Université de Montréal
Faculté des arts et des sciences
Département de sciences économiques

CAHIER 9423

UNIT ROOT TESTS IN ARMA MODELS WITH DATA DEPENDENT METHODS FOR THE SELECTION OF THE TRUNCATION LAG

Serena NG[1] and Pierre PERRON[1]

[1] Département de sciences économiques and Centre de recherche et développement en économique (C.R.D.E.), Université de Montréal.

December 1994

The authors acknowledge grants from the Social Sciences and Humanities Research Council of Canada (SSHRC). The second author would like to thank also the Fonds pour la formation de chercheurs et l'aide à la recherche du Québec (FCAR) and the National Science Foundation for financial support.

C.P. 6128, succursale Centre-ville
Montréal (Québec) H3C 3J7

Télécopieur (FAX) : (514) 343-5831
Courrier électronique (E-Mail) : econo@tornade.ERE.Umontreal.CA

Ce cahier a également été publié au Centre de recherche et développement en économique (C.R.D.E.) (publication no 2794).

Dépôt légal - 1994
Bibliothèque nationale du Québec
Bibliothèque nationale du Canada

ISSN 0709-9231

RÉSUMÉ

Cette étude analyse le choix du nombre de retards dans le contexte du test de Said-Dickey pour la présence d'une racine unitaire dans un modèle ARMA général. Il est démontré qu'une relation déterministe entre le nombre de retards et la taille de l'échantillon est dominée par des règles dépendantes des données qui tiennent compte de l'information échantillonnale. En particulier, on étudie des règles dépendantes des données qui ne sont pas contraintes à respecter la condition de borne inférieure autrement imposée par Said-Dickey. Le critère d'information d'Akaike entre dans cette catégorie. Les propriétés analytiques du nombre de retards déterminé par une classe de critères d'information sont comparées aux propriétés de ceux basés sur un test séquentiel de la signification des coefficients sur les retards additionnels. L'impact de l'une ou l'autre de ces méthodes sur les propriétés asymptotiques du test de racine unitaire est analysé et des simulations sont utilisées pour illustrer leur comportement distinctif dans le cas d'un échantillon fini. Nos résultats favorisent les méthodes fondées sur des tests séquentiels plutôt que sur des critères d'information, étant donné qu'elles démontrent moins de distorsions de niveau et ont des puissances comparables.

Mots-clés : Dickey-Fuller, Said-Dickey, critère d'information, sélection de modèle, général à spécifique.

ABSTRACT

This paper analyzes the choice of the truncation lag in the context of the Said-Dickey test for the presence of a unit root in a general autoregressive moving-average model. It is shown that a deterministic relationship between the truncation lag and the sample size is dominated by data dependent rules which take sample information into account. In particular, we study data dependent rules which are not constrained to satisfy the lower bound condition imposed by Said-Dickey. Akaike's information criterion falls into this category. The analytical properties of the truncation lag selected according to a class of information criteria are compared to those based on sequential testing for the significance of coefficients on additional lags. The asymptotic properties of the unit root test under various methods for selecting the truncation lag are analyzed, and simulations are used to show their distinctive behavior in finite samples. Our results favor methods based on sequential tests over those based on information criteria since the former show less size distortions and have comparable power.

Key words: Dickey-Fuller, Said-Dickey, information criteria, order selection, general to specific.

1. Introduction.

Testing for the presence of a unit root in a time series of data has become a common starting point of applied work in macroeconomics. Except in very special cases, one often assumes that the series to be tested is driven by serially correlated innovations and tests for the presence of a unit root using statistics which take serial dependence into account. One such statistic that has become very popular is the Augmented Dickey Fuller t test due to Dickey and Fuller (1979) and Said and Dickey (1984). Their test, hereafter referred to as t_ρ, is based on estimates from an augmented autoregression. The test is valid for stationary and invertible $ARMA$ noise functions of unknown order provided the truncation lag, k, is chosen in relationship to the sample size, T, to satisfy lower and upper bound conditions.

An issue that arises with the implementation of t_ρ is the choice of k. Work of Schwert (1989), Agiakloglou and Newbold (1991) and Harris (1992) have found the order of the autoregression to have important size and power implications. This paper provides a formal analysis of the relevance of k in the test procedure. One of our objectives is to show, via simulations, that a deterministic rule which relates k to T is inferior to a data dependent rule which takes sample information into account. Another objective is to clarify the role of the lower and the upper bound on k in the limiting behavior of the statistic t_ρ. We study the asymptotic properties of t_ρ and of the estimates from the augmented autoregression with k chosen using different data dependent rules. Among these are information based model selection rules (such as the AIC and the *Schwartz* criteria), and sequential testing for the significance of the coefficients on lags (such as F or t tests). We show that with parameter values for which size problems surface, information based rules tend to select values of k that are consistently smaller than those chosen through sequential testing for the significance of coefficients on additional lags, and the size distortions associated with the former method are correspondingly larger. Thus, the choice of the data dependent rule has bearing on the size and power of the test. These issues are of particular relevance in finite samples.

The paper is structured as follows. Section 2 puts forth the Said and Dickey framework, the role of the upper and lower bound conditions on k, and the implications for t_ρ with and without the lower bound. Section 3 provides a discussion of procedures typically used to select k. Formal definitions of 'deterministic' and 'adaptive' rules are given. The properties of t_ρ with k chosen according to information criteria and sequential testing for the significance of coefficients on lags are analyzed in Sections 4 and 5, respectively. Implications of these results are presented in Section 6. We conclude with suggestions for procedures to select k and directions for future research. Proofs of theorems are given in a technical appendix.

2. The Said-Dickey Approach.

2.1 The Test Statistic.

Suppose the Data Generating Process (DGP) for $\{y_t\}$ is given by

$$y_t = \rho y_{t-1} + u_t, \tag{2.1}$$

$$u_t = \sum_{i=1}^{p} \alpha_i u_{t-i} + e_t + \sum_{j=1}^{q} \theta_j e_{t-j},$$

where $e_t \sim i.i.d.(0, \sigma_e^2)$ with bounded fourth moment. Assuming that $\{u_t\}$ is stationary and invertible with autoregressive and moving-average polynomials that do not share common roots, $\{y_t\}$ evolves according to

$$\Delta y_t = (\rho - 1)y_{t-1} + \sum_{i=1}^{\infty} d_i u_{t-i} + e_t, \tag{2.2}$$

where the coefficients d_i $(i = 1, \cdots, \infty)$ are functions of the parameters $\{\alpha_i, \theta_j; i = 1, .., p, j = 1, ..., q\}$. The true order of the autoregression is infinity when $q > 0$. The null hypothesis of interest is $\rho = 1$, in which case a unit root is said to exist and the DGP is an $ARIMA(p, 1, q)$. Since $\Delta y_t = u_t$ under the null hypothesis, (2.2) can also be seen as an autoregression in Δy_t augmented by y_{t-1}, viz.:

$$\Delta y_t = (\rho - 1)y_{t-1} + \sum_{i=1}^{\infty} d_i \Delta y_{t-i} + e_t. \tag{2.3}$$

When the orders p and q are unknown, as is often the case in practice, Said and Dickey (1984) suggested approximating the infinite autoregression by a truncated version whose order is a function of the number of observations, T:

$$\Delta y_t = d_0 y_{t-1} + \sum_{i=1}^{k} d_i \Delta y_{t-i} + e_{tk}, \tag{2.4}$$

where $d_0 = \rho - 1$, and for future reference, we denote $d(k) = (d_1, \cdots, d_k)$. The OLS estimates are similarly defined as $\hat{d}_0 = \hat{\rho} - 1$ and $\hat{d}(k) = (\hat{d}_1, ..., \hat{d}_k)$. The order of truncation, k, is assumed to satisfy some conditions to ensure consistency of the least squares estimates. More precisely, Said and Dickey (1984) assumed

(A1) k is chosen as a function of T such that $k^3/T \to 0$ and $k \to \infty$ as $T \to \infty$,

(A2) there exist $c > 0$ and $r > 0$ such that $ck > T^{1/r}$.

Assumption (A1) is based on the work of Berk (1974) who showed consistency of the parameter estimates in an autoregression of the form (2.4) but without the level regressor, y_{t-1}, and when the process is stationary. The assumption is imposed to ensure that the number of regressors does not increase too fast as to induce excess variability in the estimators. Assumption (A2) is often an overlooked condition. It is a lower bound condition which restricts k to be at least a polynomial rate in T. It rules out values of k that are proportional to $\log T$. Intuitively, (A2) prohibits k from being too small as to provide an inadequate approximation to the true model. It is more restrictive than

(A2′) k satisfies $k^{1/2}\sum_{j=k+1}^{\infty}|d_i| \to 0$ and $k \to \infty$ as $T \to \infty$.

Assumption (A2′) was used by Berk (1974), and, in related work by Lewis and Reinsel (1985), to show consistency of the OLS estimates in an autoregression applied to a stationary process. Note that (A2′) is satisfied for any $\{u_t\}$ that is a stationary and invertible $ARMA$ process as long as $k \to \infty$ as $T \to \infty$ irrespective of the rate at which k increases. Of particular importance is the fact that unlike (A2), (A2′) allows k to grow at a logarithmic rate. In Berk (1974) and Lewis and Reinsel (1985), Assumption (A2′) is strengthened to

(A2″) k satisfies $T^{1/2}\sum_{j=k+1}^{\infty}|d_i| \to 0$ as $k \to \infty$ and $T \to \infty$

to ensure $\sqrt{T}$ consistency of $\hat{d}(k)$. Note that (A2″) implicitly rules out k growing at a $\log(T)$ rate and is basically equivalent to (A2). Consistency of $\hat{d}(k)$ may be achieved at a rate slower than $\sqrt{T}$ if (A2′) is satisfied but not (A2″).

The discussion above applies when the DGP is an infinite autoregression, as would be the case if moving average components were present. When dealing with a finite autoregression, (A2″) is automatically satisfied. In fact, k need not grow to infinity so long as it is selected to be larger than the true order. Hence, most of the results below also apply to the case of a finite autoregression. For a more specific treatment of this case, see Hall (1992).

Said and Dickey's result states that when k satisfies (A1) and (A2), the least squares estimates $\hat{d}(k)$ are $\sqrt{T}$ consistent, and the coefficient on y_{t-1} provides a basis for testing the unit root hypothesis. The limiting distribution for the t statistic on $\hat{d}_0 = (\hat{\rho} - 1)$ for testing $\rho = 1$ is such that

$$t_\rho \Rightarrow \left(\int_0^1 W(r)dW(r)\right)\left(\int_0^1 W(r)^2 dr\right)^{-1/2}, \tag{2.5}$$

where $W(r)$ is a standard Brownian motion in the space $C[0,1]$. Percentiles of this distribution are given in Fuller (1976). The result stated in (2.5) extends naturally to the inclusion of deterministic components in (2.4). In that case, the Wiener process is replaced by its detrended counterpart.

2.2 A Useful Result.

Of interest are the properties of the test statistic when k is chosen as a function of T to satisfy (A1) but not necessarily (A2), since such procedures are commonly used in applied work as we discuss below. The following Lemma considers the validity of Said and Dickey's (1984) result when the lower bound condition (A2) is not imposed.

Lemma 2.1 : *Let $\{y_t\}$ be given by (2.1). Let t_ρ be obtained from the truncated autoregression (2.4) with k chosen such that (A1) is satisfied. i) The asymptotic distribution of t_ρ continues to be given by (2.5) without (A2). ii) $\hat{d}(k) = (\hat{d}_1, \cdots, \hat{d}_k)$, is not, in general, $\sqrt{T}$ consistent for $d(k) = (d_1, \cdots, d_k)$ if (A2) or (A2′) does not hold. In that case, there exists a λ, with $|d_j| \leq C_1\lambda^j$ for some constant C_1 and $0 < \lambda < 1$, such that $\lambda^{-k}(\hat{d}_i - d_i) = O_p(1)$, $(i = 1, \ldots, k)$.*

Lemma 2.1 states that although $\sqrt{T}$ consistency of the coefficients on Δy_{t-i} is not assured without (A2), $\hat{d}_0$ is still consistent for d_0 at rate T, and t_ρ attains the same limiting distribution as defined in (2.5) with Assumption (A1) alone. The proof of consistency of $\hat{d}_0$ and $\hat{d}(k)$ under (A1) and (A2) is given in Said and Dickey (1984). The lower bound condition enters the analysis only when considering the properties of coefficients pertaining to Δy_{t-i}. Specifically, $\sqrt{T}$ consistency of $\hat{d}(k)$ requires, from Lemma 2 of Berk (1974), that

$$E\left((T-k)^{-1}\sum_{j=1}^{k}\{\sum_{t=k+1}^{T} u_{t-j}(e_{tk} - e_t)\}^2\right) \leq k(T-k)\sum_{i=k+1}^{\infty} d_i^2 \to 0. \tag{2.6}$$

Since $(e_{tk} - e_t)$ is the error in approximating an infinite autoregression by a truncated autoregression, it is larger the smaller is k ; the role of the lower bound is therefore intuitive in this context. Sufficient conditions for (2.6) to hold are provided by either (A2), as invoked by Said and Dickey (1984), or (A2″), as used by Berk (1974) and by Lewis and Reinsel (1985). Recall, however, that k growing at a logarithmic rate is ruled out by either (A2) or (A2″).

To see the ramifications of this condition, suppose that $\{u_t\}$ is an $MA(1)$ with coefficient θ, and, hence, $d_i = -(-\theta)^i$. The condition (2.6) is equivalent to requiring that $\log(k) + \log(T-k) + k\log\theta^2$ diverges to $-\infty$. Now take $k = b\log(T)$ for some constant $b > 0$. Clearly, $k^3/T \to 0$ and (A1) is satisfied, but the condition for $\sqrt{T}$ consistency is (approximately) $1 + b\log(\theta^2) < 0$. This condition fails when $|\theta| > \exp(-1/2b)$. Hence, for any fixed rule satisfying $k = b\log(T)$, there will exist a range of values of θ such that (2.6) does not hold. In that case, $\hat{d}(k)$ will not achieve consistency at rate $\sqrt{T}$, but at the slower rate of $T^{(1-a)/2}$, with $a = 1 + blog(\theta^2)$ in the case of an $MA(1)$ (or, equivalently, $|\theta|^{-k}(\hat{d}_i - d_i) = O_p(1)$ as stated in Lemma 2.1). As we will see in subsequent sections, this logarithmic rate is of special interest.

The result that the estimates for the coefficients on Δy_{t-i} might achieve consistency at a rate slower than $\sqrt{T}$ extends to the case when $\{u_t\}$ satisfies a general $ARMA(p,q)$ model, using the fact that the coefficients d_i are such that $|d_i| < C_1\lambda^i$, $0 < \lambda < 1$ for some constant C_1 (see, e.g. Fuller (1976)). The important point is that $(\hat{\rho} - 1)$ will continue to be order T consistent even without the lower bound condition. The asymptotic equivalence of t_ρ with and without (A2) follows from this result, and the result that consistency of the least squares estimates is enough to ensure the consistency of $\hat{\sigma}_k^2$ for σ_e^2.

Although all estimates from the regression (2.4) will be consistent whether or not (A2) holds, the lower bound condition on k is important. The coefficients on the stationary regressors will converge at a rate slower than $\sqrt{T}$ when the lower bound condition is not satisfied. Therefore, choices of k which satisfy (A2) will yield coefficient estimates on the stationary differences that achieve consistency at a faster rate, and can be expected to lead to unit root tests having better finite sample properties than those which do not.

3. The Selection of k.

This section consists of three parts. First, in Section 3.1, we use simulations to show that any *a priori* rule which presets the value of k is likely to result in size distortions and/or power loss unless that value of k happens to be chosen appropriately. This is so, even if k is chosen to be a fixed function of T. In Section 3.2, we discuss the specifics of two data dependent rules whereby the relationship between k and T depends on the given sample of data. In Section 3.3, we further restrict our analysis to data dependent rules which satisfy (A1) only and analyze the limiting distribution of t_ρ when such data dependent rules are used.

3.1 Rules of Thumb.

3.1.1 Fixing k.

Although the asymptotic distribution of t_ρ is derived under the assumption that k increases at an appropriate rate with T, the theoretical conditions (A1) and (A2) provide little practical guidance for choosing k. The common practice is to fix k at a value that is independent of T. Using (2.1) as the DGP, we considered numerous parameterizations of α_i and θ_j with k fixed to be 1 through 10. As the results reported in Table 1.a (moving average case) indicate, the properties of the statistic can be quite different depending on the chosen value of k. For example, when $\theta = -0.8$, fixing k to be 4 yields an exact size of 28 per cent instead of the 5 per cent nominal size, noting that the exact size worsens to 0.939 when θ is -0.95. However, size distortions are much smaller the larger is k. Although size distortions are much smaller when θ is positive, t_ρ is oversized when k is odd but undersized when k is even.

Although in autoregressive models (see Table 1.b) the exact size of the test for all choices of k is generally close to the nominal size (provided k is larger than the true order), the choice of k has implications for power. As is evident from Table 1.b, an over parameterized model is associated with lower power. Thus, while a liberal choice of k will reduce size distortions in moving average models, it will, in general, yield lower power.

We also performed similar simulations for T=200 and T=500. As expected, power increases for every value of k in both the MA and the AR cases. With respect to the size of the test, the results for the AR case are qualitatively the same as when $T = 100$. For positive moving average models, the zig-zag pattern of size distortions as k alternates between odd and even persists even when T is 500. However, for negative moving average models, size distortions increase with T for a given value of k. For example, with $\theta = -0.8$ and $k = 3$, the exact size increases from .455 to .598 as T increases from 100 to 500.

3.1.2 Choosing k as a fixed function of T.

Any rule that defines k as a deterministic function of T fits into this category. A rule often used in unit root tests is due to Schwert (1989). For given constants c and d, the truncation lag, k, is chosen according to

$$k = int\left\{c\,(T/100)^{1/d}\right\}.$$

Values of $c = 4$ and 12 and $d = 4$ were used in Schwert's extensive Monte Carlo analysis. He found that the size of the test is significantly better with $c = 12$ the closer the moving average coefficient, θ, is to -1. Problems encountered in fixing k arbitrarily will also arise if k is chosen as a deterministic function of T since one is faced with a given sample size in practice. In general, there is no way to assure that arbitrarily chosen values of c and d are adequate for a given data series unless c and d happen to be chosen correctly.

The simulations highlight the fact that conditions on k appropriate for asymptotic inference are not necessarily good practical guidelines for selecting k. Indeed, the value of k which ensures an exact size close to the nominal size and also produces high power is highly dependent on the actual DGP, i.e. the values of the AR and MA parameters. Rules of thumb ignore such sample information and is the main reason why fixing k is to be avoided as a matter of practice.

3.2 Data Dependent Rules.

3.2.1 Information Based Rules.

The order of an autoregressive process is often chosen by minimizing an objective function which trades off parsimony against reductions in the sum of squared residuals. Following Hannan and Deistler (1988), we consider an objective function of the general form

$$I_k = \log \hat{\sigma}_k^2 + kC_T/T, \tag{3.1}$$

where C_T is a sequence that satisfies $C_T > 0$, $C_T/T \to 0$. The familiar Akaike (1974) Information Criterion (AIC) is obtained as a special case with $C_T = 2$. Another popular criterion is that of Schwartz (1978) with $C_T = \log T$. Other criteria such as the Bayesian Information criterion (BIC) can be shown to fall within the class of I_k. For econometric applications, the AIC and the *Schwartz* criteria are more common and will be considered in subsequent sections.

3.2.2 Sequential Tests for the Significance of the Coefficients on Lags.

The premise of a sequential test is a general to specific modeling strategy which chooses between a model with m lags and a model with $r = m + n$ lags. Let $\hat{d}(m, r)$ denote the vector of coefficients $(\hat{d}_{m+1}, \cdots, \hat{d}_r)$ obtained by applying OLS to (2.4), with $\hat{\sigma}_r^2 = (T - r)^{-1} \sum_{t=r+1}^{T} \hat{e}_{tr}^2$, and

$$M_r = \sum_{t=r+1}^{T} (y_{t-1}, \Delta y_{t-1}, \cdots, \Delta y_{t-r})'(y_{t-1}, \Delta y_{t-1}, \cdots, \Delta y_{t-r}).$$

Let $M_r^{-1}(n)$ be the lower right hand $(n \times n)$ block of M_r^{-1}. We define the Wald test for the null hypothesis that the coefficients on the last n lags are jointly equal to 0 as

$$J(m, r) = \hat{d}(m, r)' \left(M_r^{-1}(n)\right)^{-1} \hat{d}(m, r)/\hat{\sigma}_r^2. \tag{3.2}$$

We now provide a formal definition of the procedure for choosing $\hat{k}$ from a set of possible values $\{0, 1, \cdots, kmax\}$, where $kmax$ is selected a priori.

Definition 3.1 : *The general to specific modeling strategy chooses $\hat{k}$ to be either i) $m + 1$ if, at significance level α, $J(m, r)$ is the first statistic in the sequence $J(i, i + n), \{i = kmax - 1, \cdots, 1\}$, which is significantly different from zero, or ii) 0 if $J(i, i+n)$ is not significantly different from zero for all $i = kmax - 1, kmax - 2, \cdots, 1$.*

The idea is to start with the most general model with $kmax + n - 1$ lags and test if the coefficients of the last lags are significant. If they are, $\hat{k} = kmax$; otherwise the next step is to estimate an autoregression of order $kmax - 2 + n$ and perform the joint test again. This procedure is repeated until a rejection occurs or the sequential testing leads to the boundary of zero lags. This procedure has been analyzed by Hall (1992) in the case of a pure autoregressive process.

The $J(m, r)$ statistic specializes to a t statistic on the last lag if the test is performed with $n = 1$. This special case of the general to specific procedure has been used by Perron (1989). See, also, Perron (1990) and Perron and Vogelsang (1992) for simulation results for unit root tests allowing a break in the trend function and the noise function assumed to be an $ARMA$ process. Although in principle, one can start with $kmin$ lags and gradually increase k until the next included lag is insignificant, Hall (1992) found that a specific to general approach is not, in general, asymptotically valid in the pure AR case. He also found the finite sample properties of statistics associated with a specific to general approach to be inferior to those based on a general to specific scheme in more general $ARMA$ models. In subsequent analyses, only the general to specific approach will be analyzed.

3.3 Rules Satisfying the Upper Bound Condition.

We now restrict our attention to deterministic and data dependent rules which satisfy (A1). Formal definitions for the rules considered are as follows.

Definition 3.2 : Deterministic Rules. *Let* $\widetilde{K} = (\tilde{k}_1, \tilde{k}_2, \ldots)$ *be the set of points in* $\widetilde{K}_\infty = \times_{T=1}^{\infty} S_T$*, where* $S_T = \{0, 1, \cdots, [T/2]\}$*, with* $\tilde{k}_T \to \infty$ *and* $\tilde{k}_T^3/T \to 0$ *as* $T \to \infty$.

Simply put, $\widetilde{K}$ is the collection of deterministic rules that satisfy the conditions of Lemma 2.1. Our definition of deterministic rules is adapted from Eastwood and Gallant (1991) who studied the selection of the truncation point in a univariate Fourier series expansion fitted by least squares. In our context, Schwert's rule of thumb is, for example, an element of $\widetilde{K}$.

Definition 3.3 : Adaptive Rules. *An adaptive truncation rule is a sequence of random variables* $\widehat{K}_\infty = (\hat{k}_1, \hat{k}_2, \cdots)$*. We say that an adaptive truncation rule maps into the set of deterministic rules* $\widetilde{K}$ *if there exists a deterministic rule* $\tilde{k}_T$ *such that* $\widetilde{K} = (\tilde{k}_1, \tilde{k}_2, \cdots)$ *is a subset of* $\widetilde{K}_\infty$ *and* $\hat{k}_T - \tilde{k}_T \to_p 0$.

The following Lemma considers the limiting distribution of t_ρ when it is based on adaptive rules that map into the set of deterministic rules $\widetilde{K}$.

Lemma 3.4 : *Suppose we have an adaptive truncation rule* $\widehat{K}_\infty = (\hat{k}_1, \hat{k}_2, \ldots)$ *that maps into the set of deterministic rules* $\widetilde{K}$ *stated in Definition 3.2, and let* $t_\rho(\hat{k}_T)$ *be the* t *statistic for testing* $\rho = 1$ *in regression (2.4) estimated with* $\hat{k}_T$ *lags. Then* $t_\rho(\hat{k}_T) \Rightarrow \int_0^1 W(r)dW(r)\left(\int_0^1 W(r)^2 dr\right)^{-1/2}$.

The proof is analogous to Theorem 5 of Eastwood and Gallant (1991) and is therefore omitted. The importance of Lemma 3.4 is that the limiting distribution of t_ρ is the same whether one uses a deterministic rule in $\widetilde{K}$ or an adaptive rule that maps into $\widetilde{K}$. The issue then becomes which of the selection procedures deliver better finite sample properties in testing for the presence of a unit root.

Deterministic rules are useful for analytical purposes since they help establish the properties of t_ρ under adaptive rules. However, as seen from the results reported earlier, size and power will be affected whenever k is fixed in a deterministic way unless the rule happens to be chosen correctly. Adaptive rules take sample information into account and are therefore likely to dominate deterministic rules. In the next two sections, our analysis will be further restricted to adaptive rules only.

4. Adaptive Rule 1: Information Criteria.

This section presents properties of $\hat{k}$ and t_ρ when an information criterion as defined in (3.1) is used to select the truncation lag in regression (2.4). A related issue has been studied by Hannan and Deistler (1988) in the context of stationary variables with the autoregression

$$x_t = \sum_{i=1}^{k} \delta_i x_{t-i} + e_{tk}. \tag{4.1}$$

The next Lemma summarizes a result of theirs that is relevant to our analysis.

Lemma 4.1 : *Let x_t be a stationary and invertible ARMA process with finite fourth moment and $\tilde{\sigma}_k^2 = (T-k)^{-1}\sum_{t=k+1}^{T}\tilde{e}_{tk}^2$ with $\tilde{e}_{tk}$ the OLS residuals from regression (4.1). Let C_T be a function of T such that $C_T > 0$ and $C_T/T \to 0$, and $\tilde{k}_T = \arg\min_k (\log(\tilde{\sigma}_k^2) + kC_T/T)$. Then $\lim_{T\to\infty}\tilde{k}_T/b\log T = 1$ for some constant b.*

The result that the AIC with $C_T = 2$ chooses a value of k that is proportional to $\log T$ in a univariate Gaussian ARMA model is due to Shibata (1980). Hannan and Deistler (1988) provide a unified asymptotic framework to show that the feature of $\log T$ proportionality is generic to information based rules applied, in particular, to stationary and invertible $ARMA$ models. The logarithmic rule also extends to multivariate and/or ARMAX models as Hannan and Deistler (1988) have shown. Their result is useful in studying the properties of $\hat{k}_T$ within the context of an augmented autoregression of the form (2.4) derived for an $ARIMA$ $(p, 1, q)$ process. The following Lemma shows that their result extends to this latter case.

Lemma 4.2 : *Let y_t satisfies (2.1) and define $\hat{\sigma}_k^2 = (T-k)^{-1}\sum_{t=k+1}^T \hat{e}_{tk}^2$, where $\hat{e}_{tk}$ are the least squares residuals from the augmented autoregression (2.4). Let $\tilde{\sigma}_k^2 = (T-k)^{-1}\sum_{t=k+1}^T \tilde{e}_{tk}^2$ where $\tilde{e}_{tk}$ are the OLS residuals from the restricted regression (4.1) with $x_t = \Delta y_t$. Then, $\hat{\sigma}_k^2 = \tilde{\sigma}_k^2 + o_p(T^{-1/2})$ provided k satisfies (A1).*

Lemma 4.2 implies that the difference between the residual sum of squares from an augmented autoregression and a restricted one is $o_p(T^{-1/2})$ uniformly in k. Hence, the information criteria and the corresponding values of k that minimize such criteria are asymptotically the same in both cases. Thus the *AIC* and *Schwartz* criteria, when applied to the augmented autoregression defined in (2.4), also select truncation lags that are proportional to $\log T$ under the null hypothesis of a unit root. The implication for the unit root test is summarized in the following Theorem.

Theorem 4.3 : *If k is selected using an information criterion in the class I_k as defined in (3.1), then t_ρ has a limiting distribution defined by (2.5) under the null hypothesis of a unit root.*

The order of truncation selected by the *AIC* or the *Schwartz* criteria is proportional to $\log T$. Since such a rule satisfies (A1) that $k^3/T \to 0$ and $k \to \infty$ as $T \to \infty$, it is also an adaptive rule that maps into the set of deterministic rules $\widetilde{K}$. The result therefore, follows directly from Lemmas 2.1 and 3.4. Theorem 4.3 still holds when the *DGP* is a finite instead of an infinite autoregression provided the information criterion does not asymptotically underparameterize the model; see Hall (1992). Note, however, that the information criteria is not an adaptive rule that maps into the set of deterministic rules which satisfy both (A1) and (A2).

4.1 A Special Case: An $MA(1)$.

Since the truncation lag selected from regression (2.4) when the series is an *ARIMA* (p,1,q) and the truncation lag selected on the basis of (4.1) when the series is a stationary and invertible *ARMA* process have the same asymptotic properties, we can, for simplicity, use the restricted framework to provide more insight about the properties of the truncation lag selected using information criteria. Specifically, we consider an $MA(1)$ process defined as:

$$x_t = e_t + \theta e_{t-1} = \sum_{i=1}^{\infty} \phi_i x_{t-i} + e_t,$$

where $\phi_i = -(-\theta)^i$. The true order of the autoregression is infinity for all values of $\theta \neq 0$. The estimated regression is

$$x_t = \sum_{i=1}^{k} \phi_i x_{t-i} + e_{tk}.$$

It is straightforward to show that $\tilde{\sigma}_k^2$ is approximately related to k by

$$\tilde{\sigma}_k^2 \simeq \sigma_e^2(1 - \theta^{2(k+2)}) \left(1 - \theta^{2(k+1)}\right)^{-1}.$$

Minimizing the Akaike information criterion, $\log \tilde{\sigma}_k^2 + 2k/T$, the solution is asymptotically equivalent to:

$$\tilde{k}(AIC) \approx \left(\log(T) + \log[(\theta^2 - 1)\log\theta^2] - \log 2\right) (|\log\theta^2|)^{-1}. \tag{4.2}$$

Table 2 presents the approximation to $\tilde{k}(AIC)$ provided by (4.2) for various values of $|\theta|$ and T. For small θ, ϕ_i is small and declines geometrically as i increases. One might then expect the AIC to chose a low order since extra parameters have little information content but reduce the degrees of freedom. Table 2 shows that, indeed, for $|\theta| \leq 0.4$, low values of k are selected by AIC. However, as $|\theta|$ gets large, ϕ_i will remain non-negligible even for i quite large. Increasing the length of the autoregression should, in principle, improve the approximation to the DGP. However, the k selected by AIC only increases at a logarithmic rate. Except when T becomes impracticably large, the AIC will abandon information at large lags in favor of a very parsimonious model. Hence, in practice, one can expect the chosen k to be no higher than five even with T as large as 500 when $|\theta|$ is close to 1.

5. Adaptive Rule 2: Testing for the Significance of Coefficients on Lags.

This section analyzes the properties of $\hat{k}$ and t_ρ when $\hat{k}$ is chosen by the $J(m,r)$ statistic described in Section 3.2 to sequentially test for the significance of coefficients on additional lags. The following Lemma is useful in establishing the limiting distribution of t_ρ.

Lemma 5.1 : *Let $\{y_t\}$ be generated by (2.1) and suppose that Assumptions (A1) and (A2′) hold. Let $\hat{d}(k)$ be obtained from the augmented autoregression (2.4) and let $J(k-n,k)$ be as defined in (3.2). Then $J(k-n,k)$ is asymptotically distributed as χ^2 with n degrees of freedom.*

Berk (1974) proved consistency and asymptotic normality of the coefficients in the restricted regression under (A1) and (A2″). See also Lewis and Reinsel (1985). The crucial element in the proof of Lemma 5.1 is the fact that when $\{\Delta y_t\}$ is a stationary and invertible $ARMA$ process, the coefficients $d(k)$ converge to 0 at a rate that yields an asymptotic equivalence between the Wald test that $\hat{d}(k) = d(k)$ and the Wald test that $\hat{d}(k) = 0$. Indeed, Lemma 5.1 requires Assumption (A2″) to ensure that $\sqrt{T}d(k) \to 0$, which, in turn, ensures asymptotic normality of $\sqrt{T}\hat{d}(k)$.

5.1 A Special Case: An $MA(1)$ and the t Test.

We now specialize the sequential procedure described in Section 3.2.2 to the case where $n = 1$. The square root of the statistic $J(k-1,k)$ then simplifies to a t test for the significance of the coefficient on the last lag in an autoregression of order k:

$$t_{\hat{d}_k} = \sqrt{T}\hat{d}_k(\hat{\sigma}_k^2 T M_k^{-1}(1))^{-1/2}.$$

The sequential procedure chooses a value of $\hat{k}$ if $t_{\hat{d}_k}$ is significant at some pre-specified level α in an estimated autoregression of order $\hat{k}$, while the t-statistics $t_{\hat{d}_k}$ are insignificant in estimated autoregressions of order k for all k in the range $(\hat{k}, kmax]$. We can show that if Δy_t is an $MA(1)$, i.e. $\Delta y_t = e_t + \theta e_{t-1}$,

$$\hat{d}_k \simeq \theta^{k-1}(1-\theta^2)\Big(1-\theta^{2(k+1)}\Big)^{-1},$$

$$t_{\hat{d}_k} \simeq \sqrt{T}\theta^{k-1}(1-\theta^2)\Big((1-\theta^{2k})(1-\theta^{2(k+2)})\Big)^{-1/2}.$$

The above results show that both $t_{\hat{d}_k}$ and $\sqrt{T}\hat{d}_k$ will converge to zero if k increases at a polynomial rate. Given the result of Lewis and Reinsel (1985, Theorem 4) that $t_{\hat{d}_k = d_k} \equiv (\hat{d}_k - d_k)/(\hat{\sigma}_k^2 M_k^{-1}(1))^{1/2}$ is asymptotically distributed $N(0,1)$ if k increases at a polynomial rate satisfying (A1) and (A2″), $t_{\hat{d}_k}$ can be shown to have the same asymptotic distribution under these restrictions on the rate of increase of k.

It is of some interest to note that the results above also imply that a specific to general procedure starting from any lower bound $kmin$ that tests for the significance of the coefficient on the last lag would select a $\hat{k}$ that increases to infinity at a logarithmic rate when $\{\Delta y_t\}$ contains a moving average component. Hence, such a specific to general procedure would have the same asymptotic properties as a selection rule based on an information criterion.

Note that the asymptotic normality result of Berk (1974) and Lewis and Reinsel (1985) used to prove Lemma 5.1 requires that $\hat{k}$ increases at some polynomial rate, or at least at a rate that ensures (A2″) is satisfied. A logarithmic rate is not sufficient. We now show that the truncation lag selected by a general to specific procedure will be of an order higher than $\log T$ provided $kmax$ increases at a rate faster than $\log T$. In fact, the selected truncation lag will grow at the same rate as $kmax$.

Lemma 5.2 : *If $\hat{k}$ is selected by means of the general to specific strategy described in Definition 3.1 and kmax increases at a rate such that (A1) and (A2″) are satisfied, then $\hat{k}$ increases at the same rate as kmax .*

The intuition behind the result stated in Lemma 5.2 is as follows. Under the assumptions of Lemma 5.1, $J(k-n,k)$ is asymptotically distributed as a χ^2 random variable with n degrees of freedom. The limiting probability that $J(k-n,k)$ is statistically significant is therefore α, the size of the test. For a given $\hat{k} < kmax$ to be chosen, it must be the case that all prior statistics in the sequential procedure $(J(i-n,i);\ i = kmax-1, ..., \hat{k}-1)$ are statistically insignificant. This event occurs, for large samples, with probability $\alpha(1-\alpha)^{kmax-\hat{k}}$. Since $\hat{k} \leq kmax$ and $kmax \to \infty$, this probability vanishes as $T \to \infty$ unless $\hat{k}$ increases at the same rate as $kmax$.

The importance of Lemma 5.2 is that if $kmax$ is chosen to increase at a polynomial rate, then $\hat{k}$ will also increase at a polynomial rate. This implies that Assumption (A2) or (A2″) can be satisfied with judicious choice of $kmax$, thereby ensuring the results of Lemma 5.1 hold. Lemma 5.2 allows us to state the following Theorem concerning the limiting behavior of the unit root test under this truncation lag selection rule.

Theorem 5.3 : *If kmax satisfies (A1) and (A2″) and $\hat{k}$ is chosen from the general to specific sequential procedure stated in Definition 3.1, then $t_\rho(\hat{k})$ has the same limiting distribution as (2.5).*

Since $\hat{k}$ maps into a deterministic rule in the set $\widetilde{K}$ by Lemma 5.2, the result follows from Lemma 3.4. In fact, $\hat{k}$ maps into the set of deterministic rules that satisfy (A1) and (A2) since $\hat{k}$ increases at a polynomial rate under the conditions of Theorem 5.3.

6. Finite Sample Simulations.

The results of the preceding sections can be summarized as follows. An information criterion will choose values of k that are proportional to $\log T$, a rate ruled out by (A2). However, the k selected using the $J(m, r)$ statistic to test for the significance of lags will increase at the same rate as the prespecified $kmax$, itself increasing at a polynomial rate. Because a logarithmic rate of increase is slow compared to a polynomial rate, an information criterion will choose values of k that are generally much smaller than a general to specific t test, for example. Although the log proportionality rule might fail the lower bound condition, the limiting distribution of t_ρ is unaffected. In such a case, the estimates of the coefficients on Δy_{t-i} in the augmented autoregression will be consistent at a rate slower than $\sqrt{T}$ for some $DGPs$. In the MA(1) case, a large value of $|\theta|$ is more likely to be associated with a slower rate of consistency for $\hat{d}(k)$. We now examine the implications of these results in finite samples.

The results reported below are based on 5000 simulations for different values of θ_j and α_i. For each parameterization, the selected values of k and the corresponding values of t_ρ are recorded. The simulations were performed on a 486/25 MHz PC with code compiled using the Borland C (Version 3) compiler. Random numbers are generated using the ran1() function from Press et al. (1988) with time (in seconds) as seed. We considered T = 100, 200, and 500. For a given T, different values for $kmax$ and $kmin$ are examined. We focus on results for $T = 100$ with $kmax = 10$ and $kmin = 0$ without loss of generality, and discuss results for other configurations where appropriate. The complete set of results is available on request.

We select, for presentation, results based on two information criteria: the AIC and $Schwartz$. The results for the BIC and the $Hannan - Quinn$ criteria show no appreciable difference. For the general to specific strategy, we considered the t as well as the F test, but only present results for the t test at the 5 and the 10 percent levels. In general, a tighter model is selected using a lower significance level.

6.1 Frequency Distribution of $\hat{k}$.

We first examine the number of times $k = i$ ($i = 1, \cdots, 10$) is being selected by each of the procedures during the 5000 simulations. Reported in Table 3.a (moving average) and 3.b (autoregressive) are the frequency counts. As we can see, both information criteria consistently select values of k that are less than three. While the k's selected for autoregressive models seem appropriate given that the *DGPs* considered are of order no higher than four, the information criteria yield very parsimonious models when the *DGP* is driven by a moving average process. Although the true order of autoregression is infinity in those cases, the *AIC* and *Schwartz* criteria continue to choose values of two and three for k. When θ is large, the coefficients in the autoregression die off only slowly. Truncating the autoregression at a low order will yield a more parsimonious model but at the loss of information. The cost of parsimony will be judged in terms of the size and power of t_ρ in the next subsection.

In the moving average case, the values of k selected by a general to specific modeling strategy are quite evenly distributed over the range $[2, kmax = 10]$, with some mass concentrated at $k = 1$. This result follows directly from Lemma 5.1. A further implication of the Lemma is that the chosen value of k will be closer to $kmax$ the more liberal is the size of the test. Thus, the frequency of k chosen to be five and above is higher under the 10 percent t test than under the 5 percent t test.

6.2 Size and Power.

Having confirmed that information criteria choose values of k that tend to be small, we now proceed to show that in many cases, the method used to choose k can have size and power implications. The results are reported in Tables 4.a and 4.b for $T = 100$, with the power of the test evaluated at $\rho = 0.95$ and 0.85. Turning first to moving average models (Table 4.a), we see that for positive values of θ, the size of the test is similar for all methods of selecting k. When $\theta = 0.8$, the 10 percent t test picks k to be five or smaller 40 percent of the time, whereas the *AIC* picks k to be in the same range twice as often (See Table 3.a). Although such variations in the choice of k appear to yield small size differences, power is slightly higher the more parsimonious is the model. It is well known that the *Schwartz* criterion imposes a heavy penalty for over parameterization. Thus, for positive moving average models, the *Schwartz* criterion tends to yield higher power for a given size.

The result that stands out in Table 4.a is the large size distortions when θ is negative. The problem of size distortion with unit root tests in the presence of negative moving errors is well documented (e.g., Schwert (1989)). Although Schwert used deterministic rules to select k, he also noted that the exact size depends on the choice of k. Our results confirm that the more conservative the criterion for selecting the truncation order, the larger the size distortions associated with t_ρ. For example, size distortions associated with the conservative *Schwartz* criterion are significantly larger than those associated with the 10 percent t test, the most liberal of the criterion considered. From the frequency counts, we see that the *Schwartz* criterion chooses values of k less than three 90 percent of the time, whereas the 10 percent t test chooses values of k *greater* than three with a probability of 0.9.

Table 4.b indicates that for autoregressive models, all methods produce estimates of k that are as large as the true order with high probability. Accordingly, all selection procedures produce an exact size that is close to the nominal size. The 10 percent t test tends to have lower power, however. According to the frequency counts, the t test tends to over parameterize autoregressive models. For example, the 10 percent t test selects k greater than four over 40 per cent of the time when the DGP is a fourth order autoregression. Thus, under parameterization is associated with larger size distortions and over parameterization with power loss when $T = 100$.

The size of the test for moving average models with $T = 200$ are reported in Table 4.c. Note that size distortions in negative moving average models persist as T increases. The *Schwartz* criterion continues to be associated with significantly larger size distortions than the 10 percent t test. However, in cases for which size distortion is not an issue, as in autoregressive models, the discrepancies in power across selection procedures vanish almost completely when $T = 500$. We report, in Table 4.d, the size and power for autoregressive models at $T = 200$. Compared to the results for $T = 100$, power is higher throughout, and the differences in power across selection procedures are smaller. Thus, discrepancies in power across selection procedures are small for typical sample sizes encountered in economic analyses, but size distortions are not. A t or an F test therefore has an advantage over information criterion in that they produce tests with more accurate size without much loss of power.

6.3 The Choice of k and Size Distortions.

When θ in the noise function is large and negative, y_t is close to having a common factor and behaves more like a white noise than an integrated process. The asymptotic properties of the normalized least squares estimator in this case have been shown in Nabeya and Perron (1992) and Perron (1992) to be different from those derived under standard assumptions. In view of those results, one would conjecture that there is also a discrepancy between the finite sample distribution of t_ρ and its approximate distribution as defined by (2.5). But, as we can see, the extent of size distortions varies with k. This suggests that k affects the adequacy of (2.5) as an approximating distribution. The question is, how?

Using a local asymptotic framework, Pantula (1991) parameterized θ as $-(1-T^{-\eta})$ and showed that the limit of t_ρ is given by (2.5) only if $0 < \eta < 0.25$, but diverges to $-\infty$ at rate T^η/k if $0.25 < \eta < 0.625$, with limiting distribution given by:

$$kT^{-\eta}t_\rho \Rightarrow -\left(\int_0^1 W(r)^2 dr\right)^{-1}. \tag{6.1}$$

Since $k = O(T^{1/4})$ by assumption, and (6.1) is valid for $\eta > 0.25$, the limiting distribution of t_ρ will always tend to $-\infty$. But, the larger is the rate of increase of k, the slower the rate of divergence, and the smaller are the discrepancies between the exact and the approximate distributions of t_ρ. Consequently, even though η is 0.35 when $T = 100$ and $\theta = -0.8$, size distortions are noticeably smaller at larger values of k when critical values from (2.5) are used for hypothesis testing.

To reinforce the importance of a large k when θ is large and negative, we report, in Table 5, the size of the test at selected parameter values for $T = 200$ and $T = 500$ when a different lower bound, $kmin$, is prescribed. We set $kmin$ and $kmax$ to four and twelve when $T = 200$, and to six and fourteen when $T = 500$. Evidently, the larger is $kmin$, the larger is k, and the smaller are the size distortions.

The importance of $kmin$ and $kmax$ in all selection procedures must be emphasized. If we raise the value of $kmin$ and let the information criteria select k from the range $[kmin, kmax]$, and $kmin > \log T$, the criteria will choose $kmin$ in large samples since $\log T$ is outside the permissible range. Loosely speaking, the choice of $kmin$ can be seen as a practical way of imposing the lower bound condition (A2). On the other hand, it is the choice of $kmax$ that is more important in a general to specific model selection strategy. By Lemma 5.1, the test statistic will choose $k \in [kmin, kmax]$ with declining probability as k moves away from $kmax$. Thus, the larger is $kmax$, the higher the probability that a larger k will be chosen. The larger is k, the better the size, at the expense, however, of power losses.

7. Conclusions.

This paper has analyzed issues related to the selection of the truncation lag in unit root tests of the type proposed by Dickey and Fuller (1979) and Said and Dickey (1984). We have focused on the implications of the lower bound condition on t_ρ

used in Said and Dickey (1984). Procedures which do not satisfy this condition tend to select truncation lags that are too small for some parameter values. Information based rules such as *AIC* and *Schwartz* fit into this category.

A general feature of our results is that an overly parsimonious model can have large size distortions, but an over parameterized model may have low power. However, the size problem is more severe than power loss in the sense that discrepancies in power across selection procedures diminish as T increases, but size distortions persist even for large sample sizes for some methods of selecting k. In this regard, a t or F test for the significance of lags will have an advantage over information based rules such as the *AIC* since they produce tests with more robust size properties across models.

There remains, of course, several avenues for further research that follow from the framework used in this paper. First, given the problems associated with approximating a general $ARMA$ process by a finite autoregression, one might be tempted to construct unit root tests from an estimated $ARMA(p,q)$ process whose order is selected using a consistent procedure, such as the one discussed in Dickey and Said (1981). However, in view of the problems associated with maximum likelihood estimates of processes with moving-average components, it is not evident that the latter method can provide statistical improvement. A comparison of the various estimation methods in the context of unit root tests would be useful.

The second is an extension of the results to the multivariate case whereby vector autoregressive processes are used to approximate more general multivariate linear processes. While one expects, and preliminary work suggests, the same qualitative results to hold, the analysis is not a straightforward extension because of possible cointegration among the variables.

The third topic concerns the issue of optimal lag selection. Our analysis has concentrated on two particular classes of lag length selection that are widely used in practice. None of these need be optimal. The difficulty, however, lies in finding the proper way to assess the procedures for selecting k since the purpose of estimating these autoregressions is not in obtaining a particular estimate that is as precise as possible, but rather the unit root test itself. The optimality criterion therefore needs to be based on an appropriate trade-off between type I and type II errors in the application of the unit root test.

References

Akaike, H. (1974), "A New Look at Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716-723.

Agiakloglou, C. and Newbold, P. (1992), "Empirical Evidence on Dickey-Fuller Type Tests," *Journal of Time Series Analysis* , 13, 471-483.

Banerjee, A., Lumsdaine, R. L. and Stock, J. H. (1992), "Recursive and Sequential Tests of the Unit Root and Tread Break Hypothesis," *Journal of Business & Economic Statistics*, 10, 271-287.

Berk, K. (1974), "Consistent Autoregressive Spectral Estimates," *Annals of Statistics*, 2, 489-502.

Chan, N.H. and Wei, C. (1988), "Limiting Distribution of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, 367-401.

Dickey, D. A. and Fuller, W. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

Dickey, D. A. and Said, S. E. (1981), "Testing $ARMA(p,1,q)$ versus $ARMA(p+1,q)$," *Proceedings of the Business and Economics Statistics Section*, American Statistical Association.

Eastwood, B. J. and Gallant, R. A. (1991), "Adaptive Rules for Semi nonparametric Estimators that Achieve Asymptotic Normality," *Econometric Theory*, 7, 307-340.

Fuller, W. (1976), *Introduction to Statistical Time Series*, New York: John Wiley and Sons.

Hall, A. (1992), "Testing for a Unit Root in Time Series with Pretest Data Based Model Selection," Forthcoming in *Journal of Business and Economic Statistics*.

Hannan, E. J. and Deistler, M. (1988), *The Statistical Theory of Linear Systems*, New York: John Wiley and Sons.

Harris, R.I.D. (1992), "Testing for Unit Roots Using the Augmented Dickey-Fuller Test," *Economic Letters*, 38, 381-386.

Lewis, R. and Reinsel, G. C. (1985), "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16, 393-411.

Nabeya, S. and Perron, P. (1992), "Local Asymptotic Distributions Related to the AR(1) Model with Dependent Errors," Forthcoming in *Journal of Econometrics*.

Pantula, S. G. (1991), "Asymptotic Distributions of Unit-Root Tests When the Process is Nearly Stationary," *Journal of Business and Economic Statistics*, 9, 63-71.

Perron, P. (1989), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, 57, 1361-1401.

Perron, P. (1990), "Further Evidence of Breaking Trend Functions in Macroeconomic Time Series," manuscript, Princeton University.

Perron, P. (1992), "The Adequacy of Asymptotic Approximations in the Near-Integrated Autoregressive Model with Dependent Errors," manuscript, Université de Montréal.

Perron, P. and Vogelsang, T. J. (1992), "Nonstationarity and Level Shifts With an Application to Purchasing Power Parity," *Journal of Business & Economic Statistics*, 10, 301-320.

Press, W. H., Teukolsky, S., Vetterling, W. and Flannery, B. (1988), *Numerical Recipies in C*, Cambridge: Cambridge University Press.

Said, S. and Dickey, D. A. (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599-607.

Schwartz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461-464.

Shibata, R. (1980), "Asymptotic Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process," *Annals of Statistics*, 8, 147-164.

Schwert, G. W. (1989), "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business & Economic Statistics*, 7, 147-160.

Appendix

The following notation will be used in this Appendix. Unless otherwise stated, we shall let C_1 be an arbitrary constant (not necessarily the same throughout). Let $D_T = diag\{(T-k)^{-1}, (T-k)^{-1/2}, ..., (T-k)^{-1/2}\}$, $U_t' = (y_{t-1}, X_t')$, $X_t' = (\Delta y_{t-1}, \ldots, \Delta y_{t-k})$. Let $M_k = \sum_{t=k+1}^T U_t U_t'$ and $R_k = \sum_{t=k+1}^T X_t X_t'$. Thus,

$$R_k = \sum_{t=k+1}^{T} (\Delta y_{t-1}, \Delta y_{t-2}, ..., \Delta y_{t-k})'(\Delta y_{t-1}, \Delta y_{t-2}, ..., \Delta y_{t-k}),$$

and

$$D_T M_k D_T = \begin{bmatrix} (T-k)^{-2}\sum_{t=k+1}^T y_{t-1}^2 & (T-k)^{-3/2}\sum_{t=k+1}^T y_{t-1}X_t' \\ (T-k)^{-3/2}\sum_{t=k+1}^T y_{t-1}X_t & (T-k)^{-1}R_k \end{bmatrix}.$$

Note that from Said and Dickey (1984), the limit of $D_T M_k D_T$ is block diagonal with the two blocks corresponding to the limits of $(T-k)^{-2}\sum_{t=k+1}^T y_{t-1}^2$ and $(T-k)^{-1}R_k$. We also let $M_k^{-1}(1)$ be the first diagonal element of the matrix M_k^{-1}, and $M_k^{-1}(k)$ be the lower right hand $k \times k$ block of M_k^{-1}.

For a matrix C, the matrix norm is defined by $\|C\| = \sup_{\|x\|<1} \|Cx\|$, where for a vector z, $\|z\| = (z'z)^{1/2}$. Using this norm, Lemma 3 of Berk (1974) showed that $k^{1/2}\|(T-k)R_k^{-1} - \Gamma^{-1}\| \to 0$, where Γ is a $k \times k$ matrix with typical elements $\Gamma_{ij} = E(\Delta y_{t-i}\Delta y_{t-j})$.

Proof of Lemma 2.1

The proof for consistency of the least squares estimates in the augmented autoregression (2.4) is given in Said and Dickey (1984) and will not be repeated here. It is nevertheless important to point out the two steps involved in the proof. The first step is to show that $k^{1/2}\|(D_T M_k D_T)^{-1} - M^{-1}\|$ converges to 0 for some limiting block diagonal matrix M. For this step, Assumption (A1) is sufficient and the argument follows from Lemma 3 of Berk (1974). The second step is to show that $\|D_T \sum_{t=k+1}^T U_t e_{tk}\| = O_p(k^{1/2})$. The combination of the two steps imply $T\hat{d}_0 = O_p(1)$, $\sqrt{T}(\hat{d}_i - d_i) = O_p(1)$ $(i = 1, ..., k)$, and $\hat{\sigma}_{ek}^2 \to \sigma_e^2$. Assumption (A2) is used only in this second step, and more specifically, to ensure that

$$E\left((T-k)^{-1}\sum_{j=1}^{k}\left(\sum_{t=k+1}^{T}\Delta y_{t-j}(e_{tk} - e_t)\right)^2\right) \le C_1 \cdot k(T-k)\sum_{i=k+1}^{\infty} d_i^2 \to 0, \quad \text{(A.1)}$$

as $T \to \infty$ for some constant C_1. Note that (A2″) is also enough to guarantee that (A.1) holds. However, (A.1) is sufficient but not necessary to ensure that t_ρ has the limiting distribution given by (2.5). To see this, we first express t_ρ as:

$$t_\rho = \left((T-k)^{-1} \sum_{t=k+1}^{T} y_{t-1} e_{tk}\right) \left(\hat{\sigma}_k^2 (T-k)^{-2} [M_k^{-1}(1)]^{-1}\right)^{-1/2}.$$

From Said and Dickey (1984), $T^{-2}[M_k^{-1}(1)]^{-1} \Rightarrow \sigma_e^2 \int_0^1 W(r)^2 dr$ provided (A1) holds. Consider now the numerator:

$$(T-k)^{-1} \sum_{t=k+1}^{T} y_{t-1} e_{tk} = (T-k)^{-1} \sum_{t=k+1}^{T} y_{t-1} e_t + (T-k)^{-1} \sum_{t=k+1}^{T} y_{t-1} \sum_{i=k+1}^{\infty} d_i \Delta y_{t-i}. \tag{A.2}$$

It is straightforward to show that $(T-k)^{-1} \sum_{t=k+1}^{T} y_{t-1} e_t \Rightarrow \sigma_e^2 \int_0^1 W(r) dW(r)$ provided $k \to \infty$ and $k/T \to 0$ as $T \to \infty$. Consider now the second term in (A.2). We have:

$$E[(T-k)^{-1} \textstyle\sum_{t=k+1}^{T} y_{t-1} \sum_{i=k+1}^{\infty} d_i \Delta y_{t-i}]^2$$

$$= \textstyle\sum_{i=k+1}^{\infty} \sum_{j=k+1}^{\infty} d_i d_j (T-k)^{-2} \sum_{t=k+1}^{T} \sum_{s=k+1}^{T} E[y_{t-1} \Delta y_{t-i} y_{s-1} \Delta y_{s-j}]$$

$$\le C_1 \textstyle\sum_{i=k+1}^{\infty} d_i \sum_{j=k+1}^{\infty} d_j \le C_1 \sum_{i=k+1}^{\infty} \lambda^i \sum_{j=k+1}^{\infty} \lambda^j$$

$$= C_1 (\textstyle\sum_{i=k+1}^{\infty} \lambda^i)^2 = C_1 \lambda^{2k}/(1-\lambda)^2 \to 0 \text{ as } k \to \infty.$$

The first inequality follows from Said and Dickey (1984, p.601) who state that there exists a constant C_1 such that $(T-k)^{-2} \sum_{t=k+1}^{T} \sum_{s=k+1}^{T} E[y_{t-1} \Delta y_{t-i} y_{s-1} \Delta y_{s-j}] \le C_1$. The second inequality uses the fact that Δy_t is a stationary and invertible *ARMA* process, and hence there exists λ , $0 < \lambda < 1$, such that $|d_i| < C_1 \lambda^i$ for a different constant C_1. Therefore, $(T-k)^{-1} \sum_{t=k+1}^{T} y_{t-1} e_{tk} \Rightarrow \sigma_e^2 \int_0^1 W(r) dW(r)$ under the sole condition that $k/T \to 0$ and $k \to \infty$ as $T \to \infty$. Neither (A2) nor (A2″) is needed to establish the limiting distribution of (2.5).

To consider the properties of $\hat{d}(k)$ without the lower bound condition, it can be seen, from Lemma 2 of Berk or Lemma 5.2 of Said and Dickey (1984), that consistency of $\hat{d}(k)$ still holds if

$$E\left((T-k)^{-2} \sum_{j=1}^{k} \left(\sum_{t=k+1}^{T} u_{t-j}(e_{tk} - e_t)\right)^2\right) \le C_1 k \sum_{i=k+1}^{\infty} d_i^2 \le C_1 k \lambda^{2k}/(1-\lambda^2) \to 0. \tag{A.3}$$

The condition (A.3) is satisfied for any stationary and invertible $ARMA$ process provided $k \to \infty$, which is assured under (A1). More generally, the rate at which $\hat{d}(k)$ converges to $d(k)$ can be found by writing

$$\hat{d}(k) - d(k) = \left((T-k)M_k^{-1}(k) - \Gamma^{-1}\right)(T-k)^{-1}\textstyle\sum_{t=k+1}^{T} X_t e_{tk}$$
$$+\Gamma^{-1}(T-k)^{-1}\textstyle\sum_{t=k+1}^{T} X_t e_t + \Gamma^{-1}(T-k)^{-1}\sum_{t=k+1}^{T} X_t(e_{tk} - e_t). \qquad \text{(A.4)}$$

Taking norms, the first term is $o_p(T^{-1/2})$, and the second is $O_p(k^{1/2}T^{-1/2})$ by the results of Said and Dickey (1984), whether or not (A2) is satisfied. Using (A.3), the third term is $O_p(k^{1/2}\lambda^k)$ for some λ such that $|d_i| \le C_1\lambda^i$. If (A2) or (A2″) is satisfied, the second term in (A.4) dominates since the third term is $o_p(1)$. In that case, $\|\sqrt{T}(\hat{d}(k) - d(k))\| = O_p(k^{1/2})$, and $\sqrt{T}(\hat{d}_i - d_i) = O_p(1)$, $i = 1, \ldots k$. If (A2″) is not satisfied, the third term in (A.4) dominates and $\|\lambda^{-k}(\hat{d}(k) - d(k))\| = O_p(k^{1/2})$, or $\lambda^{-k}(\hat{d}_i - d_i) = O_p(1)$ $(i = 1, \ldots, k)$.

The proof of Lemma 2.1 is completed by showing $\hat{\sigma}_k^2 \to \sigma_e^2$ without any lower bound condition. The result follows from consistency of the least squares estimates. The proof is standard and is henceforth omitted.

Proof of Lemma 4.2

Let $\hat{d}(k) = (\hat{d}_1, \ldots, \hat{d}_k)$ be obtained by applying OLS to the augmented autoregression (2.4), $\tilde{d}(k) = (\tilde{d}_1, \ldots, \tilde{d}_k)$ be obtained by applying OLS to (4.1) with $x_t = \Delta y_t$. We have $\hat{d}(k) - d(k) = M_k^{-1}(k)\sum_{t=k+1}^{T} X_t' e_{tk}$, and $\tilde{d}(k) - d(k) = R_k^{-1}\sum_{t=k+1}^{T} X_t'(e_{tk} + d_0 y_{t-1}) \equiv R_k^{-1}\sum_{t=k+1}^{T} X_t' e_{tk}$ since $d_0 = 0$ under the null hypothesis of a unit root. Hence,

$$\hat{d}(k) - \tilde{d}(k) = (M_k^{-1}(k) - R_k^{-1})\sum_{t=k+1}^{T} X_t' e_{tk}.$$

Note from Lemma 5.2 of Said and Dickey (1984) that $\|(T-k)^{-1}\sum_{t=k+1}^{T} X_t' e_{tk}\| = O_p(k^{1/2}T^{-1/2})$. By partition inversion, $(T-k)M_k^{-1}(k) = ((T-k)^{-1}R_k - A)^{-1}$, where

$$A = (T-k)^{-1}(\textstyle\sum_{t=k+1}^{T} y_{t-1}X_t)(\sum_{s=k+1}^{T} y_{s-1}X_s')(\sum_{t=k+1}^{T} y_{t-1}^2)^{-1}$$
$$= (T-k)^{-3}(\textstyle\sum_{t=k+1}^{T}\sum_{s=k+1}^{T} y_{t-1}y_{s-1}X_t X_s')/(T-k)^{-2}\sum_{t=k+1}^{T} y_{t-1}^2.$$

Note that $(T-k)^{-2}\sum_{t=k+1}^{T} y_{t-1}^2 = O_p(1)$, and by Said and Dickey (1984, p.601), each element of the numerator of A is bounded by $C_1/(T-k)$ for some constant C_1. Since A is a $k \times k$ matrix, and $E(\|A\|^2) \le C_1 k^2/(T-k)$, we have $k^{1/2}\|A\|$ converges to zero provided $k^3/T \to 0$. Thus, $\|(T-k)M_k^{-1}(k) - (T-k)R_k^{-1}\|$

$$= \|(T-k)M_k^{-1}(k)\left((T-k)^{-1}R_k - \left((T-k)M_k^{-1}(k)\right)^{-1}\right)(T-k)R_k^{-1}\|$$

$$= \|(T-k)M_k^{-1}(k)\ A\ (T-k)R_k^{-1}\| \le \|(T-k)M_k^{-1}(k)\|\|A\|\|(T-k)R_k^{-1}\|.$$

Since $\|(T-k)M_k^{-1}(k)\|$ and $\|(T-k)R_k^{-1}\|$ are $O_p(1)$ (see Said and Dickey (1984)) and $k^{1/2}\|A\| \to 0$, $k^{1/2}\|(T-k)M_k^{-1}(k) - (T-k)R_k^{-1}\| \to 0$. Combining these results, we have

$$T^{1/2}\|\hat{d}(k) - \tilde{d}(k)\| \le k^{1/2}\|(T-k)M_k^{-1}(k) - (T-k)R_k^{-1}\| k^{-1/2}T^{1/2}\|(T-k)^{-1}\sum_{t=k+1}^{T} X_t' e_{tk}\| \tag{A.5}$$

$\to 0$ as $T \to \infty$ provided $k^3/T \to 0$.

We are now in a position to prove Lemma 4.2. Using the definitions of $\hat{e}_{tk}$ and $\tilde{e}_{tk}$, we have:

$$\hat{\sigma}_k^2 = (T-k)^{-1}\sum_{t=k+1}^{T}(\Delta y_t - \hat{d}_0 y_{t-1} - \hat{d}(k)'X_t)^2$$

$$= \tilde{\sigma}_k^2 + (T-k)^{-1}\hat{d}_0^2 \sum_{t=k+1}^{T} y_{t-1}^2 - 2(T-k)^{-1}\hat{d}_0 \sum_{t=k+1}^{T} y_{t-1}\tilde{e}_{tk}$$

$$+(\hat{d}(k) - \tilde{d}(k))'\left[(T-k)^{-1}\sum_{t=k+1}^{T} X_t X_t'\right](\hat{d}(k) - \tilde{d}(k))$$

$$-2\left(\hat{d}(k) - \tilde{d}(k)\right)'(T-k)^{-1}\sum_{t=k+1}^{T} X_t \tilde{e}_{tk}$$

$$+2(T-k)^{-1}\hat{d}_0\left(\hat{d}(k) - \tilde{d}(k)\right)'\sum_{t=k+1}^{T} X_t y_{t-1}.$$

We now consider each term individually.

(i). $(T-k)^{-1}(T-k)^2\hat{d}_0^2(T-k)^{-2}\sum_{t=k+1}^{T} y_{t-1}^2 = O_p(T^{-1})$ since $(T-k)^2\hat{d}_0^2 = O_p(1)$, and $(T-k)^{-2}\sum_{t=k+1}^{T} y_{t-1}^2 = O_p(1)$.

(ii). $(T-k)^{-1}(T-k)\hat{d}_0(T-k)^{-1}\sum_{t=k+1}^{T} y_{t-1}\tilde{e}_{tk} = O_p(T^{-1})$. Since $T\hat{d}_0 = O_p(1)$, we need to show that $(T-k)^{-1}\sum_{t=k+1}^{T} y_{t-1}\tilde{e}_{tk}$ is $O_p(1)$. Using the fact that $\tilde{e}_{tk} = e_{tk} + (\tilde{d}(k) - d(k))'X_t$, we have

$$(T-k)^{-1}\sum_{t=k+1}^{T} y_{t-1}\tilde{e}_{tk} = (T-k)^{-1}\sum_{t=k+1}^{T} y_{t-1}e_{tk} + (T-k)^{-1}(\tilde{d}(k) - d(k))'\sum_{t=k+1}^{T} y_{t-1}X_t.$$

The first term is $O_p(1)$ (see the proof of Lemma 2.1). We now show that the second term vanishes. We have, for $|d_i| \le C_1\lambda^i$, with $0 < \lambda < 1$,

$$\|\tilde{d}(k) - d(k)\|\|(T-k)^{-1}\sum_{t=k+1}^{T} y_{t-1}X_t\| = \begin{cases} O_p(k^{1/2}\lambda^k)O_p(k^{1/2}) & \text{if (A2) is not satisfied;} \\ O_p(k^{1/2}T^{-1/2})O_p(k^{1/2}) & \text{if (A2) is satisfied;} \end{cases}$$

and is $o_p(1)$ since $k^2/T \to 0$ in the latter case and $k \to \infty$ with $\lambda < 1$ in the former.
(iii). For the third term, taking norms, we have

$$\|(\hat{d}(k) - \tilde{d}(k))' \left((T-k)^{-1} \sum_{t=k+1}^{T} X_t X_t'\right) (\hat{d}(k) - \tilde{d}(k))\|$$

$$\leq \|(\hat{d}(k) - \tilde{d}(k))\| \|(T-k)^{-1} \sum_{t=k+1}^{T} X_t X_t'\| \|(\hat{d}(k) - \tilde{d}(k))\|$$

$$= o_p(T^{-1/2}) \cdot O_p(1) \cdot o_p(T^{-1/2}) = o_p(T^{-1})$$

using (A.5). Hence, the third term is $o_p(T^{-1})$.
(iv). $(\hat{d}(k) - \tilde{d}(k))'(T-k)^{-1} \sum_{t=k+1}^{T} X_t \tilde{e}_{tk} = (\hat{d}(k) - \tilde{d}(k))'(T-k)^{-1} \sum_{t=k+1}^{T} X_t e_{tk}$

$$+(\hat{d}(k) - \tilde{d}(k))' \left((T-k)^{-1} \sum_{t=k+1}^{T} X_t X_t'\right) (\hat{d}(k) - d(k)).$$

Taking norms, we have, for the first term,

$$\|(\hat{d}(k) - \tilde{d}(k))'(T-k)^{-1} \sum_{t=k+1}^{T} X_t e_{tk}\| \leq \|(\hat{d}(k) - \tilde{d}(k))\| \|(T-k)^{-1} \sum_{t=k+1}^{T} X_t e_{tk}\|$$

$$= o_p(T^{-1/2}) \cdot O_p(k^{1/2} T^{-1/2}) = o_p(k^{1/2} T^{-1}).$$

For the second term, we have

$$\|(\hat{d}(k) - \tilde{d}(k))' \left((T-k)^{-1} \sum_{t=k+1}^{T} X_t X_t'\right) (\hat{d}(k) - d(k))\|$$

$$\leq \|(\hat{d}(k) - \tilde{d}(k))\| \|(T-k)^{-1} \sum_{t=k+1}^{T} X_t X_t'\| \|(\hat{d}(k) - d(k))\|$$

$$= o_p(T^{-1/2}) \cdot O_p(1) \cdot O_p(k^{1/2} T^{-1/2}) = o_p(k^{1/2} T^{-1}) \text{ if (A2) is satisfied;}$$

$$= o_p(T^{-1/2}) \cdot O_p(1) \cdot O_p(k^{1/2} \lambda^k) = o_p(k^{1/2} T^{-1/2} \lambda^k) \text{ if (A2) is not satisfied.}$$

(v). Since $T\hat{d}_0 = O_p(1)$, we consider

$$\|(\hat{d}(k) - \tilde{d}(k))(T-k)^{-2} \sum_{t=k+1}^{T} X_t y_{t-1}\|$$

$$\leq \|\hat{d}(k) - \tilde{d}(k)\| \|(T-k)^{-2} \sum_{t=k+1}^{T} X_t y_{t-1}\|$$

$$= o_p(T^{-1/2}) O_p(k^{1/2} T^{-1}) = o_p(k^{1/2} T^{-3/2}).$$

Collecting results from (i) to (v), we have

$$\hat{\sigma}_k^2 = \tilde{\sigma}_k^2 + o_p(k^{1/2} T^{-1}) \text{ if (A2) is satisfied;}$$

$$\hat{\sigma}_k^2 = \tilde{\sigma}_k^2 + o_p(k^{1/2} T^{-1/2} \lambda^k) \text{ if (A2) is not satisfied.}$$

Since $k/T \to 0$ and $k^{1/2}\lambda^k \to 0$ as $k \to \infty$ and $T \to \infty$, we have, whether or not (A2) is satisfied,

$$\hat{\sigma}_k^2 = \tilde{\sigma}_k^2 + o_p(T^{-1/2}).$$

Proof of Lemma 5.1

We first note (from the proof of Lemma 4.1) that $\hat{d}(n) = \tilde{d}(n) + o_p(1)$, where $\tilde{d}(n)$ corresponds to the OLS estimates from the restricted regression without the lagged dependent variable. Using the block diagonality of M_k, we have the following asymptotic relation:

$$J(k-n,k) = (T-k)\tilde{d}(n)'\left((T-k)R_k^{-1}(n)\right)^{-1}\tilde{d}(n) + o_p(1),$$

where $R_k^{-1}(n)$ is the lower $n \times n$ block of $R_k{}^{-1}$. We now apply the following decomposition:

$$J(k-n,k) = \sqrt{T}(\tilde{d}(n) - d(n))'\left((T-k)^{-1}R_k^{-1}(n)\right)^{-1}\sqrt{T}(\tilde{d}(n) - d(n))$$

$$+2\sqrt{T}(\tilde{d}(n) - d(n))'\left((T-k)R_k^{-1}(n)\right)^{-1}\sqrt{T}d(n)$$

$$+\sqrt{T}d(n)'\left((T-k)R_k^{-1}(n)\right)^{-1}\sqrt{T}d(n) + o_p(1).$$

By Theorem 4 of Lewis and Reinsel (1984), the first term is asymptotically distributed as χ^2 with n degrees of freedom. To complete the proof, it remains to show that the other terms vanish as $T \to \infty$. We first note that $[(T-k)R_k^{-1}(n)]^{-1} \to_p R$, say. Given that Δy_t is a stationary and invertible $ARMA$ process, a typical element of $\sqrt{T}d(n)$, say $\sqrt{T}d_{k+i}$ $(i = 1,..,n)$, is such that $|\sqrt{T}d_{k+i}| \leq C_1\sqrt{T}\lambda^{k+i}$, for some C_1 and $0 < \lambda < 1$. Hence, under the conditions of (A2″), $\sqrt{T}d(n) \to 0$. It follows that the last term converges to 0. Finally, to show that the second term also vanishes, we simply note that under the conditions of (A2″), $\sqrt{T}(\tilde{d}(n) - d(n)) = O_p(1)$.

Proof of Lemma 5.2

Since $kmax$ is assumed to increase in such a way that assumption (A2″) is satisfied, the conditions of Lemma 5.1 hold and $J(kmax, kmax+n)$ is asymptotically distributed as a χ^2 random variable with n degrees of freedom. Let $\hat{k}_T$ be the estimate of k selected by the sequential procedure described in Definition 3.1. Then:

$$\lim_{T\to\infty} P\left[\hat{k}_T \neq kmax\right] = 1 - \alpha,$$

and using the rules of conditional probability,

$$\lim_{T\to\infty} P\left[\hat{k}_T = kmax - 1 | \hat{k}_T \neq kmax\right] = \alpha.$$

This implies

$$\lim_{T\to\infty} P\left[\hat{k}_T = kmax - 1 \cap \hat{k}_T \neq kmax\right]$$

$$= \lim_{T\to\infty} P\left[\hat{k}_T = kmax - 1 | \hat{k}_T \neq kmax\right] P\left[\hat{k}_T \neq kmax\right] = \alpha(1-\alpha).$$

Now

$$\lim_{T\to\infty} P[\hat{k}_T = kmax - 2 | \hat{k}_T \neq kmax - 1 \cap \hat{k}_T \neq kmax] = \alpha$$

$$= \lim_{T\to\infty} \frac{P\left[\hat{k}_T = kmax - 2 \cap \hat{k}_T \neq kmax - 1 \cap \hat{k}_T \neq kmax\right]}{P(\hat{k}_T \neq kmax - 1 \cap \hat{k}_T \neq kmax)},$$

and

$$\lim_{T\to\infty} P(\hat{k}_T \neq kmax - 1 \cap \hat{k}_T \neq kmax)$$

$$= \lim_{T\to\infty} P(\hat{k}_T \neq kmax - 1 | \hat{k}_T \neq kmax) P(\hat{k}_T \neq kmax) = (1-\alpha)^2.$$

This implies

$$\lim_{T\to\infty} P\left[\hat{k}_T = kmax - 2 \cap \hat{k}_T \neq kmax - 1 \cap \hat{k}_T \neq kmax\right] = \alpha(1-\alpha)^2.$$

We can deduce, by recursion, that

$$\begin{aligned} &\lim_{T\to\infty} P\left[\hat{k}_T = kmin \cap \hat{k}_T \neq kmin + 1 \cdots \cap \hat{k}_T \neq kmax\right] \\ &= \lim_{T\to\infty} [\alpha(1-\alpha)^{kmax - kmin}]. \end{aligned} \tag{A.6}$$

Now suppose that $kmin$ increases to infinity at a rate slower than $kmax$. From (A.6), the application of the sequential procedure implies the probability that $\hat{k}_T$ reaches $kmin$ is zero in the limit since $kmax - kmin \to \infty$. It follows that for any given $kmin$ and $kmax$, $\hat{k}_T$ must be bounded away from $kmin$. Since $kmin$ can be any arbitrary sequence, it follows that $\hat{k}_T$ has a zero probability of increasing at a rate slower than $kmax$. With the inequality $kmin \leq kmax$, $\hat{k}_T$ must, therefore, increase at the same rate as $kmax$.

Table 1.a: Size and Power of Unit Root Tests, MA Case, T = 100 (5,000 Replications).

DGP: $y_t = \rho y_{t-1} + u_t$, $u_t = e_t + \theta e_{t-1}$,

Regression: $\Delta y_t = \delta_0 y_{t-1} + \Sigma_{i=1}^{k} \delta_i \Delta y_{t-i} + v_t$.

ρ	θ	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
1.0	.8	.123	.031	.075	.037	.062	.041	.062	.041	.051	.041
1.0	.5	.103	.047	.065	.052	.057	.054	.053	.053	.051	.048
1.0	.3	.073	.051	.055	.056	.053	.047	.050	.048	.045	.046
1.0	.0	.049	.048	.046	.044	.046	.044	.045	.044	.044	.041
1.0	−.3	.091	.062	.057	.052	.052	.048	.049	.048	.048	.045
1.0	−.5	.214	.099	.068	.055	.051	.051	.052	.050	.049	.048
1.0	−.8	.880	.640	.434	.283	.200	.132	.110	.082	.074	.060
.95	.8	.307	.044	.176	.059	.129	.064	.106	.067	.085	.064
.95	.5	.237	.074	.130	.089	.103	.086	.086	.079	.077	.072
.95	.3	.162	.088	.101	.092	.092	.085	.087	.081	.075	.069
.95	.0	.117	.111	.108	.102	.099	.093	.083	.082	.080	.068
.95	−.3	.219	.139	.122	.108	.100	.091	.094	.089	.087	.081
.95	−.5	.477	.246	.157	.115	.104	.087	.085	.081	.078	.069
.95	−.8	.997	.941	.782	.584	.444	.329	.255	.203	.164	.130
.85	.8	.788	.201	.524	.212	.379	.203	.278	.166	.207	.149
.85	.5	.708	.316	.425	.297	.308	.252	.239	.203	.191	.169
.85	.3	.598	.393	.395	.334	.306	.267	.241	.214	.196	.171
.85	.0	.510	.436	.399	.343	.316	.271	.251	.218	.193	.166
.85	−.3	.746	.540	.452	.376	.344	.287	.266	.233	.209	.176
.85	−.5	.961	.779	.614	.482	.423	.353	.315	.275	.242	.208
.85	−.8	1.000	1.000	.996	.956	.886	.759	.671	.565	.483	.399

Table 1.b: Size and Power of Unit Root Tests, AR Case, T = 100 (5,000 replications).

DGP: $y_t = \rho y_{t-1} + \Sigma_{i=1}^{4} \phi_i \Delta y_{t-i} + u_t$,

Regression: $\Delta y_t = \delta_0 y_{t-1} + \Sigma_{i=1}^{k} \delta_i \Delta y_{t-i} + v_t$.

ρ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
1.0	.6	.0	.0	.0	.058	.055	.057	.053	.053	.053	.058	.055	.055	.052
1.0	−.6	.0	.0	.0	.054	.054	.054	.052	.052	.047	.045	.043	.044	.043
1.0	.4	.2	.0	.0	.033	.052	.052	.051	.050	.052	.050	.048	.047	.046
1.0	.3	.3	.25	.14	.062	.034	.038	.051	.054	.052	.050	.047	.048	.052
.95	.6	.0	.0	.0	.391	.346	.320	.281	.260	.225	.210	.191	.178	.154
.95.	−.6	.0	.0	.0	.078	.075	.073	.068	.070	.061	.061	.058	.056	.055
.95	.4	.2	.0	.0	.125	.354	.328	.288	.275	.237	.225	.204	.194	.165
.95	.3	.3	.25	.14	.137	.650	.865	.903	.876	.814	.763	.675	.608	.522
.85	.6	.0	.0	.0	.976	.938	.883	.805	.720	.626	.560	.479	.435	.357
.85	−.6	.0	.0	.0	.252	.224	.207	.188	.171	.159	.154	.134	.129	.122
.85	.4	.2	.0	.0	.818	.933	.889	.799	.718	.626	.560	.479	.435	.357
.85	.3	.3	.25	.14	.688	.964	.991	.993	.984	.960	.908	.827	.748	.647

Table 2: Approximation to the Selected Truncation Lag Using AIC in the MA(1) Model.

DGP: $x_t = e_t + \theta e_{t-1}$,

Regression: $x_t = \Sigma_{i=1}^{k} \delta_i x_{t-i} + v_t$.

$\|\theta\|$	.2	.4	.6	.8
T=100	1	2	3	3
T=10,000	3	4	7	13
T=1,000,000	4	7	12	23

Table 3.a: Frequency Count of Selected Lag Lengths k, MA Case, T = 100.

DGP: $y_t = y_{t-1} + u_t$; $u_t = e_t + \theta e_{t-1}$,

Regression: $\Delta y_t = \delta_0 y_{t-1} + \Sigma_{i=1}^{k} \delta_i \Delta y_{t-i} + v_t$.

	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
					a) $\theta = .8$					
$t_{sig}(10)$	.001	.028	.115	.131	.149	.113	.135	.104	.121	.102
$t_{sig}(5)$	.013	.103	.210	.157	.147	.092	.097	.063	.065	.053
AIC	.016	.149	.325	.219	.150	.069	.039	.020	.009	.005
Schwartz	.151	.381	.326	.100	.035	.004	.001	.001	.000	.000
					b) $\theta = .3$					
$t_{sig}(10)$	.304	.081	.067	.055	.059	.062	.074	.079	.086	.086
$t_{sig}(5)$	.468	.071	.047	.040	.038	.039	.043	.031	.049	.039
AIC	.676	.131	.037	.013	.007	.002	.002	.001	.000	.000
Schwartz	.611	.037	.004	.001	.000	.000	.000	.000	.000	.000
					c) $\theta = -.5$					
$t_{sig}(10)$	.220	.189	.075	.061	.058	.068	.069	.077	.077	.083
$t_{sig}(5)$	.387	.212	.054	.041	.039	.041	.041	.041	.041	.039
AIC	.484	.331	.073	.025	.011	.007	.003	.001	.000	.000
Schwartz	.621	.169	.016	.003	.000	.000	.000	.000	.000	.000
					d) $\theta = -.8$					
$t_{sig}(10)$	.074	.113	.109	.116	.076	.097	.075	.092	.080	.085
$t_{sig}(5)$	.123	.162	.121	.100	.061	.065	.045	.047	.045	.041
AIC	.197	.225	.146	.091	.033	.019	.008	.003	.002	.001
Schwartz	.264	.172	.056	.019	.002	.000	.000	.000	.000	.000

Table 3.b: Frequency Count of Selected Lag Lengths k, AR case, T = 100.

DGP: $y_t = y_{t-1} + \Sigma_{i=1}^{4} \phi_i \Delta y_{t-i} + u_t$,

Regression: $\Delta y_t = \delta_0 y_{t-1} + \Sigma_{i=1}^{k} \delta_i \Delta y_{t-i} + v_t$.

	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
a) $\phi_1 = .6, \phi_2 = \phi_3 = \phi_4 = .0$										
$t_{sig}(10)$	.407	.045	.047	.051	.057	.064	.071	.078	.091	.089
$t_{sig}(5)$	.644	.037	.033	.033	.037	.042	.040	.043	.047	.044
AIC	.878	.075	.027	.010	.004	.002	.001	.001	.001	.000
Schwartz	.976	.020	.003	.000	.000	.000	.000	.000	.000	.000
b) $\phi_1 = -.6, \phi_2 = \phi_3 = \phi_4 = .0$										
$t_{sig}(10)$	.407	.049	.051	.050	.060	.069	.068	.076	.081	.090
$t_{sig}(5)$	.652	.036	.037	.035	.041	.040	.036	.040	.040	.044
AIC	.866	.086	.027	.011	.006	.003	.000	.000	.000	.000
Schwartz	.978	.018	.002	.001	.000	.000	.000	.000	.000	.000
c) $\phi_1 = .4, \phi_2 = .2, \phi_3 = \phi_4 = .0$										
$t_{sig}(10)$	.200	.242	.055	.057	.060	.058	.073	.070	.092	.089
$t_{sig}(5)$	.391	.285	.038	.039	.038	.038	.043	.035	.042	.044
AIC	.455	.443	.057	.021	.008	.003	.002	.001	.001	.000
Schwartz	.680	.277	.008	.001	.000	.000	.000	.000	.000	.000
d) $\phi_1 = .30, \phi_2 = .30, \phi_3 = .25, \phi_4 = .14$										
$t_{sig}(10)$	.008	.095	.299	.149	.059	.063	.067	.078	.084	.092
$t_{sig}(5)$	.023	.192	.388	.137	.034	.037	.042	.047	.043	.046
AIC	.026	.196	.494	.204	.035	.016	.007	.003	.001	.001
Schwartz	.082	.369	.419	.076	.005	.002	.000	.000	.000	.000

Table 4.a: Size and Power of Unit Root Tests;
MA Case, T = 100, kmax = 10.

DGP: $y_t = \rho y_{t-1} + u_t,\ u_t = e_t + \theta e_{t-1}$,

Regression: $\Delta y_t = \delta_0 y_{t-1} + \Sigma_{i=1}^{k} \delta_i \Delta y_{t-i} + v_t$.

ρ	θ	$t_{sig}(10)$	$t_{sig}(5)$	AIC	Schwartz
1.0	.80	.069	.073	.068	.071
1.0	.50	.083	.087	.082	.088
1.0	.30	.075	.077	.070	.069
1.0	.00	.063	.059	.052	.046
1.0	–.30	.097	.126	.127	.174
1.0	–.50	.116	.158	.167	.244
1.0	–.80	.304	.424	.561	.733
.95	.80	.136	.151	.146	.158
.95	.50	.164	.184	.170	.196
.95	.30	.158	.162	.152	.144
.95	.00	.153	.151	.140	.126
.95	–.30	.228	.292	.294	.393
.95	–.50	.254	.336	.377	.510
.95	–.80	.534	.704	.877	.963
.85	.80	.347	.387	.405	.451
.85	.50	.445	.510	.520	.586
.85	.30	.465	.513	.536	.505
.85	.00	.486	.540	.580	.575
.85	–.30	.555	.682	.758	.859
.85	–.50	.627	.753	.860	.936
.85	–.80	.825	.908	.996	1.000

Table 4.b: Size and Power of Unit Root Tests;
AR Case, T = 100, kmax = 10.

$$\text{DGP: } y_t = \rho y_{t-1} + \Sigma_{i=1}^{4} \phi_i \Delta y_{t-i} + u_t,$$

$$\text{Regression: } \Delta y_t = \delta_0 y_{t-1} + \Sigma_{i=1}^{k} \delta_i \Delta y_{t-i} + v_t.$$

ρ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	$t_{sig}(10)$	$t_{sig}(5)$	AIC	Schwartz
1.0	.6	.0	.0	.0	.078	.075	.066	.060
1.0	−.6	.0	.0	.0	.068	.066	.066	.060
1.0	.4	.2	.0	.0	.066	.062	.055	.047
1.0	.3	.3	.25	.14	.066	.062	.058	.052
.95	.6	.0	.0	.0	.371	.399	.404	.394
.95	−.6	.0	.0	.0	.101	.099	.087	.080
.95	.4	.2	.0	.0	.346	.336	.338	.267
.95	.3	.3	.25	.14	.822	.840	.886	.837
.85	.6	.0	.0	.0	.782	.870	.960	.972
.85	−.6	.0	.0	.0	.269	.274	.268	.256
.85	.4	.2	.0	.0	.763	.824	.899	.867
.85	.3	.3	.25	.14	.901	.937	.976	.947

Table 4.c: Size of Unit Root Tests; MA Case, T = 200.
kmax = 12.

DGP: $y_t = \rho y_{t-1} + u_t,\ u_t = e_t + \theta e_{t-1},$

Regression: $\Delta y_t = \delta_0 y_{t-1} + \Sigma_{i=1}^{k} \delta_i \Delta y_{t-i} + v_t.$

ρ	θ	$t_{sig}(10)$	$t_{sig}(5)$	AIC	Schwartz
1.0	.80	.056	.060	.059	.063
1.0	.50	.061	.064	.056	.064
1.0	.30	.061	.064	.061	.066
1.0	.00	.064	.066	.059	.057
1.0	−.30	.067	.076	.076	.102
1.0	−.50	.085	.110	.121	.168
1.0	−.80	.177	.250	.366	.557

Table 4.d: Size and Power of Unit Root Tests;
AR Case, T = 200, kmax = 12.

DGP: $y_t = \rho y_{t-1} + \Sigma_{i=1}^{4} \phi_i \Delta y_{t-i} + u_t,$

Regression: $\Delta y_t = \delta_0 y_{t-1} + \Sigma_{i=1}^{k} \delta_i \Delta y_{t-i} + v_t.$

ρ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	$t_{sig}(10)$	$t_{sig}(5)$	AIC	Schwartz
1.0	.6	.0	.0	.0	.063	.060	.057	.054
1.0	−.6	.0	.0	.0	.063	.064	.058	.056
1.0	.4	.2	.0	.0	.062	.061	.056	.048
1.0	.3	.3	.25	.14	.076	.072	.070	.059
.95	.6	.0	.0	.0	.738	.815	.897	.908
.95	−.6	.0	.0	.0	.166	.168	.160	.153
.95	.4	.2	.0	.0	.712	.709	.837	.784
.95	.3	.3	.25	.14	.979	.988	1.000	.998
.85	.6	.0	.0	.0	.955	.974	1.000	1.000
.85	−.6	.0	.0	.0	.608	.603	.738	.749
.85	.4	.2	.0	.0	.954	.975	1.000	1.000
.85	.3	.3	.25	.14	.994	.997	1.000	1.000

Table 5: Size of Unit Root Tests, MA Case;
Different choices of T, kmax and kmin.

DGP: $y_t = \rho y_{t-1} + u_t$, $u_t = e_t + \theta e_{t-1}$,

Regression: $\Delta y_t = \delta_0 y_{t-1} + \Sigma_{i=1}^{k} \delta_i \Delta y_{t-i} + v_t$.

ρ	θ	$t_{sig}(10)$	$t_{sig}(5)$	AIC	Schwartz
		a) T = 200, kmax = 12, kmin = 4.			
1.0	.80	.066	.068	.066	.056
1.0	.50	.057	.055	.050	.048
1.0	.30	.062	.060	.057	.054
1.0	.00	.056	.056	.053	.052
1.0	−.30	.061	.059	.051	.050
1.0	−.50	.061	.061	.060	.059
1.0	−.80	.168	.228	.281	.343
		b) T = 500, kmax = 14, kmin = 6.			
1.0	.80	.050	.048	.050	.046
1.0	.50	.053	.052	.048	.047
1.0	.30	.057	.057	.060	.060
1.0	.00	.052	.052	.053	.052
1.0	−.30	.065	.064	.062	.059
1.0	−.50	.059	.060	.058	.056
1.0	−.80	.104	.134	.167	.213
		c) T = 500, kmax = 14, kmin = 9.			
1.0	.80	.059	.062	.057	.058
1.0	.50	.059	.059	.058	.057
1.0	.30	.058	.058	.059	.056
1.0	.0	.056	.056	.056	.056
1.0	−.30	.052	.053	.052	.052
1.0	−.50	.058	.056	.056	.056
1.0	−.80	.093	.103	.108	.116

Université de Montréal
Département de sciences économiques
Centre de documentation
, C.P. 6128, succursale Centre-ville
Montréal (Québec)
H3C 3J7

Cahiers de recherche (Discussion papers)
1993 à aujourd'hui (1993 to date)

Si vous désirez obtenir un exemplaire, vous n'avez qu'à faire parvenir votre demande et votre paiement (5 $ l'unité) à l'adresse ci-haut mentionnée. / To obtain a copy ($ 5 each), please send your request and prepayment to the above-mentioned address.

9301 : Mercenier, Jean, "Nonuniqueness of Solutions in Applied General-Equilibrium Models with Scale Economies and Imperfect Competition : A Theoretical *Curiosum*?", 26 pages.
9302 : Lemieux, Thomas, "Unions and Wage Inequality in Canada and in the United States", 66 pages.
9303 : Lemieux, Thomas, "Estimating the Effects of Unions on Wage Inequality in a Two-Sector Model with Comparative Advantage and Non-Random Selection", 54 pages.
9304 : Harchaoui, Tarek H., "Time-Varying Risks and Returns : Evidence from Mining Industries Data", 20 pages.
9305 : Lévy-Garboua, Louis et Claude Montmarquette, "Une étude économétrique de la demande de théâtre sur données individuelles", 40 pages.
9306 : Montmarquette, Claude, Rachel Houle et Sophie Mahseredjian, "The Determinants of University Dropouts : A Longitudinal Analysis", 17 pages.
9307 : Gaudry, Marc, Benedikt Mandel et Werner Rothengatter, "A Disagregate Box-Cox Logit Mode Choice Model of Intercity Passenger Travel in Germany", 17 pages.
9308 : Fortin, Nicole M., "Borrowing Constraints and Female Labor Supply : Nonparametric and Parametric Evidence of the Impact of Mortgage Lending Rules", 38 pages.
9309 : Dionne, Georges, Robert Gagné, François Gagnon et Charles Vanasse, "Debt, Moral Hazard and Airline Safety : an Empirical Evidence", 34 pages.
9310 : Dionne, Georges, Anne Gibbens et Pierre St-Michel, "An Economic Analysis of Insurance Fraud", 40 pages.
9311 : Gaudry, Marc, "Asymmetric Shape and Variable Tail Thickness in Multinomial Probabilistic Responses to Significant Transport Service Level Changes", 26 pages.
9312 : Laferrière, Richard et Marc Gaudry, "Testing the Linear Inverse Power Transformation Logit Mode Choice Model", 29 pages.

9313 : Kollmann, Robert, "Fiscal Policy, Technology Shocks and the US Trade Balance Deficit", 38 pages.
9314 : Ghysels, Eric, "A Time Series Model with Periodic Stochastic Regime Switching", 54 pages.
9315 : Allard, Marie, Camille Bronsard et Lise Salvas-Bronsard, "C*-Conjugate Expectations and Duality", 22 pages.
9316 : Dudley, Leonard et Claude Montmarquette, "Government Size and Economic Convergence", 28 pages.
9317 : Bronsard, Camille, "L'histoire de l'économie mathématique racontée à Juliette", 17 pages.
9318 : Tremblay, Rodrigue, "The Quest for Competitiveness and Export-Led Growth", 16 pages.
9319 : Proulx, Pierre-Paul, "L'ALÉNA", 12 pages.
9320 : Proulx, Pierre-Paul, "Le Québec dans l'ALÉNA", 28 pages.
9321 : Dionne, Georges, Denise Desjardins, Claire Laberge-Nadeau et Urs Magg, "Medical Conditions, Risk Exposure and Truck Drivers' Accidents : an Analysis with Count Data Regression Models", 20 pages.
9322 : Ghysels, Eric, "Seasonal Adjustment and other Data Transformations", 28 pages.
9323 : Dufour, Jean-Marie et David Tessier, "On the Relationship between Impulse Response Analysis, Innovation Accounting and Granger Causality", 12 pages.
9324 : Dufour, Jean-Marie et Eric Renault, "Causalités à court et à long terme dans les modèles VAR et ARIMA multivariés", 68 pages.
9325 : Ghysels, Eric et Alastair Hall, "On Periodic Time Series and Testing the Unit Root Hypothesis", 36 pages.
9326 : Campbell, Bryan et Jean-Marie Dufour, "Exact Nonparametric Orthogonality and Random Walk Tests", 28 pages.
9327 : Proulx, Pierre-Paul, "Quebec in North America : from a Borderlands to a Borderless Economy : an Examination of its Trade Flows with the U.S.A. at the National and Regional Levels", 24 pages.
9328 : Proulx, Pierre-Paul, "L'ALÉNA, le Québec et la mutation de son espace économique", 36 pages.
9329 : Sprumont, Yves, "Strategyproof Collective Choice in Economic and Political Environments", 48 pages.
9330 : Cardia, Emanuela et Steve Ambler, "The Cyclical Behaviour of Wages and Profits under Imperfect Competition", 24 pages.
9331 : Arcand, Jean-Louis L. et Elise S. Brezis, "Disequilibrium Dynamics During the Great Depression", 64 pages.
9332 : Beaudry, Paul et Michel Poitevin, "Contract Renegotiation : a Simple Framework and Implications for Organization Theory", 48 pages.
9333 : Dagenais, Marcel G. et Jean-Marie Dufour, "Pitfalls of Rescaling Regression Models with Box-Cox Transformations", 16 pages.
9334 : Bonomo, Marco et René Garcia, "Disappointment Aversion as a Solution to the Equity Premium and the Risk-Free Rate Puzzles", 40 pages.

9335 : Ghysels, Eric et Offer Lieberman, "Dynamic Regression and Filtered Data Series : A Laplace Approximation to the Effects of Filtering in Small Samples", 24 pages.

9336 : Kollmann, Robert, "The Duration of Unemployment as a Signal : Implications for Labor Market Equilibrium", 19 pages.

9337 : Kollmann, Robert, "Fertility, Consumption and Bequests in a Model with Non-Dynastic Parental Altruism", 22 pages.

9401 : Mercenier, Jean et Bernardin Akitoby, "On Intertemporal General-Equilibrium Reallocation Effects of Europe's Move to a Single Market", janvier 1994, 41 pages.

9402 : Gauthier, Céline et Michel Poitevin, "Using Ex Ante Payments in Self-Enforcing Risk-Sharing Contracts", février 1994, 38 pages.

9403 : Ghysels, Eric et Joanna Jasiak, "Stochastic Volatility and Time Deformation : an Application of Trading Volume and Leverage Effects", février 1994, 37 pages.

9404 : Dagenais, Marcel G. et Denyse L. Dagenais, "GMM Estimators for Linear Regression Models with Errors in the Variables", avril 1994, 33 pages.

9405 : Bronsard, C., Fabienne Rosenwald et Lise Salvas-Bronsard, "Evidence on Corporate Private Debt Finance and the Term Structure of Interest Rates", avril 1994, 42 pages.

9406 : Dinardo, John, Nicole M. Fortin et Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992 : A Semiparametric Approach", avril 1994, 73 pages.

9407 : Campbell, Bryan et Jean-Marie Dufour, "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter", avril 1994, 32 pages.

9408 : Bollerslev, Tim et Eric Ghysels, "Periodic Autoregressive Conditional Heteroskedasticity", mai 1994, 29 pages.

9409 : Cardia, Emanuela, "The Effects of Government Financial Policies : Can We Assume Ricardian Equivalence?", mai 1994, 42 pages.

9410 : Kollmann, Robert, "Hidden Unemployment : A Search Theoretic Interpretation", mai 1994, 9 pages.

9411 : Kollmann, Robert, "The Correlation of Productivity Growth Across Regions and Industries in the US", juin 1994, 14 pages.

9412 : Gaudry, Marc, Benedikt Mandel et Werner Rothengatter, "Introducing Spatial Competition through an Autoregressive Contiguous Distributed (AR-C-D) Process in Intercity Generation-Distribution Models within a Quasi-Direct Format (QDF)", juin 1994, 64 pages.

9413 : Gaudry, Marc et Alexandre Le Leyzour, "Improving a Fragile Linear Logit Model Specified for High Speed Rail Demand Analysis in the Quebec-Windsor Corridor of Canada", août 1994, 39 pages.

9414 : Lewis, Tracy et Michel Poitevin, "Disclosure of Information in Regulatory Proceedings", juillet 1994, 38 pages.

9415 : Ambler, Steve, Emanuela Cardia et Jeannine Farazli, "Export Promotion and Growth", août 1994, 41 pages.

9416 : Ghysels, Eric et Haldun Sarlan, "On the Analysis of Business Cycles Through the Spectrum of Chronologies", août 1994, 37 pages.

9417 : Martel, Jocelyn et Timothy C.G. Fisher, "The Creditors' Financial Reorganization Decision : New Evidence from Canadian Data", août 1994, 21 pages.

9418 : Cannings, Kathy, Claude Montmarquette et Sophie Mahseredjian, "Entrance Quotas and Admission to Medical Schools : A Sequential Probit Model", septembre 1994, 26 pages.

9419 : Cannings, Kathy, Claude Montmarquette et Sophie Mahseredjian, "Major Choices : Undergraduate Concentrations and the Probability of Graduation", septembre 1994, 26 pages.

9420 : Nabeya, Seiji et Pierre Perron, "Approximations to Some Exact Distributions in the First Order Autoregressive Model with Dependent Errors", septembre 1994, 40 pages.

9421 : Perron, Pierre, "Further Evidence on Breaking Trend Functions in Macroeconomic Variables", octobre 1994, 50 pages.

9422 : Vogelsang, Timothy J. et Pierre Perron, "Additional Tests for a Unit Root Allowing for a Break in the Trend Function at an Unknown Time", novembre 1994, 57 pages.

9423 : Ng, Serena et Pierre Perron, "Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag", décembre 1994, 41 pages.