Discussion Paper No. 672

INFERENCE IN THE EXPLOSIVE FIRST-ORDER LINEAR DYNAMIC REGRESSION MODEL

By

Ian Domowitz
Northwestern University
and
Lars Muus
Cornell University

Revised
December 1985

This work was supported by National Science Foundation grants SES - 8410630 and SES - 8410200. We are grateful to Nick Kiefer, Peter Phillips, and Gene Savin for helpful comments and suggestions.

ABSTRACT

The large-sample properties of the maximum-likelihood estimator of parameters in a dynamic regression model with Gaussian errors are investigated, for the case of explosive dynamics. It is shown that a normalized version of the observed information matrix converges in law to a nondegenerate random variable, the distribution of which is derived. The maximum-likelihood estimator is distributed asymptotically as a variance mixture of normal random variables. Standard Likelihood-ratio, Wald, and Lagrange Multiplier statistics for testing linear restrictions are shown to be chi-squared distributed under the null hypothesis, subject to standard regularity conditions. This result is applied to a test of exclusion restrictions in a stochastic consumption model.

1. INTRODUCTION

The investigation of estimation and hypothesis testing in linear stochastic difference equations has been a major area of statistical and econometric research since the seminal work of Mann and Wald [17]. The general equation of interest can be written in the form,

$$Y_t = \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{k} \beta_j x_{jt} + \varepsilon_t \quad , \quad t = 1,2,\cdots \tag{1.1}$$

where the (x_{jt}) are taken to be nonstochastic regressors (or well behaved random regressors), and the (ε_t) are random error terms with certain prespecified properties. When the roots of the characteristic equation associated with (1.1) are all strictly less than one in absolute value, the limiting behavior of various estimators of the unknown parameters has been investigated exhaustively under a wide range of assumptions on the properties of the sequences (x_{jt}) and (ε_t).

Although Mann and Wald [17] considered estimation of the parameters of the model under the assumption of stable roots, they certainly recognized the possibility of at least one root greater than unity in absolute value, and the difficulties associated with this problem. In the case of no exogenous regressors and $p = 1$, White [25], Anderson [1], and Rao [21] have studied the limiting behavior of the maximum likelihood estimator when $|\alpha_1| > 1$.

Explosive processes are not uncommon in applied econometric work. In the important paper by Hall [13] on the stochastic properties of the permanent income hypothesis, the equation

$$c_t = \alpha_0 + \alpha_1 c_{t-1} + \beta' x_t + \varepsilon_t \tag{1.2}$$

is considered, where c_t is per capita consumption of nondurables and services, and x_t is a vector of exogenous variables. The hypothesis $\beta = 0$ is of interest. Consumption, however, grows exponentially (i.e., $\alpha_1 > 1$) and Hall's estimates confirm this. Hall, however, applies standard estimation and hypothesis testing techniques, with respect to inference, regardless of this violation of the standard ergodicity assumption with respect to the sequence $\{c_t\}$. Other examples within consumption analysis include the work of Davidson and Hendry [7], and Daly and Hadjimatheou [6]. Examples can be found in other fields. As shown in Domowitz and Muus [9], explosiveness is predicted by theoretical models of investment behavior, and has been investigated empirically by Blanchard [4].

The sole treatment of statistical inference based on the model (1.1), which allows for a single unstable root and the inclusion of nonstochastic regressors, is that of Fuller, Hasza and Goebel [12]. In their work, (ε_t) is a Gaussian white noise error process. Their results are derived via a complicated orthogonalization procedure, which facilitates proofs of the main theorems, but presents several complications. First, the conditions under which the results are valid are stated in terms of the orthogonalized variables, making difficult the verification of any such requirements on the primitive variables of the model and its original structure. Second, the inverse transformation used to obtain results on the original model parameters involves random scaling factors, which conceal the structure of the unconditional limiting distribution of the maximum likelihood estimator of the unknown parameter vector. Finally the use of an orthogonalization argument precludes any immediate extensions to the case of nonlinear models in the future.

In this paper, we consider maximum likelihood estimation of (1.1) for the case p = 1 and $|\alpha_1| > 1$, using an extension of LeCam's [16] work on locally asymptotically normal families of distributions. This extension involves variance mixtures of normal distributions, and is exposited in varying levels of detail and generality in Basawa and Scott [3], Domowitz [8], and Domowitz and Muus [10]. The usefulness of this framework is well illustrated for the case of pure autoregressive models by Basawa and Brockwell [2]. The regularity conditions we impose are quite standard, and are applied directly to the primitives of the regression model. The concept of random normalization factors is avoided, given our technique of proof.

Our principal results are two. From the theoretical point of view, we characterize the limiting distribution of the maximum likelihood estimator as a multivariate normal distribution, whose variance structure is that of a nondegenerate random variable. This finding is in sharp contrast to results obtained for stable (ergodic) models, in which the variance of the limiting distribution is always nonstochastic. The distribution of the parameter variance-covariance structure is derived. The major result of practical importance is that the likelihood ratio, Wald, and, suitably modified, score statistics, for tests of linear restrictions, remain chi-square distributed under the null hypothesis. The non-null distributions for sequences of local alternatives are extremely complex, however, and are beyond the scope of this paper.

We begin with an exposition of a basic proposition, which describes the framework upon which we base our results. Maximum likelihood estimation of the linear dynamic regression model with normal errors is considered in

Section 3, with an emphasis on the convergence properties of the (normalized) observed Fisher information matrix. Conditions are given, under which the information converges to a well-defined nondegenerate random variable, the distribution of which is known. We may then characterize the asymptotic distribution of the parameter estimator as a particular variance mixture of normals. This result is used in Section 4 to develop the distribution of commonly used test statistics, under the null hypothesis. Finally, in Section 5 we consider the extent to which the analysis performed by Hall [13] and others who study (1.2) remains valid, once explosiveness is taken into account. This is accomplished by a simple empirical example.

2. A Framework For Analysis

The asymptotic theory which underlies standard analyses of maximum-likelihood estimators is based on the concept of locally asymptotically normal families of distributions.[1] Crudely stated, the score tends in law to a random variable with the normal distribution, and the variance of this asymptotic distribution is a degenerate random variable, i.e., nonstochastic. In explosive dynamic linear models, the asymptotic distribution of the estimator need not be normal, and may have infinite variance.[2] More generally, however, the limiting distribution of the score is characterized by a variance that is a nondegenerate random variable, as exposited by Basawa and Scott [3]. This insight has led to the introduction of a concept of locally asymptotically mixed normal (LAMN) families of distributions. Our work on the linear regression model relies on such a conceptualization. General discussion of this framework is contained in Basawa and Scott [3], Domowitz [8], and Domowitz and Muus [10]. In this section, we simply state and discuss a basic proposition, which illustrates the fundamental notions, and provides the cornerstone of our subsequent analysis. The discussion abstracts from a precise measure-theoretic underpinning, in order to present an uncluttered exposition of the concepts involved. All events are simply assumed to take place on a well-defined probability space, (Ω, B, P). In particular, all relevant functions are taken to be measurable throughout this paper, without further comment. Given our continuity conditions, and that we work on finite-dimensional Euclidean space, this is without further loss of generality.

Let $\ell_T(\theta)$ denote the log likelihood function based on T observations of a random variable, y. The basic results are expressed in terms of the log likelihood ratio,

$$L_T(\theta_1,\theta_2) = \ell_T(\theta_2) - \ell_T(\theta_1)$$

where θ_1 and θ_2 take values on a subset of finite-dimensional Euclidean space. Let $\nabla\ell_T(\theta)$ denote the score vector, i.e., the px1 vector of derivatives of ℓ_T with respect to θ. The matrix $I_T(\theta)$ is defined to be the negative of the matrix of second derivatives of ℓ_T with respect to θ. We also define $D_T(\theta)$ to be a pxp diagonal matrix, with elements $E[\partial\ell_T(\theta)/\partial\theta_i]^2$, $i = 1,\cdots,p$, arranged along the diagonal, and zeroes elsewhere, when the expectation exists. Finally, set

$$s_T(\theta) = D_T(\theta)^{-1/2}\nabla\ell_T(\theta)$$

and

$$I_T^*(\theta) = D_T(\theta)^{-1/2} I_T(\theta)\ D_T(\theta)^{-1/2}.$$

Note that the quantity $s_T(\theta)$ is simply a normalized score vector, while the matrix $I_T^*(\theta)$ resembles the usual estimator of the information matrix, up to a normalizing constant. It will be seen that such a normalization is crucial for obtaining useful asymptotic results, similar to the analysis of maximum probability estimators by Weiss and Wolfowitz [24]. We now impose some regularity conditions.

<u>Assumption 1</u>. For every θ interior to the parameter space Θ, a compact subset of R^p, $\ell_T(\theta)$ is twice continuously differentiable, almost surely (a.s.). The expectations $E[\partial\ell_T(\theta)/\partial\theta_i]^2$, $i = 1,\cdots, p$, are finite for each finite T.

<u>Assumption 2</u>. $D_T(\theta)^{-1} \to 0$ as $T \to \infty$, uniformly in $\theta \in \Theta$.

<u>Assumption 3</u>. There exists a pxp random matrix, $I^*(\theta)$, which is a.s. finite and positive definite, such that $I_T^*(\theta) \overset{w}{\to} I^*(\theta)$, uniformly in θ, where $\overset{w}{\to}$ denotes weak convergence (convergence in distribution).

Assumption 1 is a standard regularity condition, which may hold as $E[\partial\ell_T(\theta)/\partial\theta_i]^2 \to \infty$, $T \to \infty$. Assumption 2 is essentially a consistency requirement, which must be fulfilled in any application of interest. Taken together, the first two assumptions basically are used to justify a Taylor series approximation to the log likelihood ratio.

The crucial condition in practice is Assumption 3. As of now, we know of no way of fulfilling this condition except by construction. Any such construction relies crucially on the assumed density of the observations, unlike standard asymptotic analysis. As a brief illustration, consider a first-order autoregression with no constant term, driven by independent standard normal random errors. If the autoregression is stable, I_T^* converges to unity (in probability), a standard result for which normality is not required. If the autoregression is explosive, I_T^* converges weakly to a random variable with the $\chi^2(1)$ distribution, which we shall establish as a corollary to our results in the next section.[3]

We now state the proposition central to our analysis. The proof is due to Basawa and Scott [3].[4]

Proposition 1. Under Assumptions 1-3, for all T sufficiently large, and all θ interior to Θ,

$$\text{I.} \quad L_T(\theta + D_T(\theta)^{-1/2} h_T, \theta) = h_T' s_T(\theta) - \tfrac{1}{2} h_T' I_T^*(\theta) h_T + o_p(1)$$

for all bounded sequences $\{h_T\}$ of elements of R^p, where $I_T^*(\theta)$ is a.s. positive definite.

$$\text{II.} \quad \left(s_T(\theta),\ I_T^*(\theta)\right) \xrightarrow{w} \left(I^*(\theta)^{1/2} Z,\ I^*(\theta)\right)$$

where Z is a px1 standard normal vector, independent of $I^*(\theta)$.

When a likelihood sequence $\{\ell_T\}$ satisfies the conditions of this proposition, we say that such a sequence satisfies the LAMN conditions. The first statement simply justifies a series expansion of the log likelihood ratio. The second defines the limiting distribution of the score, when the distribution of $I^*(\theta)$ is known. In the case of nonstochastic $I^*(\theta)$, the proposition provides a workable definition of the concept of locally asymptotically normal families of distributions, for most cases of interest in econometrics.

Given this result, it is hardly surprising that

$$\left(D_T(\theta)^{1/2} (\hat{\theta}_T - \theta),\ I_T^*(\theta) \right) \overset{w}{\rightarrow} \left(I^*(\theta)^{-1/2} Z,\ I^*(\theta) \right)$$

where $\hat{\theta}_T$ is the maximum likelihood estimator.[5] We omit details. Thus, the basic difficulty in calculating the limiting distribution of the estimator lies in the verification of a distribution for $I^*(\theta)$, followed by the calculation of the mixture distribution. This would be required, for example, if the usual standard errors for coefficients were desired in a regression framework. As might be imagined, such an exercise can be extremely difficult. Fortunately, the behavior of the likelihood ratio statistic under the null hypothesis can be characterized if only the conditions of the above proposition are satisfied, without expliciting calculating the mixture. To this we now turn, for the case of the dynamic linear regression model.

3. THE MODEL

Consider the nonergodic process,

$$y_t = \alpha\, y_{t-1} + \beta' x_t + \varepsilon_t \quad , t = 1, \cdots , T \tag{3.1}$$

where $y_o = 0$ and $|\alpha| > 1$; $x_t = [x_{1t} \cdots x_{kt}]'$ is a k x 1 vector of nonstochastic regressors, and (ε_t) is a Gaussian white noise process (i.e., $\varepsilon_t \sim N(0,\sigma^2)$, $\sigma^2 > 0$). The unknown parameters are $\alpha, \beta = [\beta_1 \cdots \beta_k]'$, and σ^2. Our results are valid for arbitrary y_0, but at the expense of considerably more notation, which we eschew here. The case of $|\alpha| < 1$ is also covered by the analysis, but the unit root problem, $|\alpha| = 1$, is not covered.[6]

From (3.1) we obtain by recursion,

$$y_t = \sum_{i=0}^{t-1} \alpha^i(\beta' x_{t-i} + \varepsilon_i) \tag{3.2}$$

Hence, if $\mu_t = E[y_t]$ and $\sigma_t^2 = \mathrm{Var}\,[y_t]$,

$$\mu_t = \sum_{i=0}^{t-1} \alpha^i \beta' x_{t-i} \tag{3.3}$$

$$\sigma_t^2 = \sigma^2 \frac{\alpha^{2t}-1}{\alpha^2-1} \tag{3.4}$$

Obviously, $\sigma_t^2 \to \infty$ as $t \to \infty$, reflecting the explosive nature of the process.

In order to find the limiting distribution of the maximum likelihood estimator and associated test statistics we impose the following regularity conditions on the (nonstochastic) regressors.

<u>Assumption 4.</u> (i) $x_{it} = o(\gamma^t)$ for all $|\gamma| > 1$, $i = 1,\cdots,k$.

(ii) $\sum_{t=1}^{T} x_{it}x_{jt} / (\sum_{t=1}^{T} x_{it})^{1/2} (\sum_{t=1}^{T} x_{jt}^2)^{1/2} \to q_{ij}$

for $T \to \infty$, where the kxk matrix $Q = [q_{ij}]_{ij=1,\cdots,k}$ is positive definite, with finite elements.

(iii) $\sup\limits_{1<t<T} \dfrac{x_{it}^2}{\sum_{s=1}^{T} x_{is}^2} \to 0$ for $T \to \infty$, $i = 1,\cdots,k$.

These assumptions are virtually identical to the usual regularity conditions imposed on regressors when analyzing the limiting behavior of the least-squares estimator. While (i) rules out exponential growth in the regressors, polynomial time trends are allowed within our framework. Condition (ii) is the standard identification condition, ruling out perfect multicollinearity. The last condition ensures the nonsingularity of the limiting normalized Fisher information matrix, to be defined below. Similar regularity conditions may be found in standard econometrics texts, such as Judge, et al. [15, p. 162], where they are sometimes referred to as "Grenander conditions."

The log-likelihood function (apart from a constant) corresponding to the dynamic regression model (3.1) is given by

$$\ell_T(\alpha,\beta,\sigma^2) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \alpha y_{t-1} - \beta' x_t)^2 \tag{3.5}$$

By differentiation we obtain the score vector

$$\nabla\ell_T(\alpha,\beta,\sigma^2) = \begin{bmatrix} 1/\sigma^2 \sum_{t-1}^{T} \varepsilon_T y_{t-1} \\ 1/\sigma^2 \sum_{t=1}^{T} \varepsilon_t x_t \\ -T/2\sigma^2 + 1/2\sigma^4 \sum_{t=1}^{T} \varepsilon_t^2 \end{bmatrix} \tag{3.6}$$

from which we obtain the well known maximum likelihood estimators by solving the likelihood equations $\nabla\ell_T(\alpha,\beta,\sigma^2) = 0$.

The observed Fisher information matrix is given by

$$I_T(\alpha,\beta,\sigma^2) = \begin{bmatrix} 1/\sigma^2 \sum_{t=1}^{T} y_{t-1}^{\,2} & 1/\sigma^2 \sum_{t=1}^{T} x_{1t}y_{t-1} & \cdots & 1/\sigma^2 \sum_{t=1}^{T} x_{kt}y_{t-1} & 1/\sigma^4 \sum_{t=1}^{T} \varepsilon_t y_{t-1} \\ & 1/\sigma^2 \sum_{t=1}^{T} x_{1t}^2 & \cdots & 1/\sigma^2 \sum_{t=1}^{T} x_{1t}x_{kt} & 1/\sigma^4 \sum_{t=1}^{T} \varepsilon_t x_{1t} \\ & & & & \vdots \\ & & & 1/\sigma^2 \sum_{t=1}^{T} x_{kt}^2 & 1/\sigma^4 \sum_{t=1}^{T} \varepsilon_t x_{kt} \\ & & & & -T/2\sigma^4 + 1/\sigma^6 \sum_{t=1}^{T} \varepsilon_t^2 \end{bmatrix}$$

and the diagonal normalization matrix is

$$D_T(\alpha,\beta,\sigma^2) = \begin{bmatrix} 1/\sigma^2 \sum_{t=1}^{T} Ey_{t-1}^2 & & & & \\ & 1/\sigma^2 \sum_{t=1}^{T} x_{1t}^{\,2} & & & \\ & & \ddots & & \\ & & & 1/\sigma^2 \sum_{t=1}^{T} x_{kt}^{\,2} & \\ & & & & T/2\sigma^4 \end{bmatrix}$$

Thus, we obtain the normalized Fisher information matrix

$I_T^*(\alpha,\beta,\sigma^2) = D_T(\alpha,\beta,\sigma^2)^{-1/2} I_T(\alpha,\beta,\sigma^2) D_T(\alpha,\beta,\sigma^2)^{-1/2}$ given by

$$I_T^*(\alpha,\beta,\sigma^2) = \begin{bmatrix} A_T & B_{1T} & \cdots & B_{kT} & C_T \\ & F_{11T} & \cdots & F_{kT} & H_{1T} \\ & & & & \vdots \\ & & & F_{kkT} & H_{kT} \\ & & & & J_T \end{bmatrix} \qquad (3.7)$$

where,

$$A_T = \sum_{t=1}^{T} y_{t-1}{}^2 \,/\, E[\sum_{t=1}^{T} y_{t-1}{}^2]$$

$$B_{iT} = \sum_{t=1}^{T} x_{it} y_{t-1} \,/\, (E[\sum_{t=1}^{T} y_{t-1}{}^2])^{1/2} (\sum_{t=1}^{T} x_{it}{}^2)^{1/2}$$

$$C_T = \sqrt{2} \sum_{t=1}^{T} \varepsilon_T y_{t-1} \,/\, \sigma \sqrt{T}\, (E[\sum_{t=1}^{T} y_{t-1}{}^2])^{1/2}$$

$$F_{ijT} = \sum_{t=1}^{T} x_{it}\, x_{jt} \,/\, (\sum_{t=1}^{T} x_{it}^2)^{1/2} (\sum_{t=1}^{T} x_{jt}{}^2)^{1/2}$$

$$H_{iT} = \sqrt{2} \sum_{t=1}^{T} \varepsilon_t x_{it} \,/\, \sigma \sqrt{T}\, (\sum_{t=1}^{T} x_{it}^2)^{1/2}$$

$$J_T = (2/T\sigma^2) \sum_{t=1}^{T} \varepsilon_t^2 - 1$$

The general form of the matrix I_T^* is not new in the analysis of the least-squares estimator, although our use of expectations in D_T is somewhat unusual. Such a formulation has proved convenient for handling cases in which the regressor cross-product matrix, normalized by $1/T$, does not have a finite nonsingular limit [15, p. 161]. Its primary application has been to regressions with nonstochastic regressors which may grow as polynomials in time. A similar form is used by Fuller, et al. [12] and Wei [23] for vectors of suitably transformed or orthogonalized regressors. We use expectations to provide a nonstochastic damping factor for the exponential growth in the lagged dependent variable.

The convergence of $I_T^*(\alpha,\beta,\sigma^2)$ is established by considering the limiting properties of each individual element. Intermediate lemmas and proofs are relegated to the appendix. One intermediate result is of independent interest, for two reasons. It can be shown to hold for more general situations than that considered here, and it allows us to link our results on the regression model to those concerning simple autoregressions, already in the literature.

Lemma 1. Given model (3.1) and Assumption 4,

$$\sum_{t=1}^{T-1} (y_t - \mu_t)^2 / E[\sum_{t=1}^{T} y_{t-1}^2] \xrightarrow{a.s.} W^2 \ , \text{ for } T \longrightarrow \infty$$

where $W \sim N(0,1)$.

This result is valid for lagged dependent variable structures of any order, with a single explosive root. Under suitable additional conditions, it also holds for a nonlinear regression, with separable, linearly explosive dynamics. We use it to contrast our results with those of Anderson [1], below. Given that properly normalized terms in μ_t converge to zero, the lemma is the cornerstone of our first theorem.

Theorem 1. Given model (3.1) and Assumption 4, $I_T^*(\alpha,\beta,\sigma^2) \xrightarrow{a.s.} I^*$ for $T \to \infty$, where,

$$I^* = \begin{bmatrix} W^2 & 0 & 0 & \cdots & 0 & 0 \\ & 1 & q_{12} & \cdots & q_{1k} & 0 \\ & & 1 & \cdots & q_{2k} & 0 \\ & & & & & \vdots \\ & & & & 1 & 0 \\ & & & & & 1 \end{bmatrix}$$

and $W^2 \sim \chi_1^2$, while the q_{ij} are defined in Assumption 4.

It is interesting, but not too surprising, that I^* is a block diagonal matrix. For the particular case at hand, the only random element in I^* is W^2, distributed as a central chi-square with one degree of freedom.

Derivation of this matrix satisfies Assumption 3 of Section 2 by construction. For our linear model, Assumptions 1 and 2 are rather trivially satisfied, leading to the following result.

Theorem 2. Given model (3.1) and Assumption 4, the sequence of likelihoods $\{\ell_T(\alpha,\beta,\sigma^2)\}$ defined by equation (3.5), satisfies the LAMN conditions.

The importance of this result lies in its implications for the asymptotic distribution of the score. Letting $\theta = (\alpha\ \beta'\ \sigma^2)'$, we have shown that

$$s_T(\theta) \xrightarrow{w} N(0,\ I^*(\theta))\ ,\ T \to \infty \tag{3.8}$$

where $s_T = D_T(\theta)^{-1/2}\, \nabla \ell_T$. The distribution of the score vector can, therefore, be characterized as a variance mixture of normals, the mixture distribution involving the chi-square distribution. This result forms the basis of standard test procedures, which are considered in the next section.

Before turning to issues of hypothesis testing, we consider a specialization of (3.1), which has received attention in the literature, most notably by Anderson [1]. Under the regularity conditions considered here, equation (3.8) implies

$$D_T(\theta)^{1/2}\, (\hat{\theta}_T - \theta) \xrightarrow{w} I^*(\theta)^{-1/2}\, Z \tag{3.9}$$

where $\hat{\theta}_T = (\hat{\alpha}_T\ ,\ \hat{\beta}_T'\ \hat{\sigma}_T^2)'$ is the maximum likelihood estimator, and Z is a standard normal random variable. In the case of a first-order autoregression, i.e., $\beta = 0$, we have

$$\left[E\ \textstyle\sum_{t=1}^{T}\ y_{t-1}^2\right]^{1/2} (\hat{\alpha}_T - \alpha) \xrightarrow{w} W^{-1}\ Z \tag{3.10}$$

Since the sequence of likelihoods is in the LAMN family, it follows, by definition, that W and Z are independent. The limiting distribution of the estimator, under this particular normalization is, therefore, Cauchy, as shown by Anderson [1, Theorem 2.7]. There exists, however, a random normalizing factor such that the limiting distribution of the estimator is normal

(Anderson [1], Theorem 2.8). This follows directly from (3.10) and Lemma 1, i.e.,

$$\left[\sum_{t=1}^{T} y_{t-1}^2\right]^{1/2} (\hat{\alpha}_T - \alpha) \xrightarrow{w} Z. \tag{3.11}$$

Similar results may be obtained for higher order autoregressions, with a single explosive root, as in the work of Rao [21] or Basawa and Brockwell [2].

4. Hypothesis Testing

We consider the case of a composite null hypothesis, in which a subset of the parameters, possibly including α, is constrained to a preassigned vector of values.[7] As noted by Engle [11], any well-defined linear hypothesis can be transformed to fit this description. We are interested in the properties of the standard Wald, likelihood ratio, and Lagrange multiplier statistics under the null hypothesis.

Let $\theta = (\theta_1' \; \theta_2')'$, where θ_1 is a $k_1 \times 1$ vector and θ_2 is a $k_2 \times 1$ vector ($k_1 + k_2 = k + 2$). The null hypothesis is taken to be $\theta_1 = \theta_1^o$, θ_2 unrestricted. Define $\tilde{\theta} = (\theta_1^{o'} \; \tilde{\theta}_2')'$ to be the maximum likelihood estimator under the null, and let $\hat{\theta}$ denote the vector of unrestricted estimates, $(\hat{\theta}_1' \; \hat{\theta}_2')'$. We suppress the T-subscripts on the estimators for notational simplicity at this stage. We similarly partition

$$I^*(\theta) = \begin{bmatrix} I_{11}^* & I_{12}^* \\ I_{21}^* & I_{22}^* \end{bmatrix} \qquad I_T^*(\theta) = \begin{bmatrix} I_{T11}^* & I_{T12}^* \\ I_{T21}^* & I_{T22}^* \end{bmatrix}$$

accordingly, where the dependence of the components on the elements of θ is understood. The matrices $I(\theta)$ and $I_T(\theta)$ are partitioned in precisely the same manner.

The conventional likelihood ratio statistic is written as

$$LR_T = -2[\ell_T(\tilde{\theta}) - \ell_T(\hat{\theta})] \tag{4.1}$$

and the Lagrange multiplier statistic is simply

$$LM_T = s_T(\tilde{\theta})' I_T^* (\tilde{\theta})^{-1} s_T(\tilde{\theta})' \tag{4.2}$$

which reduces to the more conventional form

$$LM_T = \nabla\ell_T(\tilde{\theta})' I_T(\tilde{\theta})^{-1} \nabla\ell_T(\tilde{\theta}) \tag{4.3}$$

in practical application. In terms of θ_1, this statistic may be written as

$$\nabla\ell_{1T}(\tilde{\theta})' \; I^{11}(\tilde{\theta}) \; \nabla\ell_{1T}(\tilde{\theta})$$

where $I^{11}(\theta)$ is the partitioned inverse of $I(\theta)$ and $\nabla\ell_{1T}$ is partitioned as θ_1. Finally, the Wald statistic is

$$W_T = (\hat{\theta}_1 - \theta_1^o)' \; [I_{11T}^* (\hat{\theta}) - I_{12T}^* (\hat{\theta}) \; I_{22T}^*(\hat{\theta})^{-1} \; I_{21T}^*(\hat{\theta})] \tag{4.4}$$

It is well known that all three test statistics have the same asymptotic distribution under the null hypothesis, in standard settings. We reaffirm this result for the nonergodic regression model.

From Proposition 1 we get the validity of the Taylor series expansion,

$$\ell_T(\theta_1,\theta_2) = \ell_T(\tilde{\theta}_1,\tilde{\theta}_2) + s_T(\tilde{\theta}_1,\tilde{\theta}_2)'(\theta-\tilde{\theta})$$

$$- \tfrac{1}{2}(\theta-\tilde{\theta})' D_T(\bar{\theta})^{-1/2} I_T(\bar{\theta})\ D_T(\bar{\theta})^{-1/2}(\theta-\tilde{\theta}) + o_p(1)$$

$$= \ell_T(\tilde{\theta}_1,\tilde{\theta}_2) - \tfrac{1}{2}(\theta-\tilde{\theta})' D_T\ (\bar{\theta})^{-1/2} I_T(\bar{\theta})\ D_T(\bar{\theta})^{-1/2}(\theta-\tilde{\theta}) + o_p(1) \quad (4.5)$$

remembering that $s_T(\tilde{\theta}_1,\tilde{\theta}_2) = 0$, and where $\bar{\theta}$ lies on the line segment between θ and $\tilde{\theta}$. From (4.5) we obtain,

$$LR_T = 2[\ell_T(\tilde{\theta}) - \ell_T(\hat{\theta})] = (\hat{\theta}-\tilde{\theta})' D_T^{-1/2}(\bar{\bar{\theta}})\ I_T(\bar{\bar{\theta}})\ D_T^{-1/2}(\bar{\bar{\theta}})(\hat{\theta}-\tilde{\theta})$$

$$+ o_p(1) \qquad (4.6)$$

where $\bar{\bar{\theta}}$ lies on the line segment between $\hat{\theta}$ and $\tilde{\theta}$. Since the functions D_T and I_T are continuous and $\hat{\theta}$ and $\tilde{\theta}$ are consistent under the null, a standard argument establishes the asymptotic equivalence between the likelihood ratio statistic and Wald statistic.

Applying a first-order Taylor series expansion to the score vector we obtain,

$$s_T(\theta) = s_T(\hat{\theta}) - D_T(\ddot{\theta})^{-1/2} I_T(\ddot{\theta})(\hat{\theta}-\theta) + o_p(1) \qquad (4.7)$$

where $\ddot{\theta}$ lies on the line segment between θ and $\hat{\theta}$. Evaluation at the point

$\theta = \tilde{\theta}$ yields,

$$s_T(\hat{\theta}) = D_t(\dot{\theta})^{-1/2} I_T(\dot{\theta}) (\hat{\theta}-\tilde{\theta}) + o_p(1) \tag{4.8}$$

where $\dot{\theta}$ lies on the line segment between $\hat{\theta}$ and $\tilde{\theta}$. From (4.6) we then obtain,

$$(\hat{\theta}-\tilde{\theta}) D_T(\bar{\bar{\theta}})^{-1/2} I_T(\bar{\bar{\theta}}) D_T(\bar{\bar{\theta}})^{-1/2} (\hat{\theta}-\tilde{\theta}) + o_p(1)$$

$$= s_T(\hat{\theta}) I_T^*(\dot{\theta})^{-1} s_T(\hat{\theta}) + o_p(1) \tag{4.9}$$

from which we obtain the asymptotic equivalence between the Wald test statistic and LM test statistic. Hence, we have established the asymptotic equivalence of all three test statistics.

From (3.8) we obtain by standard arguments, resting on the continuity of the functions D_T and I_T as well as the consistency of the estimators involved, that

$$s_T(\hat{\theta})' I_T^*(\dot{\theta})^{-1} s_T(\hat{\theta}) + o_p(1) = s_{1T}(\hat{\theta}) I_T^{*11}(\dot{\theta}) s_{1T}(\hat{\theta}) + o_p(1)$$

$$\xrightarrow{w} \chi^2(k_1)$$

Hence we have established the following result.

Theorem 2. Under Assumption 4

(i) LR_T, LM_T, and W_T are asymptotically equivalent.

(ii) $LR_T \xrightarrow{w} \chi^2(k_1)$

(iii) $LM_T \xrightarrow{w} \chi^2(k_1)$

(iv) $W_T \xrightarrow{w} \chi^2(k_1)$

under the null hypothesis.

In short, the conventional test procedures carry over to the nonergodic case. We have little to say about behavior under the alternative, however. The usual drift terms considered in an analysis of local power properties now involve nondegenerate random variables with the chi-squared distribution, even in the limit. This issue is deferred to future research.

5. An Empirical Analysis

Several examples of explosive processes have occurred in applied econometric analysis. An important example is provided by Hall [13], who studies the stochastic properties implied by the life cycle-permanent income hypothesis. On the basis of a quadratic utility function $u(c_t) = \frac{1}{2}(\alpha - c_t)^2$, $c_t < \alpha$, where α denotes the bliss level of consumption (c_t), Hall derives the estimating equation.[8]

$$\Delta c_t = \varepsilon_t \tag{5.1}$$

where (ε_t) is an uncorrelated stochastic process, and Δ denotes the difference operator, $\Delta x_t = x_t - x_{t-1}$. To investigate the empirical validity of the above result Hall [13] considers the regression,

$$c_t = \alpha_o + \alpha_1 c_{t-1} + \beta' x_{t-1} + \varepsilon_t \tag{5.2}$$

where x_{t-1} is a kx1 vector of regressors in the household's information set at time t-1. Under the null hypothesis, $\beta = 0$, and we leave α_1 unconstrained.

In the empirical implementation, the standard consumption measure (including nondurables and services per capita) is used. Results are obtained using seasonally adjusted U.S. time series data.

Since it is an empirical fact that consumption (per capita) grows exponentially rather than linearly, it is hardly surprising that Hall [13] obtains values of α_1 in excess of unity. In this case, the techniques developed in the previous sections are needed in order to test $\beta=0$. Applying seasonally adjusted aggregate data over the time period 1950:1-1983:4, we obtain the results given in Table 1. Notice that the conventional standard errors have not been reported, since these are meaningless within the framework of the linear explosive regression process. In order to carry out the orthogonality test proposed by Hall [13], we have used $x_t = (c_{t-2}, c_{t-3}, c_{t-4}, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4})$, where y denotes real personal disposable income per capita. In column 1 we report the estimates with the restriction $\beta=0$ imposed, while the unrestricted estimates are reported in column 2. As expected, $\alpha_1 > 1$ in both cases. As shown in the previous section, the ordinary hypothesis testing procedures remain valid under the maintained hypothesis $|\alpha_1| > 1$. Computing the likelihood ratio statistic LR for the hypothesis H_0: $\beta = 0$ against the alternative H_1: $\beta \neq 0$, we obtain LR = 6.47. Thus, the marginal probability of rejection is 0.25, leading us to conclude that we cannot reject the null at a 75% - significance level. As found by Hall [13], the life cycle-permanent income hypothesis cannot be rejected using the above set of exogenous variables (obviously, a different choice of x_t could lead to such a rejection). It could very well be argued that the quadratic utility function specification induces explosiveness by not taking into account that the bliss level of consumption rises over time. However, we have carried out the above empirical analysis mainly to illustrate the econometric problem and techniques discussed in the previous sections.

Table 1

Hall's consumption model 1950:1-1983:4

$$c_t = \alpha_0 + \alpha_1 c_{t-1} + \beta' x_t + \varepsilon_t$$

	(1)	(2)
constant	-1.2998	-3.4401
c_{t-1}	1.0054	1.1702
c_{t-2}	-	-0.0785
c_{t-3}	-	-0.1289
c_{t-4}	-	-0.2038
y_{t-1}	-	0.0356
y_{t-2}	-	-0.1088
y_{t-3}	-	0.0656
y_{t-4}	-	-0.0022
SSE	26117	23409
T	136	136

Note: Data are taken from the National Income and Product Accounts. SSE is the sum of squared errors.

6. SUMMARY AND CONCLUSIONS

In the present paper we have studied the limiting distribution of the maximum likelihood estimator in the linear nonergodic regression model (3.1). In particular, we have demonstrated that the limiting distribution of the vector of maximum likelihood estimators is a variance mixture of normal distributions, the mixture distribution involving the chi-squared distribution in an essential way. Furthermore, we have shown that the traditional hypothesis testing procedures carry over in the explosive case. We note the simplicity in testing parametric restrictions in the model due to the block diagonality of the limiting normalized Fisher information matrix.

Our results on asymptotic inference obviously suggest the need for theoretical and numerical investigations of the small-sample properties of such commonly used statistics. Theoretical progress for the AR(1) model, without regressors, has been made by Satchell [22]. Our results indicate that his suggestion concerning expansions around the distribution of the test statistic, rather than around the normal distribution, has merit and should be explored ([22], p. 1283). Nankervis and Savin [19] have already carried out a serious Monte Carlo study of the problem. Small sample properties are found to be quite good when the regressors are independent normal, although convergence is slow for tests associated with the autoregressive parameter. In the case of nonstochastic regressors considered here (a time trend is used in the study), their results are not quite as promising for very small samples. Better performance is obtained for larger values of the autoregressive coefficient. At T = 100, however, the F-statistic performs quite well, as predicted by the large-sample theory presented here.

Several extensions of the model studied here suggest themselves. It would be interesting to establish the appropriate regularity conditions and derive the asymptotic distribution of the maximum likelihood estimators in models containing stochastic regressors, nonlinear regression functions, and general ARMAX structures. In addition it would be of practical importance to establish the limiting properties of the general least squares estimator (not assuming normality of the error term), as well as establishing the distribution of the test statistics under (local) alternatives. These, and other extensions, will be the topic of further research.

Appendix

In the present section we shall establish the result stated in Theorem 1. We shall prove the convergence of the normalized Fisher information matrix $I_T^*(\alpha,\beta,\sigma^2)$ by considering in turn each element of the matrix.

By Assumption 4 (ii) we immediately get,

$$\sum_{t=1}^T x_{it}\, x_{jt}/(\sum_{t=1}^T x_{it}^2)^{1/2}(\sum_{t=1}^T x_{jt}^2)^{1/2} \longrightarrow q_{ij} \text{ for } T \rightarrow \infty \tag{A.1}$$

Since $\frac{2}{T\sigma^2} \sum_{t=1}^T \varepsilon_t^2 - 1 = \frac{2}{T} \chi_T^2 - 1$ we get by Kolmogorov's Strong Law of Large Numbers,

$$\frac{2}{T\sigma^2} \sum_{t=1}^T \varepsilon_t^2 - 1 \xrightarrow{a.s.} 1 \qquad \text{for } T \rightarrow \infty . \tag{A.2}$$

Since $\sqrt{2} \sum_{t=1}^T \varepsilon_t\, x_{it}/\sigma \sqrt{T}\, (\sum_{t=1}^T x_{it}^2)^{1/2} \sim N(0,\frac{2}{T})$ we get

$$\sqrt{2} \sum_{t=1}^T \varepsilon_t x_{it}/\sigma \sqrt{T}\, (\Sigma_{t=1}^T x_{it}^2)^{1/2} \xrightarrow{a.s.} 0 \text{ for } T \rightarrow \infty . \tag{A.3}$$

We immediately obtain $E[\sqrt{2} \sum_{t=1}^T \varepsilon_t\, y_{t-1}/\sigma\sqrt{T}\, (E[\sum_{t=1}^T y_{t-1}^2])^{1/2}] = 0$

and $Var\, [\sqrt{2} \sum_{t=1}^T \varepsilon_t y_{t-1}/\sigma\sqrt{T}\, (E[\sum_{t=1}^T/y_{t-1}^2])^{1/2}] = \frac{2}{T}$. Hence, by Kolmogorov's Strong Law of Large Numbers,

$$\sqrt{2} \sum_{t=1}^T \varepsilon_t x_{it}/\sigma\sqrt{T}\, (\sum_{t=1}^T x_{it}^2)^{1/2} \xrightarrow{a.s.} 0 \text{ for } T \rightarrow \infty . \tag{A.4}$$

What remains to be shown is the convergence of the elements

$$\sum_{t=1}^{T} y_{t-1}^{2}/E[\sum_{t=1}^{T} y_{t-1}^{2}] \text{ and } \sum_{t=1}^{T} x_{it}y_{t-1}/(\sum_{t=1}^{T} x_{it}^{2})^{1/2} (E[\sum_{t=1}^{T} y_{t-1}^{2}])^{1/2}.$$

This problem is resolved by proving a number of Lemmas.

Since $y_o = 0$ we have, using 3.3,

$$\sum_{t=1}^{T} y_{t-1}^{2} = \sum_{t=1}^{T-1} (y_t - \mu_t)^2 + \sum_{t=1}^{T-1} \mu_t^2 + 2 \sum_{t=1}^{T-1} \mu_t(y_t - \mu_t) \tag{A.5}$$

and, using 3.4,

$$E[\sum_{t=1}^{T} y_{t-1}^{2}] = \sum_{t=1}^{T-1} \sigma_t^2 + \sum_{t=1}^{T-1} \mu_t^2 = \omega_T(\alpha^2-1)^{-2} \tag{A.6}$$

where,

$$\omega_T = \sigma^2[\alpha^{2T} - \alpha^2 - (T-1)(\alpha^2 -1)] + (\alpha^2 - 1) \sum_{t=1}^{T-1} \mu_t^2 \tag{A.7}$$

We shall use the following result established by Cramer and Leadbetter [5], formula (4.6).

<u>Lemma A.1.</u> Let (X_t) be a sequence of random variables, such that

$E[(X_{t+1} - X_t)^2] < \delta_t^3$ all t where (δ_t)

is a sequence of positive real numbers, such that

$\sum_{t=1}^{\infty} \delta_t < \infty$. Then (X_t) converges almost surely.

Define

$$W_t = \sqrt{\alpha^2-1}\ \sigma^{-2}(\alpha^{2t}-1)^{-1}(y_t - \mu_t) \tag{A.8}$$

Then $W_t \sim N(0,1)$, cf. formula 3.4. Furthermore, define
$W_t^* = \gamma_t\ W_t$, where

$$\gamma_t = (\alpha^{2t} - 1)^{1/2}\alpha^{-t}$$

Obviously, $\gamma_t \longrightarrow 1$ for $t \to \infty$, and

$$W_t^* = \sqrt{\alpha^2-1}\ \sigma^{-2}\ \alpha^{-t}\ (y_t - \mu_t) \tag{A.9}$$

<u>Lemma A.2.</u> $\quad W_t^* \xrightarrow{a.s.} W$ for $t \to \infty$, where $W \sim N(0,1)$.

Proof: We have,

$$E[(W_{t+1}^* - W_t^*)^2] = E[\{\frac{\sqrt{\alpha^2-1}}{\sigma\ \alpha^{t+1}}\ (y_{t+1} - \mu_{t+1}) - \frac{\sqrt{\alpha^2-1}}{\sigma\ \alpha^t}\ (y_t - \mu_t)\}^2]$$

$$= E[(\frac{\sqrt{\alpha^2-1}}{\sigma\ \alpha^{t+1}}\ \varepsilon_{t+1})^2]$$

$$= (\alpha^2-1)\ \alpha^{-2(t+1)}$$

Thus, by Lemma A.1, the almost sure convergence of (W_t^*) is established, since $\gamma_t \to 1$ for $t \to \infty$, it follows that $W \sim N(0,1)$.

<u>Lemma A.3.</u> $\omega_T \alpha^{-2T} \longrightarrow \sigma^2$ for $T \to \infty$.

Proof: We have,

$$\omega_T\, \alpha^{-2T} = \alpha^{-2T}\, \sigma^2[\alpha^{2T} - \alpha^2 - (T-1)(\alpha^2-1)]$$

$$+ \alpha^{-2T}\, (\alpha^2-1)^2 \sum_{t=1}^{T-1} \mu_t^2 \qquad \text{(A.10)}$$

Obviously, $\alpha^{-2T}\sigma^2[\alpha^{2T} - \alpha^2 - (T-1)(\alpha^2-1)] \longrightarrow \sigma^2$ for $T \to \infty$.

The second term in (A.10) converges, since

$$\alpha^{-2T}(\alpha^2-1)^2 \sum_{t=1}^{T-1} \mu_t^2 = (\alpha^2 - 1)^2 \sum_{t=1}^{T-1} \mu_t^2\, \alpha^{-2T}$$

$$= (\alpha^2 - 1)^2 \sum_{t=1}^{\infty} \mu_t^2\, \alpha^{-2T}\, \mathbf{1}_{[1,T-1]}(t)$$

$$\longrightarrow 0 \quad \text{for} \quad T \to \infty$$

by dominated convergence, since by Assumption (i), there exists $K > 0$ such that

$$\sum_{t=1}^{\infty} \mu_t^2\, \alpha^{-2T} < K \sum_{t=1}^{\infty} \alpha^{-2t} < \infty$$

(cf. formula 3.3)

Lemma A.4. $\sum_{t=1}^{T-1} (y_t - \mu_t)^2 \,/\, E[\sum_{t=1}^{T} y_{t-1}^2] \xrightarrow{a.s.} W^2$ for $T \to \infty$, where $W \sim N(0,1)$.

Proof: We have

$$\begin{aligned}
\sum_{t=1}^{T-1} (y_t - \mu_t)^2 / E[\sum_{t=1}^{T} y_{t-1}^2] &= \omega_T^{-1}(\alpha^2 - 1)^2 \sum_{t=1}^{T-1} (y_t - \mu_t)^2 \\
&= \sigma^2(\alpha^2-1) \sum_{t=1}^{T-1} \alpha^{2T}\omega_T^{-1}W_t^{*2}\, \alpha^{-2t} \\
&= \sigma^2(\alpha^2-1) \sum_{t=1}^{\infty} \alpha^{2T}\omega_T^{-1}W_t^{*2}\alpha^{-2t} 1_{[1,T-1]}(t) \\
&\to (\alpha^2-1)\, W^2 \sum_{t=1}^{\infty} \alpha^{-2t} \\
&= W^2
\end{aligned}$$

by dominated convergence, since $\alpha^{2T}\omega_T^{-1}$ is bounded, $\sum_{t=1}^{\infty} \alpha^{-2t} < \infty$, and there exists a bound $K(\omega) < \infty$, such that $|W_t^*(\omega)| < K(\omega)$ for almost every realization of $(W_t^*(\omega))$.

Convergence of the second term in (4.6) is stated in the following lemma.

Lemma A.5. $\sum_{t=1}^{T-1} \mu_t^2 / E[\sum_{t=1}^{T} y_{t-1}^2] \xrightarrow{a.s.}$ for $T \to \infty$.

Proof: We have,

$$\sum_{t=1}^{T-1} \mu_t^2 / \; E[\sum_{t=1}^{T} y_{t-1}^2] = \sum_{t=1}^{T-1} \mu_t^2 \,/\, (\sum_{t=1}^{T-1} \mu_t^2 + \sum_{t=1}^{T} \sigma_t^2) \to 0 \quad \text{for } T \to \infty$$

since $\sum_{t=1}^{T-1} \sigma_t^2 / \sum_{t=1}^{T-1} \mu_t^2 = \sigma^2[\alpha^{2T} - \alpha^2 - (T-1)(\alpha^2 - 1)] \sum_{t=1}^{T-1} \mu_t^2 \to \infty$

for $T \to \infty$ by the argument of Lemma A.3.

<u>Lemma A.6.</u> $\sum_{T=1}^{T-1} \mu_t(y_t - \mu_t) / E[\sum_{t=1}^{T} y_{t-1}^2] \xrightarrow{a.s.} 0$ for $T \to \infty$.

Proof: We have,

$$\sum_{t=1}^{T-1} \mu_t(y_t - \mu_t) / E[\sum_{t=1}^{T} y_{t-1}^2] = (\alpha^2-1)^2\omega_T^{-1} \sum_{t=1}^{T-1} \mu_t(y_t - \mu_t)$$

$$= \sigma(\alpha^2 -1)^{3/2} \sum_{t=1}^{T-1} \alpha^{2T}\omega_T^{-1} W_t^* \alpha^{-2t+T}$$

$$= \sigma(\alpha^2 -1)^{3/2} \sum_{t=1}^{\infty} \alpha^{2T}\omega_T^{-1} W_t^* \alpha^{-2t+T} 1_{[1,T-1]}(t)$$

$$\to 0 \qquad \text{for} \quad T \to \infty$$

by the dominated convergence theorem using an argument similar to that applied in the proof of Lemma 3.

Combining lemmas A.4-A.6 we finally get the convergence result, stated in the following proposition.

PROPOSITION A.7. $\sum_{t=1}^{T} y_{t-1}^2 / E\left[\sum_{t=1}^{T} y_{t-1}^2\right] \to W^2$ almost surely, where

$$W^2 \sim \chi^2(1).$$

We shall, finally, establish the convergence of the term

$\sum_{t=1}^{T} x_{it} y_{t-1} / \left(\sum_{t=1}^{T} x_{it}^2\right)^{1/2} \left(E\left[\sum_{t-1}^{T} y_{t-1}^2\right]\right)^{1/2}$. We shall use the results stated in the following lemmas.

Lemma A.8. $\sum_{t=1}^{T-1} \mu_t / \left(E\left[\sum_{t=1}^{T} y_{t-1}^2\right]\right)^{1/2} \to 0$ as $T \to \infty$.

Proof: We have,

$$\sum_{t=1}^{T-1} \mu_t \Big/ \left(E\left[\sum_{t=1}^{T} y_{t-1}^2\right]\right)^{1/2} = \sum_{t=1}^{T-1} \mu_t / \left(\sum_{t=1}^{T=1} (\mu_t^2 + \sigma_t^2)\right)^{1/2}$$

$$\to 0 \text{ for } T \to \infty.$$

Lemma A.9 $\sum_{t=1}^{T-1} (y_t - \mu_t) / \left(E\left[\sum_{t=1}^{T} y_{t-1}^2\right]\right)^{1/2}$ for $T \to \infty$

$$\to \frac{(\alpha^2-1)^{1/2}}{\alpha - 1} W, \text{ where } W \sim N(0,1)$$

Proof: From formula (A.6) we obtain,

$$\sum_{t=1}^{T-1} (y_t - \mu_t) / \left(E\left[\sum_{t=1} y_{t-1}^2\right]\right)^{1/2} = \omega_T^{-1/2} (\alpha^2-1) \sum_{t=1}^{T-1} (y_t - \mu_t)$$

$$= \sum_{t=1}^{T-1} \omega_T^{-1/2} (\alpha^2-1)^{1/2} \sigma \, \alpha^t W_t^*$$

$$= (\alpha^2-1)^{1/2}\sigma \sum_{t=1}^{\infty} \alpha^{t-T}\, \omega_T^{-1/2} \alpha^T W_t^* 1_{[1,T-t]}(t)$$

$$\rightarrow \frac{(\alpha^2-1)^{1/2}}{\alpha-1} W \quad \text{for } T \rightarrow \infty$$

by dominated convergence, applying Lemma A.2 and Lemma A.3, noticing that

$$\lim_{T\to\infty} \sum_{t=1}^{T-1} \alpha^{t-T} = \frac{1}{\alpha-1}$$

Combining lemmas A.8 and A.9 we get

Lemma A.10. $\sum_{t=1}^{T-1} y_t/(E[\sum_{t=1}^{T} y_{t-1}^2])^{1/2} \xrightarrow{a.s.} \frac{(\alpha^2-1)}{\alpha-1}$ W for $T \rightarrow \infty$, where $W \sim N(0,1)$.

Proposition A.11. $\sum_{t=1}^{T} x_{it}\, y_{t-1}/(\sum_{t=1}^{T} x_{it}^2)^{1/2}\{E[\sum_{t=1}^{T} y_{t-1}^2]\}^{1/2} \xrightarrow{a.s.} 0$ for $T \rightarrow \infty$.

Proof: By Assumption 4 (ii) there exists δ_T, such that

$\delta_T \rightarrow 0$ for $T \rightarrow \infty$, and

$$\sum_{t=1}^{T} x_{it} y_{t-1}/(\sum_{t=1}^{T} x_{it}^2)^{1/2}(E[\sum_{t=1}^{T} y_{t-1}^2])^{1/2}$$

$$= \sum_{t=1}^{T} \frac{x_{it}}{(\sum_{t=1}^{T} x_{it}^2)^{1/2}} \; \frac{y_{t-1}}{(E[\sum_{t=1}^{T} y_{t-1}^2])^{1/2}}$$

$$\leq \delta_T \sum_{t=1}^{T-1} y_{t-1}/(E[\sum_{t=1}^{T} y_{t-1}^2])^{1/2}$$

$$\longrightarrow 0 \quad \text{for} \quad T \rightarrow \infty$$

by Lemma A.10.

From the above results, we have established the following theorem.

<u>Theorem 1.</u> $I_T^* (\alpha, \beta, \sigma^2) \xrightarrow{\text{a.s.}} I^*$ for $T \to \infty$, where

$$
I^* = \begin{array}{cccccc}
W^2 & 0 & 0 & 0 & 0 & 0 \\
 & 1 & q_{12} \cdots & q_{1k} & 0 & 0 \\
 & & & 1 \cdots & q_{2k} & 0 \\
 & & & & \vdots & \\
 & & & & 1 & 0 \\
 & & & & & 1
\end{array}
$$

and $W^2 \sim \chi^2(1)$.

Footnotes

1 See LeCam [16] for a rigorous treatment. A much lengthier, but more readable exposition is provided by Ibragimov and Has'minskii [14].

2 See, for example, the work of Anderson [1] and Rao [21] on pure autoregressive models.

3 Such a result also is implicit in the analysis of Anderson [1].

4 Our conditions are different than those of Basawa and Scott, but there is little extra to prove, and we simply credit them with the result. Reworked proofs are contained in Domowitz and Muus [10].

5 The existence of the estimator is assured under the stated conditions. See Domowitz and Muus [10] or Basawa and Scott [3].

6 The unit root problem requires some very specialized tools. See, for example, Phillips [20].

7 As implied by our previous assumptions, the hypothesis $\alpha = 1$ is excluded from consideration here.

8 In the Hall model, it is assumed that the real interest rate is constant and equal to the household's subjective rate of time preference. See Muus [18] for a derivation of model (5.1), as well as a discussion of the restrictions involved.

References

[1] Anderson, T.W.: "On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations," Annals of Mathematical Statistics, 30, (1959) 676-687.

[2] Basawa, I.V. and P.J. Brockwell: "Asymptotic Conditional Inference for Regular Nonergodic Models with an Application to Autoregressive Processes," Annals of Statistics 12, (1984) 161-171.

[3] Basawa, I.V., and D.J. Scott: Asymptotic Optimal Inference for Non-Ergodic Models, New York: Springer-Verlag, 1983.

[4] Blanchard, O.J.: "What Is Left of the Multiplier Accelerator?" American Economic Review, 71, Papers and Proceedings (1981) 150-154.

[5] Cramer, H., and J. Leadbetter: Stationary and Related Stochastic Processes, New York: Wiley, 1967.

[6] Daly, V. and Hadjimatheou, G.: "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Evidence for the U.K. Economy," Journal of Political Economy, 89 (1981) 596-599.

[7] Davidson, J.E.H. and Hendry, D.F.: "Interpreting Econometric Evidence: The Behavior of Consumers' Expenditure in the U.K.," European Economic Review, 16 (19810 177-192.

[8] Domowitz, I.: "New Directions in Non-Linear Estimation with Dependent Observations," Canadian Journal of Economics, 19 (1985) 1-27.

[9] Domowitz, I. and Muus, L.T.: "Decision Rules Originating from Stochastic Dynamic Programs: A General Framework for Empirical Analysis," in: Proceedings of the 1984 International Congress on Technology and Technology Exchange, Pittsburgh: International Technology Institute, 1984.

[10] Domowitz, I., and L.T. Muus: "Inference in Nonergodic Models with Econometric Applications." Unpublished manuscript, Northwestern University, 1985.

[11] Engle, R.F.: "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in Z. Griliches and M. Intriligator (eds.): Handbook of Econometrics, Volume II, Amsterdam: North Holland, 1984.

[12] Fuller, W.A., D.P. Hasza, and J.J. Goebel: "Estimation of the Parameters of Stochastic Difference Equations," Annals of Statistics, 9, (1981) 531-543.

[13] Hall, R. E.: "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," Journal of Political Economy, 86 (1981) 971-987.

[14] Ibragimov, I.A. and R.Z. Has'minskii: Statistical Estimation: Asymptotic Theory, New York: Springer-Verlag, 1981.

[15] Judge, C.C., Griffiths, W. E., Hill, R.C., and Lutkepohl, H., and Lee, T.C.: The Theory and Practice of Econometrics, New York: Wiley, 1985.

[16] LeCam, L.: "Locally Asymptotically Normal Families of Distributions," University of California Publications in Probability and Statistics 3, (1960) 27-98.

[17] Mann, H.B. and A. Wald: "On the Statistical Treatment of Linear Stochastic Difference Equations," Econometrica 11, (1943) 173-220.

[18] Muus, L.T.: "A Survey of Recent Developments in the Theory of Consumption," Discussion Paper, Cornell University, 1986.

[19] Nankervis, J.C. and N.E. Savin: "Finite Sample Distributions for First Order Stable and Nonstable Difference Equations with an Exogenous Variable," unpublished manuscript, Trinity College, Cambridge, July 1985.

[20] Phillips, P.C.B.: "Time Series Regression With Unit Roots," Cowles Foundation Discussion Paper NO. 740, April, 1985.

[21] Rao, M.M.: "Consistency and Limit Distributions of Estimators of Parameters in Explosive Stochastic Difference Equations," Annuals of Mathemtical Statistics 32, (1961), 195-218.

[22] Satchell, S.E.: "Approximation to the Finite Sample Distribution for Nonstable First Order Stochastic Difference Equations," Econometrica 52, (1984) 1271-1290.

[23] Wei, C.Z.: "Asymptotic Properties of Least-Squares Estimates in Stochastic Regression Models," Annals of Statistics 13, (1985) 1498-1508.

[24] Weiss, L. and J. Wolfowitz: Maximum Probability Estimators and Related Topics, New York: Springer-Verlag, 1974.

[25] White, J.S.: "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," Annals of Mathematical Statistics 29, (1958) 1188-1197.