xi, \xi') = \sum_{j=\max\{0, [n\xi']-t+1\}}^{[n\xi]-t} \theta_j. \quad (\text{A-6})$$

Under assumption (3.2),  $|a_{nt}(\xi, \xi')| = O(1)$  but not  $o(1)$ , as  $n \rightarrow \infty$ , for  $0 \leq \xi' < \xi \leq 1$  and  $-\infty \leq t \leq [n\xi]$ . Therefore consider equation (B.5) of DdJ, which is

$$\begin{aligned} \sum_{t=-\infty}^{[n\xi]} a_{nt}(\xi, \xi')^2 &= \sum_{t=[n\xi']+1}^{[n\xi]} a_{nt}(\xi, \xi')^2 + \sum_{t=-\infty}^{[n\xi']} a_{nt}(\xi, \xi')^2 \\ &= M_{1n} + M_{2n}. \end{aligned} \quad (\text{A-7})$$

It is immediate by assumptions (3.2) and (3.3) respectively that  $M_{1n} = O(n(\xi - \xi'))$  and  $M_{2n} = o(n(\xi - \xi'))$ . Thus, to establish the variance of the partial sum write  $a_{nt}$  for  $a_{nt}(0, 1)$  and consider Lemma 3.2 of DdJ. The argument has to be somewhat modified, because in this case the sequence of lag coefficients need not be monotone. Under assumption (3.3),  $|\theta_j| = O(j^{-1/2-\varepsilon})$  which corresponds to the case  $d = \frac{1}{2} - \varepsilon$  in the lemma, whereas  $\sum_{t=-\infty}^n a_{nt}^2 = O(n)$ , not  $O(n^{2-2\varepsilon})$ . However, the arguments of part (i) of the proof go through unchanged, since they depend only on the properties of the  $a_{nt}$ . In part (ii) of the proof, we find in place of equation (B-25) of DdJ that

$$\sum_{i=-\infty}^{r_n-1} B_n^3 g_{ni}^2 = O\left( B_n^2 \left( \sum_{i=-\infty}^{r_n-1} (r_n - i)^{-1-2\varepsilon} + \sum_{i=-\infty}^0 |i|^{-1-2\varepsilon} \right) \right) = O(B_n^2). \quad (\text{A-8})$$

Hence, equation (B.26) of DdJ becomes

$$\sum_{i=-\infty}^{r_n-1} W_{ni}^{*2} = O(n^{-1} B_n^2). \quad (\text{A-9})$$

However,  $B_n$  may be freely chosen subject to the conditions stated, and with  $B_n = n^{1/2-\eta}$  for  $0 < \eta < \frac{1}{2}$  the conclusion stated in (B.27) continues to hold.

It follows that, under the conditions of the theorem,  $\sigma_n^2 = O(n)$  but not  $o(n)$ . Since Theorem 3.1 of DdJ invokes the properties of the sequence  $b_j$  (corresponding to  $\theta_j$  here) only insofar as they lead to the conclusions of Lemmas 3.1 and 3.2 of DdJ, it follows that the theorem applies in this case, except that the right-hand member of equation (B.36) becomes simply  $\delta$ . ■

### Proof of Lemma 3.1

Consider the case  $0 < \varepsilon < 1$ . Placing the terms over a common denominator and applying the mean value theorem, note that

$$\frac{1}{k^\varepsilon} - \frac{1}{(n+k)^\varepsilon} = \varepsilon \left( \frac{\lambda(n, k, \varepsilon)n + k}{k(n+k)} \right)^\varepsilon \frac{n}{\lambda(n, k, \varepsilon)n + k} \quad (\text{A-10})$$

where  $0 < \lambda(n, k, \varepsilon) < 1$ , by the strict concavity of the power transformation. Hence,

$$\left( \frac{1}{k^\varepsilon} - \frac{1}{(n+k)^\varepsilon} \right)^2 = n\varepsilon^2 \left( \frac{\lambda(n, k, \varepsilon)n + k}{k(n+k)} \right)^{2\varepsilon} \left( \frac{n^{1/2}k^{1/2+\eta}}{\lambda(n, k, \varepsilon)n + k} \right)^2 k^{-1-2\eta} \quad (\text{A-11})$$

for  $0 < \eta < \frac{1}{2}$  where on the right-hand side, the first term in parentheses is  $O(1)$  as  $n \rightarrow \infty$  and  $o(1)$  as  $k \rightarrow \infty$ , and the second term in parentheses is  $o(1)$  in respect of both indices, assuming in each case that  $\lambda(n, k, \varepsilon)$  is bounded away from 0 as  $k \rightarrow \infty$ , for each  $n \geq 1$  and each  $\varepsilon$ . To show the latter condition holds, re-write (A-10) after multiplying through by  $k^\varepsilon(n+k)^\varepsilon$  and letting  $\delta = n/k$ , as

$$(1 + \delta)^\varepsilon - 1 = \varepsilon\delta(1 + \lambda\delta)^{\varepsilon-1}. \quad (\text{A-12})$$

Expanding the left-hand side of this equation in Taylor series to second order, and the right-hand side to first order, simplifying and rearranging, yields

$$\lambda = \frac{1}{2} + O(\delta). \quad (\text{A-13})$$

The case  $\varepsilon > 1$  holds similarly. In the case  $\varepsilon = 1$ , (A-10) holds with  $\lambda = 1$ , but the obvious cancellation in (A-11) yields the same conclusion as before. ■

## References

- Andrews, D. W. K., 1984. Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–4.
- Billingsley, P., 1968. *Convergence of Probability Measures*. New York: John Wiley & Sons.
- Chan, K. S. and Tong, H., 1985. On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Advances in Applied Probability* 17, 666–78.
- Bollerslev, T., 1986. Generalised autoregressive conditional heteroscedasticity, *Journal of Econometrics* 31, 307–21.
- Cox, D. R. and H. D. Miller, 1965. *The Theory of Stochastic Processes*. London: Methuen & Co.
- Davidson, J., 1992. A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes', *Econometric Theory* 8, 313–29.
- Davidson, J., 1993. The central limit theorem for globally non-stationary near-epoch dependent functions of mixing processes: the asymptotically degenerate case, *Econometric Theory* 9, 402–12.

- Davidson, J., 1994. *Stochastic Limit Theory*. Oxford: Oxford University Press.
- Davidson, J. and R. M. de Jong, 2000. The functional central limit theorem and weak convergence to stochastic integrals II: fractionally integrated processes, *Econometric Theory* 16, 5, 621–642.
- Davidson, J. and D. Peel, 1998. A nonlinear error correction mechanism based on the bilinear model, *Economics Letters* 58, 2, 165–170.
- Davydov, Yu. A., 1970. The invariance principle for stationary processes, *Theory of Probability and its Applications* XV, 3, 487–498.
- De Jong, R. M., 1997. Central limit theorems for dependent heterogeneous random variables. *Econometric Theory* 13, 353–67.
- De Jong, R. M. and J. Davidson, 2000. The functional central limit theorem and weak convergence to stochastic integrals I: weakly dependent processes. *Econometric Theory* 16, 5, 643–666.
- Dickey, D. A. and W. A. Fuller, 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–31.
- Engle, R. F., 1995. *ARCH: Selected Readings*. Oxford: Oxford University Press.
- Gallant, A. R. and H. White, 1988. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell
- Granger, C. W. J. and T. Teräsvirta, 1993. *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.
- Hamilton, J. D., 1994. *Time Series Analysis*. Princeton: Princeton University Press
- Hannan, E. J., 1970. *Multiple Time Series*. New York: John Wiley & Sons.
- Hansen, B. E., 1991. GARCH(1,1) processes are near-epoch dependent. *Economics Letters* 36, 181–186.
- Herrndorf, Norbert, 1984. A functional central limit theorem for weakly dependent sequences of random variables, *Annals of Probability* 12, 141–53.
- Ibragimov, I. A., 1962. Some limit theorems for stationary processes, *Theory of Probability and its Applications*, 7, 349–82
- McLeish D. L., 1975. Invariance principles for dependent variables, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 32, 165–78
- Phillips, P. C. B., 1987. Time series regression with a unit root, *Econometrica* 55, 277–301.
- Phillips, P. C. B. and P. Perron, 1988. Testing for a unit root in time series regression, *Biometrika* 75, 335–346
- Priestley, M. B., 1988. *Non-Linear and Non-Stationary Time Series Analysis*. London: Academic Press.
- Subba Rao, T., 1981. On the theory of bilinear models, *Journal of the Royal Statistical Society B* 43, 244–245.
- Tong, H., 1990. *Non-Linear Time Series. A Dynamical System Approach*. Oxford: Clarendon Press.