LEAST SQUARES REGRESSION
WITH INTEGRATED
OR DYNAMIC REGRESSORS
UNDER WEAK ERROR
ASSUMPTIONS

DONALD W. K ANDREWS

Cowles Foundation, Yale University

This paper establishes consistency of least squares estimators in (i) a multiple regression model with integrated regressors and explosive, non-mixing errors, and (ii) a dynamic linear regression model with regressors and errors that may have infinite variances In the former context, the asymptotic distribution of the least squares estimator also is obtained, in certain cases

1. INTRODUCTION

This paper establishes the consistency of least squares (LS) estimators in two different linear regression contexts In the first context the regressors are integrated, and in the second, the regressors include lagged values of the dependent variable In the integrated regressor model, asymptotic distribution results also are obtained In both contexts, the results are established under weak error assumptions In particular, for the integrated regressor case, the errors may be explosive and non-mixing, and in the dynamic regression case, the errors may have infinite variances

We first discuss the regression model with integrated regressors The econometrics literature recently has exhibited a burst of activity concerning

I thank Steven N Durlauf, Peter C B Phillips, James H Stock, and several referees for very helpful discussions and comments I acknowledge gratefully the support of the Alfred P Sloan Foundation, through a Research Fellowship, and the National Science Foundation, through Grant No SLS-8419789 This paper was drafted in September 1985 and revised in October 1986
models with error corrections, co-integrated variables, and integrated regressors. The close connections between such models are discussed in Granger et al. [16, 17, 19], Stock [39], and Phillips and Durlauf [37]. These papers also present numerous asymptotic results for such models, and discuss their application to various macroeconomic relations, such as the consumption function and the term structure of interest rates. Also, see Phillips [33, 34, 36] and Cavanagh [11] for recent results for closely related autoregressive models with unit roots.

This paper contributes to the above body of literature by establishing the consistency, and for some cases the asymptotic distribution, of least squares (LS) estimators in multiple time series regressions with integrated regressors, and errors that may have (i) arbitrary dependence over time, (ii) nonzero means, (iii) deterministic time trend, and/or (iv) exploding variances. In contrast, the results of Stock [39] and Phillips and Durlauf [37] require the errors to be asymptotically weakly dependent (i.e., satisfy some mixing condition), and to have zero means and uniformly bounded variances. Since the errors are unobserved, it is particularly comforting to be able to weaken these assumptions and still obtain consistency.

The consistency results given here do not attain as fast a rate of convergence as those given by Stock [39] and Phillips and Durlauf [37] under their stronger assumptions on the errors. Nevertheless, the rate of convergence results established here are the best possible given our assumptions.

In the terminology of co-integrating models, the estimated parameters considered here constitute the co-integrating vector. The assumptions on the errors allow one to consider more general co-integration models (in some respects, at least) than those considered by Granger and associates. For example, one could have an underlying equilibrium condition that involves both integrated variables, and stationary, trending, or exploding variance variables. In this case, if the latter variables are lumped into the error term, then the paper shows that the co-integrating vector still is estimated consistently (under conditions on the growth rates of the trend and the variances of the nonintegrated variables).

We mention that the LS estimators considered here are for the regression equation specified in levels rather than in differences, because the LS estimators of the former exhibit faster rates of convergence, and hence, are asymptotically more efficient, than those of the latter, under the error assumptions of Stock [39] and Phillips and Durlauf [37]. This paper shows that this gain in efficiency is not achieved at the expense of inconsistency under broader assumptions on the errors.

The proof of the results for integrated regressor models makes use of the results of Phillips and Durlauf [37] concerning the weak convergence of the sample second moment matrix of the integrated regressors to a function of the multivariate Wiener process on [0, 1].
Next, we discuss the dynamic linear regression model. Consistency of LS estimators for this model has received considerable attention in the literature for the case where the errors have finite variances, e.g., see Robinson [37a], Anderson and Taylor [2], Crowder [12], Nelson [32], and Phillips [34]. In many contexts, however, models with infinite variance errors and regressors are being considered for economic processes. For example, there is evidence that the following economic variables exhibit fatter-than-normal tails and potentially have infinite variances: changes in the logarithms of stock market prices (see Fama [14] and Mandelbrot [29]), and speculative commodity prices (Mandelbrot [28, 29]), changes in interest rates (Roll [38] and Mandelbrot [29]), exchange rates (Mandelbrot [29]), U.S. Treasury Department cash flows (Granger and Orr [18]), and various other macroeconomic time series (Burman [10]). Such variables may enter a regression model as regressors or as unobserved components of the error. In addition, there is direct evidence of thick-tailed errors in various economic regression models. These include earnings function regressions (White and MacDonald [41]), production function regressions (Zeckhauser and Thompson [44]), money demand functions (Krasner [26]), and other macroeconomic relations (Zeckhauser and Thompson [44]). In consequence, it is desirable to extend the consistency results mentioned above to situations where the errors and regressors have infinite variances.

Bierens [9] considers the infinite variance case. He proves consistency of weighted M-estimators, but his results do not encompass LS estimation. This paper complements the results of Bierens, and the LS results mentioned above, by establishing consistency of LS in dynamic linear regression models with infinite variance errors and regressors.

The autoregressive (AR) model is a special case of the dynamic linear regression model considered here. LS estimators have been shown to be consistent in the AR model with infinite variance errors, see Kanter and Steiger [25], Yohai and Maronna [43], and Hannan and Kanter [21]. (Also see Gross and Steiger [20] and An and Chen [1] for consistency of the least absolute deviations estimator in the AR model with infinite variance errors.) In fact, the consistency proof given here for the dynamic regression model is an extension of the consistency proof of Hannan and Kanter for the simpler AR model.

We mention that general methods of proving consistency of estimators (e.g., see Huber [24] and Bates and White [8]) do not apply in the present context, because they require the sum of squared residuals, the sum of squared residuals minus the sum of squared errors, or the derivative of the sum of squared residuals with respect to the parameters, to have a well-defined expectation. None of these properties hold in the case considered here.

The remainder of the paper is organized as follows: Section 2 presents the results for the integrated regressor model, Section 3 does the same for the dynamic regression model, and Section 4 provides proofs for the results of Sections 2 and 3.
2. RESULTS FOR THE INTEGRATED REGRESSOR MODEL

Following Phillips and Durlauf [37], the multiple regression model we consider is given by

\[ y_t = Ax_t + u_t, \]
\[ x_t = x_{t-1} + e_t, \quad \text{for } t = 1, \ldots, T, \]  \hspace{1cm} (1)

where \( x_0 \) is a given random variable (and hence, may be a constant), and \( A \) is an \( n \times m \) matrix of unknown parameters. Of the integrated regressors \( \{x_t\} \) we make the following assumptions:

C1 \( \mathbb{E} u_t = 0, \sup_{r \geq 1} \mathbb{E} \|e_t\|^\beta < \infty \), for some \( \beta > 2 \), where \( \| \| \) denotes the Euclidean norm.

C2 \( T^{-1} \mathbb{E} S^2_t \rightarrow \infty \) \( \Omega \) positive definite, where \( S_T \equiv \sum_{t=1}^{T} e_t \)

C3 \( \{e_t\} \) are strong mixing with mixing numbers \( \{\alpha(k)\} \) \( k = 1, 2, \ldots \) that satisfy

\[ \sum_{k=1}^{\infty} \alpha(k)^{1-2\beta} < \infty \]

For the definition of the strong mixing, see Andrews [3–5], Herrndorf [23], or White [40]. (Note that the mixing condition used by Phillips and Durlauf [37] and McLeish [31] is very close to that of strong mixing, but is slightly weaker.) The summability condition of C3 is implied by \( \alpha(k) = O(k^{-\zeta}) \) as \( k \rightarrow \infty \), for some \( \zeta > \beta/(\beta - 2) \). The strong mixing assumption placed on the regressor innovations could be replaced by any of a number of other mixing conditions, provided a functional central limit theorem can be obtained. For example, one could employ \( p \)-mixing (see Herrndorf [22] and Phillips [36]) or \( \phi \)-mixing (see Phillips and Durlauf [37]).

The assumptions C1–C3 are quite similar to those of Phillips and Durlauf [37] but do not require the condition \( ET^{-1}(S_{T+h} - S_{h})(S_{T+k} - S_{k})' \rightarrow \Omega \) as \( \min(k, T) \rightarrow \infty \). This is due to our application of a recent functional central limit theorem of Herrndorf [23], rather than that of McLeish [31], which is used by Phillips and Durlauf [37].

The assumptions C1–C3 allow sufficient heterogeneity and dependence over time to permit the regressors to be generated by a wide variety of multivariate ARIMA processes that are integrated of order one and have non-identically distributed innovations (see Gorodetskii [15] and Withers [42]), though not for all such processes (see Andrews [3,4]).

The only assumption placed on the errors in the model of equation (1) is:

C4 \( \lim_{T \rightarrow \infty} T^{-\gamma} \sum_{t=1}^{T} \mathbb{E} u_t^2 < \infty \), for some \( \gamma < 2 \)
This condition implies that the second moments of the errors cannot blow up too quickly. It places no restriction on the dependence of the errors over time, or on the dependence between the regressors and the errors. In terms of the means and variances of the errors, C4 holds if

\[ \mu_i = O(t^{b/2}), \quad \sigma_i^2 = O(t^b), \quad \text{for some } b < 1, \]

where \( \mu_i = \max_{i \leq n} \mathbb{E} u_i \), \( \sigma_i^2 = \max_{i \leq n} \text{var}(u_i) \), and \( u_i = (u_{i1}, \ldots, u_{in}) \). Thus, the errors may exhibit considerable drift and growth of variances. The latter cannot grow at rate \( t \), however, because this is the growth rate of the variances of the integrated regressors.

The weak assumption placed on the errors allows the results of this paper to be applied to models that contain both integrated regressors of order one and other regressors. Consider the model

\[ y_t = Ax_t + Bz_t + \tilde{u}_t, \quad t = 1, \ldots, T, \]

where \( A, x_t, \) and \( x_0 \) are as above, and \( B \) is an \( n \times l \) matrix of unknown parameters. If the regressors \( z_t \) and the errors \( \tilde{u}_t \) have second moments that satisfy the growth condition of C4, then the model of equation (3) can be written as the model in equation (1) with \( u_t = Bz_t + \tilde{u}_t \). Hence, the results given below, for estimation of \( A \), are applicable even if the underlying model of interest contains nonintegrated regressors that may have nonzero means, drift, and/or exploding variances. As mentioned in the Introduction, this allows one to consider co-integration models where both integrated and nonintegrated variables enter the equilibrium condition.

Having consistently estimated the matrix \( A \), often one can use a second step procedure to consistently estimate \( B \) (provided \( z_t, \tilde{u}_t \) satisfy more standard and more restrictive assumptions than C4). See Granger and Engle [17] for a similar procedure.

As discussed in Phillips and Durlauf [37], the model of equation (1) (or the model of equation (3)) may represent a reduced form, or semireduced form, of a simultaneous equations system. In this case, none of the standard exogeneity conditions need hold for the regressors \( x_t \), since their variances blow up at a faster rate than those of the errors. Furthermore, such a model may exhibit any amount of serial correlation in the errors.

We consider two LS estimators of the matrix \( A \). The first is the LS estimator for the model of equation (1), and the second is the LS estimator of \( A \) with a fitted constant vector. Define

\[ A^* = \left( \frac{1}{T} \sum_{T} y_t x_t \right) \left( \frac{1}{T} \sum_{T} x_t x_t \right)^{-1}, \]

\[ \hat{A} = \left( \frac{1}{T} \sum_{T} y_t (x_t - \bar{x}) \right) \left( \frac{1}{T} \sum_{T} (x_t - \bar{x})(x_t - \bar{x}) \right)^{-1}, \]

(4)
where \( \bar{x} = T^{-1} \sum_{t=1}^{T} x_t \). For these LS estimators we have the following consistency results:

**THEOREM 1.** Under assumptions C1–C4, \( \forall \xi < 1 - \Gamma/2 \),

(a) \( T^{\gamma} (A^* - A) \overset{p}{\to} 0 \) as \( T \to \infty \), and

(b) \( T^{\gamma} (\hat{A} - A) \overset{p}{\to} 0 \) as \( T \to \infty \),

where \( \Gamma \) is the infimum of all values \( \gamma \) such that C4 holds.

Comments: (1) The maximal rate of convergence obtained in Theorem 1 for the special case of independent identically distributed (i.i.d.), finite variance errors (for which \( \Gamma = 1 \)) is \( T^{1/2 - \delta} \), for arbitrarily small \( \delta > 0 \) This is considerably slower than the rate \( T^{\Gamma - \delta} \) obtained by Stock [39] and Phillips and Durlauf [37] with i.i.d., mean zero, finite variance errors. Nevertheless, the rate \( T^{1/2 - \delta} \) is the best possible in the above context, unless the restriction to mean zero errors is imposed. In fact, as shown in Theorem 2 below, the rates of convergence given in Theorem 1 are the best possible under the given assumptions, for any given \( \Gamma \in (0, 2) \) These best rate results hold even if the regressors and errors are assumed to be mutually independent. (Thus, the slower rate obtained under our assumptions is not due to the lack of an assumption restricting the dependence between the regressors and errors, but rather, is due to the weak assumptions on the errors themselves.)

(2) Corresponding to the LS slope estimator \( \hat{A} \) is a LS constant vector estimator \( \hat{u} \) given by

\[
\hat{u} = \bar{y} - \hat{A}\bar{x} = (A - \hat{A})\bar{x} + \hat{u}
\] (5)

Under our assumptions, \( \hat{u} \) does not necessarily converge in probability. Even if \( \bar{u} \) converges in probability to some constant, \((A - \hat{A})\bar{x}\) may diverge. Further, since \( \{(y_t, x_t); t = 1, 2, \ldots\} \) are not necessarily strong mixing in the present context, even the zero-one result for the LS estimator (see Andrews [5]) does not apply to (\( \hat{u}, \hat{A} \)).

(3) The results of Theorem 1 can be extended easily and without change to the class of near-integrated regressors, as defined in Phillips [33]. Such regressors satisfy: \( x_t = B_T x_{t-1} + v_t \), \( t = 1, \ldots, T \), where \( B_T = \exp(T^{-1}C) \) and \( C \) is a constant matrix. See Phillips [33] for motivation for, and applications of, this class of regressors.

Next, we consider a special case of the model analyzed above. We assume the errors contain a time trend that is more dominant than the explosiveness of the error variances. (That is, the key parameter \( \Gamma \) is determined by the magnitude of the trend rather than the magnitude of the error variances.)
In particular, we assume that C1–C3 and the following assumption hold.

C5a. \( E[u_t] = t^\theta h + g(t) \) for some \( 0 \leq \theta < 1/2 \), where \( g(\cdot) \) is an \( R^s \)-valued function defined on the positive integers such that \( ||g(t)|| = o(t^\theta) \) as \( t \to \infty \), and \( h \) is a \( n \)-vector not equal to \( 0 \).

C5b. \[ \Gamma_1 = \inf \left\{ \gamma > 0 \mid \lim_{T \to \infty} \sum_{i=1}^T \frac{E(u_t - E(u_t))(u_t - E(u_t))}{(T-1)^2} < \infty \right\} < 1 + 2\gamma \]

Assumptions C1–C3 and C5 allow arbitrary dependence of the errors over time, arbitrary dependence between the regressors and errors, and heterogeneous, trending errors with exploding variances. Nevertheless, the asymptotic distribution of the LS estimator still can be obtained.

**THEOREM 2.** Under assumptions C1–C3 and C5, we have

(a) \[ T^{1/2 - \gamma}(A^* - A) \xrightarrow{d} \mathcal{L} \int_0^1 v^2 W(v) dv \frac{d\tau}{\tau^{1/2}} (\Omega^{1/2} \int_0^1 W(v)dv \frac{d\tau}{\tau^{1/2}})^{-1} \quad \text{as } T \to \infty, \]

and

(b) \[ T^{1/2 - \gamma}(\hat{A} - A) \xrightarrow{d} \mathcal{L} \int_0^1 v^2 W(v) dv \frac{d\tau}{\tau^{1/2}} - \mathcal{L} \int_0^1 W(v) dv \frac{d\tau}{\tau^{1/2}} \left( \frac{d\tau}{\tau^{1/2}} \right)^{-1} \quad \text{as } T \to \infty, \]

where \( \{W(v); \tau \in [0, 1]\} \) is a multivariate Wiener process taking values in \( C([0, 1]) \).

Comments

1. Assumption C5 implies that Assumption C4 holds with \( \Gamma = 1 + 2\gamma \). Hence, the normalization factor \( T^{1/2 - \gamma} \) of Theorem 2 equals \( T^{1/2 - \Gamma/2} \), and the rate of convergence results of Theorem 1 are shown to be best possible. This holds even if the regressors and errors are assumed to be independent.

2. Theorem 2 shows the result of erroneously ignoring a time trend in regressions with integrated regressors. Consistency still obtains (provided the trend is not too large), but the rate of convergence is much slower than, and the limit distribution is different from, that in the standard case with no time trend (cf. Phillips and Durlauf [37], Theorem 4.1). In consequence, standard error estimates and hypothesis tests that are constructed assuming no time trend, no longer are valid when a time trend exists. For related results, see Durlauf and Phillips [13] and Phillips [35].

3. **CONSISTENCY RESULTS FOR THE DYNAMIC REGRESSION MODEL**

In this section, we consider the dynamic linear regression model.

\[ y_t = Y_1 \delta + W_t \theta + \epsilon_t, \quad t = \ldots, 1, 2, \ldots, T, \ldots, \]

(6)
where the errors $e_t$ have zero means, but possibly infinite variances, and
where $Y_t = (y_{t-1}, \ldots, y_{t-p})'$, $W_t = (w_{t-1}, \ldots, w_{t-K})'$, $\delta = (\delta_{1t}, \ldots, \delta_{pt})'$, and $\theta = (\theta_{11}, \ldots, \theta_{K1})$ are vectors of lagged dependent variables, weakly exogenous regressors, and unknown parameters, respectively.

The least squares estimators $(\hat{\delta}, \hat{\theta})$ of $(\delta, \theta)$ are solutions to

$$\inf_{\delta \in \mathbb{R}^p, \hat{\theta} \in \mathbb{R}^K} \sum_{t=1}^{T} (y_t - Y_t'\delta - W_t'\hat{\theta})^2$$

(where the random variables indexed by $t = -p + 1, \ldots, 1, \ldots, T$ are assumed to be observed) It is well known that LS estimators always exist, and are unique if $\sum_{i=1}^{T} G_i G_i'$ is of full rank $p + K$, where $G_i = (Y_i', W_i')'$.

The maximal moment exponent of a sequence of random variables or vectors (rv's) $\{X_t\}$ is defined to be

$$g = \sup_{t \geq 1} \left\{ \zeta > 0 : \sup_{t \geq 1} E[|X_t|^\zeta] < \infty \right\}$$

The strong mixing numbers $\{x(k)\}$ of a sequence of rv's are said to be of size $s (<0)$ if $x(k) = o(k^s)$ as $k \to \infty$. This definition of “size” is stronger than that of McLeish [30], but is considerably simpler. The theorem below also holds using his weaker definition of size.

The following assumptions are used to establish consistency of $(\hat{\delta}, \hat{\theta})$:

A1 The roots of the equation $1 - \sum_{j=1}^{p} \delta_j z^j = 0$ are outside the unit circle

A2 The errors and regressors $\{e_t, W_t\}$ form a strong mixing sequence of rv's of size $-r/(1-r)$, for some $r > 1$, and have maximal moment exponent $\alpha > 1$.

The errors have conditional means zero, i.e., $E(e_t | W_t, \ldots, W_{t+r}, \ldots, e_0, \ldots, \epsilon_{t-1}) = 0$ a.s. $\forall t$, and satisfy the identification condition:

$$V(M) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E \delta^2_i I(e_t^2 \leq M) > 0,$$

for some $M_1 < \infty$ if $Ee_t^2 = \infty$ for some $t$, and for $M_1 = \infty$ otherwise, where $I(\cdot)$ denotes the indicator function.

A3 The regressors $\{W_t\}$ satisfy the identification condition

$$V_t(M_2) = \frac{1}{T} \sum_{i=1}^{T} E \tilde{W}_t \tilde{W}_t' I(||\tilde{W}_t||^2 \leq M_2)$$

has eigenvalues bounded away from zero for $T$ large, for some $M_2 < \infty$ if $E||W_t||^2 = \infty$ for some $t$, and for $M_2 = \infty$ otherwise, where $\tilde{W}_t = (W_t', W_{t-1}, \ldots, W_{t-p})'$.

(By definition, the case $r = 1$ requires $\{W_t\}$ to be $m$-dependent, for some $m$ finite.)
Under Assumption A1 there exist coefficients \( \{ \varnothing_m \} \) such that

\[
\sum_{m=0}^{\infty} \varnothing_m z^m = (1 - \sum_{j=1}^{p} \delta_j z^j)^{-1},
\]

and \( |\varnothing_m| \) converges to zero exponentially fast as \( m \to \infty \). Hence,

\[
y_i = \sum_{m=0}^{\infty} \varnothing_m (W_{i-m} \theta + e_{i-m}) \text{ a.s.}
\]

(where the sum converges absolutely a.s. under Assumptions A2 and A3, since

\[
E \left| \sum_{m=0}^{\infty} \varnothing_m (W_{i-m} \theta + e_{i-m}) \right| \leq \sum_{m=0}^{\infty} |\varnothing_m| \left[ \sup_i E|W_i \theta| + \sup_t E|\xi_t| \right] < \infty.
\]

Assumptions A1 and A4 imply that the dynamic model is not explosive. Assumption A2 allows considerable heterogeneity of the errors. If the errors are identically distributed, however, then A2 simplifies. In this case, \( \alpha \) is the maximal moment exponent of the single rv \( e_t \), and the condition \( V(M_1) > 0 \) is automatically satisfied provided \( e_t \) is not identically zero. Without the condition \( V(M_1) > 0 \), the model is not necessarily identified, because we can have \( y_t \equiv 0 \) with probability one.

Assumption A3 allows considerable heterogeneity and dependence over time of the regressors. It also allows the regressors, like the errors, to have infinite variances. It is possible to replace the strong mixing assumption of A3 by some other assumption of asymptotic weak dependence that yields the strong law of large numbers (SLLN) with fast rate of convergence for the rvs \( \{W_t\} \).

Assumption A4 places some restriction on the heterogeneity of the regressors and errors. It is automatically satisfied under Assumptions A1–A3, if \( \{W_t, e_t\} \) is stationary and \( \sup_{n \geq 1} \left[ E|\xi_t W_{t-n}|^2 + E|\xi_t e_{t-n}|^2 \right] < \infty \).

The condition on the eigenvalues of \( V_p(M_2) \) ensures sufficient variability of the regressors to identify \( \theta \). To see why \( p \) lags of the regressors arise in the identification condition, rewrite model of equation (6) as an ARMAX model.

\[
y_t^* = Y_t^* \delta + \left[ \sum_{j=0}^{p} \delta_j W_{i-j} \right] \theta + e_t^*,
\]
where

\[ y_t^* = \sum_{j=0}^{p} \delta_j y_{t-j}, \quad Y_t^* = (y_{t-1}^*, \ldots, y_{t-p}^*), \quad \epsilon_t^* = \sum_{j=0}^{p} \delta_j y_{t-j}, \]

and \( \delta_0 \equiv -1 \). Within the class of ARMAX models, the parameter \( \theta \) is identified only if \( \sum_{j=0}^{p} \delta_j W_{t-j} \) has sufficient variability. This must hold for any true parameter \( \delta \). Hence the identification condition depends on \( p \) lags of \( W_t \).

In the case of independent regressors, this condition reduces to the same condition with \( \hat{W}_t \) replaced by \( W_t \).

The main result of this section is the following:

**THEOREM 3** Under assumptions A1–A4, the LS estimators \( (\hat{\delta}, \hat{\theta}) \) of \( (\delta, \theta) \) are strongly consistent. Moreover,

\[ T^T (\hat{\delta}, \hat{\theta}) - (\delta, \theta) \xrightarrow{T \to \infty} 0 \text{ a.s., } \forall t < \min\{1 - r/2, 1/2\}. \]

Consistency of the dynamic regressor parameters \( \delta \), but not the exogenous regressor parameters \( \theta \), can be established for the case where \( x < 1 \) (i.e., the case where the errors have undefined means), by a method similar to that given below, if we assume that the errors are stochastically dominated by a random variable in the domain of attraction of a stable law with exponent \( x \), \( \theta \) lies in a compact set, and the exogenous regressors are independent over time. Since the latter assumption is overly restrictive in the context of the dynamic regression model, we do not pursue this result here.

### 4. PROOFS

Before proving Theorem 1, we state several definitions and a lemma. Let \( X_{t}(\tau) = T^{-1/2} \Omega^{-1/2} S_{[\tau t]}, \) for \( \tau \in [0, 1] \), where \( S_{T} = \sum_{Y} v_{i} \) for \( T \geq 1 \), \( S_{0} = 0 \), and \( \lceil T \tau \rceil \) denotes the largest integer less than or equal to \( T \tau \). As above, \( W(\cdot) = \{W(\tau) : \tau \in [0, 1]\} \) is the multivariate Wiener process that takes values in \( C[0,1]^n \) (see Phillips and Durlauf [37]).

The following lemma is the univariate functional central limit theorem of Herrndorf ([23], Corollary 1) extended to the multivariate setting by the argument of Phillips and Durlauf ([37], Theorem 2.1).

**LEMMA 1.** Under assumptions C1–C3, \( X_{t}(\cdot) \Rightarrow W(\cdot) \) as \( T \to \infty \), where \( \Rightarrow \) denotes weak convergence of the corresponding probability measures.
Proof of Theorem 1. Under assumptions C1–C3, Lemma 3.1b of Phillips and Durlauf [37] (using Lemma 1 above in place of their Theorem 2.1) gives

\[ T^{-2} \sum_{t=1}^{T} x_t x'_t = \Omega^{1/2} \int_{0}^{1} W(t)W(t') dt \Omega^{1/2} \quad \text{as } T \to \infty, \tag{9} \]

where the random limit matrix is positive definite with probability one. The continuous mapping theorem applied to equation (9) yields

\[ T^2 \left( \sum_{t=1}^{T} x_t x'_t \right)^{-1} = O_p(1) \quad \text{as } T \to \infty. \tag{10} \]

Similarly, by Lemma 3.1c of Phillips and Durlauf [37], Lemma 1 above, and the continuous mapping theorem, we get

\[ T^2 \left( \sum_{t=1}^{T} (x_t - \bar{x})(x_t - \bar{x}') \right)^{-1} = O_p(1) \quad \text{as } T \to \infty. \tag{11} \]

Consider arbitrary elements of the matrices \( \sum_{t=1}^{T} u_t x_t, \sum_{t=1}^{T} u_t \bar{x}, \sum_{t=1}^{T} x_t x'_t, \) and \( \sum_{t=1}^{T} u_t u'_t. \) For notational simplicity, omit subscripts and denote them by \( \sum_{t=1}^{T} u_t x_t, \)
\( \sum_{t=1}^{T} u_t \bar{x}, \) \( \sum_{t=1}^{T} x_t^2, \) and \( \sum_{t=1}^{T} u_t^2, \) respectively. By Assumption C4 and Markov’s inequality, \( \forall \delta > 0, \forall \varepsilon > 0, \) and \( \forall \gamma > 1, \)

\[ P \left( T^{-\gamma - \delta} \sum_{t=1}^{T} u_t^2 > \varepsilon \right) \leq T^{-\delta} \left( T^{-\gamma} \sum_{t=1}^{T} E u_t^2 \right)^{1/2} \leq T^{-\delta \gamma} \to 0. \tag{12} \]

Thus, \( T^{-\gamma - \delta} \sum_{t=1}^{T} u_t^2 = o_p(1) \) as \( T \to \infty. \) This result, equation (9), and the Cauchy–Schwartz inequality give: \( \forall \delta > 0 \) and \( \forall \gamma > 1, \)

\[ T^{-1 - \gamma/2 - \delta} \left| \sum_{t=1}^{T} u_t x_t \right| \leq T^{-1 - \gamma/2 - \delta} \left( \sum_{t=1}^{T} u_t^2 \right)^{1/2} \left( \sum_{t=1}^{T} x_t^2 \right)^{1/2} = o_p(1) \]

as \( T \to \infty, \tag{13} \)

and

\[ T^{-1 - \gamma/2 - \delta} \left| \sum_{t=1}^{T} u_t \bar{x} \right| \leq T^{-1 - \gamma/2 - \delta} \left( \sum_{t=1}^{T} u_t^2 \right)^{1/2} \left( \sum_{t=1}^{T} \bar{x}_t^2 \right)^{1/2} = o_p(1) \]

as \( T \to \infty. \tag{14} \)
Treating $\sum_T u_i x_i$, etc., as matrices once again, equations (10) and (13) combine to yield

$$T^{1 - \gamma / 2 - \delta} \left( \sum_T u_i x_i \right) \left( \sum_T x_i x_i' \right)^{-1} = o_p(1) \quad (15)$$

as $T \rightarrow \infty$, $\forall \delta > 0$ and $\forall \gamma > \Gamma$, which establishes part (a). Similarly, equations (11), (13), and (14) give

$$T^{1 - \gamma / 2 - \delta} \left( \sum_T u_i x_i - \sum_T u_i \bar{x} \right) \left( \sum_T (x_i - \bar{x})(x_i - \bar{x})' \right)^{-1} = o_p(1) \quad (16)$$

as $T \rightarrow \infty$, $\forall \delta > 0$ and $\forall \gamma > \Gamma$, which establishes part (b).

Proof of Theorem 2 We write $T^{1 / 2 - \gamma}(A^* - A) = H_{1T} + H_{2T}$, where

$$H_{1T} = T^{-3 / 2 - \gamma} \sum_T (E u_i) x_i' \left( T^{-2} \sum_T x_i x_i' \right)^{-1}$$

and

$$H_{2T} = T^{1 / 2 - \gamma} \sum_T (u_i - E u_i) x_i' \left( \sum_T x_i x_i' \right)^{-1} \quad (17)$$

By Theorem 1, $T^{-1 / 2 + \epsilon} H_{2T} \xrightarrow{T \rightarrow \infty} 0$ a.s., $\forall \epsilon < 1 - \Gamma / 2$. Hence, $H_{2T} \xrightarrow{T \rightarrow \infty} 0$ a.s., if $(-1 / 2 + \delta) + (1 - \Gamma / 2) > 0$. The latter holds under assumption CS(b).

Take $X_T(\tau)$ as above, and define $A_T(\tau) = ([T \tau]/T)^y$ and $G_T(\tau) = (g([T \tau])/T)^{y'}$, for $\tau \in [0, 1]$. By algebra, we have

$$H_{1T} = \left( T^{-3 / 2 - \gamma} \sum_T h^T x_i' + T^{-3 / 2 - \gamma} \sum_T g(t) x_i' \right) \left( T^{-2} \sum_T x_i x_i' \right)^{-1}$$

$$= \left( h \int_0^1 A_T(\tau) X_T(\tau)' d\tau \Omega^{1 / 2} + \int_0^1 G_T(\tau) X_T(\tau)' d\tau \Omega^{1 / 2} + o_p(1) \right)$$

$$\times \left( \Omega^{1 / 2} \int_0^1 X_T'(\tau) X_T(\tau)' d\tau \Omega^{1 / 2} \right)^{-1} \quad (18)$$
As above, $X_T(\cdot) \Rightarrow W(\cdot)$ as $T \to \infty$, and we claim that

\[
(1) \quad \sup_{\tau \in [0, 1]} A_T(\tau) - \tau^\alpha \xrightarrow{T \to \infty} 0, \text{ and}
\]

\[
(2) \quad \sup_{\tau \in [0, 1]} \|G_T(\tau)\| \xrightarrow{T \to \infty} 0
\]

Results (i) and (ii) imply that $A_T(\cdot) \Rightarrow A(\cdot)$ and $G_T(\cdot) \Rightarrow 0$ as $T \to \infty$, where $A(\tau) = \tau^\alpha$ for $\tau \in [0, 1]$ and $0$ is the function on $[0, 1]$ that is identically equal to an $n$-vector of zeros (and where weak convergence is defined using the product topology on $D[0, 1]^n$ generated by the Skorohod metric topology on $D[0, 1]$, see Phillips and Durlauf [37]). Since $A_T(\cdot)$ and $G_T(\cdot)$ are non-random, $(X_T(\cdot), A_T(\cdot), G_T(\cdot)) \Rightarrow (W(\cdot), A(\cdot), 0)$ as $T \to \infty$. Hence, equation (18) and the continuous mapping theorem yield the result of part (a).

To show claim (i), we have for $\alpha \in [0, 1/2]$,

\[
\sup_{\tau \in [0, 1]} |A_T(\tau) - \tau^\alpha| = \sup_{\tau \in [0, 1]} (\tau^\alpha - [T\tau]^\beta/T^\alpha) = 1/T^\alpha \xrightarrow{T \to \infty} 0, \quad (19)
\]

since the supremum is attained by a sequence of $\tau$ values that approaches $1/T$ from below.

To show claim (ii), note that

\[
\sup_{\tau \in [0, 1]} \|G_T(\tau)\| = \sup_{r \in T} \|g([T\tau])\|/T^\alpha = \sup_{r \in T} \|g(t)/T^\alpha = \|g(m_T)/T^\alpha, \quad (20)
\]

for some non-decreasing sequence of positive integers $\{m_T\}$. Suppose $m = \sup_{i \geq 1} m_i < \infty$. Then, $\|g(m_T)/T^\alpha \leq \|g(m_i)/T^\alpha \xrightarrow{T \to \infty} 0$. Alternatively, suppose $m_i \uparrow \infty$ as $T \to \infty$. Then, $\|g(m_T)/T^\alpha \leq \|g(m_T)/m_T^\alpha \xrightarrow{T \to \infty} 0$, using Assumption C5(a). This finishes the proof of part (a).

The proof of part (b) is analogous, noting that

\[
T^{-3/2-\alpha} \sum_{t=1}^{T} hT^\beta X_T = h \int_0^1 A_T(\tau) d\tau \int_0^1 X_T(\tau) d\tau \Omega^{1/2} + o(1), \quad (21)
\]

and

\[
\int_0^1 \tau^\alpha d\tau = 1/(\alpha + 1).
\]

As mentioned in Section 3, the proof of Theorem 3 extends that of Hannan and Kanter [21] for AR models. Before giving the proof, we introduce some notation and state a lemma. Let $v = (v_1, \ldots, v_{p+k})$, be a $p + K$ vector, and take $X_{\gamma} \sim \begin{pmatrix} X_\gamma \\ W_\gamma \end{pmatrix}$.
LEMMA 2. Under assumptions A1–A4, for any constant $c > 0,$

$$\lim_{T \to \infty} \left( \inf_{v: \|v\|_1 = c} \frac{1}{T} \sum_{t=1}^{T} X_{tv}^2 \right) > 0 \text{ a.s.}$$

Comment: If the regression function includes a constant, as is usually the case, then

$$\lim_{T \to \infty} \inf_{v: \|v\|_1 = c} T^{-\kappa} \sum_{t=1}^{T} X_{tv}^2 = 0, \forall \kappa > 1.$$  

Hence, the exponent $\kappa = 1$ in Lemma 2 is as large as possible. This differs from the purely autoregressive model, where the result of Lemma 2 holds for all $\kappa < 2/\alpha$ even if $\alpha < 2,$ see Hannan and Kanter [21]. It explains the slower rate of convergence in our Theorem 2, than in Hannan and Kanter's result.

Proof of Theorem 3 The normal equations for the LS estimator are

$$\sum_{t=1}^{T} (Y_t - Y_t\hat{\delta} - W_t\hat{\theta}) \begin{pmatrix} Y_t \\ W_t \end{pmatrix} = 0. \quad (22)$$

From equation (6), we have

$$\sum_{t=1}^{T} (Y_t - Y_t\delta - W_t\theta) \begin{pmatrix} Y_t \\ W_t \end{pmatrix} = \sum_{t=1}^{T} \epsilon_t \begin{pmatrix} Y_t \\ W_t \end{pmatrix}. \quad (23)$$

Combining equations (22) and (23) gives

$$T^{-\gamma} \sum_{t=1}^{T} \left[ Y_t(\hat{\delta} - \delta) + W_t(\hat{\theta} - \theta) \right] \begin{pmatrix} Y_t \\ W_t \end{pmatrix} = T^{-\gamma} \sum_{t=1}^{T} \epsilon_t \begin{pmatrix} Y_t \\ W_t \end{pmatrix}, \quad (24)$$

where we take $\gamma > \max\{\gamma/\alpha, 1/2\}.$ Let $\hat{\delta} = T^{1-\gamma/2} \left( \hat{\delta} - \delta \right)/\Delta,$ where $\Delta = \left\| (\hat{\delta} - \delta, \hat{\theta} - \theta) \right\|^{1/2}.$ Then, equation (24) post-multiplied by $T^{-1-\gamma/2} \hat{\delta}/\Delta$ can be written as

$$T^{-1} \sum_{t=1}^{T} X_{tv}^2 = \left[ T^{-\gamma} \sum_{t=1}^{T} \epsilon_t \begin{pmatrix} Y_t \\ W_t \end{pmatrix} \right] T^{-1-\gamma/2} \hat{\delta}/\Delta \quad (25)$$

Under the assumptions, the sequence of r.v’s $\{\epsilon_t W_t\}$ has maximal moment exponent $\alpha,$ and zero means. Hence, by the SLLN with fast rate of convergence for strong mixing r.v’s (e.g., see Lemma 1 of Andrews [7] which uses the three
series theorem of McLeish [30] and a result of Loeve [27]), we have

\[ T^{-\gamma} \sum_{i=1}^{T} \epsilon_i W_i \overset{\mathcal{L}}{\rightarrow} 0 \text{ a.s., } \forall \gamma > \max \{r/2, 1/2\} \quad (26) \]

Also, by Assumptions A2 and A4, the sequence of rv’s \( \{\epsilon_i y_i, j\} \) satisfies \( E\epsilon_i y_i, j \overset{\mathcal{L}}{\rightarrow} 0 \) and \( |\epsilon_i y_i, j| \leq Q, \forall i \), where \( E\epsilon_i^2 < \infty, \forall \epsilon_i < z \), for \( j = 1, \ldots, p \). In consequence, a result of Loeve ([27], Theorem 29.6.1, pp 387) implies that

\[ T^{-1/2} \sum_{i=1}^{T} \epsilon_i y_i, j \overset{\mathcal{L}}{\rightarrow} 0 \text{ a.s., } \forall \epsilon_i < \min \{z, 2\}, \]

and so,

\[ T^{-1} \sum_{i=1}^{T} \epsilon_i y_i \overset{\mathcal{L}}{\rightarrow} 0 \text{ a.s., } \forall \gamma > \max \{r/2, 1/2\} \quad (27) \]

Since \( \|T^{-(1-\gamma)/2} \hat{g}/\Delta\| = 1 \), equations (25), (26), and (27) imply that

\[ \lim_{t \to \infty} T^{-t} \sum_{i=1}^{T} X_{in}^2 = 0 \text{ a.s.} \]

Lemma 2 now yields \( \||\hat{g}|| \overset{T^{-1} \to \infty}{\rightarrow} 0 \text{ a.s., which gives the desired result.} \]

**Proof of Lemma 2**  Let \( \tilde{Z}_{tv} = Z_{tv} + B_{tv} \), where

\[ B_{tv} = B_{1tv} - B_{3tv}, \quad B_{1tv} = \sum_{j=1}^{p} y_j W_{t-j} \beta_j, \]

\[ B_{3tv} = \sum_{i=0}^{p} \delta_i W_{t-j}^{v*}, \quad v^* = \{v_{p+1}, \ldots, v_{p+k}\}, \]

\[ \delta_0 = -1, \text{ and } Z_{tv} = \sum_{j=1}^{p} y_{j-t} \beta_{j-t}. \]

By algebra, we have

\[ X_{tv} = \tilde{Z}_{tv} + \sum_{i=1}^{l} \delta_i X_{i-t} \quad (28) \]

The triangle inequality and the Cauchy–Schwarz inequality give

\[ \left( \frac{1}{T} \sum_{i=1}^{T} X_{iv}^2 \right)^{1/2} \geq \left( \frac{1}{T} \sum_{i=1}^{T} \left[ \sum_{i=1}^{l} \delta_i X_{i-t} \right]^2 \right)^{1/2} - \left( \frac{1}{T} \sum_{i=1}^{T} Z_{tv}^2 \right)^{1/2}, \]
and
\[
\frac{1}{T} \left( \sum_{j=1}^{T} \left( \frac{\sum_{i=1}^{T} \delta_i X_{i-j} - \nu}{\left( \sum_{i=1}^{T} \delta_i \right) \left( \sum_{i=1}^{T} X_{i-j}^2 \right)} \right) \right)^2 \leq \frac{1}{T} \sum_{i=1}^{T} \left( \frac{\sum_{i=1}^{T} \delta_i^2}{\left( \sum_{i=1}^{T} \delta_i \right) \left( \sum_{i=1}^{T} X_{i-j}^2 \right)} \right) \leq p \left( \sum_{i=1}^{T} \delta_i^2 \right) \left( \frac{1}{T} \sum_{i=1}^{T} X_{i-j}^2 \right) + o(1),
\]
respectively, where \(o(1)\) holds uniformly in \(\nu\) for \(\|\nu\| = \epsilon\). Hence, it suffices to show
\[
D = \lim_{T \to \infty} \inf_{\|\nu\| = \epsilon} \frac{1}{T} \sum_{i=1}^{T} Z_{\nu i}^2 > 0 \text{ a.s.} \tag{30}
\]

We have
\[
D = \lim_{T \to \infty} \inf_{\|\nu\| = \epsilon} \left[ \frac{1}{T} \sum_{i=1}^{T} Z_{\nu i}^2 + \frac{1}{T} \sum_{i=1}^{T} B_{\nu i}^2 \right],
\]

because the cross-product term \(\frac{2}{T} \sum_{i=1}^{T} Z_{i} B_{\nu i}\) is the average of mean zero strong mixing \(r\nu\)'s, and hence, converges to zero a.s. by the SLLN.

Next, we write
\[
D = \min \{D_1, D_2\}, \tag{32}
\]

where
\[
D_1 = \lim_{T \to \infty} \inf_{\|\nu\| = \epsilon} \frac{1}{T} \sum_{i=1}^{T} \left[ Z_{\nu i}^2 \right],
\]

\[
D_2 = \lim_{T \to \infty} \inf_{\|\nu\| = \epsilon} \frac{1}{T} \sum_{i=1}^{T} \left[ Z_{\nu i}^2 \right].
\]

\(v^1 \equiv (v_1, \ldots, v_p)'\), and \(b\) is a constant in \([0, c)\).

Now, we have
\[
D_1 \geq \lim_{T \to \infty} \inf_{\|\nu\| = \epsilon} \frac{1}{T} \sum_{i=1}^{T} Z_{\nu i}^2 = \lim_{T \to \infty} \inf_{\|\nu\| = \epsilon} \sum_{j=1}^{p} v_j^2 \left( \frac{1}{T} \sum_{i=1}^{T} \delta_i^2 \right),
\]

where the equality holds because the cross-product terms are averages of mean zero strong mixing \(r\nu\)'s, and hence, converge to zero a.s. by the SLLN.
Also, by the SLLN and A2,

\[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} e_t^2 \geq \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} e_t^2 1(e_t^2 \leq M_1) \]

\[ = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E e_t^2 1(e_t^2 \leq M_1) > 0 \text{ a.s} \quad (34) \]

Thus, \( D_1 > 0 \) a.s.

In addition,

\[ D_2 \geq \lim_{T \to \infty} \inf_{\|v\| \leq c} \frac{1}{T} \sum_{t=1}^{T} [B_{t_{1v}} - B_{2tv}]^2 1(\|\vec{\omega}_t\| \leq M_2) \]

\[ \geq H_1(b) + H_2(b), \quad (35) \]

where

\[ H_1(b) = \lim_{T \to \infty} \inf_{\|v\| \leq c} \frac{2}{T} \sum_{t=1}^{T} B_{t_{1v}} B_{2tv} 1(\|\vec{\omega}_t\| \leq M_2), \]

and

\[ H_2(b) = \lim_{T \to \infty} \inf_{\|v\| \leq c} \frac{1}{T} \sum_{t=1}^{T} B_{2tv}^2 1(\|\vec{\omega}_t\| \leq M_2) \]

We can make \(|H_1(b)| \) arbitrarily small by choosing \( b \) sufficiently small, since

\[ |H_1(b)| \leq b \cdot \left[ 2pM_2^2cK \frac{3}{2} \|\| \sum_{i=0}^{p} \delta_i \| \right]. \quad (36) \]

Also, \( H_2(b) \) is bounded away from zero for \( b \leq \sqrt{3c/2} \), since

\[ H_2(b) \geq \lim_{T \to \infty} \inf_{\|v\| \leq c} (\delta^* \otimes v^*) \left[ \frac{1}{T} \sum_{t=1}^{T} \vec{W}_t \vec{W}_t^* 1(\|\vec{W}_t\| \leq M_2) \right] (\delta^* \otimes v^*) \quad (37) \]

\[ \geq \lim_{T \to \infty} (\delta^* \otimes v^*) \left[ \frac{1}{T} \sum_{t=1}^{T} E \vec{W}_t \vec{W}_t^* 1(\|\vec{W}_t\| \leq M_2) \right] (\delta^* \otimes v^*) \]

\[ \geq \lim_{T \to \infty} (\delta^* \otimes v^*) \left[ \frac{1}{T} \sum_{t=1}^{T} E \vec{W}_t \vec{W}_t^* 1(\|\vec{W}_t\| \leq M_2) \right] (\delta^* \otimes v^*) \]

\[ \geq d > 0, \quad (38) \]
for some constant \( d \) that does not depend on \( b \) where \( \delta^* = (-1, \delta') \), \( \psi \) is a vector that minimizes the right-hand-side of equation (37), the equality holds by the SLLN and the fact that \( ||\psi|| = \epsilon/2 \), and the third inequality holds by A3.

Equations (36) and (38) imply that \( D_2 > 0 \) a.s., and the proof is complete.

NOTES

1. In this regard, the error assumption used here is similar to those used in Andrews [6] to obtain consistency of LS slope estimates in linear regression with regressors and errors that are non-explosive, but have undefined means. The method of proof in the two papers, however, is quite different.

REFERENCES

11. Cavanagh, C. Roots local to unity. Manuscript, Department of Economics, Harvard University, 1985
28 Mandelbrot, B The variation of certain speculative prices *Journal of Business* 36 (1963) 394–419
29 Mandelbrot, B The variation of some other speculative prices *Journal of Business* 40 (1967) 393–413
33 Phillips, P C B Regression theory for near-integrated time series Cowies Foundation Discussion Paper No. 781, Yale University, 1986
40 Stock, J H Asymptotic properties of least squares estimators of cointegrating vectors Manuscript, Kennedy School of Government, Harvard University, 1985