A TEST FOR CHANGES IN A POLYNOMIAL TREND FUNCTION FOR A DYNAMIC TIME SERIES

Pierre Perron
Princeton University and C.R.D.E.

Econometric Research Program
Research Memorandum No. 363

March 1991
Revised July 1991

My thanks to Gregory Chow and Serena Ng for useful comments on an earlier draft. This research was supported by grants from the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada and the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec.

Econometric Research Program
Princeton University
203 Fisher Hall
Princeton, NJ 08544-1021, USA

ABSTRACT

This paper considers a test for structural change in the coefficients of a polynomial trend function in a time series of data characterized by an autoregressive noise component. The specifications analyzed are extensions of a procedure originally proposed by Gardner (1969) and MacNeill (1978). Our test is valid whether or not the noise component contains a unit root. The limiting distribution and local asymptotic power function are derived and appropriate critical values are tabulated. An extensive simulation experiment is performed to assess the size and power of the test in the case where the polynomial trend function contains only a constant or a constant and a trend. Interesting non–monotonic power properties are uncovered. In particular, the power function of the test eventually decreases to zero as the magnitude of the structural change increases. This feature is likely to be common to most test procedures for structural change when lagged dependent variables are included as regressors. We provide an explanation for this phenomenon based on the results of Perron (1989, 1990a). A modification involving the use of data in differenced form helps to mitigate this problem for many cases of practical interest. The test is applied to postwar quarterly real GNP (or GDP) series for the G–7 countries.

Key Words : Cumulative sum test, unit root process, autoregressive model, structural change, segmented trends.

Pierre Perron
Department of Economics
Princeton University
Princeton, NJ, 08544
(609) 258–4014

1. INTRODUCTION.

The issue of structural change has attracted, for understandable reasons, a lot of attention in both the statistics and econometrics literature. Much of the work has focused on developing procedures to test for structural change under restrictive assumptions. These restrictions usually involve one or more of the following : i) independently and identically distributed data, ii) non–trending data , iii) no unit roots in the univariate representation of each series. For a review of the literature the reader is referred to the annotated bibliographies of Shaban (1980) and Hackl and Westlund (1989) and the surveys of Zacks (1983), Krishnaiah and Miao (1988) and Deshayes and Picard (1986).

The nature of most macroeconomic variables is such that none of these restrictions is appealing given that many variables exhibit both a tendency to increase over time and strong serial correlation. Furthermore, it has been argued that many macroeconomic time series are better characterized as having a unit root in their univariate time series representation (see Nelson and Plosser (1982)). It is therefore of interest to consider testing procedures that allow a relaxation of these assumptions.

It is only recently that some advances have been made to tackle the issue of structural change in series where one or more of the above mentioned restrictions are not imposed. Andrews (1990) considered the Wald, the Likelihood ratio and the Lagrange Multiplier (LM) statistics for a general regression model involving data that are possibly dependent and heterogeneously distributed. Hansen (1990a) also considered LM procedures in such a context. Kramer, Ploberger and Alt (1988) extended the CUSUM test of Brown, Durbin and Evans (1975) to the case where serial correlation is present. However, none of these studies provides procedures valid for cases where the data are trending and/or contain a unit root. Kim and Siegmund (1989) considered the case where the data are trending but did not permit serial correlation and unit roots. Recent papers by Chu (1989) and Chu and White (1990) discussed tests that are valid in a univariate context where the variables can be trended and serially correlated but without a unit root. On the other hand, Banerjee, Lumsdaine and Stock (1989) discussed a procedure that is valid for testing for a change in the slope of the trend function in a time series of data characterized by the presence of a unit root. A contribution which relaxed many restrictions is that of Hansen (1990b) who considered testing for structural change in regression models with cointegration.

This paper adds a contribution by proposing and discussing the properties of a procedure designed to test for a structural change in the trend function of a univariate time series allowing for the presence of serial correlation with or without a unit root. This work is motivated by our recent investigations (Perron (1989), (1990a,b)) where we argued that many macroeconomic time series are likely to be characterized by stationary fluctuations around a trend function with a structural change rather than by a unit root process with a time–invariant drift.

Our approach is in the class of tests for structural change based on the behavior of cumulative sums introduced by Page (1955). It is an extension of the tests proposed by Gardner (1969) and extended by MacNeill (1978). More specifically it is based on the behavior of cumulative sums of estimated residuals in an autoregression where a polynomial trend function is included. The test procedure to be discussed has some optimality properties in the case where the series is normal and exhibits no serial correlation. Indeed, as shown by Gardner and MacNeill, it is the likelihood ratio test where a Bayesian prior is imposed on the possible structural change. This prior specifies at most one change (of either sign) in the values of the parameters with uniform prior probabilities on all possible time periods for the change.

The outline of the paper is as follows. Section 2 describes the model and the test statistics. Section 3 derives the asymptotic distributions under the hypothesis that a unit root is present as well as under the hypothesis that it is not present. The different rates of convergence of the original statistic proposed by MacNeill (1978) with and without a unit root imply that the statistic is useful only when the presence or absence of a unit root is known. When no such prior information is available, a modified framework is necessary to test for a structural change. Section 4 discusses such a transformation and derives the appropriate limiting distributions. Though the limiting distributions of the statistics are different with and without a unit root, their rates of convergence are shown to be the same and proper inference can be carried using the maximal critical values. These are tabulated for polynomial trend functions of various order. The local asymptotic power function of the test is discussed in Section 5.

Section 6 contains an extensive analysis of the finite sample properties of the procedure for the case of a change in the mean or a change in the slope of the trend function. Some interesting non–monotonic properties are uncovered. In particular, it is shown that the power of the test eventually decreases as the magnitude of the change increases. Sections 7 discusses alternative specifications that alleviate this problem. Section

8 presents an empirical application for testing for a change in the slope of the trend function of real GNP (or GDP) series for the G–7 countries analyzed in Campbell and Mankiw (1989), Banerjee, Lumsdaine and Stock (1990) and Perron (1990b). Finally, Section 9 contains concluding comments and an appendix contains the proof of the various theorems stated throughout the text.

Our finding that the power decreases as the magnitude of the structural change increases may also hold for many testing procedures dealing with data sets that may exhibit serial correlation over time. An example of interest is the dynamic CUSUM test of Brown, Durbin and Evans (1975) as extended by Kramer, Ploberger and Alt (1988). The properties of this procedure, analyzed in Perron (1991), are consistent with our present findings. These results call for further work concerning the properties of a wide class of tests for structural change in the context of dynamic models.

2. THE DATA–GENERATING PROCESS AND THE STATISTICS.

The basic process of interest is the following statistical model describing a given series $\{y_t\}$ as the sum of a polynomial trend function of order p (N_t) and a noise function X_t characterized by an autoregressive process of order k :

$$y_t = N_t + X_t \ , \tag{1}$$

$$N_t = \Sigma_{i=0}^{p} \beta_{i,t} t^i \ , \tag{2}$$

$$X_t = \Sigma_{j=1}^{k} \alpha_j X_{t-j} + e_t . \tag{3}$$

We denote the autoregressive polynomial by $A(L) = 1 - \alpha_1 L - \dots - \alpha_k L^k$. It is assumed throughout that the equation $A(z) = 0$ contains at most one real valued unit root and that the remaining roots lie strictly outside the unit circle. The errors $\{e_t\}$ are assumed to be i.i.d. $(0, \sigma_e^2)$ with finite fourth moment. The requirement that the noise component be an autoregressive process of finite order can be relaxed without affecting many of the results to be presented. We choose to restrict ourselves to this class of processes for ease of exposition and to keep the technical details of the proofs to a minimum. Note also that we specify the autoregressive parameters to be time invariant. Some of the procedures discussed will be consistent against time–varying α_i's but we omit this generalization as we wish to focus on the properties of the tests under the alternative of a time–varying trend function.

Under the null hypothesis, the coefficient of the trend function $\beta_{i,t}$ are assumed to be time–invariant, i.e. we have $H_0 : \beta_{i,t} = \beta_i$, for all t (i = 0, ..., p). Under the alternative hypothesis the coefficients $\beta_{i,t}$ can change at some dates. Though the power and consistency results of the tests to be presented are valid under a general class of alternative hypotheses, we shall consider the following special case of a one time change in the coefficients at a given date T_B which we assume, without loss of generality, to be some proportion of the sample size T, i.e. $T_B = \lambda T$. Throughout λ is treated as an unknown variable. For this special case we have :

$$H_1^* : N_t = \Sigma_{i=0}^{p} [\beta_i t^i + 1(t > \lambda T)(t - \lambda T)^i \delta_i] \ , \qquad (i = 0, \dots, p) \tag{4}$$

where $1(t > \lambda T) = 1$ if $t > \lambda T$ and 0 otherwise. The trend function specified by (4) is one where both segments are joined at the time of break unless $\delta_0 \neq 0$. The results are qualitatively similar if we impose the more usual specification $\beta_{i,t} = \beta_i + 1(t > \lambda T)\delta_i$. It is also useful, for later discussions, to write the model as follows :

$$y_t = \Sigma_{i=0}^{p} \gamma_{i,t} t^i + \Sigma_{j=1}^{k} \alpha_j y_{t-j} + e_t. \tag{5}$$

The coefficients $\gamma_{i,t}$ are functions of the original coefficients $\beta_{i,t}$ and α_i via the identity :

$$A(L)\Sigma_{i=0}^{p} \beta_{i,t} t^i \equiv \Sigma_{i=0}^{p} \gamma_{i,t} t^i. \tag{6}$$

For example consider the case of a trendless process ($p = 0$). Under the null hypothesis of no structural change, $A(1)\beta_0 = \gamma_0$. In the case of a first–order polynomial in t : $\gamma_1 = A(1)\beta_1$ and $\gamma_0 = A(1)\beta_0 + \Phi\beta_1$ where $\Phi = \Sigma_{j=1}^{k} j\alpha_j$ is the mean lag coefficient. Note that if a unit root is present then $A(1) = 0$ and $\gamma_1 = 0$. This generalizes since $\gamma_{p,t} = 0$ whenever $A(1) = 0$. Moreover, when a unit root is present, none of the coefficients $\gamma_{i,t}$ is a function of the intercept of the trend function $\beta_{0,t}$. These observations will prove useful when discussing the consistency and power of the tests.

To derive the test statistics consider first the following regression estimated by OLS using a sample of size T:

$$y_t = \Sigma_{i=0}^{p} \tilde{\beta}_i t^i + \tilde{e}_{p,t}, \qquad (t = 1, ..., T) \tag{7}$$

where we denote the estimated residuals by $\tilde{e}_{p,t}$ to highlight the fact that they are obtained from a regression involving a polynomial in time of order p. The test statistic considered by Gardner (1969) and MacNeill (1978), denoted by $QS_T(p)$ (the Q statistic from the static regression with a polynomial time trend of order p) is given by :

$$QS_T(p) = T^{-2}\tilde{\sigma}^{-2}\Sigma_{t=1}^{T-1}(\Sigma_{j=1}^{t}\tilde{e}_{p,j})^2, \tag{8}$$

where $\tilde{\sigma}^2 = T^{-1}\Sigma_{t=1}^{T}\tilde{e}_{p,t}^2$. MacNeill derived the asymptotic distribution of the statistic $QS_T(p)$ under the assumption that no serial correlation is present, i.e. $\alpha_j = 0$ ($j = 1, ..., k$).

In the next section we consider the limiting behavior of the statistic $QS_T(p)$ allowing for the presence of serial correlation in the noise function with and without a unit root and propose a non–parametric correction that allows, in the stationary case, valid asymptotic inference using the critical values tabulated in MacNeill (1978).

An alternative to using the estimated residuals from regression (7) is to consider the following regression, estimated from a sample of size T + k + 1, where lags of the data are introduced as regressors :

$$y_t = \Sigma_{i=0}^{p} \hat{\gamma}_i t^i + \Sigma_{j=1}^{k} \hat{\alpha}_j y_{t-j} + \hat{e}_{p,t} \qquad (t = 1, ..., T). \tag{9}$$

The test statistic $QD_T(p)$ (the Q statistic from the dynamic regression with a polynomial time trend of order p) is then defined as :

$$QD_T(p) = T^{-2} \hat{\sigma}^{-2} \Sigma_{t=1}^{T-1} (\Sigma_{j=1}^{t} \hat{e}_{p,j})^2 , \tag{10}$$

where, similarly, $\hat{\sigma}^2 = T^{-1} \Sigma_{t=1}^{T} \hat{e}_{p,t}^2$. The limiting distribution of $QD_T(p)$ is considered in section 4 for both the case where a unit root is present and the case where it is not. Note that the statistic $QD_T(p)$ could have been constructed using a two–step procedure by detrending the data in the first step. The asymptotic distribution remains identical if there is no unit root but is different if one is present. Given the simplicity of constructing $QD_T(p)$ using (9) we shall only consider this version.

3. THE LIMITING DISTRIBUTION OF $QS_T(p)$.

In this Section, we consider the limiting distribution of the statistic $QS_T(p)$ under the null hypothesis of no structural change in the coefficients of the polynomial trend function. We discuss separately the case where the autoregressive polynomial contains a unit root and the case where it does not. It is useful first to define some notation. Let the (p + 1 by p + 1) matrix D be defined by :

$$D = \begin{bmatrix} 1 & 1/2 & \cdots & 1/(p+1) \\ 1/2 & 1/3 & \cdots & 1/(p+2) \\ \cdots & \cdots & \ddots & \cdots \\ 1/(p+1) & 1/(p+2) & & 1/(2p+1) \end{bmatrix}. \qquad (11)$$

Note that D is symmetric and that D^{-1} exists. Note also that D is the limit of the appropriate normalization of the second moment matrix of the regressors in equation (7) (see the Appendix for detail). Now define the (p+1) vector of random variables $[Z_p(0), \ldots, Z_p(p)]$ by the relation :

$$\begin{bmatrix} Z_p(0) \\ Z_p(1) \\ \vdots \\ Z_p(p) \end{bmatrix} = D^{-1} \begin{bmatrix} W(1) \\ W(1) \ -\int_0^1 W(s)\,ds \\ \\ W(1) \ - \ p\int_0^1 s^{\,p-1} W(s)ds \end{bmatrix}, \qquad (12)$$

where W(s) is the unit Wiener process defined on C(0, 1), the space of real–valued continuous functions on the interval (0, 1). Note that each element $W(1) - m\int_0^1 s^{m-1} W(s)ds$ is the limit in distribution of the quantity $T^{-(m+1/2)}\Sigma_{t=1}^T t^m e_t$ when e_t is a martingale difference process with unit variance. Finally, define the quantity $B_p(r)$ as:

$$B_p(r) = W(r) - \Sigma_{i=0}^p Z_p(i) r^{i+1}/(i+1)\,. \qquad (13)$$

To understand the nature of the process $B_p(r)$ it is useful to consider some special cases. Consider first p = 0, i.e. when only a constant is included as a regressor in equation (7). Then $B_0(r) = W(r) - rW(1)$, the standard Brownian Bridge. In the case of a first–order polynomial trend (p = 1), we have (see also Kulperger (1987) and MacNeill

(1978)) :

$$B_1(r) = W(r) + 2[W(1) - 3\int_0^1 W(s)ds]r - 3[W(1) - 2\int_0^1 W(s)ds]r^2 . \qquad (14)$$

We are now in a position to state the limiting distribution of the statistic $QS_T(p)$ as T increases to infinity. In the following theorem, and throughout the text, $\Rightarrow$ denotes weak convergence in distribution.

THEOREM 1 : *Let $\{y_t\}_1^T$ be a stochastic process defined by (1) – (3) with $\beta_{i,t} = \beta_i$, and $QS_T(p)$ be defined by (8). Then, if the autoregressive polynomial does not contain a unit root (i.e. $A(1) > 0$), we have as $T \to \infty$:*

$$QS_T(p) \Rightarrow (h_x(0)/\sigma_x^2) \int_0^1 B_p(r)^2 dr ,$$

where $h_x(0) = \sigma_e^2 A(1)^{-2}$ is (2π times) the spectral density function of the process X_t evaluated at frequency zero, and $\sigma_x^2 = E(X_t)^2$ is its variance.

Remark 1 : The result of Theorem 1 is valid under more general conditions on the noise function X_t. Indeed it holds if X_t satisfies "mixing type" conditions such that a weak convergence result can be applied to their partial sums, i.e. $T^{-1/2}\Sigma_{t=1}^{[Tr]}X_t \Rightarrow h_x(0)^{1/2}W(r)$ (see, e.g., Herrndorf (1984)). In that case, $h_x(0)$ and σ_x^2 are defined as $h_x(0) = \lim_{T\to\infty} T^{-1}E(S_T^2)$ where $S_T = \Sigma_{t=1}^T X_t$, and $\sigma_x^2 = \lim_{T\to\infty} T^{-1}\Sigma_{t=1}^T E(X_t^2)$.

Theorem 1 is an extension of a result of MacNeill (1978) who considered the case where X_t is uncorrelated. In that case, $h_x(0) = \sigma_x^2$ and the limiting distribution reduces to $\int_0^1 B_p(r)^2 dr$ which is free of nuisance parameters. MacNeill (1978) has tabulated the critical values of this limiting distribution for p = 0, 1, ..., 5 (see also Section 4).

The results of Theorem 1 can also be used to provide an asymptotically valid test for structural change that relies on the critical values tabulated by MacNeill (1978). Let $\hat{h}_x(0)$ be a consistent estimator of $h_x(0)$. Examples include the autoregressive spectral density estimators and the nonparametric kernel based estimators which use the estimated residuals from (7). In the latter case, the estimator may take the form :

$$\hat{h}_x(0) = T^{-1}\Sigma_{t=1}^{T}\tilde{e}_{p,t}^2 + 2T^{-1}\Sigma_{\tau=1}^{m}\omega(m,\tau)\Sigma_{t=\tau+1}^{T}\tilde{e}_{p,t}\tilde{e}_{p,t-\tau}\,.$$

Certain regularity conditions on the window $\omega(m,\tau)$ and the rate of increase of the truncation lag m relative to the sample size T are necessary in order to ensure that $\hat{h}_x(0)$ is consistent (see Andrews (1991) for a general treatment). For later use, we simply note that a necessary condition is that $m/T \to 0$ as $T \to \infty$. With such a consistent estimator we have:

$$QS_T^*(p) \equiv (\tilde{\sigma}^2/\hat{h}_x(0))QS_T(p) \Rightarrow \int_0^1 B_p(r)^2 dr. \tag{15}$$

We consider now the case where a unit root is present in the autoregressive polynomial describing X_t, i.e. $A(1) = 0$. For this case, it is easier to adopt a different notation. Let $W_p^*(r)$ be the continuous time residuals from a projection of the Wiener process $W(r)$ on the functions $\{1, r, \ldots, r^p\}$. In the leading cases where $p = 0$ or 1, we have (e.g., Park and Phillips (1988)) :

$$W_0^*(r) = W(r) - \int_0^1 W(s)ds\,,$$

$$W_1^*(r) = W(r) - 4[\int_0^1 W(s)ds - (3/2)\int_0^1 sW(s)ds] + 6r[\int_0^1 W(s)ds - 2\int_0^1 sW(s)ds]\,.$$

Using this notation the result is stated in the following theorem.

THEOREM 2 : *Let $\{y_t\}_1^T$ be generated by (1) – (3) with $\beta_{i,t} = \beta_i$, and assume that the autoregressive polynomial contains a unit root ($A(1) = 0$). Let $QS_T(p)$ be defined by (8), then as $T \to \infty$:*

$$T^{-1}QS_T(p) \Rightarrow \int_0^1 [\int_0^r W_p^*(s)ds]^2 dr \,/\, \int_0^1 W_p^*(r)^2 dr\,.$$

Remark 2 : The result of Theorem 2 holds under more general conditions than those stated. Indeed, the same result applies if the first–differences of X_t satisfies the mixing conditions stated in Remark 1. Note also that the limiting distribution is independent of nuisance parameters and, hence, appropriate percentage points can be tabulated.

The main feature of Theorem 2 is that the statistic $QS_T(p)$ diverges as T increases. This implies that when the noise component X_t contains a unit root and the critical values from the asymptotic distribution under stationarity are used, the statistic could lead to a rejection of no structural change even if no structural change is present.

It is of interest to note that if a unit root is present the nonparametric transformation $QS_T^*(p)$ defined by (15) is also unbounded and cannot serve as the basis for tests of hypotheses. To see this, first note that $T^{-2}\tilde{\sigma}^2 QS_T(p) \Rightarrow \sigma^2 \int_0^1 [\int_0^r W_p^*(s)ds]^2 dr$ (see the Appendix for detail). Extending a result of Phillips (1991, p. 432), we have, when $\hat{h}_x(0)$ is a kernel based estimator, that $(mT)^{-1}\hat{h}_x(0) \Rightarrow \kappa\sigma^2 \int_0^1 W_p^*(r)^2 dr$. The constant κ is defined by $\kappa = \int_{-1}^1 K(s)ds$ where $K(\tau/m) = \omega(m,\tau)$ is the kernel used in constructing $\hat{h}_x(0)$. For example, with the Bartlett triangular window $\omega(m,\tau) = 1 - \tau/(m + 1)$, $K(s) = 1 - |s|$ and $\kappa = 1$. Combining these results we have $(m/T)QS_T^*(p) \equiv (m/T)(\tilde{\sigma}^2/\hat{h}_x(0))QS_T(p) \Rightarrow \int_0^1 [\int_0^r W_p^*(s)ds]^2 dr / \kappa \int_0^1 W_p^*(r)^2 dr$. Since the consistency of $\hat{h}_x(0)$ in the stationary case requires $m/T \to 0$ as $T \to \infty$, this implies that $QS_T^*(p)$ diverges as $T \to \infty$ when a unit root is present.

The above results imply that the statistic $QS_T(p)$ is useful only if we have prior knowledge whether a unit root is present or not in the noise component. Note that a pre–test procedure is not feasible here for the following reason. As documented in Perron (1989, 1990a), tests for a unit root will be biased (even asymptotically) against non–rejection of the unit root hypothesis if the trend function of the data exhibits a structural change. Given that prior knowledge of the presence or absence of a unit root is rarely, if ever, available there is a need to consider an alternative framework so that the test will remain valid irrespective of the presence or absence of a unit root. Such a procedure can be achieved using the statistic $QD_T(p)$ defined by (10) which is discussed in the next section.

4. THE LIMITING DISTRIBUTION OF $QD_T(p)$.

In this Section we consider the limiting distribution of the statistic $QD_T(p)$ constructed using cumulative partial sums of estimated residuals from a regression involving lags of the data as well as the polynomial trend function. We first consider the limiting distribution under the hypothesis that the noise function X_t does not contain a unit root.

THEOREM 3 : *Suppose that $\{y_t\}_{-k}^{T}$ is generated by (1) – (3) with $\beta_{i,t} = \beta_i$, and assume that the autoregressive polynomial does not contain a unit root ($A(1) > 0$). Let $QD_T(p)$ be defined by (10), then as $T \to \infty$:*

$$QD_T(p) \Rightarrow \int_0^1 B_p(r)^2 dr \ ,$$

where $B_p(r)$ is defined in (13).

Remark 3 : a) The proof of Theorem 3 involves a simple modification and an extension of a result in Kulperger (1987). It shows that the introduction of lags of the data in the regression effectively eliminates the dependency of the asymptotic distribution on nuisance parameters. In fact, the asymptotic distribution of $QD_T(p)$ under the model given by (1) – (3) is the same as the asymptotic distribution of $QS_T(p)$ when the data are generated by (1) and (2) with X_t being an i.i.d. sequence. b) The conditions under which Theorem 3 holds could presumably be relaxed substantially. It is indeed likely, following the work of Berk (1974) and Said and Dickey (1984), that the same result would hold if the noise component X_t was a finite order ARMA process if the number of lags in the autoregression (9) increases at a suitable rate as the sample size increases. The proof of such a theoretical extension is not undertaken in the present study. c) It is of interest to note that the asymptotic distribution is the same if the residuals are constructed using a two step procedure where the data is first detrended and the residuals are estimated from an autoregression using these detrended variables.

Percentage points of the limiting distribution in Theorem 3 have been tabulated by MacNeill (1978) for $p = 0, \ldots, 5$. For convenience we reproduce them in Table I. The case where the noise component X_t contains a unit root yields a different result stated in the following Theorem.

THEOREM 4 : *Suppose that $\{y_t\}_{-k}^{T}$ is generated by (1) – (3) with $\beta_{i,t} = \beta_i$, and assume that the autoregressive polynomial contains a unit root ($A(1) = 0$). Let $QD_T(p)$ be defined by (10), then as $T \to \infty$:*

$$QD_T(p) \Rightarrow \int_0^1 [B_p(r) + H(p)\int_0^r W_p^*(s)ds]^2 dr \ ,$$

where $B_p(r)$ is as defined in (13), $W_p^*(s)$ is the continuous time residuals from a projection of a Wiener process $W(r)$ on the functions $\{1, r, ..., r^p\}$ and $H(p)$ is the limiting distribution of $T(\hat{\alpha} - 1)$ where $\hat{\alpha}$ is the estimate, from (9), of the sum of the autoregressive coefficients $\alpha = \Sigma_{i=1}^{k}\alpha_i$. More precisely $H(p)$ is defined by :

$$H(p) = \int_0^1 W_p^*(s)dW(s) \ / \int_0^1 W_p^*(s)^2 ds \ .$$

Remark 4 : a) As in the stationary case the conditions under which the result holds could be relaxed to allow the noise component X_t to follow a finite order ARMA process if the number of lags included in regression (9) increases at a suitable rate as the sample size increases to infinity. b) It is of interest to note that, unlike the stationary case, the asymptotic distribution is different if the residuals are constructed using a two step procedure where the data is first detrended and the residuals are estimated from an autoregression using these detrended variables. c) The limiting distribution stated in Theorem 4 is a function of the order of the polynomial trend function included as regressors but is otherwise free of nuisance parameters. Hence, percentage points can be tabulated.

Theorems 3 and 4 show the asymptotic distribution of the statistic $QD_T(p)$ to be different in the cases where a unit root is present and where it is not. Accordingly, the critical values are different. However, unlike the results concerning the statistic $QS_T(p)$, the rates of convergence are the same whether or not there is a unit root. This implies that when there is no prior information on whether a unit root is present or absent a valid test would result by using the larger of the two sets of critical values. The test based on these maximal critical values will then have an asymptotic size that is no greater than the prespecified nominal size for both cases and will have an asymptotic size equal to the nominal size for one of the two cases.

The above mentioned procedure requires knowledge of the critical values of the asymptotic distribution stated in Theorem 4. To this effect we used a simulation procedure based on partial sums of N(0, 1) variables as approximations to the Wiener process. Integrals are approximated by normalized sums of 1,000 steps and 10,000 replications are generated to obtain the critical values. The relevant percentage points are presented in Table II for p = 0, ..., 5. As can be seen by comparing Tables I and II, the maximal critical values are always those corresponding to the asymptotic distribution under the hypothesis that a unit root is present in the noise component. This is useful as only one set of critical values are needed, namely those of Table II.

A comparison of the critical values in Tables I and II also provides a rough guide about the extent to which the test will be undersized if the data is actually generated by a process without a unit root but critical values for the unit root case are used. For p = 0, the test will not be much undersized since the critical values for both cases are quite similar. When p is larger the extent to which the test will be undersized increases. For example, when p = 1 and if no unit root is present, a 10% size test will lead to an asymptotic size of between 2.5 and 5% . The difference is substantial when p = 5. For example, a 10% size test will lead to an asymptotic size of below 1%. The extent to which the size is affected in finite samples will be discussed in more detail in Section 6 for cases p = 0 and p = 1.

It is useful to note that the modification which makes the procedure discussed above possible is the introduction of at least one lag of the data in the polynomial regression equation. Indeed not all lags need to be introduced. If only one lag is included a nonparametric correction is, however, needed to account for the remaining (stationary) serial correlation in the residuals. Such a correction is similar to that discussed in Section 3. However, we shall not pursue this approach in more detail in the following discussion.

The statistic $QD_T(p)$ offers a simple procedure to test for structural change in the trend function while allowing a series to be characterized by the presence or absence of a unit root. Its implementation is particularly simple as it requires only the residuals from the same regression as one would use to test for a unit root using the Dickey–Fuller (1979) methodology (in the leading cases where p = 0 or 1). The next two sections consider issues related to power and consistency in both an asymptotic context and in finite samples.

5. THE LOCAL ASYMPTOTIC POWER FUNCTION OF $QD_T(p)$.

In this Section, we study the consistency properties and the local asymptotic power function of the test statistic $QD_T(p)$ under a special class of sequences of local alternatives. The data–generating process is defined by (1) – (3). For simplicity of exposition and interpretation, the coefficients $\beta_{i,t}$ (i = 0, ..., p) are assumed to exhibit a change at a single date $T_B = \lambda T$. To study consistency, we specifically assume that N_t is generated by (4) under the alternative hypothesis of a structural change. The analysis can readily be extended to a more general class of alternative specifications, but this simple case is sufficient to illustrate the major features of interest. We start with the following Theorem concerning the issue of consistency.

THEOREM 5 : *Let $\{y_t\}$ be generated by (1) – (3) with coefficients $\beta_{i,t}$ (i = 0, ..., p) of the trend function specified by (4).*

(i) If the autoregressive polynomial does not contain a unit root ($A(1) > 0$), $QD_T(p) \to \infty$ as $T \to \infty$ provided $\delta_i \neq 0$ for at least any one i (i = 0, ..., p). Hence, the statistic is consistent against a structural change in any of the coefficients of the trend function.

(ii) If the autoregressive polynomial does contain a unit root ($A(1) = 0$), $QD_T(p) \to \infty$ if and only if $\delta_i \neq 0$ for at least one i (i = 1, ..., p) and the test is then consistent. If the only non–zero δ_i is δ_0, $QD_T(p) \Rightarrow \int_0^1 [B_p(r) + H(p)\int_0^r W_p^*(s)ds]^2 dr$ (see Theorem 4) and the test is not consistent.

Theorem 5 states that if the noise component is stationary $QD_T(p)$ will be a consistent test against a structural change in any of the parameters of the trend function. However, if the noise component contains a unit root, the test will not be consistent against an alternative hypothesis of a change in intercept. The intuition behind this result is quite straightforward when looking at the DGP in the framework described by equations (5) and (6). As we stated earlier none of the coefficients in equation (5) is a function of β_0, the intercept of the trend function when a unit root is present. Hence, the estimates and the residuals of regression (9), on which the test is based, will not be affected in large samples by a change in the intercept. Stated differently, a change in intercept causes only a one time outlier in the first–difference representation of the data and this outlier will accordingly not affect the statistic sufficiently to make it diverge and be consistent. This point, though straightforward, will prove to be of some importance in analyzing issues of changes in the level of dynamic time series.

We now turn to the analysis of the local asymptotic power function of the test. To that effect we still maintain the data–generating process as specified by (1) – (3) with the trend function now satisfying the following sequence of local alternatives :

$$H^*_{1,T} : N_t = \Sigma^p_{i=0}[\beta_i t^i + 1(t > \lambda T)(t - \lambda T)^i \delta_i / T^{i+1/2}] . \qquad (i = 0, \ldots, p) \qquad (16)$$

To motivate the normalization specified by (16), note that we can then write (3) as :

$$N_t = \Sigma^p_{i=0}\beta_i t^i + \Sigma^p_{i=0} 1(t > \lambda T)(\delta_i / T^{1/2})(t/T - \lambda)^i . \qquad (17)$$

Hence, each component $(\delta_i/T^{1/2})(t/T - \lambda)^i$ converges to zero at the rate $T^{1/2}$, as is usual in local asymptotic analyses. To state our result, we need to define additional notation. Let $Z^*_p \equiv [Z^*_p(0), Z^*_p(1), \ldots, Z^*_p(p)]$ be the vector of random variables defined by :

$$Z^*_p = D^{-1} V^*_p , \qquad (18)$$

where D is defined in (11) and V^*_p is the (p+1) vector $[V^*_p(0), V^*_p(1), \ldots, V^*_p(p)]$ with typical elements defined by (for m = 0, ..., p) :

$$V^*_p(m) = W(1) - m\int_0^1 s^{m-1} W(s)ds + A(1)\Sigma^p_{i=0}(\delta_i/\sigma_e)\int_0^{1-\lambda}(r + \lambda)^m r^i dr . \qquad (19)$$

Note that $V^*_p(m)$ differs from the typical element on the right hand side of (12) by the third term only. Note also that this last term vanishes as A(1) approaches 0. When p = 0, we have :

$$Z^*_0(0) = W(1) + A(1)(\delta_0/\sigma_e)(1 - \lambda) ,$$

and when p = 1, we have :

$$Z^*_1(0) = (1/12)\Big\{2\int_0^1 W(s)ds - 6W(1) + A(1)(\delta_0/\sigma_e)(1 - \lambda)(1 - 3\lambda)/12$$

$$+ A(1)(\delta_1/\sigma_e)\lambda(1-\lambda)^2/12\Big\},$$

$$Z_1^*(1) = (1/12)\Big\{W(1)/2 - \int_0^1 W(s)ds + A(1)(\delta_0/\sigma_e)\lambda(1-\lambda)/2$$
$$+ A(1)(\delta_1/\sigma_e)(1-\lambda)^2(1+2\lambda)/12\Big\}.$$

Corresponding to the elements $B_p(r)$ defined in (13), we also define the following variable :

$$B^*_{p,\delta}(r) = W(r) - \Sigma_{i=0}^{p} Z_p^*(i) r^{i+1}/(i+1)$$
$$+ 1(r > \lambda)A(1)\Sigma_{i=0}^{p}(\delta_i/\sigma_e)(r-\lambda)^{i+1}/(i+1) \ , \qquad (20)$$

where $1(r > \lambda) = 1$ if $r > \lambda$ and 0 otherwise. $B^*_{p,\delta}(r)$ differs from the variable $B_p(r)$ discussed in Section 3 by the inclusion of $Z_p^*(i)$ instead of $Z_p(i)$ and by the last component which vanishes as A(1) approaches 0. We are now in a position to state the following result concerning the local asymptotic power function of the test.

THEOREM 6 : *Suppose that $\{y_t\}$ is generated by (1) – (3) with coefficients $\beta_{i,t}$ specified by the sequence of local alternatives (16), then as $T \to \infty$:*

i) If the autoregressive polynomial has no unit root ($A(1) > 0$) :

$$QD_T(p) \Rightarrow \int_0^1 B^*_{p,\delta}(r)^2 dr \ ;$$

ii) If the autoregressive polynomial has a unit root ($A(1) = 0$) :

$$QD_T(p) \Rightarrow \int_0^1 [B_p(r) + H(p)\int_0^r W_p^*(s)ds]^2 dr \ ,$$

where $B^*_{p,\delta}(r)$ is defined by (20), $B_p(r)$ by (13) and the variables $H(p)$ and $W_p^*(s)$ are as defined in Theorem 4.

Theorem 6 has several interesting implications. Consider first the case where the noise component is stationary. The statistic $QD_T(p)$ has a non–degenerate asymptotic local power function. The power increases as any δ_i increases in absolute value. However, for a given value of δ_i, the local limiting distribution approaches the limiting null distribution as A(1) approaches zero. We would therefore expect the test to have lower power, in finite samples, if the noise component corresponds to a more persistent process.

This statement is reinforced by the result concerning the case where a unit root is present. Here the test has a degenerate asymptotic local power function in the sense that the limiting distribution under the sequence of alternatives specified by (16) is the same as the limiting distribution under the null hypothesis. Hence, the asymptotic local power function of $QD_T(p)$ is equal to the size of the test when a unit root is present in the process generating the data.

Some intuition for the degenerate asymptotic local power function in the unit root case can be obtained using the following argument. Consider the model (1) – (3) with the coefficients satisfying the sequence of local alternatives specified by (16). If a unit root is present, we can write the data–generating process as :

$$y_t = \Sigma_{i=0}^{p-1}\gamma_i t^i + 1(t > \lambda T)\Sigma_{i=1}^{p-1}(\eta_i/T^{3/2})(t/T - \lambda)^i + \Sigma_{j=1}^{k}\alpha_j y_{t-j} + e_t , \tag{21}$$

where γ_i and η_i are defined by the relations (see (6)):

$$A(L)\Sigma_{i=0}^{p}\beta_i t^i = \Sigma_{i=0}^{p}\gamma_i t^i , \tag{22}$$

$$A(L)\Sigma_{i=0}^{p}(\delta_i/T^{i+1/2})(t - \lambda T)^i = \Sigma_{i=0}^{p}(\eta_i/T^{i+1/2})(t - \lambda T)^i . \tag{23}$$

What transpires from equation (21) is that the coefficients on the components of the trend function converge to the value under the null hypothesis at rate $T^{3/2}$ when there is a unit root instead of at rate $T^{1/2}$ in the stationary case. Given the rate of convergence of the estimates in regression (9), this rate of approach to the null values is too fast to affect the limiting distribution.

Of course, for this unit root case, one could define a sequence of local alternatives different from that in (16) by specifying $N_t = \Sigma_{i=0}^{p}[\beta_i t^i + 1(t > \lambda T)(t - \lambda T)^i\delta_i/T^{i-1/2}]$. This would allow a non–degenerate local asymptotic distribution unless the only non–zero non–centrality parameter is δ_0. This shows again that the test is consistent against changes in any coefficient of the trend function except the intercept. The derivation of an asymptotic power function under this modified sequence of alternative is straightforward adopting the methods used in this paper. We refrain from providing such a generalization given that our aim is not in obtaining approximations to the power functions but rather to provide a framework for qualitative comparison of the power functions as the sum of the autoregressive coefficients varies.

The above discussion suggests the following predictions about the finite sample power of the statistic $QD_T(p)$ against alternatives of a one–time change in some coefficients of the trend function. First, the test is likely to have respectable power if the noise component does not exhibit too much persistence, i.e. $A(1)$ is not too close to 0. As $A(1)$ approaches 0 the power will be small against changes in any component of the trend function whether it be the intercept or any other coefficients (such as the rate of growth in a first–order polynomial trend function). Secondly, the power will be even lower in the case where a unit root is present. In that case, the power for a change in any coefficient (except the intercept) will increase slowly as the sample size increases given that the test is consistent but has zero local asymptotic power. In the case of a change in intercept the power will not increase given the inconsistency of the test against such alternatives. These and other features are documented in the next section which presents a simulation study of the finite sample properties.

6. A SIMULATION ANALYSIS OF THE FINITE SAMPLE PROPERTIES.

For the analysis of the finite sample properties of the test $QD_T(p)$ we concentrate on the leading cases where p is 0 or 1. When p = 0, the test is applied to detect a possible change in the mean of a trendless series, and when p = 1 the test can be used for either a change in the intercept or the slope of the trend function (or both). We start with a discussion of the finite sample size of the test. The design of the experiment is as follows. We generate samples of length 100 from the following special case of model (1) – (3) :

$$y_t = \alpha y_{t-1} + e_t , \tag{24}$$

where $e_t \sim$ i.i.d. N(0,1) and $y_0 = 0$. Model (24) simply specifies an AR(1) with constant mean 0. There is no loss in generality in specifying the parameters β_i to be zero under the null hypothesis. We generated 10,000 replications of the process (24), and for each one we calculated the statistics $QD_T(0)$ and $QD_T(1)$ as specified in (10) and (9) with k = 1. The experiment was performed for 20 values of α, namely $\alpha = -0.9,(.1), 1.0$. The critical values used are those corresponding to the asymptotic distribution for the case $\alpha = 1$ (Table II).

Figure 1 presents the exact size of the test for nominal sizes of 1, 5 and 10% for the case where p = 0, i.e. when only a mean is estimated. As can be seen from this figure, the exact size of the test is below the nominal size for all values of α. However, the extent to which the test is undersized is not severe especially if α is below 0.6. When α is between .7 and .9 the discrepancies are somewhat larger. As expected the exact size of the test shows a large change from $\alpha = 0.9$ to $\alpha = 1.0$. Figure 2 presents similar results for the case where p = 1, i.e. when $QD_T(p)$ is constructed from a regression with a first–order polynomial in t. Again, the test is conservative for all values of α. However, the extent to which the test is undersized is more important. When the process is stationary, a test with a 5% nominal size has an exact size of approximately 1%. This feature is to be expected given the larger relative discrepancies between the asymptotic distribution of $QD_T(p)$ in the stationary and unit root cases for higher values of p. This reduction in the exact size (for stationary processes) seems to be a price that one has to pay in order not to impose any a priori restriction about the presence or absence of a unit root in the noise function.

We now turn to the analysis of the power of the test. Consider first the case where p = 0 and the test is to detect a change in mean. Under the alternative hypothesis the data are generated by the following special case of the model (1) through (4):

$$y_t = 1(t > \lambda T)\delta_0 + X_t\,, \qquad (t = 1, ..., T) \tag{25}$$

$$X_t = \alpha X_{t-1} + e_t, \tag{26}$$

where $e_t \sim$ i.i.d. N(0,1) and $X_0 = 0$. The model consisting of (25) and (26) specifies an autoregressive process with a changing mean at date λT, the magnitude of the change being δ_0. We consider only one value of the sample size, namely T = 100. Again we consider 20 values of α, $\alpha = -0.9$, (0.1), 1.0 and the power of the test is evaluated at 20 different values of δ_0, $\delta_0 = 1$, (1), 20. For each of these cases 5,000 replications are used and the experiment is performed for $\lambda = 0.25$, 0.50 and 0.75. The nominal size of the test is 5% using the critical values from the asymptotic distribution with a unit root (Table II).

The power functions graphed in Figure 3 (a, b and c) are for cases $\lambda = 0.25$, 0.50 and 0.75 respectively. Several features stand out from these graphs. First, for a given value of the change in mean, δ_0, the power is lower when α is closer to 1 (as expected given the local asymptotic power function derived in the previous section). The increase in power is rapid going from $\delta_0 = 0$ to $\delta_0 = 1$, reaching one for negative values of α. This initial increase is less rapid as α increases, i.e. the slope of the power function at $\delta_0 = 0$ decreases as α increases. Again, this feature is well explained by the local asymptotic power result. When $\alpha = 1$, the power does not increase as δ_0 increases in accord with the fact that the test $QD_T(0)$ is inconsistent in this case. Also, the power is larger for a change occurring late in the sample. However, the most striking feature is that *for a given value of α the power eventually decreases to zero as the magnitude of the change in mean (δ_0) increases.*

The striking feature that the power function eventually decreases to zero as δ_0 increases is not an implication of any of the asymptotic results discussed previously. Yet the intuition behind it is quite simple. As documented in Perron (1990a), a change in mean causes a bias in the least–squares autoregressive estimator obtained from regression (9) with p = 0. This bias is such that the estimate of α is attracted to the value 1. In the limit, as δ_0 increases and with a large sample, the estimator of the autoregressive parameter converges to one, irrespective of the true value of α (in the range permitted here). Therefore as δ_0 increases, the fitted process behaves like a random walk with an outlier at time λT reflecting the change in mean, in which case the test has no power. It is then clear that such a change cannot be detected. As seen from the graphs in Figure 3, the power of the test decreases quite rapidly once δ_0 reaches a certain level. This threshold level is smaller and the rate of decrease is greater as α approaches one.

We now turn to the simulation results concerning the behavior of the statistic $QD_T(1)$ where a first–order polynomial in t is included in the regression (9). We first consider the power of the test against a change in the intercept of the trend function. For this case, the setup of the experiment is similar in almost every aspect as the one above. In particular, the data are again generated by (25) and (26). The only difference is that we consider 14 values of δ_0, namely $\delta_0 = 1, (1), 14$. The results are presented in Figure 4 (a, b and c) for the cases $\lambda = 0.25$, 0.5 and 0.75, respectively. They are qualitatively similar to those for the statistic $QD_T(0)$ discussed above. Again, the power of the test eventually decreases to zero as the magnitude of the change in mean increases, and, for a given value of δ_0, the power is lower as α approaches one. The results again reflect the fact that the statistic is inconsistent against a change in mean if $\alpha = 1$ as the test is biased for all values of δ_0. The differences from the case where no trend is estimated are that the power is uniformly lower and that the power seems to be lowest for a change occurring at mid–sample ($\lambda = 0.5$).

The third simulation experiment concerns the power of the statistic $QD_T(1)$ for detecting a change in the slope of the trend function. Here the data are generated by :

$$y_t = 1(t > \lambda T)(t - \lambda T)\delta_1 + X_t , \qquad (t = 1, \ldots, T) \qquad (27)$$

where X_t is again given by (26). The model (27) specifies a joint segmented trend function with initial slope and mean 0 until time λT after which the slope changes to δ_1. The specifications are as above ($T = 100$, $\lambda = 0.25, 0.5, 0.75$, $\alpha = -0.9, (.1), 1.0$ and 5,000 replications) except that δ_1 takes 10 different values specified by $\delta_1 = 0.05, (0.05), 0.5$.

The results presented in Figure 5 (a, b and c) are for cases $\lambda = 0.25$, 0.5 and 0.75, respectively. They are, somewhat surprising at first sight, similar to the power results for a change in mean. Again, as expected from the local asymptotic power function derived in Section 5, the power is lower for a given δ_1 as α approaches one. Again, for a fixed α, the power eventually decreases as the magnitude of the change in slope increases. This feature is not implied by any of the asymptotic results presented earlier. However, these results can be used to provide an intuitive explanation for this behavior. As documented in Perron (1989), a change in slope will create a bias in the least–squares estimator of the autoregressive coefficient in the regression model (9) with $p = 1$ (and $k = 1$). This bias is, again, such that the estimator is attracted to the value one. In fact, for any fixed change in slope δ_1, the limit of the estimator of the autoregressive parameter is one as the sample size increases. In this case the fitted model behaves like a unit root process with a change in drift (or slope of the trend function). However, as documented in Theorem 6, the test has a degenerate asymptotic local power function in that unit root case. Hence, the power should be low when the change in slope is large (even though the test is consistent).

A final simulation experiment concerning the power of the statistic $QD_T(1)$ is performed to analyze the consistency property of the test against a change in slope when $\alpha = 1$. As stated in Theorem 5, $QD_T(1)$ is consistent in that case. However, the simulation results presented in Figure 5 do not seem to support this fact, the power being barely above the size in the most favorable cases. It is also of interest to see how fast the power approaches 1 in the case of a consistent test with a degenerate local asymptotic power function. To this effect we simulated the process (27) with $\alpha = 1$, $\lambda = 0.5$ and $\delta_1 = 0.5$. The experiment was repeated for sample sizes T = 100, 200, 500, 1000, 2000 and 5000. The results are presented in Table III for tests with size 1, 2.5, 5 and 10%. As can be seen from the results, the power increases very slowly as the sample size increases. With T = 500, a 5% test has a power of only .12. Even at T = 2000, the power is only .860. It is only when T = 5000 that the power of the test is one. The slow convergence documented is likely to hold also for stationary processes with a large autoregressive parameter.

These simulation results have the following implications. For sample size of common lengths, the test will not be able to detect a change in either the mean or the slope *of any magnitude* if the underlying noise component shows some persistence, e.g. a value of α greater than, say, 0.7. If the process shows less persistence, e.g. α less than 0.7, the test will be good at detecting *small changes in the mean or the slope but not large changes.*

These features are particularly troublesome when analyzing macroeconomic data since in most cases the underlying noise component is likely to exhibit some degree of persistence, even under the hypothesis that a change in the trend function is present (see, Perron (1989, 1990b) for examples). This calls for two different further topics of investigation. First, to see if some modifications are possible to allow greater power when the noise component is strongly positively correlated and/or the magnitude of the change is large. When considering a change in mean no such modification appears to be available. However, in the case of a change in slope, we discuss in the next section a modification which appears helpful for cases of practical interest. A second topic of investigation is to find a theoretical framework which can explain this non–monotonic power function and to provide a better approximation to these finite sample properties. This will be undertaken in a subsequent study.

7. TESTING FOR A BREAKING TREND USING FIRST–DIFFERENCED DATA.

In this Section, we consider an alternative specification of the testing procedure that alleviates, in some ways, the power problems discussed in the last section. The following analysis is motivated by the fact that for many empirical applications of interest with macroeconomic data, the noise component is strongly correlated implying that we are in a region where the statistic $QD_T(p)$ has basically no power even if the structural change is large. Our analysis is specialized to the case of a change in the slope of the trend function. The data–generating process is assumed to be of the form :

$$y_t = N_t + X_t = \beta_0 + \beta_{1,t}t + X_t , \tag{28}$$

where X_t is defined by (2). The null hypothesis specifies that $\beta_{1,t} = \beta_1$ for all t and we consider the following alternative hypothesis :

$$H_1^{**} : N_t = \beta_0 + \beta_1 t + 1(t > \lambda T)(t - \lambda T)\delta_1 . \tag{29}$$

Consider first the case where a unit root is present in the noise component. We define the first–differences of the data by $dy_t = y_t - y_{t-1}$. With the coefficients α_i^* defined by the relation $A(L) = (1 - L)A^*(L)$, we have an AR(k–1) in first–differences:

$$dy_t = \gamma_{0,t} + \Sigma_{i=1}^{k-1}\alpha_i^* dy_{t-i} + e_t , \tag{30}$$

where $\gamma_{0,t}$ is defined by the relation $A(L)\beta_{1,t}t = \gamma_{0,t}$. Hence, by considering the data in first–differences we transform a case with a unit root into a case where the noise component is stationary and the change in the slope of y_t corresponds to a change in the mean of dy_t. When the original noise component is stationary we have :

$$dy_t = \gamma_{0,t} + \Sigma_{j=1}^{k}\alpha_j dy_{t-j} + u_t , \tag{31}$$

where $u_t = e_t - e_{t-1}$ and $\gamma_{0,t} = A(1)\beta_1 + 1(t > \lambda T)A(1)\delta_1$. In particular, (31) shows that if the noise component is stationary, first–differencing the data induces a moving–average unit root in the residuals. For empirical implementations we consider the following estimated regression :

$$dy_t = \hat{\beta}_1 + \Sigma_{j=1}^{k}\hat{\alpha}_j dy_{t-j} + \hat{u}_t\ , \tag{32}$$

and use the estimated residuals $\hat{u}_t$ to construct the statistic $QF_T(1)$ (the statistic Q using first–differenced data applied to test for a change in a first order polynomial trend):

$$QF_T(1) = T^{-2}\hat{\sigma}^{-2}\Sigma_{t=1}^{T-1}(\Sigma_{j=1}^{t}\hat{u}_j)^2, \tag{33}$$

where $\hat{\sigma}^2 = T^{-1}\Sigma_{t=1}^{T}\hat{u}_t^2$. We have the following result concerning the asymptotic distribution of $QF_T(1)$ under the null hypothesis of no structural change.

THEOREM 7 : *Suppose that* $\{y_t\}$ *is generated by (28) with* $\beta_{1,t} = \beta_1$ *for all t. Consider the statistic* $QF_T(1)$ *defined by (33) constructed using data in first–differences. Let* $B_0(r)$ *be defined by (13). Then as* $T \to \infty$*:*

a) if there is a unit root in the noise component $(A(1) = 0)$: $QF_T(1) \Rightarrow \int_0^1 B_0(r)^2 dr$.

b) if there is no unit root in the noise component $(A(1) > 0)$: $QF_T(1) \Rightarrow 0$.

c) if the the largest root of A(L) is modeled as local to unity, i.e. $A(L) = (1 - \alpha_T L)A^*(L)$ *where* $\alpha_T = 1 - c/T$ *and the parameters of* $A^*(L)$ *do not depend upon T :*

$$QF_T(1) \Rightarrow \int_0^1 [B_0(r) - c\int_0^r J_c^*(s)ds]^2 dr\ ,$$

where $J_c^*(r) = J_c(r) - \int_0^1 J_c(r)dr$, *the demeaned version of the Ornstein–Uhlenbeck diffusion process defined by* $J_c(r) = \int_0^r exp(c(r - s))dW(s)$ *where* $W(s)$ *is the Wiener process defined on* $C(0,1)$.

Theorem 7 has the following implications. First if there is a unit root in the noise component, the asymptotic distribution of $QF_T(1)$ is the same as the asymptotic distribution of $QD_T(0)$ in the stationary case. Hence, the appropriate asymptotic critical values are those of the first column of Table 1. When the noise component is stationary, the limit of $QF_T(1)$ is zero. Hence, in keeping with our earlier approach, an asymptotically valid test can be constructed using the critical values under the unit root case. This implies a test with zero asymptotic size in the stationary case. This shows a discontinuity in the asymptotic distribution which is not present in the finite sample distribution. Given that this version of the test procedure is intended to apply to processes with strong correlation in the noise component, it may be better to consider an asymptotic framework where the largest root of the autoregressive polynomial A(L) is local to unity. This is done in part (c), which shows a non–degenerate local asymptotic distribution in the stationary case which reduces to the limit distribution stated in part (a) when $c = 0$ (i.e. when a unit root is present). In this context, the statistic $QF_T(1)$ will have a non–zero asymptotic size if the critical values from the distribution in the unit root case are used. As before the asymptotic size under stationarity will be less than the size of the test under a unit root.

The following results consider the behavior of the test under the alternative hypothesis. We discuss consistency against the alternative specified by (29) and the local asymptotic power function under a sequence of alternatives defined by (16) with $i = 1$.

THEOREM 8 : *Suppose that $\{y_t\}$ is generated by (28) and consider the statistic $QF_T(1)$ defined by (33) constructed using data in first–differences. Suppose that the rejection region is constructed using the critical values of the limiting null distribution in the unit root case (Theorem 7, part (a)). Then as $T \to \infty$:*

a) If N_t is specified by (29) : $QF_T(1) \Rightarrow \infty$;

b) If N_t is specified by (16) with $i = 1$: i) if a unit root is present ($A(1) = 0$), $QF_T(1) \Rightarrow \int_0^1 B_0(r)^2 dr$; ii) $QF_T(1) \Rightarrow 0$ if a unit root is not present in the autoregressive polynomial ($A(1) > 0$); iii) if the the largest root of $A(L)$ is modeled as local to unity, i.e. $A(L) = (1 - \alpha_T L)A^*(L)$ where $\alpha_T = 1 - c/T$ and the parameters of $A^*(L)$ do not depend upon T, $QF_T(1) \Rightarrow \int_0^1 [B_0(r) - c\int_0^r J_c^*(s)ds]^2 dr$, with $B_0(r)$ and $J_c^*(r)$ as defined in Theorem 7.

Part (a) of Theorem 8 applies whether or not the noise component of the process contains a unit root. It states that the test is consistent against a change in the slope of the trend function (even in the stationary case where the asymptotic size of the test is zero). Part (b) considers the local asymptotic power function of the test. It shows that it is degenerate in all cases considered. If no unit root is present the asymptotic local power of the test is zero, unless the asymptotic framework used is the nearly integrated one. If a unit root is present the asymptotic local power is equal to the size of the test. This implies that we should expect the finite sample power function to be low for small changes in the slope of the trend but does not have much implications with respect to the finite sample power for large changes. These finite sample issues are analyzed in the following simulation experiment.

To allow proper comparisons with the statistic $QD_T(1)$ we generated data in exactly the same way as we did in the last section, namely using (27) as the data–generating process. Exactly the same specifications were used except that we considered a wider range of values for δ_1, namely from 0 to 1.0 (again in steps of .05). The results are presented in Figure 6 for the three cases corresponding to a change at different points in the sample, namely, $\lambda = 0.25$, 0.50 and 0.75.

The results are quite striking. First, as expected from the degenerate local asymptotic result the power is very low (essentially zero) for low values of δ_1 , say less than 0.3. This is contrary to the behavior of the statistic $QD_T(1)$ where the power is quite high in that part of the parameter space if the autoregressive coefficient is not too large. More interestingly, the power increases rapidly to reach a value of one as δ_1 is increased further. This behavior is again the opposite to that of $QD_T(1)$. Hence, as the statistic based on levels of the data, $QD_T(1)$, is not able to detect large changes in the slope of the trend function, the statistic based on data in first–differences, $QF_T(1)$, is able to do so even if the noise component is stationary. Conversely, $QD_T(1)$ is, in general, good at detecting small changes while $QF_T(1)$ is not. The only part of the parameter space where neither statistic is able to detect small to medium changes in the slope of the trend function is when the autoregressive parameter is large, say greater than 0.7.

The theoretical and simulation analyses described above suggest that, in the case of a change in the slope of the trend function, the statistics $QD_T(1)$ and $QF_T(1)$ should be used in conjunction as they have complementary properties. If the underlying noise component contains a strong correlation, only the statistic $QF_T(1)$ will be useful and for large change only. The next section uses the statistic $QF_T(1)$ to test for the presence of a change in slope in several real GNP (or GDP) series.

8. EMPIRICAL APPLICATIONS.

This Section analyzes an international data set of postwar quarterly real GNP or GDP series. The type (GNP or GDP) and the sampling period of the series used, listed in Table IV, were dictated by data availability and a desire to obtain results that are comparable with previous studies. The countries analyzed are : USA, Canada, Japan, France, Germany, Italy and the United Kingdom. The series for Canada was obtained from the Cansim data bank and the series for Japan and France from the IFS data tape. The remaining series (U.S.A., U.K., Germany and Italy) are from Data Resources Inc. and are the same as those used in Campbell and Mankiw (1989). All series are seasonally adjusted and at annual rates, except for the USA and the United Kingdom which are at quarterly rates. The aim of the analysis is to test whether these series are characterized by the presence of a change in slope as argued in Perron (1990b).

To apply the test we used a data–dependent method to select the order of the autoregression. This was done using a sequential t–statistic on the coefficient of the last lag of the estimated autoregression. We started with a maximum order which we arbitrarily set at kmax = 8. If the coefficient on the eighth lag is not significant, using a two–tailed 5% test with the critical values from the normal distribution, we estimated an autoregression of order 7 and repeated the test on the coefficient of its last lag, and so on until a rejection is found. If none is significant we set $k = 0$. If $k = 8$ yields a coefficient on the last lag which is significant we checked whether the residuals exhibit any remaining correlation using the Box–Pierce statistic. In no case for which $k = 8$ was selected was there evidence of further serial correlation.

We first applied the statistic $QD_T(1)$ using the levels of the data and failed to reject the null hypothesis of no structural change for all series. This result is not surprising given that the underlying noise component of the series is most likely to exhibit strong positive serial correlation. Hence we constructed, for each series, the statistic $QF_T(1)$ using data in first–differences. The test should be able to reject no structural change if the change in slope is large enough. The results are presented in Table IV. The null hypothesis of no structural change can be rejected for all countries except the USA. The rejection is at the 10% level for Germany, Japan and the U.K., at the 5% level for Canada, at the 2.5% level for Italy and at the 1% level for France. Again, these rejections hold irrespective of the presence or absence of a unit root in the noise component.

9. CONCLUSIONS.

This paper has considered in detail the behavior of a common test for structural change introduced by Gardner (1969) and MacNeill (1978). In a first step, we have extended this testing procedure to the case where serial correlation is possibly present in the data. In particular, we have allowed for the possibility that the noise component of the series be characterized by the presence of a unit root. These extensions are such that the proper testing procedure implies the need to incorporate lags of the data in the regression defining the residuals on which the test is based. It was shown how the introduction of such lagged dependent variables induces a peculiar behavior of the power function of the test in finite samples. More specifically, the power was shown to eventually decrease to zero as the magnitude of the change in the coefficient increases. In the case of a change in the slope of the trend function, it was shown how a test based on a regression in first–differences allows one to partially circumvent such a drawback for cases of practical interest.

This non–monotonic behavior of the power function is likely to be common to most tests for structural change that permits the possibility of serially correlated data. For example, the dynamic version of the CUSUM test of Brown, Durbin and Evans (1975) for the standard linear regression model, as extended by Kramer, Ploberger and Alt (1988), also share this property (see Perron (1991)). Our conjecture is that all tests that permit the possibility that the noise component may contain serial correlation of the unit root type will share this property. If an a priori restriction that the possible serial correlation be of a stationary nature is imposed, it may be possible to circumvent this problem. However, such a test would require a correction for serial correlation that is not based on the introduction of lagged dependent variables in the data. Of course, these conjectures are at this time highly speculative and we hope to report additional results in the near future. In any event, this paper has demonstrated that care must be exercised when applying tests for structural change in a dynamic context.

The test presented in this paper is easy to implement and can be useful, for example, as a diagnostic in the application of unit root tests of the type proposed by Dickey and Fuller (1979). However, as our power analysis demonstrated, care must be exercised when interpreting the results. In the case where a change in the slope of the trend function is suspected it would be highly desirable to use both the version of the Q statistic from the dynamic regression in levels and that from the regression in first–differences.

REFERENCES

Andrews, D. W. K. (1990): "Tests for Parameter Instability and Structural Change With Unknown Change Point," Cowles Foundation Discussion paper No. 943, Yale University.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Banerjee, A., R. L. Lumsdaine, and J. H. Stock (1989): "Recursive and Sequential Tests for a Unit Root: Theory and International Evidence," mimeo, Harvard University.

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

Bhat, B. R. and K. S. Chandra (1988): "Least Squares Estimation of the Parameters of a Stochastic Equation with Polynomial Regression Component," *Communications in Statistics*, 17(10), 3427–3446.

Billingsley, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

Brown, R. L., J. Durbin, and J. M. Evans (1975): "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society*, Series B, 37, 149–192.

Campbell, J. Y. and N. G. Mankiw (1989): "International Evidence on the Persistence of Economic Fluctuations," *Journal of Monetary Economics*, 23, 319–333.

Chu, C.–S. J. (1989): "New Tests for Parameter Constancy in Stationary and Nonstationary Regression Models," Mimeo, University of California at San Diego.

Chu, C.–S. J. and H. White (1990): "Testing for Structural Change in Some Simple Time Series Models," unpublished manuscript, University of California at San Diego.

Deshayes, J. and D. Picard (1986): "Off–line Statistical Analysis of Change Point Models Using Non–parametric and Likelihood Methods," in M. Basseville and A. Beneviste (eds), *Detection of Abrupt Changes in Signals and Dynamical Systems* (Lecture Notes in Control and Information Sciences 77), 103–168, Berlin: Springer.

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

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

Gardner, L. A. (1969): "On Detecting Changes in the Mean of Normal Variates," *Annals of Mathematical Statistics*, 40, 116–126.

Hackl, P. and A. H. Westlund (1989): "Statistical Analysis of "Structural Change": An Annotated Bibliography," *Empirical Economics*, 14, 167–192.

Hansen, B. E. (1990a): "Lagrange Multiplier Tests for Parameter Instability in Non–Linear Models," mimeo, University of Rochester.

Hansen, B. E. (1990b): "Tests for Parameter Instability in Regressions with I(1) Processes," mimeo, University of Rochester.

Herrndorf, N. (1984): A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *Annals of Probability*, 12, 141–153.

Kim, H. J. and D. Siegmund (1989): "The Likelihood Ratio Test for a Change–point in Simple Linear Regression," *Biometrika*, 76, 409–423.

Kramer, W., W. Ploberger, and R. Alt (1988): "Testing for Structural Change in Dynamic Models," *Econometrica*, 56, 1355–1369.

Krishnaiah, P. R. and B. Q. Miao (1988): "Review about Estimation of Change Points," in *Handbook of Statistics*, 7, ed. by P. R. Krishnaiah and C. R. Rao. New York: Elsevier.

Kulperger, R. J. (1987): "Some Remarks on Regression Residuals with Autoregressive Errors and Their Residual Processes," *Journal of Applied Probability*, 24, 668–678.

MacNeill, I. B. (1978): "Properties of Sequences of Partial Sums of Polynomial Regression Residuals with Applications to Tests for Change of Regression at Unknown Times," *Annals of Statistics*, 6, 422–433.

Nelson, C. R. and C. I. Plosser (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139–162.

Ouliaris, S., J. Y. Park, and P.C.B. Phillips (1989): "Testing for a Unit Root in the Presence of a Maintained Trend," in *Advances in Econometrics and Modelling*, ed. by B. Raj. Dordrecht: Kluwer, 7–28.

Page, E. S. (1955): "A Test for a Change in a Parameter Occurring at an Unknown Point," *Biometrika*, 42, 523–526.

Park, J. Y. and P. C. B. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes: Part I," *Econometric Theory*, 4, 468–497.

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

Perron, P. (1990a): "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business and Economic Statistics*, 8, 153–162.

Perron, P. (1990b): "Further Evidence on Breaking Trend Functions in Macroeconomic Variables," Econometric Research Program Memorandum No. 350, Princeton University.

Perron, P. (1991): "The Finite Sample Power Function of Dynamic CUSUM Tests," manuscript in preparation, Princeton University.

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

Phillips, P. C. B. (1991): "Spectral Regression for Cointegrated Time Series," in W. Barnett, J. Powell and G. Tauchen (eds), *Nonparametric and Semiparametric Methods in Economics and Statistics*, Cambridge: Cambridge University Press, 413–435.

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

Shaban, S. A. (1980): "Change Point Problem and Two–Phase Regression: an Annotated Bibliography," *International Statistical Review*, 48, 83–93.

Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113–144.

Zacks, S. (1983): "Survey of Classical and Bayesian Approaches to the Change–point Problem: Fixed Sample and Sequential Procedures of Testing and Estimation," in M. H. Rizvi et al. (eds), *Recent Advances in Statistics*, New York: Academic Press, 245–269.

APPENDIX : PROOF OF THE THEOREMS

In the proof of the Theorems, we can without loss of generality set $\beta_i = 0$ (i = 0, ..., p). Henceforth we consider, under the null hypothesis, $\{y_t\}$ as a zero mean stationary AR(k) process as specified by (3), i.e. $y_t = X_t$. Throughout, W(r) denotes the standard Wiener process defined on C[0,1], the space of all real–valued continuous function on the unit interval. Weak convergence of the associated probability measures is denoted by '$\Rightarrow$', and '$\to$' is used to denote convergence in probability. The strategy of the proofs is as follows. We construct stochastic processes lying on D[0,1], the space of real–valued functions on the interval [0,1] that are right continuous and have finite left limit. These are shown to converge weakly to some random variable lying in C[0,1] using basic functional central limit theorems for partial sums and the continuous mapping theorem (see, e.g., Billingsley (1968), Theorem 5.1). The latter is used frequently and hence references to it are suppressed. Unless otherwise indicated Σ denotes the sum from 1 to T.

Proof of Theorem 1 : The proof follows closely that of Kulperger (1987). It is nevertheless useful to discuss the main steps as they are needed in the proof of other theorems. Under the null hypothesis of no structural change we have :

$$\tilde{e}_{p,t} = X_t - \Sigma_{i=0}^{p} \tilde{\beta}_i t^i , \tag{A.1}$$

where $\tilde{\beta}_i$ (i = 0, ..., p) are defined by the normal equations :

$$D_T \begin{bmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \\ \vdots \\ \tilde{\beta}_p \end{bmatrix} = \begin{bmatrix} \Sigma X_t \\ \Sigma t X_t \\ \vdots \\ \Sigma t^p X_t \end{bmatrix}$$

with D_T a (p + 1) by (p + 1) matrix defined by :

$$D_T = \begin{bmatrix} T & \Sigma t & \cdots & \Sigma t^p \\ \Sigma t & \Sigma t^2 & \cdots & \Sigma t^{p+1} \\ \vdots & & \ddots & \vdots \\ \Sigma t^p & \Sigma t^{p+1} & \cdots & \Sigma t^{2p} \end{bmatrix} . \tag{A.2}$$

Let D be defined by (11). Asymptotically, the normal equations satisfy :

$$D\begin{bmatrix} T^{1/2} & \tilde{\beta}_0 \\ T^{3/2} & \tilde{\beta}_1 \\ & \vdots \\ T^{p+1/2} & \tilde{\beta}_p \end{bmatrix} = \begin{bmatrix} T^{-1/2} & \Sigma X_t \\ T^{-3/2} & \Sigma t X_t \\ & \vdots \\ T^{-p-1/2} & \Sigma t^p X_t \end{bmatrix} .$$

Since X_t ia a stationary AR(k) process it satisfies the conditions for the application of a functional central limit theorem allowing dependence. From Lemma 1 (g) of Sims, Stock and Watson (1990), we have :

$$T^{-(j+1/2)}\Sigma t^j X_t \Rightarrow \sigma(W(1) - j\int_0^1 r^{j-1}W(r)dr) = \sigma\int_0^1 r^j dW(r) \text{ , } (j = 0, \ldots, p) \qquad \text{(A.3)}$$

where $\sigma^2 = \lim_{T\to\infty} T^{-1}(\Sigma X_t)^2 = 2\pi f_x(0)$ where $f_x(0)$ is the spectral density function of X_t evaluated at frequency 0 (in the AR(k) case, $2\pi f_x(0) = \sigma_e^2/A(1)^2$). Since D is non–singular:

$$(T^{1/2}\tilde{\beta}_0, \ldots, T^{p+1/2}\tilde{\beta}_p) \Rightarrow \sigma(Z_p(0), \ldots, Z_p(p)) \text{ ,} \qquad \text{(A.4)}$$

as defined in (12). From (A.1), we have :

$$T^{-1/2}\Sigma_{j=1}^t \tilde{e}_{p,j} = T^{-1/2}\Sigma_{j=1}^t X_j - \Sigma_{i=0}^p T^{i+1/2}\tilde{\beta}_i T^{-1}\Sigma_{j=1}^t (j/T)^i \text{ .} \qquad \text{(A.5)}$$

Define the following processes on D[0,1] for $(j-1)/T \leq r < j/T$: $H_T^1(r) = T^{-1/2}\Sigma_{j=1}^{[Tr]}\tilde{e}_{p,j}$ and note the following convergence results (for a proof of (A.6) see, e.g., Herrndorf (1984)):

$$H_T^2(r) \equiv T^{-1/2}\Sigma_{j=1}^{[Tr]} X_j \Rightarrow \sigma W(r) \text{ ,} \qquad \text{(A.6)}$$

$$T^{-1}\Sigma_{j=1}^{[Tr]} f(j/T) \Rightarrow \int_0^r f(s)ds \text{ ,} \qquad \text{(A.7)}$$

provided $f([Tr]/T) \Rightarrow f(r)$ uniformly. For example, in the case $f(j/T) = (j/T)^i$ we have

$$H_{T,i}^3(r) \equiv T^{-1}\Sigma_{j=1}^{[Tr]}(j/T)^i \Rightarrow r^{i+1}/(i+1) \text{ . } (i = 0, \ldots, p). \qquad \text{(A.8)}$$

We write :

$$\tilde{\sigma}^2 QS_T(p) = T^{-2}\Sigma_{t=1}^{T-1}(\Sigma_{j=1}^{t}\tilde{e}_{p,j})^2 = \int_0^1 H_T^1(r)^2 dr$$

$$= \int_0^1 [H_T^2(r) - \Sigma_{i=0}^{p} T^{i+1/2}\tilde{\beta}_i H_{T,i}^3(r)]^2 dr$$

$$\Rightarrow \sigma^2 \int_0^1 [W(r) - \Sigma_{i=0}^{p} Z_p(i) r^{i+1}/(i+1)]^2 dr \equiv \sigma^2 \int_0^1 B_p(r)^2 dr \,,$$

using (A.4), (A.6) and (A.8). It remains to show that $\tilde{\sigma}^2 \to \sigma_x^2$. We have :

$$\tilde{\sigma}^2 = T^{-1}\Sigma\tilde{e}_{p,t}^2 = T^{-1}\Sigma(X_t - \Sigma_{i=0}^{p}\tilde{\beta}_i t^i)^2$$

$$= T^{-1}\Sigma X_t^2 - 2T^{-1}\Sigma_{i=0}^{p} T^{i+1/2}\tilde{\beta}_i T^{-i-1/2}\Sigma t^i X_t$$

$$+ T^{-1}\Sigma_{i=0}^{p}\Sigma_{j=0}^{p} T^{i+1/2}\tilde{\beta}_i T^{j+1/2}\tilde{\beta}_j T^{-1}\Sigma(t/T)^{i+j}$$

$$= T^{-1}\Sigma X_t^2 + o_p(1) \text{ in view of (A.3), (A.4) and (A.8),}$$

$$\to \sigma_x^2 \,.$$

Proof of Theorem 2 : Since the autoregressive polynomial has a unit root, we can write $X_t = X_{t-1} + w_t$ where w_t is a stationary AR(k–1) process with autoregressive polynomial $(1 - L)^{-1}A(L)$. Denote by $y_{t-j}^{p,j}$ (t = 1, ..., T; j = 0, ..., k) the residuals from a projection of y_{t-j} on $\{1, t, ..., t^p\}$. Under the null hypothesis, $y_t^{p,0} = X_t^{p,0} = \tilde{e}_{p,t}$. Define the stochastic process $H_T^4(r) = T^{-1/2}\tilde{e}_{p,[Tr]} = T^{-1/2}\tilde{e}_{p,j-1}$ with $(j-1)/T \le r < j/T$. Using arguments in Ouliaris, Park and Phillips (1989) we have :

$$H_T^4(r) \Rightarrow \sigma_w W_p^*(r) \,, \qquad \text{(A.9)}$$

where $W_p^*(r)$ is the projection residual of a Wiener process W(r) on the subspace generated by the polynomial function $\{1, r, ..., r^p\}$ in $L^2[0,1]$, the Hilbert space of square integrable functions on [0,1]. $\sigma_w^2 = \lim_{T\to\infty} T^{-1}[S_{w,T}^2] = 2\pi f_w(0)$ where $S_{w,T} = \Sigma_{t=1}^T w_t$ and $f_w(0)$ is the spectral density function of w_t evaluated at frequency 0. We write :

$$T^{-4}\Sigma_{t=1}^{T-1}(\Sigma_{j=1}^{t}\tilde{e}_{p,j})^2 = \int_0^1(\int_0^r H_T^4(s)ds)^2dr \Rightarrow \sigma_w^2\int_0^1(\int_0^r W_p^*(s)ds)^2dr\ , \qquad (A.10)$$

using (A.9). Consider now the sum of squared residuals :

$$T^{-1}\tilde{\sigma}^2 = T^{-2}\Sigma\tilde{e}_{p,t}^2 = \int_0^1 H_T^4(r)^2dr \Rightarrow \sigma_w^2\int_0^1 W_p^*(r)^2dr\ . \qquad (A.11)$$

Theorem 2 follows using (A.10) and (A.11).

Proof of Theorem 3 : We consider the data–generating process written as :

$$y_t = \Sigma_{i=0}^p \gamma_i t^i + \Sigma_{j=1}^k \alpha_j y_{t-j} + e_t\ . \qquad (A.12)$$

(A.12) is equivalent to :

$$y_t^{p,0} = \Sigma_{j=1}^k \alpha_j y_{t-j}^{p,j} + e_t^p\ . \qquad (A.13)$$

Denoting by $\hat{\alpha}_j$ the least–squares estimator of α_j from (9), the estimated residuals from that regression can be defined as :

$$\begin{aligned}\hat{e}_{p,t} &= y_t^{p,0} - \Sigma_{j=1}^k \hat{\alpha}_j y_{t-j}^{p,j} \\ &= y_t^{p,0} - \Sigma_{j=1}^k \alpha_j y_{t-j}^{p,j} - \Sigma_{j=1}^k(\hat{\alpha}_j - \alpha_j)y_{t-j}^{p,j} \\ &= e_t^p - \Sigma_{j=1}^k(\hat{\alpha}_j - \alpha_j)y_{t-j}^{p,j}\ , \qquad (A.14)\end{aligned}$$

using (A.13). Now define the stochastic process $H_T^5(r) = T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{e}_{p,t}$ (for $(j-1)/T \le r < j/T$). From (A.14), we have :

$$H_T^5(r) = T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t^p - \Sigma_{j=1}^k T^{1/2}(\hat{\alpha}_j - \alpha_j)T^{-1}\Sigma_{t=1}^{[Tr]}y_{t-j}^{p,j}$$

$$= T^{-1/2}\Sigma_{t=1}^{[Tr]} e_t^p + o_p(1) ,$$

since $T^{1/2}(\hat{\alpha}_j - \alpha_j) = O_p(1)$, and $T^{-1}\Sigma_{t=1}^{[Tr]} y_{t-j}^{p,j} \to 0$ (a.s) (see, e.g., Bhat and Chandra (1988), Lemma 2.2 and Theorem 4.1; see also parts (d) and (e) of Lemma A.1 below). Using developments exactly analogous to those in Theorem 1 (with $e_t \sim$ i.i.d. $(0, \sigma_e^2)$ instead of the AR process X_t), we have :

$$H_T^5(r) \Rightarrow \sigma_e B_p(r). \tag{A.15}$$

The proof follows by noting that :

$$QD_T(p) = \hat{\sigma}^{-2} T^{-2} \Sigma_{t=1}^{T-1} (\Sigma_{j=1}^{t} \hat{e}_{p,j})^2 = \hat{\sigma}^{-2} \int_0^1 H_T^5(r)^2 dr \Rightarrow \int_0^1 B_p(r)^2 dr ,$$

using (A.15) and the fact the $\hat{\sigma}^2 \to \sigma_e^2$ (a.s), the proof of which is omitted.

Proof of Theorem 4 : We write (A.13) as

$$y_t^{p,0} = \alpha y_{t-1}^{p,1} + \Sigma_{j=2}^{k} d_j \Delta^* y_{t-j+1}^{p,j-1} + e_t^p, \tag{A.16}$$

where $\alpha = \Sigma_{j=1}^{k} \alpha_j$, $d_j = - \Sigma_{i=j}^{k} \alpha_i$ (j = 2, ..., k) and Δ^* denotes the difference operator applied to both the subscript and superscript, i.e. $\Delta^* y_{t-j+1}^{p,j-1} = y_{t-j+1}^{p,j-1} - y_{t-j}^{p,j}$. The estimated residuals from (9) are defined in terms of the least–squares estimates $\hat{\alpha} = \Sigma_{j=1}^{k} \hat{\alpha}_j$ and $\hat{d}_j = - \Sigma_{i=j}^{k} \hat{\alpha}_i$, by :

$$\hat{e}_{p,t} = y_t^{p,0} - \hat{\alpha} y_{t-1}^{p,1} - \Sigma_{j=2}^{k} \hat{d}_j \Delta^* y_{t-j+1}^{p,j-1}$$

$$= e_t^p - (\hat{\alpha} - 1) y_{t-1}^{p,1} - \Sigma_{j=2}^{k} (\hat{d}_j - d_j) \Delta^* y_{t-j+1}^{p,j-1} ,$$

using (A.16) and $\alpha = 1$ with a unit root. The stochastic process $H_T^5(r)$ is expressed as:

$$H_T^5(r) = T^{-1/2}\Sigma_{t=1}^{[Tr]} \hat{e}_{p,t} = T^{-1/2}\, \Sigma_{t=1}^{[Tr]} e_t^p - T(\hat{\alpha} - 1) T^{-3/2} \Sigma_{t=1}^{[Tr]} y_{t-1}^{p,1}$$

$$-\Sigma_{j=2}^{k}T^{1/2}(\hat{d}_j - d_j)T^{-1}\Sigma_{t=1}^{[Tr]}\Delta^* y_{t-j+1}^{p,j-1} \qquad (A.17)$$

$$= T^{-1/2}\,\Sigma_{t=1}^{[Tr]}e_t^p - T(\hat{\alpha}-1)T^{-3/2}\Sigma_{t=1}^{[Tr]}y_{t-1}^{p,1} + o_p(1),$$

since $T^{1/2}(\hat{d}_j - d_j) = O_p(1)$ (see Fuller (1976), for the cases p =0, 1) and it can be shown that $T^{-1}\Sigma_{t=1}^{[Tr]}\Delta^* y_{t-j+1}^{p,j-1} \to 0$ (j = 2, ..., k). Using results from Ouliaris, Park and Phillips (1989) and Dickey and Fuller (1979) we have :

$$T(\hat{\alpha}-1) \Rightarrow (1 - d_2 - \ldots - d_k)\int_0^1 W_p^*(s)dW(s)/\int_0^1 W_p^*(s)^2 ds \;. \qquad (A.18)$$

Using (A.9) and $y_{t-1}^{p,1} = y_t^{p,0} + o_p(1)$, $T^{-3/2}\Sigma_{t=1}^{[Tr]}y_{t-1}^{p,1} = \int_0^r [H_T^4(s) + o_p(1)]ds \Rightarrow \sigma_w \int_0^r W_p^*(s)ds$. Note that $\sigma_w^2 = \sigma_e^2(1 - d_2 - \ldots - d_k)^{-2}$ in the present notation. Using this last result, (A.15) and (A.18), we have :

$$H_T^5(r) \Rightarrow \sigma_e B_p(r) - \sigma_e\left(\int_0^1 W_p^*(s)dW(s)/\int_0^1 W_p^*(s)^2 ds\right)\int_0^r W_p^*(s)ds \,. \qquad (A.19)$$

The proof of the Theorem follows noting that $QS_T(p) = \hat{\sigma}^{-2}T^{-2}\Sigma_{t=1}^{T-1}(\Sigma_{j=1}^{t}\hat{e}_{p,j})^2 = \hat{\sigma}^{-2}\int_0^1 H_T^5(r)^2 dr$ and using (A.19) and the fact that $\hat{\sigma}^2 \to \sigma_e^2$ (a.s.), the proof of which follows arguments similar to those used by Dickey and Fuller (1979) in the case p = 1.

Remark A.1: The proof of Theorem 3 is similar to that of Kulperger (1987) who considers the case where the series is detrended in a prior step instead of using the regression equation (9). Our result shows the limiting distribution to be invariant to the detrending procedure in the stationary case. However, such is not the case when the autoregressive polynomial contains a unit root. Here the limiting distribution of $QD_T(p)$ is different whether one uses regression (9) or detrend the data prior to estimating the autoregression.

Proof of Theorem 5 : Follows easily from the proof of Theorem 6.

Proof of Theorem 6 (i): Again, without loss of generality, we set $\beta_i = 0$ (i = 0, ..., p) and note, using (1) – (3) and (16), that $\{y_t\}$ can be expressed as :

$$y_t = \Sigma_{j=1}^k \alpha_j y_{t-j} + e_t^*$$

where

$$e_t^* = e_t + \Sigma_{i=0}^p A(L)1(t > \lambda T)(\delta_i/T^{1/2+i})(t - \lambda T)^i , \tag{A.20}$$

or in detrended form as :

$$y_t^{p,0} = \Sigma_{j=1}^k \alpha_j y_{t-j}^{p,j} + e_t^{*p} , \tag{A.21}$$

where e_t^{*p} (t = 1, ..., T) are the residuals from a projection of e_t^* on $\{1, t, ..., t^p\}$. From (A.14) the estimated OLS residuals from (9) are defined as:

$$\hat{e}_{p,t} = y_t^{p,0} - \Sigma_{j=1}^k \alpha_j y_{t-j}^{p,j} - \Sigma_{j=1}^k (\hat{\alpha}_j - \alpha_j) y_{t-j}^{p,j} = e_t^{*p} - \Sigma_{j=1}^k (\hat{\alpha}_j - \alpha_j) y_{t-j}^{p,j}, \tag{A.22}$$

using (A.21) where $\hat{\alpha}_j$ (j = 1, ..., k) are the OLS estimates of α_j from (9). We organize the proof of the Theorem starting with a Lemma concerning the limit of various elements.

LEMMA A.1 : *Let $\{y_t\}$ be generated by (1) – (3) and (16) with $A(1) > 0$ and, from (9), denote by $\hat{\alpha}_j$ (j = 1, ..., k) the OLS estimates of α_j, $\hat{e}_{p,t}$ the estimated residuals and $\hat{\sigma}^2$ the estimate of the variance of the residuals. Let $Z_p^*(m)$, $V_p^*(m)$ (m = 0, ..., p) and $B_{p,\delta}^*(r)$ be defined by (18), (19) and (20) respectively.*

a) Let e_t^* be defined by (A.20) : $T^{-m-1/2}\Sigma_{t=1}^T t^m e_t^* \Rightarrow \sigma_e V_p^*(m)$ (m = 0, ..., p).

b) Let $\hat{\tau}_i$ (i = 0, ..., p) be the estimated coefficients from an OLS regression of e_t^* on $\{1, t, ..., t^p\}$, then : $(T^{1/2}\hat{\tau}_0, T^{3/2}\hat{\tau}_1, ..., T^{p+1/2}\hat{\tau}_p) \Rightarrow \sigma_e(Z_p^*(0), Z_p^*(1), ..., Z_p^*(p))$.

c) Let e_t^{*p} be defined by (A.21), then : $T^{-1/2}\Sigma_{t=1}^{[Tr]} e_t^{*p} \Rightarrow \sigma_e B_{p,\delta}^*(r)$.

d) Let $y_{t-j}^{p,j}$ (j = 0, ..., k; t = 1, ..., T) be the residuals from a projection of y_{t-j} on $\{1, t,$

..., $t^p\}$, *then* $T^{-1/2}\Sigma_{t=1}^{[Tr]} y_{t-j}^{p,j} \Rightarrow (\sigma_e/A(1))B^*_{p,\delta}(r)$.

e) $\hat{\alpha}_j \rightarrow \alpha_j$ *(in probability) for* $j = 1, ..., k$.

f) $T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{e}_{p,t} \Rightarrow \sigma_e B^*_{p,\delta}(r)$.

g) $\hat{\sigma}^2 \rightarrow \sigma_e^2$ *(in probability).*

The proof of the Theorem follows easily using parts (f) and (g). We have :

$$QS_T(p) = \hat{\sigma}^{-2}T^{-2}\Sigma_{t=1}^{T-1}(\Sigma_{j=1}^{t}\hat{e}_{p,j})^2 = \hat{\sigma}^{-2}\int_0^1(T^{-1/2}\Sigma_{j=1}^{[Tr]}\hat{e}_{p,j})^2 dr \Rightarrow \int_0^1 B^*_{p,\delta}(r)^2 dr .$$

as required. The proof of the various parts of the Lemma follows.

Proof of part (a). Using (A.20), we have :

$$T^{-m-1/2}\Sigma_{t=1}^{T}t^m e_t^* = T^{-m-1/2}\Sigma_{t=1}^{T}t^m e_t + T^{-m-i-1}\Sigma_{t=1}^{T}t^m\Sigma_{i=0}^{p}A(L)1(t > \lambda T)\delta_i(t-\lambda T)^i.$$

As a special case of (A.3), $T^{-m-1/2}\Sigma_{t=1}^{T}t^m e_t \Rightarrow \sigma_e(W(1) - m\int_0^1 r^{m-1}W(r)dr)$. The second term can be expressed as :

$$T^{-m-i-1}\Sigma_{t=1}^{T}t^m\Sigma_{i=0}^{p}A(L)1(t > \lambda T)\delta_i(t-\lambda T)^i$$
$$= A(1)\Sigma_{i=0}^{p}\delta_i T^{-1}\Sigma_{t=\lambda T+1}^{T}(t/T-\lambda)^i(t/T)^m + o(1)$$
$$\Rightarrow A(1)\Sigma_{i=0}^{p}\delta_i\int_0^{1-\lambda}(r+\lambda)^m r^i\, dr ,$$

using (A.7). The results from part (a) then follows.

Proof of part (b). The estimates $\hat{\tau}_i$ (i = 0, ..., p) are defined by the normal equations :

$$D_T \begin{bmatrix} \hat{\tau}_0 \\ \hat{\tau}_1 \\ \vdots \\ \hat{\tau}_p \end{bmatrix} = \begin{bmatrix} \Sigma e_t^* \\ \Sigma t\, e_t^* \\ \vdots \\ \Sigma t^p e_t^* \end{bmatrix},$$

where D_T is as defined in (A.2). With D defined as in (11), the normal equations asymptotically satisfy the relations :

$$\begin{bmatrix} T^{1/2} & \hat{\tau}_0 \\ T^{3/2} & \hat{\tau}_1 \\ & \vdots \\ T^{p+1/2} & \hat{\tau}_p \end{bmatrix} = D^{-1} \begin{bmatrix} T^{-1/2} & \Sigma e_t^* \\ T^{-3/2} & \Sigma t\, e_t^* \\ & \vdots \\ T^{-p-1/2} & \Sigma t^p e_t^* \end{bmatrix}. \tag{A.23}$$

The proof of part (b) follows using (A.23) and part (a) of the Lemma.

Proof of part (c) : The residuals e_t^{*p} are defined by the OLS regression :

$$e_t^* = \Sigma_{i=0}^p \hat{\tau}_i t^i + e_t^{*p}. \tag{A.24}$$

Then :

$$T^{-1/2}\Sigma_{t=1}^{[Tr]} e_t^{*p} = T^{-1/2}\Sigma_{t=1}^{[Tr]} e_t^* - \Sigma_{i=0}^p T^{1/2+i}\hat{\tau}_i T^{-1}\Sigma_{t=1}^{[Tr]}(t/T)^i$$

$$= T^{-1/2}\Sigma_{t=1}^{[Tr]} e_t + \Sigma_{i=0}^p A(1)\delta_i T^{-1}\Sigma_{t=\lambda T+1}^{[Tr]}(t/T-\lambda)^i$$

$$- \Sigma_{i=0}^p T^{1/2+i}\hat{\tau}_i T^{-1}\Sigma_{t=1}^{[Tr]}(t/T)^i + o_p(1) \tag{A.25}$$

$$\Rightarrow \sigma_e W(r) + 1(r>\lambda)\Sigma_{i=0}^p A(1)\delta_i (r-\lambda)^{i+1}/(i+1) - \sigma_e \Sigma_{i=0}^p Z_p^*(i) r^{i+1}/(i+1)$$

$$\equiv \sigma_e B_{p,\delta}^*(r),$$

using (A.6), (A.7), (A.20) and part (b) of the Lemma.

Proof of part (d). Note first that from (1) and (16) and using the simplifying assumption that $\beta_i = 0$ (i = 0, ..., p), y_t can be expressed as :

$$y_t = X_t + \Sigma_{i=0}^{p} 1(t > \lambda T)(\delta_i / T^{1/2+i})(t - \lambda T)^i . \tag{A.26}$$

Let $y_t^{p,0}$ be defined by the OLS regression :

$$y_t = \Sigma_{i=0}^{p} \hat{\nu}_i t^i + y_t^{p,0}. \tag{A.27}$$

The estimates $\hat{\nu}_i$ are given asymptotically by :

$$\begin{bmatrix} T^{1/2} & \hat{\nu}_0 \\ T^{3/2} & \hat{\nu}_1 \\ & \vdots \\ T^{p+1/2} & \hat{\nu}_p \end{bmatrix} = D^{-1} \begin{bmatrix} T^{-1/2} & \Sigma y_t \\ T^{-3/2} & \Sigma t y_t \\ & \vdots \\ T^{-p-1/2} & \Sigma t^p y_t \end{bmatrix} .$$

Now, using (A.26),

$$T^{-m-1/2} \Sigma t^m y_t = T^{-m-1/2} \Sigma t^m X_t + \Sigma_{i=0}^{p} \delta_i T^{-1} \Sigma_{t=\lambda T+1}^{T} (t/T)^m (t/T - \lambda)^i$$

$$\Rightarrow \sigma(W(1) - m\int_0^1 r^{m-1} W(r) dr) + \Sigma_{i=0}^{p} \delta_i \int_0^{1-\lambda} (r + \lambda)^m r^i \, dr \equiv \sigma V_p^*(m)$$

for m = 0, ..., p, using (A.3) and (A.7), where $\sigma^2 = 2\pi f_x(0) = \sigma_e^2 / A(1)^2$. Hence,

$$(T^{1/2} \hat{\nu}_0, \ldots, T^{1/2+p} \hat{\nu}_p) \Rightarrow \sigma(Z_p^*(0), \ldots, Z_p^*(p)) , \tag{A.28}$$

where $Z_p^*(i)$ (i = 0, ..., p) is defined in (18). Using (A.26) and (A.27) :

$$y_t^{p,0} = X_t + \Sigma_{i=0}^{p} 1(t > \lambda T)(\delta_i / T^{1/2+i})(t - \lambda T)^i - \Sigma_{i=0}^{p} \hat{\nu}_i t^i , \tag{A.29}$$

and

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}y_t^{p,0} = T^{-1/2}\Sigma_{t=1}^{[Tr]}X_t$$

$$+ \Sigma_{i=0}^{p}\delta_i T^{-1}\Sigma_{t=\lambda T+1}^{[Tr]}(t/T - \lambda)^i - \Sigma_{i=0}^{p}T^{1/2+i}\hat{\nu}_i T^{-1}\Sigma_{t=1}^{[Tr]}(t/T)^i$$

$$\Rightarrow \sigma W(r) + 1(r > \lambda)\Sigma_{i=0}^{p}\delta_i(r-\lambda)^{i+1}/(i+1) - \sigma\Sigma_{i=0}^{p}Z_p^*(i)r^{i+1}/(i+1).$$

$$\equiv (\sigma_e/A(1))B_{p,\delta}^*(r),$$

using (A.3), (A.7) and (A.28), with $Z_p^*(i)$ and $B_{p,\delta}^*(r)$ defined in (18) and (20). The proof is exactly analogous for $T^{-1/2}\Sigma_{t=1}^{[Tr]}y_{t-j}^{p,j}$ $(j = 1, ..., k)$.

Proof of part (e). Let $\hat{A} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_k)$, $A = (\alpha_1, \alpha_2, ..., \alpha_k)$, $V_T = T^{-1}\Sigma_{t=1}^{T}W_tW_t'$ and $E_T = T^{-1}\Sigma W_t'e_t^{*p}$ with $W_t' = (y_{t-1}^{p,1}, ..., y_{t-k}^{p,k})$. Using (A.21), we have $\hat{A} - A = V_T^{-1}E_T$. The proof can be achieved by showing that $T^{-1}\Sigma_{t=1}^{T}y_{t-i}^{p,i}y_{t-j}^{p,j}$ $(i, j = 1, ..., k)$, a typical element of V_T, is bounded and that $T^{-1}\Sigma_{t=1}^{T}y_{t-i}^{p,i}e_t^{*p}$ $(i = 1, ..., k)$, a typical element of E_T, converges to 0. Denote by $\hat{\nu}_i^m$ the i–th OLS estimate in a regression of y_{t-m} on $\{1, t, ..., t^p\}$. Using (A.26) we have :

$$T^{-1}\Sigma_{t=1}^{T}y_{t-i}^{p,i}y_{t-j}^{p,j}$$

$$= T^{-1}\Sigma_{t=1}^{T}(X_{t-i} + \Sigma_{n=0}^{p}1(t-i > \lambda T)(\delta_n/T^{1/2+n})(t - \lambda T - i)^n - \Sigma_{n=0}^{p}\hat{\nu}_n^i t^n)$$

$$\cdot (X_{t-j} + \Sigma_{m=0}^{p}1(t-j > \lambda T)(\delta_m/T^{1/2+m})(t - \lambda T - j)^m - \Sigma_{m=0}^{p}\hat{\nu}_m^j t^m)$$

$$= T^{-1}\Sigma_{t=1}^{T}X_{t-i}X_{t-j} + T^{-1}\Sigma_{m=0}^{p}\delta_m T^{-1/2-m}\Sigma_{t=\lambda T+1}^{T-j}(t-\lambda T)^m X_{t-i+j}$$

$$+ T^{-1}\Sigma_{n=0}^{p}\delta_n T^{-1/2-n}\Sigma_{t=\lambda T+1}^{T-i}(t-\lambda T)^n X_{t-j+i}$$

$$- T^{-1}\Sigma_{m=0}^{p}T^{1/2+m}\hat{\nu}_m^j T^{-m-1/2}\Sigma_{t=1}^{T}t^m X_{t-i}$$

$$- T^{-1}\Sigma_{n=0}^{p}T^{1/2+n}\hat{\nu}_n^i T^{-n-1/2}\Sigma_{t=1}^{T}t^n X_{t-j}$$

$$+ T^{-1}\Sigma_{n=0}^{p}\Sigma_{m=0}^{p}T^{n+1/2}\hat{\nu}_n^i T^{m+1/2}\hat{\nu}_m^j T^{-1}\Sigma_{t=1}^{T}(t/T)^{n+m}$$

$$+ T^{-1}\Sigma_{n=0}^{p}\Sigma_{m=0}^{p}\delta_m\delta_n T^{-1}\Sigma_{t=\lambda T+1}^{T}(t/T-\lambda)^{m+n}$$

$$- T^{-1}\Sigma_{n=0}^{p}\Sigma_{m=0}^{p}\delta_n T^{1/2+m}\hat{\nu}_m^j T^{-1}\Sigma_{t=\lambda T+1}^{T}(t/T)^{m}(t/T-\lambda)^{n}$$

$$- T^{-1}\Sigma_{n=0}^{p}\Sigma_{m=0}^{p}\delta_m T^{1/2+n}\hat{\nu}_n^j T^{-1}\Sigma_{t=\lambda T+1}^{T}(t/T)^{n}(t/T-\lambda)^{m} + o_p(1)$$

$$= T^{-1}\Sigma_{t=1}^{T}X_{t-i}X_{t-j} + o_p(1).$$

using (A.3), (A.7) and (A.28). Hence,

$$T^{-1}\Sigma_{t=1}^{T}y_{t-i}^{p,i}y_{t-j}^{p,j} = O_p(1) \tag{A.30}$$

since $T^{-1}\Sigma_{t=1}^{T}X_{t-i}X_{t-j} = O_p(1)$ (see Fuller (1976)). Consider now a typical element of E_T. Using (A.29) and (A.24), we have :

$$T^{-1}\Sigma_{t=1}^{T}y_{t-i}^{p,i}e_t^{*p}$$

$$= T^{-1}\Sigma_{t=1}^{T}(X_{t-i} + \Sigma_{n=0}^{p}1(t-i>\lambda T)(\delta_n/T^{1/2+n})(t-\lambda T-i)^n - \Sigma_{n=0}^{p}\hat{\nu}_n^i t^n)e_t^{*p}$$

$$= T^{-1}\Sigma_{t=1}^{T}X_{t-i}e_t^{*} - T^{-1}\Sigma_{m=0}^{p}T^{1/2+m}\hat{\tau}_m T^{-1/2-m}\Sigma_1^T t^m X_{t-i}$$

$$+ T^{-1}\Sigma_{n=0}^{p}\delta_n T^{-1/2-n}\Sigma_{t=\lambda T+1}^{T}(t-\lambda T)^n e_t^{*}$$

$$- T^{-1}\Sigma_{n=0}^{p}\Sigma_{m=0}^{p}\delta_n T^{1/2+m}\hat{\tau}_m T^{-1}\Sigma_{t=\lambda T+1}^{T}(t/T)^{m}(t/T-\lambda)^{n}$$

$$- T^{-1}\Sigma_{n=0}^{p}T^{1/2+n}\hat{\nu}_n^i T^{-1/2-n}\Sigma_{t=1}^{T}t^n e_t^{*}$$

$$+ T^{-1}\Sigma_{n=0}^{p}\Sigma_{m=0}^{p}T^{1/2+m}\hat{\tau}_m T^{1/2+n}\hat{\nu}_n^i T^{-1}\Sigma_{t=1}^{T}(t/T)^{n+m}$$

$$\to 0. \tag{A.31}$$

using (A.3), (A.7), (A.20) (A.28), part (b) of the Lemma and the fact that $T^{-1}\Sigma_{t=1}^{T}X_{t-1}e_t \to 0$ (see Fuller (1976)).

Proof of part (f) : From (A.22) we have :

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{e}_{p,t} = T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t^{*p} - \Sigma_{j=1}^{k}(\hat{\alpha}_j - \alpha_j)T^{-1/2}\Sigma_{t=1}^{[Tr]}y_{t-j}^{p,j} \tag{A.32}$$

$$= T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t^{*p} + o_p(1)\text{ , using parts (d) and (e),}$$

$$\Rightarrow \sigma_e B^*_{p,\delta}(r)\text{ as required, using part (c) of the Lemma.}$$

Proof of part (g) : From (A.22) we have :

$$\hat{\sigma}^2 = T^{-1}\Sigma_{t=1}^{T}\hat{e}_{p,t}^2 = T^{-1}\Sigma_{t=1}^{T}(e_t^{*p} - \Sigma_{j=1}^{k}(\hat{\alpha}_j - \alpha)y_{t-j}^{p,j})^2$$

$$= T^{-1}\Sigma_{t=1}^{T}(e_t^{*p})^2 - 2\Sigma_{j=1}^{k}(\hat{\alpha}_j - \alpha_j)T^{-1}\Sigma_{t=1}^{T}y_{t-j}^{p,j}e_t^{*p}$$

$$+ \Sigma_{j=1}^{k}\Sigma_{i=1}^{k}(\hat{\alpha}_j - \alpha_j)(\hat{\alpha}_i - \alpha_i)T^{-1}\Sigma_{t=1}^{T}y_{t-j}^{p,j}y_{t-i}^{p,i}$$

$$= T^{-1}\Sigma_{t=1}^{T}(e_t^{*p})^2 + o_p(1)\text{, using part (e), (A.30) and (A.31),}$$

$$= T^{-1}\Sigma_{t=1}^{T}(e_t + \Sigma_{i=0}^{p}A(L)1(t > \lambda T)(\delta_i/T^{1/2+i})(t - \lambda T)^i - \Sigma_{i=0}^{p}\hat{\tau}_i t^i)^2 + o_p(1),$$

$$= T^{-1}\Sigma_{t=1}^{T}e_t^2 + 2T^{-1}A(1)\Sigma_{i=0}^{p}\delta_i T^{-1/2-i}\Sigma_{t=\lambda T+1}^{T}(t - \lambda T)^i e_t$$
$$- 2T^{-1}\Sigma_{i=0}^{p}T^{1/2+i}\hat{\tau}_i T^{-1/2-i}\Sigma_{t=1}^{T}t^i e_t$$
$$- 2T^{-1}A(1)\Sigma_{i=0}^{p}\Sigma_{j=0}^{p}\delta_i T^{1/2+j}\hat{\tau}_j T^{-1}\Sigma_{t=\lambda T+1}^{T}(t/T)^j(t/T - \lambda)^i$$
$$+ T^{-1}A(1)^2\Sigma_{i=0}^{p}\Sigma_{j=0}^{p}\delta_i\delta_j T^{-1}\Sigma_{t=\lambda T+1}^{T}(t/T - \lambda)^{i+j}$$
$$+ T^{-1}\Sigma_{i=0}^{p}\Sigma_{j=0}^{p}T^{1/2+i}\hat{\tau}_i T^{1/2+j}\hat{\tau}_j T^{-1}\Sigma_{t=1}^{T}(t/T)^{i+j} + o_p(1)$$

$$= T^{-1}\Sigma_{t=1}^{T} e_t^2 + o_p(1)$$

using a special case of (A.3), (A.7) and part (b) of the Lemma. Hence, $\hat{\sigma}^2 \to \sigma_e^2$.

Remark A.2: It is easy to prove consistency (Theorem 5 (i)) by noting that if $N_t = \Sigma_{i=0}^{p}[\beta_i t^i + 1(t > \lambda T)(t - \lambda T)^i \delta_i]$, (A.25) becomes :

$$T^{-1/2}\Sigma_{t=1}^{[Tr]} e_t^{*p} =$$

$$T^{-1/2}\Sigma_{t=1}^{[Tr]} e_t + \Sigma_{i=0}^{p} A(1)\delta_i T^{-1/2}\Sigma_{t=\lambda T+1}^{[Tr]}(t - \lambda T)^i$$

$$- \Sigma_{i=0}^{p} T^{1/2+i}\hat{\tau}_i T^{-1}\Sigma_{t=1}^{[Tr]}(t/T)^i + o_p(1). \qquad (A.33)$$

Suppose that q is the largest index associated with a non–zero value of δ_i, i.e. $\delta_q \neq 0$ and $\delta_j = 0$ for $j > q$. Note then that the second term in (A.33) diverges at rate $T^{q+1/2}$. Hence, form (A.32), $T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{e}_{p,t}$ diverges at rate $T^{q+1/2}$. This establishes the rate of divergence of the numerator of $QD_T(p)$ as T^{2q+1} since $QD_T(p) = \hat{\sigma}^{-2}\int_0^1(T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{e}_{p,t})^2 dr$. Consider now, the behavior of $\hat{\sigma}^2$. We have, $T^{-1}\Sigma_{t=1}^{T}\hat{e}_{p,t}^2 = T^{-1}\Sigma_{t=1}^{T}(e_t^{*p} - \Sigma_{j=1}^{k}(\hat{\alpha}_j - \alpha_j)y_{t-j}^{p,j})^2$. Let us concentrate on the behavior of $T^{-1}\Sigma_{t=1}^{T}(e_t^{*p})^2$. We have, using (A.33) :

$$T^{-1}\Sigma_{t=1}^{T}(e_t^{*p})^2 = T^{-1}\Sigma_{t=1}^{T}(e_t + \Sigma_{i=0}^{p}A(1)\delta_i 1(t > \lambda T)(t - \lambda T)^i - \Sigma_{i=0}^{p}\hat{\tau}_i t^i)^2 + o_p(1).$$

It is easy to see that the second term dominates the others as T increases. Hence, $T^{-1}\Sigma_{t=1}^{T}(e_t^{*p})^2$ diverges at the same rate as $T^{-1}\Sigma_{t=\lambda T+1}^{T}(\Sigma_{i=0}^{p}A(1)\delta_i(t - \lambda T)^i)^2$ which diverges at rate T^{2q}. Therefore $\hat{\sigma}^2$ diverges at rate T^{2q}. Finally, combining the above results, $QD_T(p)$ diverges at rate T showing consistency.

Proof of Theorem 6 (ii) : From the proof of parts (a), (b) and (c) of Lemma A.1, we obtain when $A(1) = 0$:

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t^{*p} \Rightarrow \sigma_e B_p(r),$$

where e_t^{*p} is defined in (A.24) and $B_p(r)$ in (13). From the appropriate modification to (A.17) :

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{e}_{p,t} = T^{-1/2}\,\Sigma_{t=1}^{[Tr]}e_t^{*p} - T(\hat{\alpha}-1)T^{-3/2}\Sigma_{t=1}^{[Tr]}y_{t-1}^{p,1}$$

$$- \Sigma_{j=2}^{k}T^{1/2}(\hat{d}_j - d_j)T^{-1}\Sigma_{t=1}^{[Tr]}\Delta^* y_{t-j+1}^{p,j-1}\,.$$

The proof is completed by showing that the limit of $T(\hat{\alpha}-1)$, $T^{-3/2}\Sigma_{t=1}^{[Tr]}y_{t-1}^{p,1}$, $\hat{d}_j$ (j= 2, ..., k) and $T^{-1}\Sigma_{t=1}^{[Tr]}\Delta^* y_{t-j+1}^{p,j-1}$ are unaffected by specifying $\beta_{i,t}$ by (16) instead of $\beta_{i,t} = \beta_i$ for all t. The proof of these assertions is tedious but straightforward and basically follows from the fact that $T^{-m-3/2}\Sigma_{t=1}^{[Tr]}t^m y_t$ has a limit that is independent of δ_i (i = 0, ..., p). To see this note that, from (1) and (16) (assuming again for simplicity that $\beta_i = 0$):

$$T^{-m-3/2}\Sigma_{t=1}^{[Tr]}t^m y_t = T^{-m-3/2}\Sigma_{t=1}^{[Tr]}t^m[\Sigma_{i=0}^{p}1(t > \lambda T)(\delta_i/T^{1/2+i})(t-\lambda T)^i + X_t]$$

$$= \Sigma_{i=0}^{p}\delta_i T^{-2}\Sigma_{\lambda T+1}^{[Tr]}(t/T)^m(t/T-\lambda)^i + T^{-3/2+m}\Sigma_{t=1}^{[Tr]}t^m X_t$$

$$\Rightarrow \sigma\int_0^1 r^m W(r)dr$$

using, e.g., Sims, Stock and Watson (1990, Lemma 1(a)) and the fact that the first term converges to 0 using (A.7). Here $\sigma^2 = 2\pi f_{\Delta x}(0)$.

Remark A.3: To prove consistency consider the behavior of the quantity $T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t^*$, where $e_t^* = e_t + \Sigma_{i=0}^{p}A(L)1(t > \lambda T)\delta_i(t-\lambda T)^i$. Again, denote by q the integer such that $\delta_q \neq 0$ and $\delta_j = 0$ for j > q. We have :

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t^* = T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t + T^{-1/2}\Sigma_{t=1}^{[Tr]}\Sigma_{i=0}^{q}A(L)1(t > \lambda T)\delta_i(t-\lambda T)^i$$

$$= T^{-1/2}\Sigma_{t=1}^{T}e_t + T^{-1/2}\Sigma_{i=0}^{q}\Sigma_{t=\lambda T+1}^{[Tr]}(1-L)A^*(1)\delta_i(t-\lambda T)^i + o(1)$$

where A*(L) is defined by the relation $A(L) = (1 - L)A^*(L)$ and where transitional effects from $t = \lambda T$ to $t = \lambda T + k$ are subsumed under the term o(1). Using the fact that $(1 - L)t^i = it^{i-1} + o(t^{i-1})$, we have :

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t^* = T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t + A^*(1)\Sigma_{i=1}^{q} i\delta_i T^{-1/2}\Sigma_{t=1}^{[Tr]-\lambda T} t^{i-1} + o(1).$$

If $q = 0$, we have $T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t^* \Rightarrow \sigma W(r)$ since the second term vanishes. Hence, the test is not consistent if the change only affects the intercept (a complete proof would show that the limit of all other terms remain unaffected in that case). On the other hand, if $q > 0$, $T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t^*$ diverges at rate $T^{q-1/2}$. Hence, $T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{e}_{p,t}$ also diverges at rate $T^{q-1/2}$ and the numerator of $QD_T(p)$ diverges at rate T^{2q-1} (see the proof of Lemma A.1(c) in particular). In a manner similar to the development in Remark A.2, it can be shown that $\hat{\sigma}^2$ diverges at rate T^{2q-2} and the statistic $QD_T(p)$ diverges at rate T and is therefore consistent provided $q > 0$.

Proof of Theorem 7 : Part (i) follows from Theorem 3 since dy_t is a stationary AR process of finite order with i.i.d. errors as specified in (30). Since only a constant is estimated, the case with $p = 0$ applies. To prove part (ii), we note following the proof of Theorem 3 (see (A.14)) that :

$$\hat{u}_t = u_t - T^{-1}\Sigma_{t=1}^{T}u_t - \Sigma_{j=1}^{k}(\hat{\alpha}_j - \alpha_j)(dy_{t-j} - T^{-1}\Sigma_{t=1}^{T}dy_{t-j}).$$

Noting that $u_t = e_t - e_{t-1}$, we have :

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{u}_t = T^{-1/2}(e_{[Tr]} - e_0) - rT^{-1/2}(e_T - e_0)$$

$$- \Sigma_{j=1}^{k}(\hat{\alpha}_j - \alpha_j)[T^{-1/2}(y_{[Tr]-j} - y_{-j}) - rT^{-1/2}(y_{T-j} - y_{-j})].$$

Hence, $T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{u}_t \Rightarrow 0$ provided $(\hat{\alpha}_j - \alpha_j)$ is bounded as T increases. As in the proof of Lemma A.1 (e), let $\hat{A} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_k)$, $A = (\alpha_1, \alpha_2, ..., \alpha_k)$, $V_T = T^{-1}\Sigma_{t=1}^{T}W_tW_t'$ and $E_T = T^{-1}\Sigma W_t'u_t$ with $W_t' = (dy_{t-1}^{p,1}, ..., dy_{t-k}^{p,k})$, where $dy_{t-j}^{p,j} = dy_{t-j} - T^{-1}\Sigma_{t=1}^{T}dy_{t-j}$. We have $\hat{A} - A = V_T^{-1}E_T$. The proof can be achieved by showing that $T^{-1}\Sigma_{t=1}^T dy_{t-i}^{p,i} dy_{t-j}^{p,j}$ (i, j = 1, ..., k), a typical element of V_T, is bounded and that $T^{-1}\Sigma_{t=1}^T dy_{t-i}^{p,i} u_t$ (i = 1, ..., k), a typical element of E_T, converges to 0. We have $T^{-1}\Sigma_{t=1}^T dy_{t-i}^{p,i} u_t = T^{-1}\Sigma_{t=1}^T y_{t-i}(e_t - e_{t-1}) - T^{-1}\Sigma_{t=1}^T y_{t-i-1}(e_t - e_{t-1}) - T^{-1}(y_{T-i} - y_{-i})T^{-1}(e_T - e_0) \to 0$ as $T \to \infty$. Consider now, a typical element of V_T, it is straightforward to show that $T^{-1}\Sigma_{t=1}^T dy_{t-i}^{p,i} dy_{t-j}^{p,j} \to 2cov_y(j-i) - cov_y(j-i-1) - cov_y(j-i+1)$, where $cov_y(k)$ denotes the covariance function of y_t at lag k. Hence, $(\hat{\alpha}_j - \alpha_j) \to 0$ as $T \to 0$ and $T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{u}_t \Rightarrow 0$. What remains to be shown is that $\hat{\sigma}^2$ has a non–degenerate limit as T increases. We have :

$$\hat{\sigma}^2 = T^{-1}\Sigma_{t=1}^T \hat{u}_t^2 = T^{-1}\Sigma_{t=1}^T [u_t - T^{-1}\Sigma_{t=1}^T u_t - \Sigma_{j=1}^k(\hat{\alpha}_j - \alpha_j)(dy_{t-j} - T^{-1}\Sigma_{t=1}^T dy_{t-j})]^2$$

$$= T^{-1}\Sigma_{t=1}^T(u_t - T^{-1}\Sigma_{t=1}^T u_t)^2 - 2\Sigma_{j=1}^k(\hat{\alpha}_j - \alpha_j)T^{-1}\Sigma_{t=1}^T(dy_{t-j} - T^{-1}\Sigma_{t=1}^T dy_{t-j})u_t$$

$$+ T^{-1}\Sigma_{t=1}^T[\Sigma_{j=1}^k(\hat{\alpha}_j - \alpha_j)(dy_{t-j} - T^{-1}\Sigma_{t=1}^T dy_{t-j})]^2 .$$

It is easy to show that the second and third term converge to zero while the first converges to $2\sigma_e^2$. Hence, combining the above elements we have $QF_T(1) \Rightarrow 0$.

To prove part (iii) we first note that the autoregressive polynomial can be written as $A(L) = (1 - (1 - c/T)L)A^*(L)$. Using the simplification $\beta_i = 0$ (i = 0, ..., p), we have :

$$(1 - L)A^*(L)y_t = e_t - (c/T)J_{t-1} , \tag{A.34}$$

where J_t is a near–integrated process defined by $J_t = (1 - c/T)J_{t-1} + e_t$. Hence, we can write the data–generating process as :

$$dy_t = \Sigma_{j=1}^{k-1}\alpha_j^* dy_{t-j} + u_t \tag{A.35}$$

where $$u_t = e_t - (c/T)J_{t-1}. \tag{A.36}$$

After some rearrangements using (A.34) and (A.35), we have,

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{u}_t = T^{-1/2}\Sigma_{t=1}^{[Tr]}(u_t - T^{-1}\Sigma_{t=1}^{T}u_t)$$

$$- \Sigma_{i=1}^{k}(\hat{\alpha}_i^* - \alpha_i^*)T^{-1/2}\Sigma_{t=1}^{[Tr]}(dy_{t-i} - T^{-1}\Sigma_{t=1}^{T}dy_{t-i}).$$

$$= T^{-1/2}\Sigma_{t=1}^{[Tr]}(u_t - T^{-1}\Sigma_{t=1}^{T}u_t) + o_p(1) ,$$

given that $\hat{\alpha}_i^* \to \alpha_i^*$ (with $\alpha_k^* = 0$) and $T^{-1/2}\Sigma_{t=1}^{[Tr]}dy_{t-i}$ is bounded (the proof is straightforward and omitted). Using (A.36),

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{u}_t = T^{-1/2}\Sigma_{t=1}^{[Tr]}e_t - rT^{-1/2}\Sigma_{t=1}^{T}e_t - c[T^{-3/2}\Sigma_{t=1}^{[Tr]}J_{t-1} - rT^{-3/2}\Sigma_{t=1}^{T}J_{t-1}]$$

$$\Rightarrow \sigma_e[W(r) - rW(1) - c(\int_0^r J_c(s)ds - r\int_0^1 J_c(s)ds)] \equiv \sigma_e[B_0(r) - c\int_0^r J_c^*(s)ds],$$

using (A.3) and Lemma 1(b) of Phillips (1987). The results follows upon verification that $\hat{\sigma}^2 \to \sigma_e^2$, the proof of which is omitted.

Proof of Theorem 8 : Part (a) when a unit root is present follows from Theorem 5 (i). To prove part (b, i) note that, under (28), $dy_t = \beta_1 + \delta_1/T^{3/2} + \Delta X_t$ with ΔX_t a stationary AR(k–1) process. The local asymptotic power function of the test must therefore converge to the size of the test using Theorem 6 (i), where a normalization by $T^{1/2}$ is used to obtain a non–degenerate local asymptotic power function. If the normalization in (28) was $T^{1/2}$ we would have $QF_T(1) \Rightarrow \int_0^1 B_{0,\delta}^*(r)^2 dr$.

To prove consistency when no unit root is present, note that the data–generating process can be written in this case as :

$$dy_t = \Sigma_{i=1}^{k}\alpha_i dy_{t-i} + u_t \tag{A.37}$$

where $$u_t = e_t - e_{t-1} + 1(t > \lambda T)A(1)\delta_1 . \tag{A.38}$$

This implies that :

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{u}_t = T^{-1/2}[e_{[Tr]} - e_0] - rT^{-1/2}[e_T - e_0]$$

$$- \Sigma_{j=1}^{k}(\hat{\alpha}_j - \alpha_j)\{T^{-1/2}[y_{[Tr]-j} - y_{-j}] - rT^{-1/2}[y_{T-j} - y_{-j}]\}$$

$$+ T^{-1/2}([Tr] - T\lambda)A(1)\delta_1 .$$

The first two terms converge to 0 as $T \to \infty$. By inspection of the third term we note that $T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{u}_t$ diverges at rate $T^{1/2}$. It is easy to show that under (A.37) $\hat{\sigma}^2$ is bounded. Therefore $QF_T(1)$ diverges at rate T and the test is consistent.

Consider now the local asymptotic power function of the test under a sequence of alternatives defined by (28). In the stationary case, dy_t is specified by (A.37) with $u_t = e_t - e_{t-1} + 1(t > \lambda T)A(1)\delta_1/T^{3/2}$, and therefore :

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{u}_t = T^{-1/2}[e_{[Tr]} - e_0] - rT^{-1/2}[e_T - e_0]$$

$$- \Sigma_{j=1}^{k}(\hat{\alpha}_j - \alpha_j)\{T^{-1/2}[y_{[Tr]-j} - y_{-j}] - rT^{-1/2}[y_{T-j} - y_{-j}]\}$$

$$+ T^{-1/2}([Tr] - T\lambda)A(1)\delta_1/T^{3/2} .$$

The last term now converges to zero. Given that $T^{-1/2}y_{[Tr]} = T^{-1/2}1(t > \lambda T)([Tr] - \lambda T)\delta_1/T^{3/2} + T^{-3/2}X_t$, we have $T^{-1/2}y_{[Tr]} \to 0$. Since the first term converges to 0, the proof is completed by showing that $(\hat{\alpha}_j - \alpha_j)$ is bounded, the proof of which is omitted.

Consider now the local asymptotic power function in the near–integrated case. After some rearrangements, we can write the data–generating process as :

$$dy_t = \Sigma_{i=1}^{k-1}\alpha_i^* dy_{t-i} + u_t , \tag{A.39}$$

$$u_t = e_t - (c/T)J_{t-1} - (c/T)LA^*(L)1(t > \lambda T)(t - \lambda T)\delta_1/T^{3/2} . \tag{A.40}$$

As before we have :

$$T^{-1/2}\Sigma_{t=1}^{[Tr]}\hat{u}_t = T^{-1/2}\Sigma_{t=1}^{[Tr]}(u_t - T^{-1}\Sigma_{t=1}^{T}u_t)$$

$$- \Sigma_{i=1}^{k}(\hat{\alpha}_i^* - \alpha_i^*)T^{-1/2}\Sigma_{t=1}^{[Tr]}(dy_{t-i} - T^{-1}\Sigma_{t=1}^{T}dy_{t-i}).$$

$$= T^{-1/2}\Sigma_{t=1}^{[Tr]}(u_t - T^{-1}\Sigma_{t=1}^{T}u_t) + o_p(1) ,$$

since $\hat{\alpha}_i^* \to \alpha_i^*$ and $T^{-1/2}\Sigma_{t=1}^{[Tr]}dy_{t-i}$ is bounded (the proof is straightforward and omitted). Using (A.40), $T^{-1/2}\Sigma_{t=1}^{[Tr]}u_t \Rightarrow \sigma_e[W(r) - c\int_0^r J_c(s)ds]$ since $c\delta_1 T^{-3}\Sigma_{T\lambda+1}^{[Tr]}LA^*(L)(t - \lambda T)$ $\to 0$ as $T \to \infty$. The proof is completed showing that $\hat{\sigma}^2 \to \sigma_e^2$ under (A.39) and (A.40).

Table I : Percentage Points of the Asymptotic Distribution of $QD_T(p)$; $p = 0, \ldots, 5$.

Stationary Case.

%	p					
	0	1	2	3	4	5
0.01	.0248	.0173	.0138	.0117	.0102	.0090
0.025	.0304	.0203	.0159	.0132	.0114	.0101
0.05	.0366	.0234	.0180	.0148	.0127	.0111
0.10	.0460	.0279	.0209	.0170	.0143	.0125
0.50	.1189	.0555	.0375	.0285	.0231	.0194
0.90	.3473	.1192	.0715	.0506	.0389	.0314
0.95	.4614	.1479	.0860	.0597	.0452	.0362
0.975	.5806	.1775	.1007	.0688	.0516	.0409
0.99	.7435	.2177	.1205	.0810	.0600	.0472
Mean	.1667	.0667	.0429	.0317	.0253	.0210

Source : MacNeill (1978), Table 2.

Table II : Percentage Points of the Asymptotic Distribution of $QD_T(p)$; $p = 0, \ldots, 5$.

Unit root case.

%	p					
	0	1	2	3	4	5
0.01	.0302	.0443	.0339	.0284	.0250	.0239
0.025	.0508	.0497	.0378	.0313	.0270	.0259
0.05	.0671	.0549	.0417	.0338	.0293	.0278
0.10	.0872	.0623	.0467	.0375	.0318	.0301
0.50	.1870	.0989	.0687	.0533	.0436	.0403
0.90	.3881	.1581	.1007	.0736	.0586	.0552
0.95	.4816	.1850	.1133	.0812	.0634	.0617
0.975	.5947	.2092	.1266	.0881	.0692	.0701
0.99	.7562	.2487	.1438	.1004	.0757	.0895
Mean	.2198	.1067	.0720	.0549	.0446	.0424

Table III : Power Function of $QD_T(1)$, Unit root and Breaking Trend.

DGP : $y_t = 1(t > 50)\delta_1 + y_{t-1} + e_t$ (t = 1,..., 100)

$\delta_1 = 0.5$, $e_t \sim N(0,1)$, $y_0 = 0$.

T	Size 0.10	0.05	0.025	0.01
100	.062	.028	.014	.005
200	.086	.043	.022	.009
500	.221	.129	.080	.033
1000	.551	.414	.303	.182
2000	.925	.860	.791	.674
5000	1.000	1.000	1.000	1.000

Note : The number of replications was 5,000 for T = 100, 200, 500 and 1000; 2,000 for T = 2,000; and 1,000 for T = 5,000.

TABLE IV : Empirical Results Using $QF_T(1)$.

Real GNP (or GDP) series for the G–7 countries.

Series	Sample	k	$QF_T(1)$
USA (GNP)	47:1–86:3	1	0.139
Canada (GDP)	47:1–89:1	8	0.549[b]
Italy (GDP)	60:1–85:1	1	0.583[c]
Germany (GNP)	60:1–86:2	8	0.406[a]
Japan (GNP)	57:1–88:4	7	0.446[a]
U.K. (GDP)	57:1–86:3	8	0.367[a]
France (GDP)[1]	65:1–88:3	1	0.893[d]
France (GDP)[2]	65:1–88:3	1	0.816[d]

Notes : a, b, c and d denote significance at the 10, 5, 2.5 and 1% level, respectively. For France, [1] denotes the case where no one–time dummy was included for the May 68 strike (68:2), while [2] denotes the case where such a dummy was included.

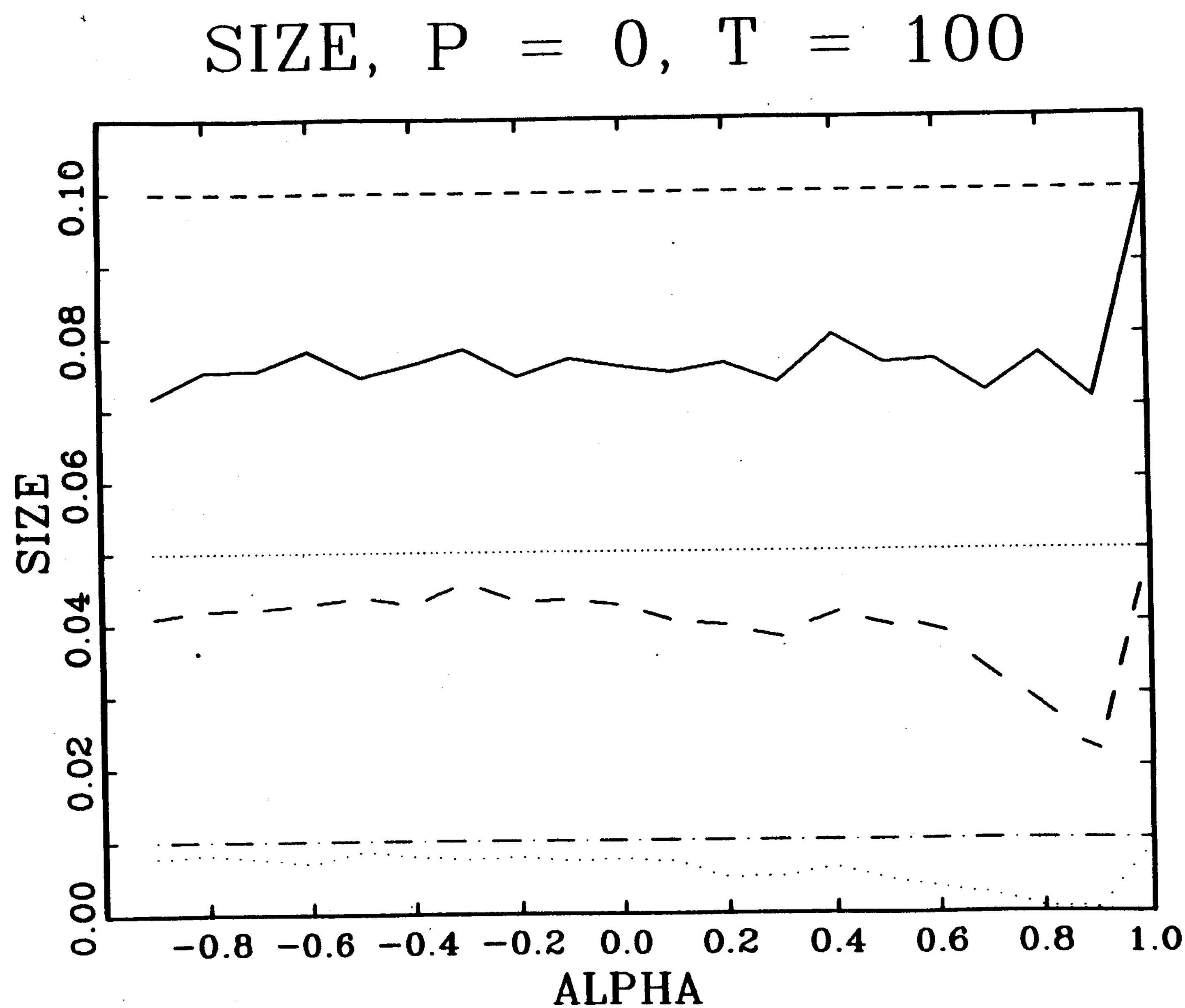


Figure 1: Exact size of $QD_T(0)$, $T=100$. The different curves are for the following nominal asymptotic sizes: (...) 1%, (– – –) 5%, and (———) 10%.

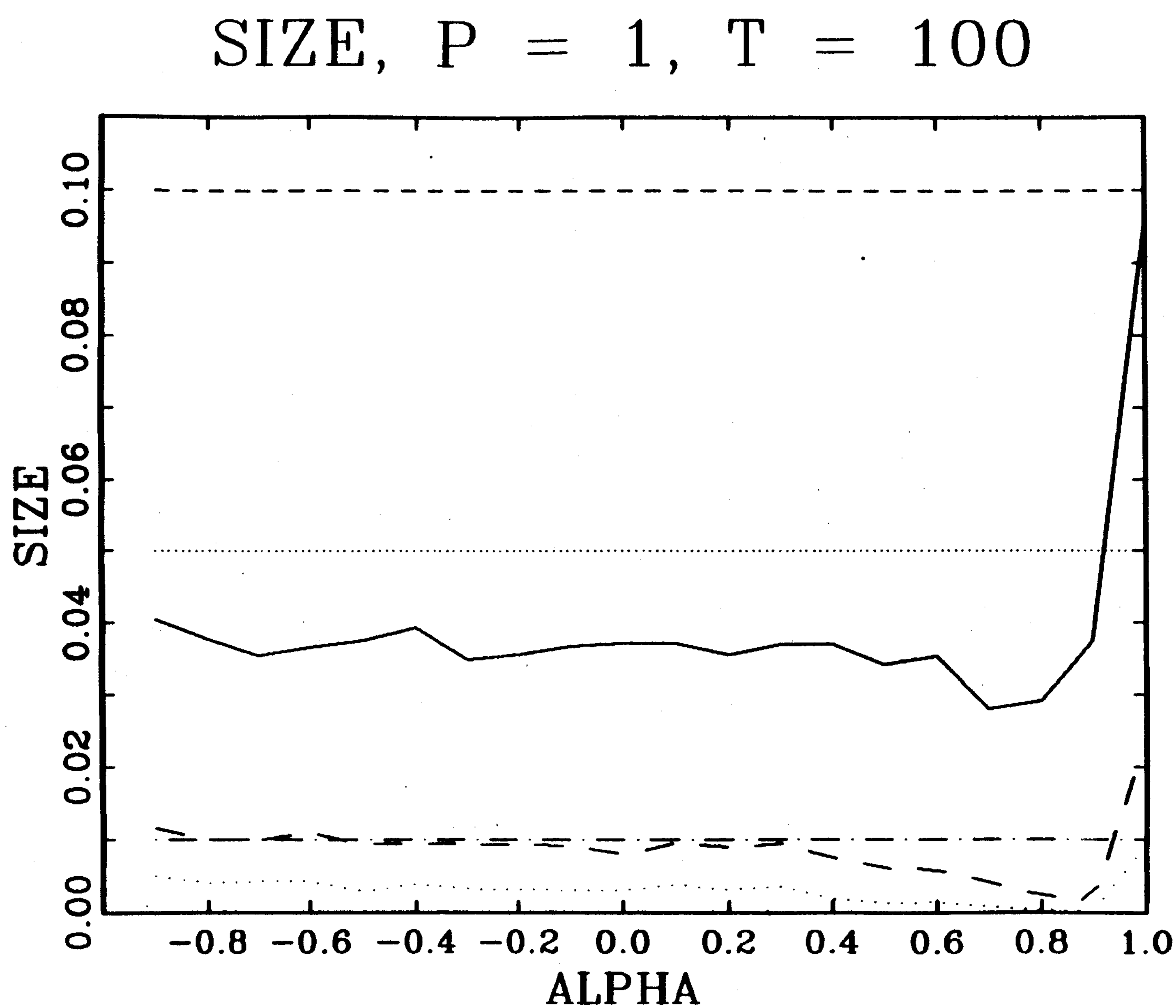
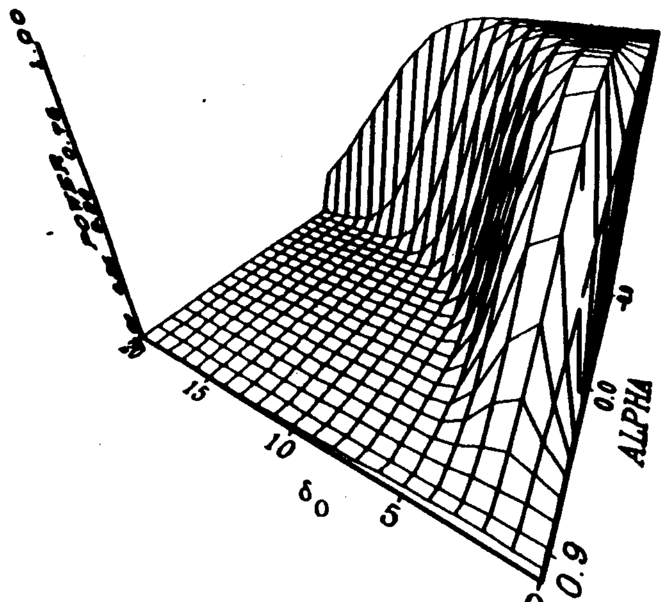
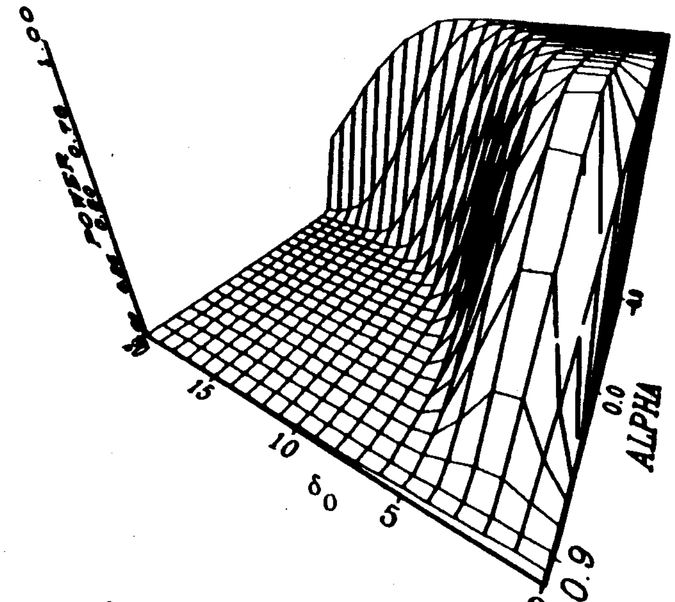


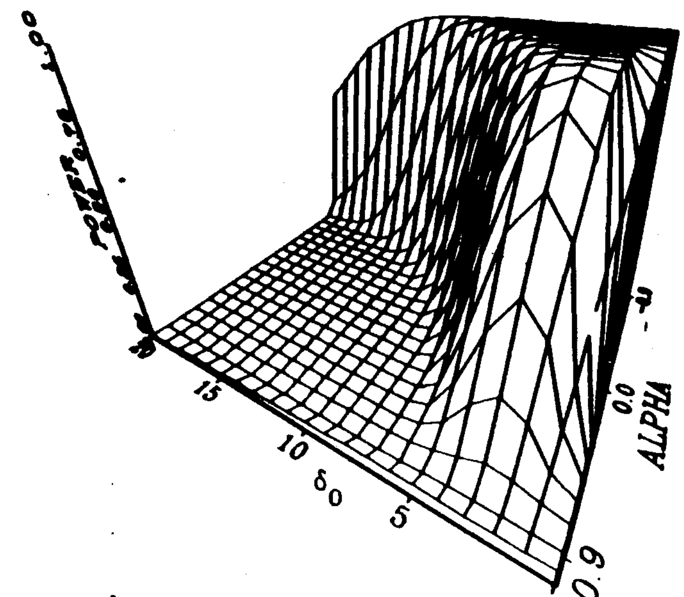
Figure 2: Exact size of $QD_T(1)$, T=100. The different curves are for the following nominal asymptotic sizes: (...) 1%, (– – –) 5%, and (———) 10%.



(a) λ = 0.25

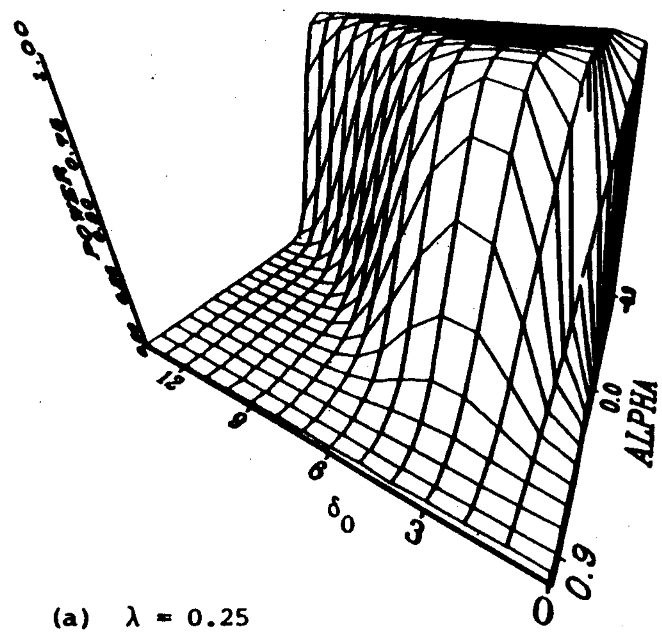


(b) λ = 0.50

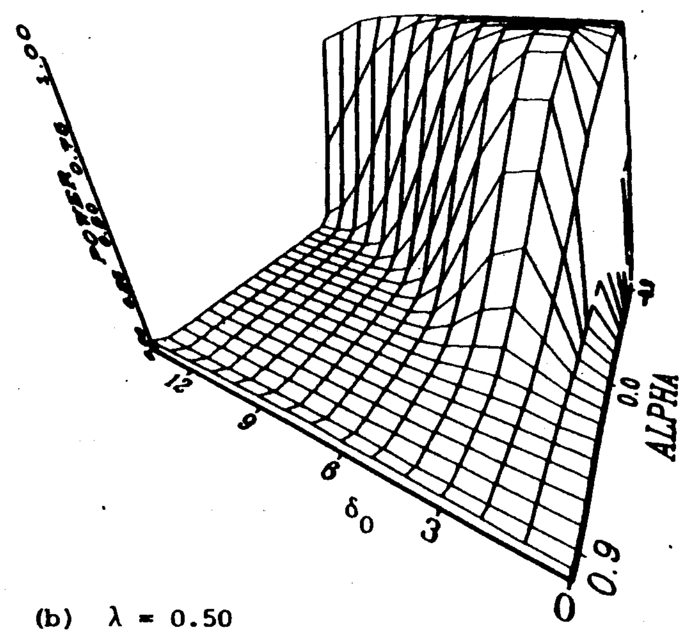


(c) λ = 0.75

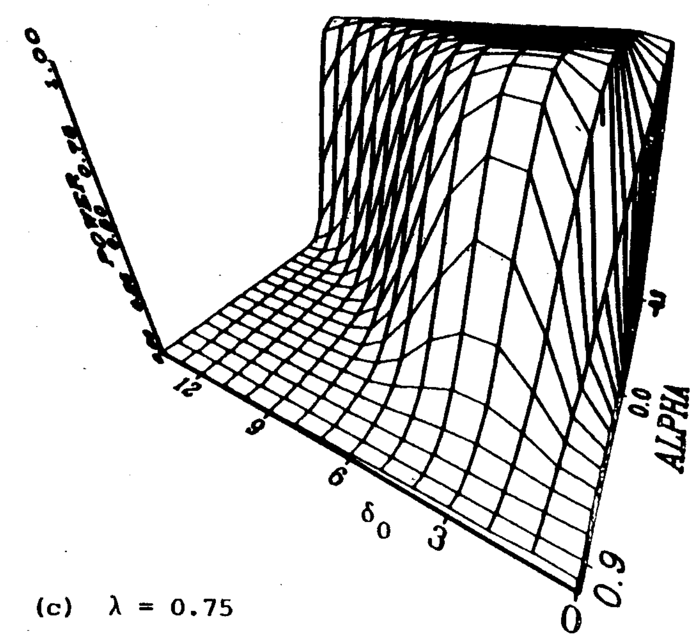
Figure 3: Power function of $QD_T(0)$ against a change in mean, T=100, 5% nominal size.



(a) λ = 0.25

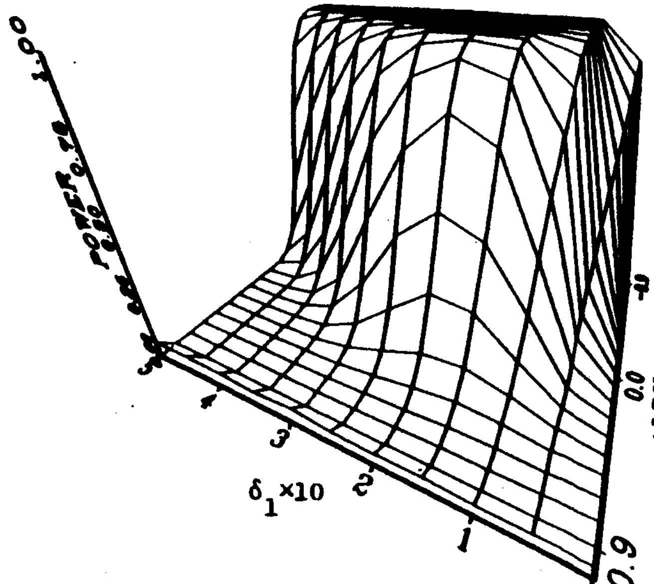


(b) λ = 0.50

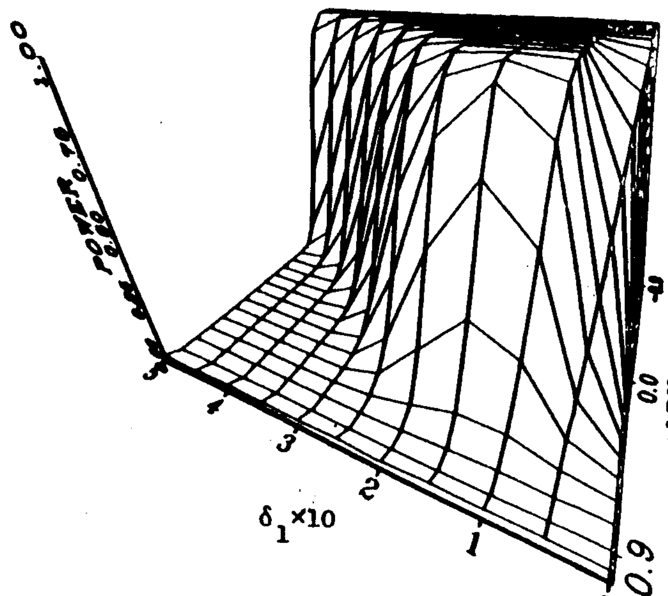


(c) λ = 0.75

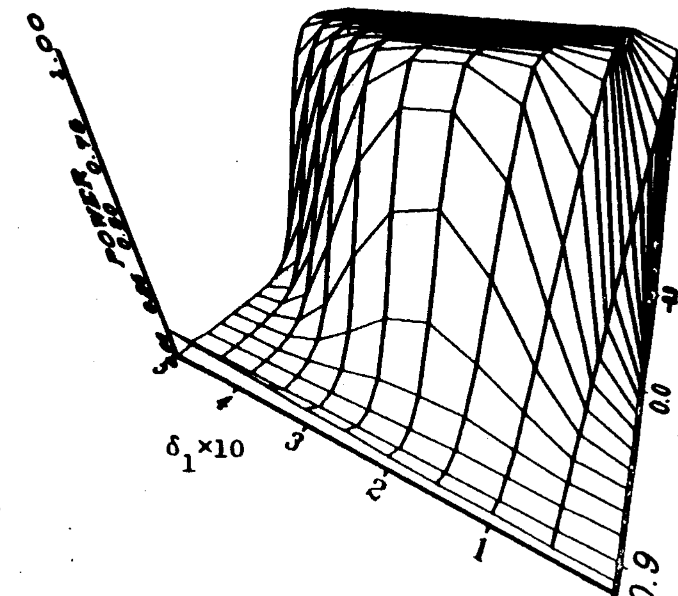
Figure 4: Power function of $QD_T(1)$ against a change in mean, T=100, 5% nominal size.



(a) $\lambda = 0.25$

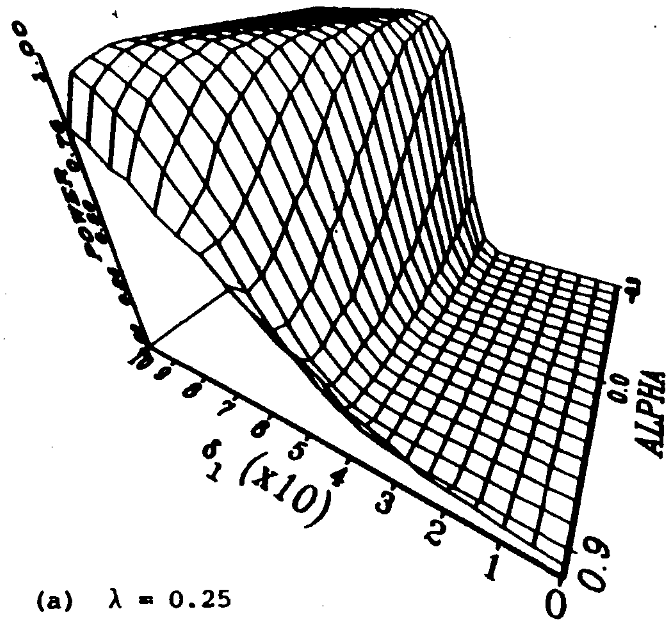


(b) $\lambda = 0.50$

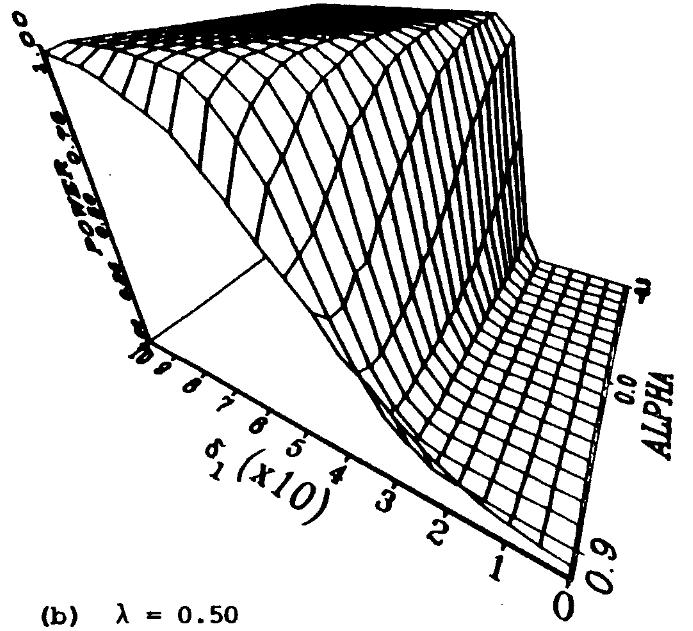


(c) $\lambda = 0.75$

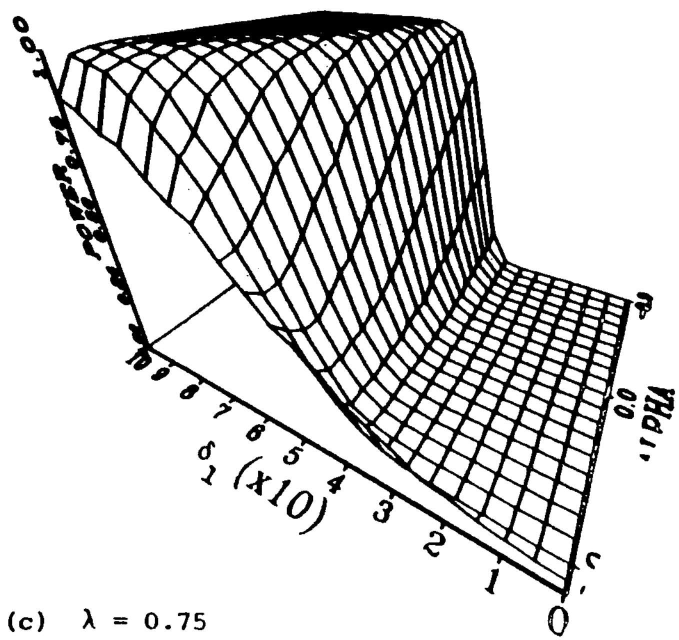
<u>Figure 5</u>: Power function of $QD_T(1)$ against a changing trend, $T=100$, 5% nominal Size.



(a) λ = 0.25



(b) λ = 0.50



(c) λ = 0.75

Figure 6: Power function of $QF_T(1)$ against a changing trend, T = 100, 5% nominal size.