Likelihood-Based Inference in Trending Time Series Models with a Root Near Unity *

Zhijie Xiao†
Department of Economics
University of Illinois at Urbana-Champaign

September 16, 1999

Abstract

This paper considers likelihood-based estimation and hypothesis tests for autoregressive time series models with unknown deterministic trends and general disturbance distributions. Asymptotic analysis on the M estimators for both the trend coefficients and the autoregressive coefficient is provided. Unit root tests based on M estimation are developed. Asymptotic distributions of these estimators and tests involve nonlinear equation systems of Brownian motions even for the simple case of least squares regression. Local power analysis is conducted and extensions of the Neyman-Pearsen test are studied. Unit root tests based on M-estimation coupled with quasi-differencing are analyzed. A Monte Carlo experiment is conducted to study the finite sample performance of these estimators and testing procedures.

*Paper to be presented at the Cowles Foundation Conference "New Developments in Time Series Econometrics," October 22 - 23, 1999. My thanks go to Peter Phillips and Roger Koenker for helpful comments. The paper was typed by the authors in SW2.5 and computations were performed in GAUSS.

†Address Correspondence: Department of Economics, University of Illinois at Urbana-Champaign, 1206 South Sixth Street, Champaign, IL 61820, USA. Phone: (217) 333-4520. Fax: (217) 244-6678. E-mail address: zxiao@uiuc.edu.

1. Introduction

In the past decade, econometricians have focused a great deal of attention on the development of hypothesis testing and estimation procedures in autoregressive time series models where the largest root is near unity. Most of these procedures are based on least square methods in linear regression models and have likelihood interpretations when the data are Gaussian. In the absence of Gaussianity, asymptotic results of these procedures generally still hold but these methods are less efficient than methods that exploit the distributional information. Monte Carlo evidence indicates that the least squares estimator can be very sensitive to certain type outliers, and inference procedures based on the least square estimation may have poor performance (see, e.g. Lucas 1994). In empirical analysis, many applications in nonstationary time series involve financial data like exchange rates whose distributions are heavy-tailed. It is therefore important to consider estimation and test procedures which are robust to departures from Gaussianity and can be applied to nonstationary time series. The present paper addresses some of these issues.

There has been some study on nonstationary time series regression with non-normal innovations, including Phillips (1995), Cox and Llatas (1991), Knight (1991), Lucas (1995), and Rothenberg and Stock (1997) etc. among others. In particular, Phillips (1995) studied robust cointegrating regressions and developed Fully Modified LAD and M-estimators for cointegrating regressions. Lucas (1995) considered unit root tests based on M-estimators. Cox and Liatas (1991) and Rothenberg and Stock (1997) studied robust estimation and inference for nearly integrated autoregressive models without deterministic trends.

Many macroeconomic time series, such as real GNP, consumption, money, and prices, display a tendency toward growth over time. Consequently, most empirical analyses in nonstationary time series literature considers unit root time series with a deterministic trend. One traditional way of modelling trending time series is to consider regressions of the following form:

$$y_t = \gamma' x_t + \alpha y_{t-1} + u_t, \tag{1.1}$$

where x_t is a deterministic trend of known form. However, as argued in Schmidt and Phillips (1992), the parameterization in (1.1) is not convenient in interpreting the deterministic component. For instance, considering the leading case that $x_t = (1, t)'$, we have

$$y_t = \gamma_0 + \gamma_1 t + \alpha y_{t-1} + u_t. \tag{1.2}$$

Such an equation has the property that the meanings of the parameters γ_0 and γ_1 differ under the null and the alternative. Under the null of a unit root, the parameters γ_0 and γ_1 represent trend and quadratic trend, respectively. However, under the alternative, γ_0 and γ_1 determine level and trend. This problem also surfaces in the unit root tests and "extra" deterministic trend component has to be introduced to remove the nuisance parameters. The introduction of surplus trend variables results in some inefficiency in the regression and reduces the power of, say, the Dickey-Fuller test from its already low level. To avoid the problem caused by the confusion over the meanings of parameters, researchers have considered as an alternative to (1.1) the following data generating process:

$$y_t = \gamma' x_t + y_t^s, \; t = 1, ..., n, \tag{1.3}$$

$$y_t^s = \alpha y_{t-1}^s + u_t, \; t = 1, ..., n. \tag{1.4}$$

This parameterization allows for the same trend component under both the null and the alternative hypothesis and is now widely used in time series analysis.

Combining (1.3) and (1.4) gives the nonlinear regression model

$$y_t = \gamma' \Delta x_t + (1-\alpha)\gamma' x_{t-1} + \alpha y_{t-1} + u_t. \tag{1.5}$$

Compared with (1.1), regression (1.5) incorporates both the null and the alternative models in a nonlinear equation.

This paper considers likelihood-based estimation and hypothesis tests for autoregressive time series model (1.3) with unknown deterministic trend and non-Gaussian innovations. Asymptotic analysis on the M-estimators, including the maximum likelihood estimators, for both the trend coefficients and the autoregressive coefficient is provided. Unit root tests against alternatives with an unknown local parameter are developed based on M estimation. Asymptotic distributions of these estimators and tests are complicated and involve nonlinear equation systems of Brownian motions even for the simple case of least squares regression. Local power analysis is conducted to show that these tests have non-trivial power against n-local alternatives. In addition, as a natural extension of the Neyman-Pearsen test, the likelihood ratio test for a unit root against a point alternative is studied and asymptotic power functions and power envelopes are derived. Parallel to the Gaussian case, unit root tests based on M-estimation coupled with quasi-differencing are analyzed. A Monte Carlo experiment is conducted to study the finite sample performance of these estimators and testing procedures.

This paper is organized as follows: Section 2 introduces the model. An asymptotic analysis of the M-estimators is given in Section 3. Section 4 studies unit root tests based on the M estimators. The asymptotic analysis of the unit root tests against a point alternative is given in Section 5. Section 6 studies the quasi-differencing M detrended unit root tests. Some Monte Carlo results are reported in Section 7 and Section 8 concludes.

A word on notation: the symbol "$\Rightarrow$" signifies weak convergence, "$\equiv$" signifies equality in distribution, and "$:=$" signifies definitional equality. "L" denotes lag operator. $\Delta = 1 - L$ is the difference operator and Δ_c signifies quasi-difference, which is defined by $\Delta_c = 1 - (1 + c/n)L$. $I(k)$ denotes integration of order k. All limits are taken as $T \to \infty$, unless otherwise specified.

2. The Model

We consider the autoregression model introduced in Section 1 that the observed time series y_t can be written as the sum of a deterministic trend d_t and a stochastic component y_t^s :

$$y_t = d_t + y_t^s, \; t = 1, ..., n. \tag{2.1}$$

$$y_t^s = \alpha y_{t-1}^s + u_t, \; t = 1, ..., n. \tag{2.2}$$

The deterministic trend d_t depends on unknown parameters and is specified as

$$d_t = \gamma' x_t, \tag{2.3}$$

where γ is a vector of the trend coefficient and x_t is a deterministic trend of known form. The leading case of the deterministic component is the linear time trend where $x_t = (1, t)$. In general, the trend function x_t may be more complex than a simple time polynomial. For example, time polynomials with sinusoidal factors and piecewise time polynomials may be used. The latter corresponds to a class of models with structural breaks in the deterministic trend. y_t^s is the stochastic component of y_t which can be represented by an autoregressive process. $\{u_t\}$ is the unobserved innovation process which is assumed to be stationary with mean zero. y_0^s is a constant (more generally, we can assume that it is a random variable of finite variance).

We want to study the estimation and hypothesis testing for this model when the parameter α is closed to one. In order to obtain large sample approximations, we employ the local-to-unity asymptotic theory investigated by Phillips (1987, 1988), Chan and Wei (1987), and others. Thus we consider the parameter space in a shrinking neighborhood of unity and reparameterize α so that

$$\alpha = 1 + \frac{c}{n}. \tag{2.4}$$

Combining (2.1), (2.2), (2.3), and (2.4), we have

$$\Delta y_t = \gamma' \Delta x_t - c\gamma' \left(\frac{x_{t-1}}{n} \right) + c \left(\frac{y_{t-1}}{n} \right) + u_t. \tag{2.5}$$

To simplify the exposition, we start with the simple case that u_t are unobserved iid errors with mean zero and unit variance. If we denote the log density of u as $f(u)$, then the conditional log density of y_t is given as follows:

$$f\left(\Delta y_t - c(y_{t-1}/n) - \gamma' \Delta x_t + c\gamma'(x_{t-1}/n)\right),$$

and the joint log density of the random sample is given by

$$\sum_{t=1}^{n} f\left(\Delta y_t - c(y_{t-1}/n) - \gamma' \Delta x_t + c\gamma'(x_{t-1}/n)\right).$$

Writing it as a function of the parameters, the above expression delivers the log likelihood function and we denote it as $L(c, \gamma)$, i.e.,

$$L(c, \gamma) = \sum_{t=1}^{n} f\left(\Delta y_t - c(y_{t-1}/n) - \gamma' \Delta x_t + c\gamma'(x_{t-1}/n)\right).$$

The maximum likelihood estimators of c and γ can then be found by maximizing $L(c, \gamma)$ with respect to c and γ. More generally, if we consider some criterion function φ, the M estimators of c and γ are obtained from a similar optimization problem with f replaced by φ.

Lucas (1995) considered unit root tests based on the M-estimator of model (1.1). Cox and Llatas (1991) studied the asymptotic behavior of the M-estimator of c for the case without a deterministic trend, and Rothenberg and Stock (1997) considered model (1.3), but the asymptotic analysis was only conducted for the simple case without deterministic trends. However, in the presence of an unknown deterministic component, the system that determines the M estimator becomes more complicated and is generally nonlinear.

3. Asymptotic Analysis of The M-Estimators

We are interested in the asymptotic behavior of the M-estimators of c and γ in model (2.5), defined as the solution of the following extreme problem:

$$\begin{bmatrix} \widehat{c} \\ \widehat{\gamma} \end{bmatrix} = \arg\max \left[\sum_{t=1}^{n} \varphi \left(\Delta y_t - c(y_{t-1}/n) - \gamma' \Delta x_t + c\gamma'(x_{t-1}/n) \right) \right] \tag{3.1}$$

for some criterion function φ. Taking $\varphi(u) = u^2$, (3.1) gives the ordinary least squares (OLS) estimator of c and γ. The maximum likelihood estimator corresponds to the case when φ is the true log density function. Although we pay particular attention to the maximum likelihood estimator, our analysis in this section will be given in a general way so that the M-estimator is covered, treating the maximum likelihood estimator as a special case of particular interest (notice that some simplifications happen in the asymptotic results when φ is the true log density).

We want to examine the asymptotic distribution of the estimator $(\widehat{c}, \widehat{\gamma})$. Under regularity conditions, the estimator $(\widehat{c}, \widehat{\gamma})$ can also be defined as a solution to the following equation system, which is the first-order condition of the extremum problem (3.1):

$$(NL) \begin{cases} \sum_{t=1}^{n} \varphi' \left(\Delta y_t - c(y_{t-1}/n) - \gamma' \Delta x_t + c\gamma'(x_{t-1}/n) \right) \Delta_c x_t = 0, \\ \sum_{t=1}^{n} \varphi' \left(\Delta y_t - c(y_{t-1}/n) - \gamma' \Delta x_t + c\gamma'(x_{t-1}/n) \right) y_{t-1}^{s} = 0. \end{cases}$$

(NL) is a nonlinear equation system. Generally there is no analytic solution even for the simple case when φ is the log density of a normal distribution. Indeed, as will become clear later, in a n-shrinking neighborhood of unity, the limiting distributions of the M estimators are jointly determined by a nonlinear equation system of Brownian motions and the limiting trending functions.

To study the asymptotic distributions of the M estimators, it is convenient for us to make the following assumptions on u_t and the criterion function φ.

ASSUMPTION A: u_t are iid with mean zero and variance one. $\varphi(\cdot)$ possesses derivatives φ' and φ''. $[u, \varphi'(u)]$ has k-th moments for some $k > 2$, $E[\varphi'(u_t)] = 0$, and φ'' is Lipschitz continuous.

Assumption A is a standard condition in asymptotic analysis of maximum likelihood estimators or M-estimators. The moment conditions on u and $\varphi'(u)$ are needed to establish the weak convergence results. We may also replace the moment condition on $\varphi'(u)$ by boundedness conditions of the derivatives of φ, because the latter and the moment condition on u imply the corresponding condition on φ'.

Denote $[\cdot]$ as the greatest lesser integer function. Then under Assumption A, as n goes to ∞, $n^{-1/2}\sum_1^{[nr]} u_t$ converges weakly to a standard Brownian motion $W_1(r)$, and thus $n^{-1/2}y^s_{[nr]}$ converges weakly to the corresponding Ornstein-Uhlenbeck process $J_c(r) = \int_0^r e^{c(r-s)}dW_1(s)$. The limiting distributions of $\widetilde{c}$ and $\widetilde{\gamma}$ will also be dependent on the weak limit of the partial sums of $\varphi'(u_t)$. We assume that the moments $\omega^2 = var[\varphi'(u_t)]$, $\delta = -E[\varphi''(u_t)]$, and $\rho = -E[u_t\varphi'(u_t)]$ exist, then $n^{-1/2}\sum_1^{[nr]}\varphi'(u_t) \Rightarrow \omega W_\varphi(r)$, where W_φ is a standard Brownian motion.

Notice that $W_1(r)$ and $W_\varphi(r)$ are correlated Brownian motions. To deal with the correlation between W_1 and W_φ explicitly, following the previous literature, we construct the random variable $v_t = \varphi'(u_t) + \rho u_t$. Then, by construction, v_t are iid with variance $(\omega^2 - \rho^2)$ and are uncorrelated with u_t. The partial sum process $n^{-1/2}\sum_1^{[nr]} v_t$ converges weakly to $\sqrt{\omega^2 - \rho^2}W_2(r)$, where $W_2(r)$ is a standard

Brownian motion independent of $W_1(r)$. $B_\varphi(r) = \omega W_\varphi(r) = \sqrt{\omega^2 - \rho^2} W_2(r) - \rho W_1(r)$.

For asymptotic analysis of the deterministic trend, we assume that there are standardizing matrices D_n and $F_n = n^{-1} D_n$ such that $D_n^{-1} x_{[nr]} \rightarrow X(r)$ and $F_n^{-1} \Delta x_{[nr]} \rightarrow g(r)$, as $n \rightarrow \infty$, uniformly in $r \in [0,1]$. In the case of a linear trend, $D = diag[1, n]$ and $X(r) = (1, r)'$. If x_t is a general p-th order polynomial trend, $D = diag[1, n, ..., n^p]$ and $X(r) = (1, r, ..., r^p)$.

If we denote the M estimators of c and γ as $\widetilde{c}$ and $\widetilde{\gamma}$, the limiting distributions of $\widetilde{c}$ will be dependent on asymptotic behavior of the random variables $n^{-1} \sum_{t=1}^n \varphi'(\Delta y_t - \widetilde{\gamma}' \Delta x_t)(y_{t-1} - \widetilde{\gamma}' x_{t-1})$ and $n^{-1} \sum_{t=1}^n \varphi'(\Delta y_t - \widetilde{\gamma}' \Delta x_t) \Delta_{\widetilde{c}} x_t$. An analysis of these two items is given in the Appendix. To derive the limiting distribution of $(\widetilde{c}, \widetilde{\gamma})$, we assume that the following conditions hold.

Assumption C: $\widetilde{c} = c + O_p(1)$, and $n^{-1/2} D_n(\widetilde{\gamma} - \gamma) = O_p(1)$.

Assumptions similar to C are standard in the development of M-estimator asymptotics. It is related to Assumption (b) in Theorem 5.1 of Phillips (1995) and the assumption on $\widetilde{\varepsilon}_t - \varepsilon_t$ in Theorem 1 of Lucas (1995). Notice that

$$\begin{aligned} \widetilde{u}_t &= \Delta y_t - \widetilde{c}(y_{t-1}/n) - \widetilde{\gamma}' \Delta x_t + \widetilde{c} \widetilde{\gamma}'(x_{t-1}/n) \\ &= u_t - (\widetilde{c} - c)(y_{t-1}^s/n) - (\widetilde{\gamma} - \gamma)' \Delta_c x_t + (\widetilde{c} - c)(\widetilde{\gamma} - \gamma)'(x_{t-1}/n). \end{aligned}$$

Under Assumption C, $\widetilde{u}_t - u_t$ satisfies the conditions in Lucas (1995) and Phillips (1995). Denote the limit of $n^{-1/2} D_n(\widetilde{\gamma} - \gamma)$ by ξ_c and the limit of $\widetilde{c}$ by η_c, the limiting distributions of the M estimators $\widetilde{c}$ and $\widetilde{\gamma}$ are given in the following theorem.

Theorem 1: Given models (2.1), (2.3), (2.2), and (2.4), for all c in a compact set, under Assumptions A and C, the limiting distributions of nonlinear regression estimators $\tilde{\gamma}$ and $\tilde{c}$ are jointly determined by the following equations:

$$\eta_c \int X_\eta(r) \underline{J}_{c\xi}(r) dr = \int X_\eta(r) d\widetilde{S}_c(r),$$

$$\eta_c \int \underline{J}_{c\xi}(r)^2 dr = \int \underline{J}_{c\xi}(r) d\widetilde{S}_c(r),$$

where

$$\begin{aligned} X_\eta(r) &= g(r) - \eta_c X(r), \ \underline{J}_{c\xi}(r) = J_c(r) - \xi_c' X(r), \\ \widetilde{S}_c(r) &= \frac{\rho}{\delta} W_1(r) + \int_0^r \underline{J}_c(s) ds - \frac{\sqrt{\omega^2 - \rho^2}}{\delta} W_2(r), \\ \underline{J}_c(r) &= c J_c(r) - \xi_c' g(r). \end{aligned}$$

Remark 1: If φ is the true log density for u_t, $\rho = 1$ and $\omega^2 = \delta \geq 1$. The departure from Gaussianity in the data is completely determined by the parameter ω^2. When the data is generated by a Gaussian process, $\omega^2 = 1$ and $W_2(\cdot)$ disappears from the limiting distribution. As ω^2 increases, the underlying distribution becomes more and more non-Gaussian.

Remark 2: In the stationary case, similar nonlinear regression estimators can be obtained from (1.5). However, under regularity conditions, closed-form solutions of the limiting distributions for these estimators can be derived and it can be shown that they are first order equivalent to the one-step Newton-Raphson estimators.

Remark 3: Monte Carlo evidence in Section 7 indicates that the maximum likelihood estimator based on this nonlinear regression generally provides a more efficient way than the OLS regression in estimating the deterministic trend when there is a large autoregressive root in y_t^s. Because the true value of the local parameter c is unknown, this maximum likelihood estimator of the deterministic trend can not achieve the efficiency level that applies when the local parameter is known.

REMARK 4: From the proof of Theorem 1, it can be derived that the M estimator for c can be written as

$$\widetilde{c} = \frac{n^{-1}\sum_{t=1}^{n}\varphi'(\Delta y_t - \widetilde{\gamma}'\Delta x_t)\,(y_{t-1} - \widetilde{\gamma}' x_{t-1})}{n^{-2}\sum_{t=1}^{n}\varphi''(\Delta y_t - \widetilde{\gamma}'\Delta x_t)\,(y_{t-1} - \widetilde{\gamma}' x_{t-1})^2} + o_p(1).$$

If we take $\varphi(u) = u^2$, the estimators are least squares regression estimators and Theorem 1 gives the results in Phillips and Xiao (1998). These estimators have likelihood interpretations when the process is actually Gaussian. However, even if the time series are not normal, the asymptotic results still hold. Denoting the nonlinear least squares regression estimators as $\widetilde{\gamma}_{ls}$ and $\widetilde{c}_{ls}$, we summarize the limiting distribution of the nonlinear least square estimators in the following corollary for convenience of later analysis.

COROLLARY 1: Given models (2.1), (2.3), (2.2), and (2.4), for all c in a compact set, the limiting distributions of nonlinear least squares regression estimators $\widetilde{\gamma}_{ls}$ and $\widetilde{c}_{ls}$ are jointly determined by the following equations:

$$\begin{aligned} \xi_c^* &= [\textstyle\int X_\eta(r) X_\eta(r)' dr]^{-1} [\textstyle\int X_\eta(r) dV_\eta(r)], \\ \eta_c^* &= \left[\textstyle\int \underline{J}_{c\xi}(r)^2 dr\right]^{-1} \left[\textstyle\int \underline{J}_{c\xi}(r) dU_\xi(r)\right], \end{aligned}$$

where

$$X_c(r) = g(r) - cX(r),\ V_\eta(r) = W(r) - (\eta_c^* - c)\int_0^r J_c(s)ds,$$
$$U_\xi(r) = W(r) - \xi_c^{*\prime}\int_0^r X_c(s)ds,$$

and $W(r)$ is a standard Brownian motion.

4. Unit Root Tests Based on M Estimators

This Section considers unit root tests based on M estimators. We are interested in the alternative hypothesis that α is less than unity. For alternatives that are distant from unity, the proposed tests will be consistent and will reject H_0 with probability close to one in large samples. Thus we consider alternatives with a root near unity. Then, the null hypothesis is $H_0 : c = 0$, and we are interested in alternative hypothesis $H_1 : c < 0$. Of course, more generally we may consider the null hypothesis that $c = c_0$.

In our discussion, we assume that φ is the true log density when considering the likelihood ratio tests. In other cases, φ can be more general criterion functions satisfying Assumption A. We start with the likelihood ratio test for the hypothesis that y_t has a unit root. Let $\widehat{\gamma}$ be the restricted maximum likelihood estimator of γ under the null hypothesis of $c = 0$ and $\widetilde{c}$ and $\widetilde{\gamma}$ be the unrestricted maximum likelihood estimators of c and γ analyzed in Section 3, the likelihood ratio test for the null hypothesis of a unit root rejects H_0 for small values of

$$L(0, \widehat{\gamma}) - L(\widetilde{c}, \widetilde{\gamma}).$$

The limiting distribution of the likelihood ratio statistic depends on the limiting distributions of both the unrestricted maximum likelihood estimator and the restricted maximum likelihood estimator. Under the null hypothesis of $c = 0$, $n^{-1/2} y^s_{[nr]}$ converges weakly to $W_1(r)$ and, as a result of Theorem 1, the limiting distributions of the unrestricted nonlinear regression estimators $\widetilde{\gamma}$ and $\widetilde{c}$ are jointly determined by the following equations:

$$\begin{cases} \eta_0 \int X_{\eta_0}(r) \underline{W}_{\xi_0}(r) dr = \int X_{\eta_0}(r) d\widetilde{S}_0(r), \\ \eta_0 \int \underline{W}_{\xi_0}(r)^2 dr = \int \underline{W}_{\xi_0}(r) d\widetilde{S}_0(r), \end{cases}$$

where

$$\underline{W}_{\xi_0}(r) = W_1(r) - \xi_0' X(r), \ \ \widetilde{S}_c(r) = \frac{1}{\omega^2} W_1(r) - \xi_0' X(r) - \frac{\sqrt{\omega^2 - 1}}{\omega^2} W_2(r).$$

For the restricted estimator, notice that, under H_0, the log likelihood is simply

$$L(0, \gamma) = \sum \varphi(\Delta y_t - \gamma' \Delta x_t).$$

The restricted MLE of γ satisfies the following first order condition:

$$\sum_{t=1}^{n} \varphi'(\Delta y_t - \widehat{\gamma}' \Delta x_t) \Delta x_t = 0.$$

When x_t contains a constant term, the corresponding element in Δx_t is zero and, as a result, the regressor in the restricted model actually has smaller dimension than x_t. To avoid the singularity problem in deriving the limiting distributions of $\widehat{\gamma}$, we express this explicitly by rewriting the trend component in the restricted model as $\gamma' \Delta x_t = \beta' \breve{x}_t$, where $\breve{x}_t = S x_t$ for some eliminator matrix S that eliminates redundant rows of x_t. Therefore $\breve{x}_t$ is usually of smaller dimension than x_t and the log likelihood can be rewritten as

$$\sum \varphi(\Delta y_t - \beta' \breve{x}_t).$$

Assume that $G_n^{-1} \breve{x}_{[nr]} \to \underline{X}(r)$ as $n \to \infty$, uniformly in $r \in [0, 1]$. Then $\widehat{\beta}$ has the following asymptotic distribution under the null of a unit root.

LEMMA 1: Under H_0 and Assumption A,

$$n^{1/2} G_n(\widehat{\beta} - \beta) \Rightarrow \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_0(r), \tag{4.1}$$

where

$$S_0(r) = \frac{\rho}{\delta} W_1(r) - \frac{\sqrt{\omega^2 - \rho^2}}{\delta} W_2(r).$$

REMARK 5: If x_t does not contain a constant term, say, $x_t = t$, then $\breve{x}_t$ is actually of the same dimension as x_t, and $\underline{X}(r) = g(r)$.

REMARK 6: If φ is log normal, $S_0(r)$ is simply $W_1(r)$ and the limit distribution (4.1) reduces to the standard result of detrending.

The asymptotic null distribution of the likelihood ratio statistic is summarized in Theorem 2.

THEOREM 2: Under Assumptions A and C and the null hypothesis $H_0 : c = 0$,

$$\begin{aligned} & 2\left[L(0,\widehat{\gamma}) - L(\widetilde{c},\widetilde{\gamma})\right] \qquad (4.2) \\ \Rightarrow\ & \delta \int \left[\eta_0 \underline{W}_{\xi_0}(r) + \xi_0' g(r)\right]^2 dr - 2\delta \int \left[\eta_0 \underline{W}_{\xi_0}(r) + \xi_0' g(r)\right] dS_0(r) \\ & +\delta \int dS_0(r) \underline{X}(r)' \left[\int \underline{X}(r) \underline{X}(r)' dr\right]^{-1} \int \underline{X}(r) dS_0(r). \end{aligned}$$

REMARK 7: We can also construct likelihood ratio tests for the null hypothesis $c = c_0$ vs. the alternative $c \neq c_0$. The principles of proofs are the same and the asymptotic results are similar to those in Theorem 2.

We may also construct test statistics directly based on the M estimator of the unknown parameter. To test the null hypothesis of a unit root (i.e., $c = 0$) against the alternative of near stationarity, we can consider the M estimator of c and reject H_0 for small value of $\widetilde{c}$. This can be treated as a generalization of the Dickey-Fuller or Phillips-Perron tests under M-estimation. From the result in Section 3, this is asymptotically equivalent to rejecting H_0 if

$$Z_M = \frac{n \sum_{t=1}^{n} \varphi'(\Delta y_t - \widetilde{\gamma}' \Delta x_t) \left(y_{t-1} - \widetilde{\gamma}' x_{t-1}\right)}{\sum_{t=1}^{n} \varphi''(\Delta y_t - \widetilde{\gamma}' \Delta x_t) \left(y_{t-1} - \widetilde{\gamma}' x_{t-1}\right)^2} < CV_\tau, \qquad (4.3)$$

for some critical value CV_τ.

THEOREM 3: Under the null hypothesis, the statistic Z_M defined by (4.3) converges weakly to

$$\left[\int \underline{W}_{\xi_0}(r)^2 dr\right]^{-1} \int \underline{W}_{\xi_0}(r) d\widetilde{S}_0(r), \tag{4.4}$$

where

$$\widetilde{S}_0(r) = \frac{\rho}{\delta} W_1(r) - \xi_0' X(r) - \frac{\sqrt{\omega^2 - \rho^2}}{\delta} W_2(r).$$

To construct asymptotically valid tests for a unit root, we need to know the distributions given in (4.2) of Theorem 2, or (4.4) of Theorem 3. In the case that φ is the log density function, we can calculate the critical values by simulating W_1 and W_2. More generally, the distribution of u_t may not be known so that ω^2, ρ, and δ must be estimated. The asymptotic null distribution is unaffected if the parameters are replaced by their consistent estimates. Thus, a robust estimate of the null distribution can be obtained by simulating the distribution with the unknown parameters replaced by their consistent estimates. Such robust tests could be inconvenient in practical analysis since the critical values will have to be calculated each time. An alternative way (see, say, Hansen 1995; Lucas 1997) is to generate conservative critical values based on normal innovations. In the special case where φ is log normal, the distribution of the estimators and testing procedures can be considerably simplified - see later discussions in this Section.

Now we consider the behavior of the likelihood ratio statistic under the local alternative hypothesis with local parameter c. Using the results of Theorem 1, the limiting distribution of (4.2) can be derived and thus the power function can be obtained. We summarize the asymptotic results in the following theorem, which shows that the likelihood ratio test has non-trivial power against the local alternative.

THEOREM 4: Under Assumptions A and C and the local alternative,

$$
\begin{aligned}
& 2\left[L(0,\widehat{\gamma})-L(\widetilde{c},\widetilde{\gamma})\right] \\
\Rightarrow\ & \omega^2\int\left[(\eta_c-c)\,\underline{J}_{c\xi}(r)+\xi_c' X_c(r)\right]^2 dr \\
& +2\omega^2\int\left[(\eta_c-c)\,\underline{J}_{c\xi}(r)+\xi_c' X_c(r)\right]dW_\varphi(r)+2c\omega^2\int J_c(r)dW_\varphi(r) \\
& +\omega^2\int d\underline{S}_c(r)\underline{X}(r)'\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)dS_c(r)-c^2\omega^2\int J_c(r)^2 dr \\
& +2c\omega^2\int J_c(r)\underline{X}(r)'dr\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)dS_c(r),
\end{aligned}
$$

where

$$
\begin{aligned}
S_c(r) &= \frac{\rho}{\delta}W_1(r)-\frac{\sqrt{\omega^2-\rho^2}}{\delta}W_2(r)+c\int_0^r J_c, \\
\underline{S}_c(r) &= \frac{\rho}{\delta}W_1(r)-\frac{\sqrt{\omega^2-\rho^2}}{\delta}W_2(r)-c\int_0^r J_c.
\end{aligned}
$$

REMARK 8: Similar results can be obtained on the test Z_M.

If $\{u_t\}$ are Gaussian, the maximum likelihood estimators are just least squares estimators and their limiting distributions are given in Corollary 1. Under the null hypothesis, these limiting variates have the following simplified representation. Denoting the limiting variates when $c=0$ by ξ_0 and η_0 respectively, we have:

COROLLARY 2: Under H_0, the limiting distributions of nonlinear least squares regression estimators $\widetilde{\gamma}_{ls}$ and $\widetilde{c}_{ls}$ are jointly determined by the following equations:

$$
\begin{aligned}
\xi_0 &= \left[\int X_{\eta_0}(r)X_{\eta_0}(r)'dr\right]^{-1}\left[\int X_{\eta_0}(r)dV_{\eta_0}(r)\right], \\
\eta_0 &= \left[\int \underline{W}_{\xi_o}(r)^2 dr\right]^{-1}\left[\int \underline{W}_{\xi_0}(r)d\underline{W}_{\xi_0}(r)\right],
\end{aligned}
$$

where

$$V_{\eta_0}(r) = W(r) - \eta_0 \int_0^r W(s)ds,$$

and $W(r)$ is a standard Brownian motion.

REMARK 9: In the special case that $\widetilde{c}$ is the nonlinear least squares estimator, the limiting null distribution of Z_M is $\left[\int \underline{W}_{\xi_0}(r)^2 dr\right]^{-1} \int \underline{W}_{\xi_0}(r) d\underline{W}_{\xi_0}(r)$.

Under the normality assumption, the likelihood ratio test can be constructed based on the least squares regression. It can be shown that, after dropping the asymptotically negligible terms, the likelihood ratio test rejects H_0 for small values of

$$LR_c = \frac{\sum_{t=2}^n \left[\Delta y_t - \widetilde{c}_{nl}(y_{t-1}/n) - \widetilde{\gamma}'_{nl}\Delta x_t + \widetilde{c}_{nl}\widetilde{\gamma}'_{nl}(x_{t-1}/n)\right]^2}{\sum_{t=2}^n \left[\Delta y_t - \widehat{\gamma}'\Delta x_t\right]^2}.$$

The restricted maximum likelihood estimator $\widehat{\gamma}$ is simply the least squares estimator of the following regression on the differenced data:

$$\Delta y_t = \gamma' \Delta x_t + \Delta y_t^s. \tag{4.5}$$

Using $\breve{x}_t$ defined earlier in this section, (4.5) can be rewritten as $\Delta y_t = \beta' \breve{x}_t + \Delta y_t^s$. The least squares estimator $\widehat{\beta}_{ls}$ has the following asymptotic distribution under the null and the alternative hypothesis.

LEMMA 2:

(1) Under the null hypothesis that $c = 0$,

$$n^{1/2} G_n(\widehat{\beta}_{ls} - \beta) \Rightarrow \left[\int \underline{X}(r)\underline{X}(r)' dr\right]^{-1} \int \underline{X}(r) dW_1(r).$$

(2) Under the alternative hypothesis,

$$n^{1/2} G(\widehat{\beta}_{ls} - \beta) \Rightarrow \left[\int \underline{X}(r)\underline{X}(r)' dr\right]^{-1} \int \underline{X}(r) dJ_c(r).$$

The asymptotic distributions of LR_c are given in the following theorem.

THEOREM 5:

(1) Under H_0,

$$
\begin{aligned}
& n(LR_c - 1) \\
\Rightarrow\ & \int \left[\eta_0 \underline{W}_{\xi_0}(r) + \xi_0' g(r)\right]^2 dr - 2\int \left[\eta_0 \underline{W}_{\xi_0}(r) + \xi_0' g(r)\right] dW(r)^2 dr \\
& + \int dW(r)\underline{X}(r)' \left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1} \int \underline{X}(r) dW(r). \qquad (4.6)
\end{aligned}
$$

(2) Under H_c,

$$
\begin{aligned}
n(LR_c - 1) \Rightarrow\ & \xi_c' \left\{\int g(r)\left[X_\eta'(r) - (\eta_c + c)X'(r)\right] dr\right\} \xi_c + \int \widetilde{R}_c(r)^2 dr \\
& + \int dS_c(r)\underline{X}(r)' \left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1} \int \underline{X}(r) dS_c(r) \\
& -2\xi_c' \int X_\eta(r) dW(r) - 2\eta_c \int J_c(r) d\underline{W}_\xi(r) \\
& -c\int J_c(r) dW(r) - c\int J_c(r) dS_c(r),
\end{aligned}
$$

where

$$
\widetilde{R}_c(r) = \eta_c J_c(r) - (\eta_c + c)X'(r)\xi_c, \text{ and } \underline{W}_\xi(r) = W(r) - \xi_c' X(r).
$$

The limiting null distribution of LR_c (4.6) depends only on the limiting deterministic trend function. For the leading case of a linear trend $x_t = (1, t)'$, we calculated the critical values based on a direct Monte Carlo experiment using 30,000 iterations with sample size equals 1000. The 5% and 10% level critical values are -12.452 and -10.562 respectively.

5. Unit Root Test Against a Point Alternative

Even for the simplest case where $x_t = 0$ (or d_t is known) and thus y_t^s is observable, there is no uniformly optimal estimator for c or uniformly optimal test for H_0. Under regularity conditions, the random variables $\sum_{t=1}^{n} \varphi'(\Delta y_t^s) \left(\frac{y_{t-1}^s}{n}\right)$ and $\sum_{t=1}^{n} \varphi''(\Delta y_t^s) \left(\frac{y_{t-1}^s}{n}\right)^2$ have a nondegenerate limiting distribution and are asymptotically jointly sufficient statistics for the local parameter c. Notice that the asymptotic sufficient statistic is two dimensional and we can not find a uniformly best estimate for c or a uniformly most powerful test for H_0 even asymptotically. Cox and Llatas (1991) studied the optimality of the MLE for this case and showed that the optimal criterion function is a linear combination of the least squares score and the true score function where the linear combination depends on the unknown parameter c. Since φ is the log density of u_t, asymptotic admissible tests could be constructed based on a linear combination of $\sum_{t=1}^{n} \varphi'(\Delta y_t^s) \left(\frac{y_{t-1}^s}{n}\right)$ and $\sum_{t=1}^{n} \varphi''(\Delta y_t^s) \left(\frac{y_{t-1}^s}{n}\right)^2$.

If we consider a unit root test against the simple point alternative $c = \overline{c} < 0$, then, in the case d_t is known, a most powerful test can be constructed based on the likelihood ratio statistic $L(0) - L(\overline{c})$ by the Neyman-Pearson Lemma. As a result, the asymptotic local power function can be calculated and a power envelope can be obtained as we change the values of $\overline{c}$ (see Rothenberg and Stock (1997) for a discussion).

When d_t is unknown, say $d_t = \gamma' x_t$ = a linear time trend (or, more generally, a polynomial trend), the trend coefficient γ has to be estimated in order to construct a feasible test. However, the use of an estimated γ changes the limiting distribution and there is no most powerful test for the unit root hypothesis. In this case, a natural generalization of the Neyman-Pearsen test for the null of $c = 0$ against the point alternative $c = \overline{c}$ is to reject for small values of the likelihood ratio $L(0,\widehat{\gamma}) - L(\overline{c},\overline{\gamma})$, where $\widehat{\gamma}$ and $\overline{\gamma}$ are the maximum likelihood estimators for γ under the null and the alternative hypothesis respectively. Elliott, Rothenberg, and Stock (1996) studied this test for the Gaussian case with a linear trend and constructed a power envelope based on such tests. In this section, we explore the asymptotic properties of such tests for cases with non-Gaussian innovations.

The asymptotic behavior of $L(0,\widehat{\gamma})$ has been analyzed in Section 4. Lemma 3 below gives the limiting distribution of $\overline{\gamma}$, for all c in a compact set.

LEMMA 3: Under Assumption A,

$$n^{-1/2}D_n(\overline{\gamma}-\gamma) \Rightarrow \left[\int X_{\overline{c}}(r)X_{\overline{c}}(r)'dr\right]^{-1}\int X_{\overline{c}}(r)d\overline{S}_c(r),$$

where

$$\overline{S}_c(r) = \frac{\rho}{\delta}W_1(r) - \frac{\sqrt{\omega^2-\rho^2}}{\delta}W_2(r) - (\overline{c}-c)\int_0^r J_c,\ X_{\overline{c}}(r) = g(r) - \overline{c}X(r).$$

By an asymptotic expansion and using the results of Lemmas 2 and 3, the asymptotic distribution of the likelihood ratio statistic can be derived.

THEOREM 6:

For all c in a compact set,

$$\begin{aligned}
&2\left[L(0,\widehat{\gamma}) - L(\overline{c},\overline{\gamma})\right] \\
\Rightarrow\ &\delta\int\left[(\overline{c}-c)\,J_c(r) + X_{\overline{c}}(r)'\left[\int X_{\overline{c}}(r)X_{\overline{c}}(r)'dr\right]^{-1}\int X_{\overline{c}}(r)d\overline{S}_c(r)\right]^2 dr \\
&-2\delta\int\left[(\overline{c}-c)\,J_c(r) + X_{\overline{c}}(r)\left[\int X_{\overline{c}}(r)X_{\overline{c}}(r)'dr\right]^{-1}\int X_{\overline{c}}(r)d\overline{S}_c(r)\right]dS_0(r) \\
&+\delta\int d\underline{S}_c(r)\underline{X}(r)'\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)dS_c(r) - c^2\delta\int J_c(r)^2 dr
\end{aligned}$$

$$-2c\delta \int J_c(r) dS_0(r) + 2c\delta \int J_c(r) \underline{X}(r)' dr \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_c(r).$$

Under the null of $c = 0$ and the point alternative $c = \overline{c}$, we obtain the limiting null distribution and the power envelope as a corollary of Theorem 6.

Corollary 3:

(1) Under $c = 0$,

$$\begin{aligned} & 2\left[L(0,\widehat{\gamma}) - L(\overline{c},\overline{\gamma})\right] \\ \Rightarrow\ & \delta \int \left[\overline{c} W_1(r) + X_{\overline{c}}(r)' \left[\int X_{\overline{c}}(r) X_{\overline{c}}(r)' dr \right]^{-1} \int X_{\overline{c}}(r) d\overline{S}_0(r) \right]^2 dr \\ & -2\delta \int \left[\overline{c} W_1(r) + X_{\overline{c}}(r)' \left[\int X_{\overline{c}}(r) X_{\overline{c}}(r)' dr \right]^{-1} \int X_{\overline{c}}(r) d\overline{S}_0(r) \right] dS_0(r) \\ & +\delta \int dS_0(r) \underline{X}(r)' \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_0(r). \end{aligned}$$

(2) Under $c = \overline{c}$

$$\begin{aligned} & 2\left[L(0,\widehat{\gamma}) - L(\overline{c},\overline{\gamma})\right] \\ \Rightarrow\ & -\delta \int dS_0(r) X_{\overline{c}}(r)' \left[\int X_{\overline{c}}(r) X_{\overline{c}}(r)' dr \right]^{-1} \int X_{\overline{c}}(r) dS_0(r) \\ & +\delta \int d\underline{S}_{\overline{c}}(r) \underline{X}(r)' \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_{\overline{c}}(r) - \overline{c}^2 \delta \int J_{\overline{c}}(r)^2 dr \\ & -2\overline{c}\delta \int J_{\overline{c}}(r) dS_0(r) + 2\overline{c}\delta \int J_{\overline{c}}(r) \underline{X}(r)' dr \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_{\overline{c}}(r). \end{aligned} \tag{5.1}$$

In the special case that φ is log normal, the likelihood ratio test rejects H_0 for small values of

$$LR_c = \frac{\sum_{t=2}^{n} \left[\Delta y_t - \overline{c}(y_{t-1}/n) - \overline{\gamma}' \Delta x_t + \overline{c}\overline{\gamma}'(x_{t-1}/n) \right]^2}{\sum_{t=2}^{n} \left[\Delta y_t - \widehat{\gamma}' \Delta x_t \right]^2}. \tag{5.2}$$

The limiting distributions of (5.2) can be derived as a corollary of Theorem 6. In particular, when x_t is a linear trend, the results reduce to the distribution given by Elliott, Rothenberg, and Stock (1996). They also derived the power envelope based on distribution (5.1), when φ is log normal and x_t is a linear trend. Obviously, in the case of non-Gaussian innovations, this distribution depends on the parameter ω^2. Monte Carlo evidence indicates that substantial power increase occurs as the parameter ω^2 increases. This suggests a potential efficiency gain from using the distributional information in the unit root tests.

6. QD M-Detrended Unit Root Tests

The M estimation can be coupled with quasi-differencing (QD) to construct a nearly efficient M-detrended unit root test. Again, like the unit root tests proposed in the previous sections, the efficient M-detrended unit root tests can be constructed based on either the likelihood ratio statistic or the estimated c directly. We analyze the test directly based on the estimated c. Other tests can be analyzed in a similar way.

For some appropriate choice $\overline{c}$, we calculate the maximum likelihood estimators for γ under the hypothesis $c = \overline{c}$,

$$\overline{\gamma} = \arg\max \sum_{t=1}^{n} \varphi\left(\Delta_{\overline{c}} y_t - \gamma' \Delta_{\overline{c}} x_t\right), \tag{6.1}$$

and construct the detrended y_t based on $\overline{\gamma}$,

$$\overline{y}_t^s = y_t - \overline{\gamma}' x_t.$$

Re-estimating c based on the M estimator of the autoregressive coefficient of $\overline{y}_t^s$, i.e.,

$$\check{c} = \arg\max \sum_{t=1}^{n} \varphi\left(\Delta y_t - c(y_{t-1}/n) - \overline{\gamma}' \Delta x_t + c\overline{\gamma}'(x_{t-1}/n)\right),$$

the efficient M-detrended unit root test can then be constructed based on $\overline{Z}_M = \breve{c}$.

Notice that the partial sum process based on $\overline{y}_t^s$ has the following asymptotic behavior:

$$\frac{1}{\sqrt{n}}\overline{y}_{[nr]}^s \Rightarrow J_c(r) - X(r)' \left[\int X_{\overline{c}} X_{\overline{c}}'\right]^{-1} \int X_{\overline{c}} d\overline{S}_c := \overline{J}_{\overline{c}}(r).$$

We can derive the limiting null distribution and the power function of $\overline{Z}_M = \breve{c}$.

Theorem 7:

(1) Under $c = 0$,

$$\breve{c} \Rightarrow \left[\int \underline{W_{\overline{c}}}^2\right]^{-1} \left[\int \underline{W_{\overline{c}}} dS_0 - \int \underline{W_{\overline{c}}} X_c' \left[\int X_{\overline{c}} X_{\overline{c}}'\right]^{-1} \int X_{\overline{c}} d\overline{S}_0\right]. \tag{6.2}$$

(2) For all c in a compact set,

$$\breve{c} \Rightarrow c + \left[\int \overline{J}_{\overline{c}}^2\right]^{-1} \left[\int \overline{J}_{\overline{c}} dS_0 - \int \overline{J}_{\overline{c}} X_c' \left[\int X_{\overline{c}} X_{\overline{c}}'\right]^{-1} \int X_{\overline{c}} d\overline{S}_0\right].$$

Remark 10: In the case that $\varphi(u) = u^2$, it can be easily verified that under the null

$$\breve{c} \Rightarrow \left[\int \underline{W_{\overline{c}}}^2\right]^{-1} \int \underline{W_{\overline{c}}} d\underline{W_{\overline{c}}},$$

which is exactly the limit distribution of the quasi-differencing detrended Phillips Z_α test (Phillips and Xiao 1998). If we choose x_t to be a constant term or a linear trend, we obtain the limiting result of Elliott, Rothenberg, and Stock (1996).

Remark 11: For time series with general serially correlated residuals, a nonparametrically modified estimator, say $\widehat{c}^+$, can be used and the same limiting results follow.

7. Monte Carlo Results

We conducted a simulation experiment to examine the sampling performance of the M estimators and the testing procedures based on them. In particular, we compared the finite sample performance of different estimators of the deterministic trend coefficient, and compared the power properties of unit root tests based on different detrending procedures. The model used for data generation was the following:

$$(DGP)\begin{cases} y_t = \gamma' x_t + y_t^s, \\ y_t^s = \alpha y_{t-1}^s + u_t, \; t = 1, ..., n, \end{cases}$$

where the true value of γ is 0, and $\{u_t\}$ is an iid sequence of t-distributions with three degrees of freedom. We standardized u_t so that it has unity variance. $y_0^s = 0$. Two sample sizes were considered: $n = 100$, $n = 200$.

We first examined the estimation of deterministic trends, i.e. γ. We considerd the leading case of a linear time trend, i.e., $x_t = (1, t)$. Notice that since the intercept term is not consistently estimable, we focused our attention on the estimation of the coefficient of t. We compared the following estimators of the deterministic trend coefficient:

(1) Ordinary least squares estimator of the trend coefficient, denoted as OLS.

(2) Least squares estimator of the trend coefficient based on the quasi-differenced data, denoted as QDLS.

(3) M estimator of the trend coefficient based on the quasi-differenced data, i.e. $\overline{\gamma}$ in (6.1), denoted as QDM.

(4) MLE estimator of the trend coefficient based on the nonlinear regression, i.e. $\widetilde{\gamma}$ in (3.1), denoted as NLM.

We considered different data sets generated by (DGP) with $\alpha = 1, 0.95, 0.9, 0.85,$ and 0.8, and sample sizes of 100 and 200. The prespecified $\bar{c}$ in quasi-differencing is -10 (other choices of prespecified $\bar{c}$ were also tried). Table A reports the estimation bias for the four estimators in different cases, and Table B reports the variances of these estimators. Notice that these estimators are unbiased and the mean squared errors are dominated by the variances. We also depicted the simulation densities of these estimators corresponding to different data sets. The information about these graphics is given below:

(1) Figure 1: simulation densities of these estimators when the sample size is 100 and the true $c = 0$.

(2) Figure 2: simulation densities when $n = 100$ and the true c is -10.

(3) Figure 3: simulation densities when $n = 200$ and the true c is 0.

(4) Figure 4: simulation densities when $n = 200$ and the true c is -10.

Some general conclusions can be drawn from these Monte Carlo result. The estimators QDM and NLM, which used the distributional information, have much better sampling performance than OLS and QDLS. Comparing QDM and NLM, we can see that NLM is effectively using an estimated c but QDM uses a prespecified c. The Monte Carlo results indicate that the estimator $\tilde{\gamma}$ using an estimated c in quasi-differencing has in general pretty good sampling performance and has avoided the additional issue of choosing a prespecified $\bar{c}$. As we can anticipate, when the true c value is close to the prespecified $\bar{c}$, QDM has slightly better performance. In other cases, NLM gives relatively better results. The evidence is very clear in the four figures.

Our next purpose is to compare the unit root test based on different detrending process. Again, we considered the above four methods of estimating deterministic trends in the construction of the unit root tests. In particular, for purpose of comparison, we compared unit root tests based on the estimated c coupled with the four different detrending procedures, i.e.,

Table A: Estimation of the Deterministic Trend (Bias)

		OLS	QDLS	QDM	NLM
$n = 100$	$\alpha = 1$	-0.0010215	-0.0006520	-0.00100501	-0.0010499
	$\alpha = 0.95$	0.0003870	0.0003634	0.00010376	0.0000627
	$\alpha = 0.90$	0.0004275	0.0003949	-0.00000908	-0.0000471
	$\alpha = 0.85$	0.0003171	0.0002694	-0.00006837	-0.0000662
	$\alpha = 0.80$	0.0002477	0.0001965	-0.00007931	-0.0000246
$n = 200$	$\alpha = 1$	0.0004712	0.0004742	-0.0015339	-0.0007481
	$\alpha = 0.95$	0.0001856	0.0001727	0.0000656	-0.000102
	$\alpha = 0.90$	0.0000221	0.0000161	0.0000122	0.00000251
	$\alpha = 0.85$	0.0000179	0.0000153	-0.0000459	-0.0000664
	$\alpha = 0.80$	0.0001281	0.0000954	-0.0000559	-0.0000325

Table B: Estimation of the Deterministic Trend (Variance)

		OLS	QDLS	QDM	NLM
$n = 100$	$\alpha = 1$	0.1158501	0.1104706	0.1056588	0.0090731
	$\alpha = 0.95$	0.0295969	0.0271542	0.0202029	0.0189686
	$\alpha = 0.90$	0.0159903	0.0141669	0.0100533	0.0106364
	$\alpha = 0.85$	0.0108151	0.0099818	0.0066257	0.0070036
	$\alpha = 0.80$	0.0081582	0.0078926	0.0054088	0.0052918
$n = 200$	$\alpha = 1$	0.0763741	0.0751235	0.0731166	0.0618703
	$\alpha = 0.95$	0.0145821	0.0133685	0.0096584	0.0092456
	$\alpha = 0.90$	0.0057442	0.0055467	0.0036155	0.0038017
	$\alpha = 0.85$	0.0038246	0.0035625	0.0021288	0.0023156
	$\alpha = 0.80$	0.0024754	0.0022922	0.0017279	0.0016296

Table C: Size-Adjusted Empirical Power (Demeaned Case)

	OLS demeaned	QD demeaned	QDM demeaned	NL demeaned
$\alpha = 0.975$	0.0976	0.1002	0.1260	0.1266
$\alpha = 0.95$	0.1792	0.1822	0.2544	0.2540
$\alpha = 0.925$	0.2982	0.3076	0.4682	0.4660
$\alpha = 0.90$	0.4546	0.4688	0.7086	0.7082
$\alpha = 0.875$	0.6278	0.6416	0.8896	0.8898
$\alpha = 0.85$	0.7874	0.7962	0.9672	0.9666
$\alpha = 0.825$	0.8968	0.9016	0.9908	0.9912
$\alpha = 0.80$	0.9578	0.9606	0.9980	0.9982

Table D: Size-Adjusted Empirical Power (Detrended Case)

	OLS detrended	QD detrended	QDM detrended	NL detrended
$\alpha = 0.975$	0.0624	0.0638	0.0786	0.0778
$\alpha = 0.95$	0.1048	0.1068	0.1592	0.1584
$\alpha = 0.925$	0.1936	0.1950	0.3166	0.3160
$\alpha = 0.90$	0.3192	0.3222	0.5434	0.5406
$\alpha = 0.875$	0.4804	0.4822	0.7684	0.7690
$\alpha = 0.85$	0.6474	0.6522	0.9134	0.9126
$\alpha = 0.825$	0.7974	0.8024	0.9736	0.9726
$\alpha = 0.80$	0.8982	0.9006	0.9926	0.9918

(1) The Dickey-Fuller test based on OLS detrending, denoted by OLS.

(2) The QD detrended DF test based on least square regression on the quasi-differenced data, denoted by QD.

(3) The $\overline{Z}_M$ (6.2) test based on M estimation plus quasi-differencing, denoted by QDM.

(4) The Z_M (4.3) test based on the nonlinear M-estimation, denoted by NL.

To provide a power comparison among the different tests, size-corrected power is reported. Both the demeaned test and the detrended test are examined. In particular, Table C reports the empirical power of the demeaned tests and Table D reports the power of the tests when a linear time trend is removed. Figure 5 depicts the power functions of the demeaned tests and Figure 6 depicts those of the detrended tests. These Monte Carlo results indicate that the testing procedures using distributional information have substantially improved power properties.

8. Conclusion

We studied likelihood-based estimation and tests in an autoregressive time series model with a near unit root and an unknown deterministic trend. In particular, we extended the nonlinear least squares regression in Phillips and Xiao (1998) to the general case with innovations that may be non-normal. Asymptotic analysis on M estimators and related testing procedures were studied. In practice, even if the exact distribution of the innovations is unknown, if the data has similar tail behavior as the density function used in the estimation, inference based on these methods should have good sampling properties. Adaptive estimation method (e.g. Seo 1996; and Beelders 1998) can be used to provide feasible procedures on estimation and tests in our model.

9. Proofs

Proof of Theorem 2

By an asymptotic expansion, it can be shown that, under H_0 and Assumption A, the likelihood ratio statistic

$$\begin{aligned} & L(0,\widehat{\gamma}) - L(\widetilde{c},\widetilde{\gamma}) \\ = & \sum_t \varphi(\Delta y_t - \widehat{\beta}'\breve{x}_t) - \sum_t \varphi(\Delta y_t - \widetilde{c}(y_{t-1}/n) - \widetilde{\gamma}'\Delta x_t + \widetilde{c}\gamma'(x_{t-1}/n)) \end{aligned}$$

has the following approximation:

$$\frac{\widetilde{c}}{n}\sum_{t=1}^{n}\varphi'(u_t)\left(y_{t-1}-\widetilde{\gamma}'x_{t-1}\right)-\frac{\widetilde{c}}{n}\sum_{t=1}^{n}\varphi''(u_t)\left(y_{t-1}-\widetilde{\gamma}'x_{t-1}\right)(\widetilde{\gamma}-\gamma)'\Delta x_t$$
$$-\frac{\widetilde{c}^2}{2n^2}\sum_{t=1}^{n}\varphi''(u_t)\left(y_{t-1}-\widetilde{\gamma}'x_{t-1}\right)^2+\sum_{t=1}^{n}\varphi'(u_t)(\widetilde{\gamma}-\gamma)'\Delta x_{t-1}$$
$$-\sum_{t=1}^{n}\varphi'(u_t)(\widehat{\beta}-\beta)'\breve{x}_t-\frac{1}{2}\sum_{t=1}^{n}\varphi''(u_t)\left[(\widetilde{\gamma}-\gamma)'\Delta x_t\right]^2$$
$$+\frac{1}{2}\sum_{t=1}^{n}\varphi''(u_t)\left[(\widehat{\beta}-\beta)'\breve{x}_t\right]^2+o_p(1). \tag{9.1}$$

By a calculation of limits of the components in (9.1), it can be derived that

$$2(L(0,\widehat{\gamma})-L(\widetilde{c},\widetilde{\gamma}))$$

converges weakly to

$$2\eta_0\int \underline{W}_{\xi_0}(r)dB_\varphi(r)+2\delta\eta_0\xi_0'\int g(r)\underline{W}_{\xi_0}(r)dr-\delta\eta_0^2\int \underline{W}_{\xi_0}(r)^2dr$$
$$+2\xi_0'\int g(r)dB_\varphi(r)-\delta\xi_0'\int g(r)g(r)'dr\xi_0$$
$$-2\int dB_\varphi(r)\underline{X}(r)'\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)dS_0(r)$$
$$-\delta\int dS_0(r)\underline{X}(r)'\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)dS_0(r),$$

and the above summation equals

$$\delta\int\left[\eta_0\underline{W}_{\xi_0}(r)+\xi_0'g(r)\right]^2dr-2\delta\int\left[\eta_0\underline{W}_{\xi_0}(r)+\xi_0'g(r)\right]dS_0(r)$$
$$+\delta\int dS_0(r)\underline{X}(r)'\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)dS_0(r).$$

■

PROOF OF THEOREM 3

Notice that under the local alternative hypothesis, $L(0,\widehat{\gamma})-L(\widetilde{c},\widetilde{\gamma})$ has the following expansion

$$\frac{(\widetilde{c}-c)}{n}\sum_{t=1}^{n}\varphi'(u_t)\left(y_{t-1}-\widetilde{\gamma}'x_{t-1}\right)-\frac{(\widetilde{c}-c)}{n}\sum_{t=1}^{n}\varphi''(u_t)\left(y_{t-1}-\widetilde{\gamma}'x_{t-1}\right)(\widetilde{\gamma}-\gamma)'\Delta_c x_t$$

$$-\frac{(\widetilde{c}-c)^2}{2n^2}\sum_{t=1}^{n}\varphi''(u_t)\left(y_{t-1}-\widetilde{\gamma}'x_{t-1}\right)^2+\sum_{t=1}^{n}\varphi'(u_t)(\widetilde{\gamma}-\gamma)'\Delta_c x_{t-1}$$
$$+\sum_{t=1}^{n}\varphi'(u_t)\frac{c}{n}y^s_{t-1}-\sum_{t=1}^{n}\varphi'(u_t)(\widehat{\beta}-\beta)'\breve{x}_t-\frac{1}{2}\sum_{t=1}^{n}\varphi''(u_t)\left[(\widetilde{\gamma}-\gamma)'\Delta_c x_t\right]^2$$
$$-\frac{c}{n}\sum_{t=1}^{n}\varphi''(u_t)y^s_{t-1}(\widehat{\beta}-\beta)'\breve{x}_t+\frac{1}{2}\sum_{t=1}^{n}\varphi''(u_t)\left[(\widehat{\beta}-\beta)'\breve{x}_t\right]^2+\frac{1}{2}\sum_{t=1}^{n}\varphi''(u_t)\left(\frac{c}{n}y^s_{t-1}\right)^2$$
$$+o_p(1).$$

The proof follows similar steps to that of Theorem 2. ■

Proof of Theorem 6

Notice that

$$\begin{aligned}L(\overline{c},\overline{\gamma}) &= \sum\varphi\left(\Delta_{\overline{c}}y_t-\overline{\gamma}'\Delta_{\overline{c}}x_t\right)\\ &= \sum\varphi\left(u_t-(\overline{c}-c)(y^s_{t-1}/n)-(\overline{\gamma}-\gamma)'\Delta_{\overline{c}}x_t\right),\end{aligned}$$

and the following asymptotics can be verified

$$\begin{aligned}\sum_{t=1}^{n}\varphi'(u_t)\frac{c}{n}y^s_{t-1} &\Rightarrow c\int J_c dB_\varphi,\\ \sum_{t=1}^{n}\varphi'(u_t)(\widehat{\beta}-\beta)'\breve{x}_t &\Rightarrow \int dS_c\underline{X}'\left[\int\underline{X}\underline{X}'\right]^{-1}\int\underline{X}dB_\varphi,\\ \frac{c}{n}\sum_{t=1}^{n}\varphi''(u_t)y^s_{t-1}(\widehat{\beta}-\beta)'\breve{x}_t &\Rightarrow -c\delta\int J_c(r)\underline{X}(r)'dr\left[\int\underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int\underline{X}(r)dS_c(r),\\ \sum_{t=1}^{n}\varphi''(u_t)\left[(\widehat{\beta}-\beta)'\breve{x}_t\right]^2 &\Rightarrow -\delta\int dS_c(r)\underline{X}(r)'\left[\int\underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int\underline{X}(r)dS_c(r),\\ \sum_{t=1}^{n}\varphi''(u_t)\left(\frac{c}{n}y^s_{t-1}\right)^2 &\Rightarrow -c^2\delta\int J_c(r)^2dr,\end{aligned}$$

$$\begin{aligned}&\sum_{t=1}^{n}\varphi'(u_t)\left(\frac{(\overline{c}-c)y^s_{t-1}}{n}+(\overline{\gamma}-\gamma)'\Delta_{\overline{c}}x_t\right)\\ \Rightarrow &\int\left[(\overline{c}-c)J_c+X_{\overline{c}}\left[\int X_{\overline{c}}X_{\overline{c}}'\right]^{-1}\int X_{\overline{c}}d\overline{S}_c\right]dB_\varphi,\end{aligned}$$

$$\sum_{t=1}^{n} \varphi''(u_t) \left(\frac{(\overline{c} - c) y^s_{t-1}}{n} + (\overline{\gamma} - \gamma)' \Delta_{\overline{c}} x_t \right)^2$$

$$\Rightarrow \quad -\delta \int \left[(\overline{c} - c) \, J_c + X_{\overline{c}}' \left[\int X_{\overline{c}} X_{\overline{c}}' \right]^{-1} \int X_{\overline{c}} d\overline{S}_c \right]^2 .$$

The result of Theorem 6 can be obtained by an asymptotic expansion. ■

10. References

Beelders, O., 1998, "Adaptive Estimation and Unit Root Tests," Working Paper, Emory University.

Canjels, N. And M. Watson, 1997, "Estimating deterministic trends in the presence of serially correlated errors," *Review of Economics and Statistics, 79,* 184-200.

Chan, N.H. and C.Z Wei, 1987, "Asymptotic inference for nearly nonstationary AR(1) processes," *Annals of Statistics*, **15**, 1050–1063. *Forecasting and Modeling in Statistics and Econometrics.* Springer–Verlag, 1996.

Cox, D. and Llatas, I., 1991, "Maximum likelihood type estimation for nearly non-stationary autoregressive time series," Annals of Statistics 19, 1109-1128.

Dickey, D.A. and W.A. Fuller, 1979, "Distribution of estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, **74**, 427–431.

Dufour, J. and M. L. King, 1991, "Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary and nonstationary AR(1) errors," *Journal of Econometrics*, **47**, 115–143.

Elliott, G., T. J. Rothenberg and J. H. Stock, 1996, "Efficient tests for an autoregressive unit root," *Econometrca*, **64**, 813–836.

Grenander, U. and M. Rosenblatt, 1957, *Statistical Analysis of Stationary Time Series.* New York: John Wiley.

Lucas, A., 1994, "An Outlier Robust Unit Root Test with an Application to the extended Nelson-Plosser Data," *Journal of Econometrics* 66, 153-174.

Lucas, A., 1995, "Unit Root tests based on M estimators," *Econometric Theory* 11, 331-346.

King M. L., 1988, "Towards a theory of point optimal testing," *Econometric Reviews*, **6**, 169–218.

Phillips, P.C.B., 1987a, "Time series regression with a unit root," *Econometrica*, **5**5, 277–302.

Phillips, P.C.B., 1987b, "Towards a unified asymptotic theory of autoregression." *Biometrika*, **74**, 535–547.

Phillips, P.C.B., 1988, "Regression Theory for Near-Integrated Time Series," *Econometrica*, **5**6, 1021-1044.

Phillips, P.C.B., 1996, Robust Nonstationary Regression, *Econometric Theory* 12.

Phillips, P.C.B., 1998, "The invalidity of critical values of tests for unit roots against trend alternatives," in progress.

Phillips, P.C.B. And C.C. Lee, 1996, "Efficiency gains from quasi-differencing under nonstationarity." In P.M. Robinson and M. Rosenblatt (eds.), *Athens*

Conference on Applied Probability and Time Series: Essays in Memory of E.J. Hannan, Springer–Verlag: New York.

Phillips, P.C.B. and P. Perron, 1988, "Testing for unit roots in time series regression," *Biometrika*, **75**, 335–346.

Phillips, P.C.B. and V. Solo, 1992, "Asymptotics for linear processes," *Annals of Statistics*, **20**, 971–1001.

Phillips, P.C.B., and Z. Xiao, 1998, "A Primer on Unit Root Testing", Journal of Economic Survey, Vol. 12, No. 5.

Rothenberg, T., and J. Stock, 1997, "Inference in a nearly integrated autoregressive model with nonnormal innovations", Jornal of Econometrics 80, 269-286.

Sargan, J. D. and A. Bhargava, 1983, "Testing residuals from least squares regression for being generated by the Gaussian random walk," *Econometrica*, **51**, 153–174.

Schmidt, P. and P.C.B. Phillips, 1992, "Testing for a unit root in the presence of deterministic trends," *Oxford Bulletin of Economics and Statistics*, **54**, 257–288.

Seo, B. "Essays on Time Series Econometrics," Ph.D. Dissertation, University of Rochester.

Stock, J.H., 1995, "Unit roots, structural breaks and trends." In R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics,* Vol. 4, 2739-2841, Amsterdam: North–Holland.

Xiao, Z., and P.C.B. Phillips, 1999, "Efficient Detrending in Cointegrating Regressions", *Econometric Theory,* Vol.15, 519-548.