April 23, 2010

1 Introduction

In the past decade, econometricians have focused a great deal of attention on the development of estimation and testing procedures in autoregressive time series models where the largest root is unity. Most of these procedures are based on least square methods and have likelihood interpretations when the data are Gaussian. In the absence of Gaussianity, these methods are less efficient than methods that exploit the distributional information. Indeed, Monte Carlo evidence indicates that the least square estimator can be very sensitive to certain type of outliers and that inference procedures based on least squares estimation may have poor performance.

Many applications in nonstationary economic time series involve data that are affected by infrequent but important events such as oil shocks, wars, natural disasters, and changes in policy regimes, indicating the presence of nonGaussian behavior in macroeconomic time series (see Balke and Fomby, 1994). It is well-documented that financial time series such as interest rate and exchange rate have heavy-tailed distributions. In such cases, it is important to consider estimation and inference procedures that are robust to departures from Gaussianity and can be applied to nonstationary time series.

For this reason, in the recent 10 years, researchers have devoted a lot of effort in the development of more efficient and robust inference procedures in nonstationary time series. One way to achieve asymptotic efficiency and robustness is the use of adaptive estimation based on nonparametric technique, see, e.g. Seo (1996) and Beelders (1998). Under appropriate regularity assumptions, tests based on adaptive estimation using nonparametric kernel methods can be constructed, although these procedures may be complicated and thus practically difficult to use.

An alternative approach is the use of M estimation. A partial list along this direction includes: Cox and Llatas (1991), Knight (1991), Phillips (1995), Lucas (1995a), Lucas (1995b), Rothenberg and Stock (1997), Juhl (1999), Xiao (2001), Koenker and Xiao (2003) and Thompson (2004) among others. In particular, Phillips (1995) studies robust cointegration regressions. Cox and Liatas (1991), Lucas (1995a and 1995b), Rothenberg and Stock (1997), Xiao (2001) and Thompson (2004) studied M-estimation and likelihood-based inference for various models of unit root (or local unit root) time series. In the aforementioned studies, the criterion functions in M-estimation are assumed to be known and the associated inferences are generally efficient only when the true likelihood functions are used.

In computing critical values, Lucas (1995b) simulates the data by generating a time series from the data generating process with i.i.d. standard normal innovations. More recently, Thompson (2004) showed that the Lucas' approach is asymptotically incorrect unless the errors are in fact normal. Based on this finding, Thompson (2004) proposed a new method to compute critical values based on a polynomial approximation which leads to correct critical values, but one still needs to specify the error distribution. In practice, the error distributions are unknown and, therefore, it is important to use a criterion function (or a density function) that has similar characteristic to the data distribution. In this paper, we compute critical values using the same polynomial approximation as in Thompson (2004), but instead of sticking on a particular criterion function, we consider a class of density functions that captures a wide range of distributions used in economic applications, and select a density function using a data-dependent approach and propose a unit root test based on this (partially) adaptive estimation.

The present paper is somehow related to the work of Shin and So (1999) where a fully adaptive (nonparametric) method is used to identify the criterion function. Although fully adaptive estimators has the theoretically attractive property of asymptotic efficiency, as suggested by Bickel (1982, p.664), Shin and So (1999) recognize that "the adaptive method is not necessarily easy to implement" (Shin and So, 1999, page 8). Moreover, the model studied in Shin and So (1999) assumes that there is no serial correlation in the error term, which turns out to be a very restrictive assumption. On the other hand, partially adaptive estimation is a more practical goal because it avoids the difficulty of nonparametric estimation of score functions. (also see similar arguments in Potscher and Prucha (1986), Hogg and Lenth (1984), McDonald and Newey (1988), and Phillips (1994)).

Thus, this paper tries to provide an intermediate class of unit root testing procedures that are more efficient than the traditional OLS-based methods in the presence of heavy-tailed distributions without the need of specifying a density function and, on the other hand, simpler than unit root tests based on fully adaptive estimation using nonparametric methods. In particular, we propose a test based on partially adaptive estimation of the augmented Dickey-Fuller (ADF) model. A data-dependent procedure is used to select an appropriate criterion function for estimation and computation of critical values.

Giving the well documented characteristic of heavy-tail behavior in economic and financial data, we consider a partially adaptive estimator based on the family of student-t distributions (Postcher and Prucha 1986). The family of student-t represents an important dimension of the space of distributions, including the normal distribution as a limiting case and the Cauchy distribution as a special case. Its adaptation parameter will depend on the scale and thickness parameters, which can be easily estimated from the data using the approach proposed by Potscher and Prucha (1986). There is no doubt that other classes of distributions can also be analyzed similarly. For example, one may consider the Pearson type IV distribution [Pearson (1895)], which also incorporates excess kurtosis in a simple way. (See, e.g. Premaratne and Bera (2005) for studies on the Pearson type IV distributions).

Monte-Carlo experiments are conducted to investigate the finite-sample performance of the partially adaptive test. Comparing to the conventional ADF test and some robust tests, the Monte Carlo results indicate that there is little loss in using the proposed unit root test when the innovations are Gaussian, and the power gains from using our partially adaptive test is substantial when there are non-Gaussian innovations.

We apply the proposed test to several important macroeconomic time series that exhibit non-Gaussian features. In particular, we re-examined the unit root property of nominal interest rate, real exchange rate, and real GDP. Traditional OLS-based tests, such as the ADF test, cannot reject the unit root hypothesis in these series. On the other hand, non-Gaussian behavior in interest rate, real exchange rate, and real GNP has been largely reported in the literature as being caused by asymmetric innovations or presence of outliers (e.g., Falk and Wang, 2003; Blanchard and Watson, 1986; Bidarkota, 2000; Balke and Fomby, 1994; and Scheinkman and LeBaron, 1989). A descriptive analysis of our data also confirms that U.S. nominal interest rate, real GNP, and real exchange rate are featured with nonGaussian characteristics. When we apply the partially adaptive test to these series, we rejected the unit root hypothesis in real GNP, supporting the literature of transitory fluctuations about trend. We were unable to reject the null of unit root in real exchange rates, implying that, as reported in Falk and Wang (2003), the purchasing power parity hypothesis may not hold in the long run even if tail heaviness are accounted for. We also found no evidence against unit root in nominal interest rate, which supports the findings in Rose (1988) and raises doubt about economic results predicted by the CCAPM and optimal monetary policy models.

The outline of the paper is as follows. Section 2 gives some important preliminaries. In particular, we study an ADF-type test for a unit root based on M estimation. Limiting distributions of the estimator and it t-statistic are derived. The partially adaptive unit root test is introduced in Section 3. Section 4 presents the results of our Monte Carlo simulations. In section 5, we discuss the relevance of the test and conduct an empirical study. Section 6 concludes. Proofs are provided in the Appendix. For notation, we use $\Rightarrow$ to signify weak convergence, L for lag operator, $\equiv$ for equality in distribution, $:=$ for definitional equality, and $[nr]$ to signify the integer part of nr.

2 The Model, Assumptions, and Preliminary Limit Theory

The subject of this paper is a time series y_t represented by the following model:

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

where α is the largest autoregressive root, and the residual term u_t is serially correlated. In the above model, the autoregressive coefficient α plays an important role in measuring persistency in economic and financial time series. Under regularity conditions, if $\alpha = 1$, y_t contains a unit root and is persistent; and if $|\alpha_1| < 1$, y_t is stationary. The high persistency in many economic and financial time series suggests that the coefficient α is near unity.

Following Dickey and Fuller (1979), we parameterize u_t as a stationary AR(k) processs

$$A(L)u_t = \varepsilon_t, \tag{2}$$

where $A(L) = \sum_{i=0}^{k} a_i L^i$ is a k-th order polynomial of the lag operator L, $a_0 = 1$, and ε_t is an iid sequence. Combining (1) and (2), we obtain the well-known Augmented Dickey-Fuller (ADF) regression model

$$\Delta y_t = \rho y_{t-1} + \sum_{j=1}^{k} \psi_j \Delta y_{t-j} + \varepsilon_t. \tag{3}$$

In the presence of a unit root ($\alpha = 1$), $\rho = 0$ in the ADF regression (3).

More generally, we may include a deterministic trend component in the ADF regression, and study the estimation in the following regression

$$\Delta y_t = \gamma' x_t + \rho y_{t-1} + \sum_{j=1}^{k} \psi_j \Delta y_{t-j} + \varepsilon_t. \tag{4}$$

where x_t is a deterministic component of known form and γ is a vector of unknown parameters. The leading cases of the deterministic component are (i) a constant term $x_t = 1$, and (ii) a linear time trend $x_t = (1, t)'$.

We want to test the unit root hypothesis ($\rho = 0$, or $\alpha = 1$) based on estimators of ρ (or α). In the simple case where ε_t is normally distributed, given observations on y_t, the maximum likelihood estimators of ρ (or α) and $\{\psi_j\}_{j=1}^{k}$ are simply the least squares estimators obtained by minimizing the residual sum of squares. In the absence of Gaussianity in ε_t, it is possible to follow the idea of Huber (1964) for the location problem in order to obtain more robust estimators. In this direction, Relles (1968), Huber (1973) introduced a class of M estimators which generally have good properties over a wide range of distributions. The M-estimators are obtained from solving the extreme problem by replacing the quadratic criterion function in OLS estimation with some general criterion function φ. In the case that φ is the true log density function of the residuals, the M-estimator is the maximum likelihood estimator.

To introduce the proposed unit root test based on partially adaptive estimation, we first consider M estimation of the ADF model in this section. In section 3, we study the problem of selecting the criterion function adaptively. The M-estimator for $(\gamma, \rho, \{\psi_j\}_{j=1}^k)$ is defined as the solution of the following extreme problem:

$$(\widehat{\gamma}, \widehat{\rho}, \left\{\widehat{\psi}_j\right\}_{j=1}^k) = \arg\max Q(\gamma, \rho, \{\psi_j\}_{j=1}^k) \tag{5}$$

where

$$Q(\gamma, \rho, \{\psi_j\}_{j=1}^k) = \sum_{t=k+1}^{n} \varphi\left(\Delta y_t - \gamma' x_t - \rho y_{t-1} - \sum_{j=1}^{k} \psi_j \Delta y_{t-j}\right)$$

for some criterion function φ. When φ is the true log density function of ε, $Q(\gamma, \alpha, \{\psi_j\}_{j=1}^k)$ is the log likelihood function and the estimator $(\widehat{\gamma}, \widehat{\rho}, \left\{\widehat{\psi}_j\right\}_{j=1}^k)$ given by (5) is the maximum likelihood estimator.

Lucas (1995b) studied similar regression models. Since those models do not include a deterministic trend in the regression and thus have slightly different limiting distributions, we give the limiting distribution of $(\widehat{\gamma}, \widehat{\rho}, \left\{\widehat{\psi}_j\right\}_{j=1}^k)$ in this section for completeness only.

For convenience of asymptotic analysis, we assume that there is a standardizing matrix F_n such that $F_n^{-1} x_{[nr]} \to X(r)$ as $n \to \infty$, uniformly in $r \in [0,1]$, where $X(r)$ is a vector of limiting trend functions. In the case of a linear trend, $F_n = diag[1, n]$ and $X(r) = (1, r)'$. If x_t is a general p-th order polynomial trend (i.e. $x_t = (1, t, \cdots, t^p)$), $F_n = diag[1, n, ..., n^p]$ and $X(r) = (1, r, ..., r^p)$.

Following the previous literature in M estimation, we make the following assumptions on ε_t and the criterion function φ for the convenience of asymptotic analysis.

Assumption A1 The roots of $A(L)$ all lie outside the unit circle, and $\{\varepsilon_t\}$ are i.i.d. random variables with mean zero and variance $\sigma^2 < \infty$.

Assumption A2 $\varphi(\cdot)$ possesses derivatives φ' and φ''. $[\varepsilon, \varphi'(\varepsilon)]$ has k-th moments for some $k > 2$, $E[\varphi'(\varepsilon_t)] = 0$, $0 < E[\varphi''(\varepsilon_t)] = \mu_\psi < \infty$, and φ'' is Lipschitz continuous.

Assumption A3 $\widetilde{\varepsilon}_t - \varepsilon_t = o_p(1)$ uniformly for all t.

Assumptions A1 - A3 are standard conditions in asymptotic analysis of M estimators. These assumptions are needed to establish the weak convergence results. Assumption $A3$ is a consistency requirement as in Knight (1989,1991) and it is not needed if φ' is the derivative of a convex function with a unique minimum. Assumptions similar to $A3$ are also standard in the development of M estimator asymptotics. It is related to Assumption (b) in Theorem 5.1 of

Phillips (1995), Assumption C in Xiao (2001), and the same as the assumption on $\widetilde{\varepsilon}_t - \varepsilon_t$ in Theorem 1 of Lucas (1995).

We denote $[\cdot]$ as the greatest lesser integer function. Then under Assumptions A1-A3, as n goes to ∞, $n^{-1/2}\sum_1^{[nr]} u_t$ converges weakly to a Brownian motion $B_u(r) = \omega_u W_1(r) = BM(\omega_u^2)$, where $\omega_u^2 = \sigma^2/A(1)^2$ is the long run variance[1] of u_t, denoted as lrvar(u_t). The limiting distributions of $(\widehat{\gamma}, \widehat{\rho}, \left\{\widehat{\psi}_j\right\}_{j=1}^k)$ will also be dependent on the weak limit of the partial sums of $\varphi'(\varepsilon_t)$. Denoting ω_φ^2 =var$[\varphi'(\varepsilon_t)]$, and $\delta = -E[\varphi''(\varepsilon_t)]$, then $n^{-1/2}\sum_1^{[nr]}\varphi'(\varepsilon_t) \Rightarrow B_\varphi(r) = \omega_\varphi W_\varphi(r) = BM(\omega_\varphi^2)$. In fact, under Assumption A1, the partial sums of the vector process $(u_t, \varphi'(\varepsilon_t))$ follow a bivariate invariance principle (see, e.g., Phillips and Durlauf (1986, Theorem 2.1, 474-476, and 486-489); Wooldridge and White (1988, Corollary 4.2); and Hansen 1992):

$$n^{-1/2}\sum_{t=1}^{[nr]}(u_t, \varphi'(\varepsilon_t))^\top \Rightarrow (B_u(r), B_\varphi(r))^\top = BM(\Sigma)$$

where

$$\Sigma = \begin{bmatrix} \omega_u^2 & \sigma_{u\varphi} \\ \sigma_{u\varphi} & \omega_\varphi^2 \end{bmatrix}$$

is the (long-run) covariance matrix of the bivariate Brownian motion.

Denote $(\gamma, \rho)' = \theta$, $(\gamma, \rho, \psi_1, \cdots, \psi_k)' = \Pi$, and $(x_t', y_{t-1}, \Delta y_{t-1}, \cdots, \Delta y_{t-k})' = Z_t$, then we can re-write the regression (4) in matrix form as

$$\Delta y_t = \Pi' Z_t + \varepsilon_t,$$

and the M estimator $\widehat{\Pi}$ maximizes

$$Q(\Pi) = \sum_t \varphi\left(\Delta y_t - \Pi' Z_t\right).$$

Finally we introduce the standardization matrix: $D_n = diag\{\sqrt{n}F_n, n\}$, and $G_n = diag\{D_n, \sqrt{n}I_k\}$, where I_k is a k-dimensional identity matrix, the limiting distributions of the M estimators $(\widehat{\gamma}, \widehat{\rho}, \left\{\widehat{\psi}_j\right\}_{j=1}^k)'$ are given in the following theorem.

Theorem 1 *Given model (1), (2), under Assumptions A1-A3 and the unit root assumption, the limiting distribution of M-estimator $\widehat{\Pi} = (\widehat{\gamma}, \widehat{\rho}, \widehat{\psi}_1, \cdots, \widehat{\psi}_k)'$ is given by*

$$G_n(\widehat{\Pi} - \Pi) \Rightarrow \frac{1}{\delta}\begin{pmatrix} \int \overline{B}_u(r)\overline{B}_u(r)'dr & 0_{2\times k} \\ 0_{k\times 2} & \Gamma \end{pmatrix}^{-1}\begin{pmatrix} \int \overline{B}_u(r)dB_\varphi(r) \\ \Phi \end{pmatrix}.$$

[1] The long run variance of a time series u_t is defined as lrvar$(u_t) = \sum_{h=-\infty}^{\infty}\mathrm{E}(u_t u_{t+h})$. It equals to 2π multiply the spectral density of u_t at zero frequency.

where , $\overline{B}_u(r) = (X(r)', B_u(r))'$, $\Phi = [\Phi_1, \cdots, \Phi_k]^\top$ *is a* k*-dimensional normal variate that is independent of* $\int \overline{B}_u(r)dB_\varphi(r)$, *and*

$$\Gamma = \begin{bmatrix} \gamma_u(0) & \cdots & \gamma_u(k-1) \\ \vdots & \ddots & \\ \gamma_u(k-1) & & \gamma_u(0) \end{bmatrix}$$

where $\gamma_u(h)$ *is the autocovariance function of* u_t.

Theorem 1 indicates that the limiting distribution of the parameters $\widehat{\psi}_1, \cdots, \widehat{\psi}_k$ are independent with the limiting distribution of $\widehat{\gamma}, \widehat{\rho}$. In particular,

$$D_n \begin{bmatrix} \widehat{\gamma} - \gamma \\ \widehat{\rho} \end{bmatrix} \Rightarrow \frac{1}{\delta} \frac{\omega_\varphi}{\omega_u} \left(\int \overline{W}_1(r)\overline{W}_1(r)'dr \right)^{-1} \int \overline{W}_1(r)dW_\varphi(r).$$

The unit root hypothesis corresponds to $\mathrm{H}_0 : \rho = 0$. We consider testing H_0 based on the t-statistic of $\widehat{\rho}$, and estimate the covariance matrix[2] by

$$\widehat{\Omega} = \left[\sum_{t=1}^{n} \varphi'' \left(\widehat{\varepsilon}_t \right) Z_t Z_t' \right]^{-1} \left[\sum_{t=1}^{n} \varphi' \left(\widehat{\varepsilon}_t \right)^2 Z_t Z_t' \right] \left[\sum_{t=1}^{n} \varphi'' \left(\widehat{\varepsilon}_t \right) Z_t Z_t' \right]^{-1}. \quad (6)$$

This is a heteroskedasticity consistent type covariance matrix estimator as in White (1980).

If we consider the t-ratio statistic of $\widehat{\rho}$:

$$t_{\widehat{\rho}} = \frac{\widehat{\rho}}{se(\widehat{\rho})} \quad (7)$$

then $t_{\widehat{\rho}}$ is simply the M regression counterpart of the well-known ADF (t-ratio) test for the unit root hypothesis. The limiting distribution of $t_{\widehat{\rho}}$ is given in the following theorem.

Theorem 2 *Under the assumptions of Theorem 1, the limiting distribution of the* t*-ratio statistic* $t_{\widehat{\rho}}$ *is given by*

$$\left(e' \left[\int \overline{W}_1(r)\overline{W}_1(r)'dr \right] e \right)^{-1/2} e' \int \overline{W}_1(r)dW_\varphi(r)$$

[2]Notice that in the special case when φ is the true log density function of ε, $Q(\Pi)$ is the log likelihood function and the estimator $\widehat{\Pi}$ is the maximum likelihood estimator. In this case, $\omega_\varphi^2 = \delta$, and we may use the following covariance matrix estimator

$$\left[\sum_{t=1}^{n} \varphi' \left(\widehat{\varepsilon}_t \right)^2 Z_t Z_t' \right]^{-1}.$$

where $\overline{W}_1(r) = (X(r)', W_1(r))'$, e is a collecting vector, that is, there is one coordinate equal to one that picks the element corresponding to the asymptotic distribution of $t_{\widehat{\rho}}$, and all the other coordinates equal zero. The above limiting distribution can also be rewritten as

$$\left(\int \underline{W}_1(r)^2 dr \right)^{-1/2} \int \underline{W}_1(r) dW_\varphi(r)$$

$\underline{W}_1(r) = W_1(r) - \int_0^1 W_1(s) X'(s) ds \left(\int_0^1 X(s) X(s)' ds \right)^{-1} X(r)$ *is the Hilbert projection in $L_2[0,1]$ of $W_1(r)$ onto the space orthogonal to X.*

Notice that W_1 and W_φ are correlated Brownian motions, the limiting distribution of $t_{\widehat{\rho}}$ is not standard and depend on nuisance parameters. However, we can decompose $\int B_u(r) dB_\varphi(r)$ (see, e.g. Hansen and Phillips (1990) and Phillips (1995)) as

$$\int B_u dB_{\varphi.u} + \lambda_{\omega\psi} \int B_u dB_u,$$

where $\lambda_{u\varphi} = \sigma_{u\varphi}/\omega_u^2$ and $B_{\varphi.u}$ is a Brownian motion with variance

$$\sigma_{\varphi.u}^2 = \omega_\varphi^2 - \sigma_{u\varphi}^2/\omega_u^2$$

and is independent with B_u. Using the above decomposition, the limiting distribution of the t-statistic $t_{\widehat{\rho}}$ can be decomposed as a simple combination of two independent well-known distributions. In addition, related critical values are tabulated in the literature and thus are ready for us to use in applications. We summarize this result in the following corollary.

Corollary 3 *Under the assumptions of Theorem 1, the limiting distribution of the t-ratio statistic $t_{\widehat{\rho}}$ can be decomposed into a mixture of the Dickey-Fuller (DF) distribution and a standard normal distribution that is independent with the DF distribution, i.e.*

$$t_{\widehat{\rho}} \Rightarrow \sqrt{1-\lambda^2} N(0,1) + \lambda \left(\int \underline{W}_1(r)^2 dr \right)^{-1/2} \int \underline{W}_1 dW_1, \tag{8}$$

where the weights are determined by λ :

$$\lambda^2 = \frac{\sigma_{u\varphi}^2}{\omega_\varphi^2 \omega_u^2}.$$

The standard normal distribution comes from

$$\left(\int \underline{W}_1(r)^2 dr \right)^{-1/2} \int \underline{W}_1(r) dW_{\varphi.1}(r),$$

since $W_1(r)$ and $W_{\varphi.1}(r)$ $(\sigma_{\varphi.u}^{-1} B_{\varphi.u}(r))$ are standard Brownian motions and are independent with each other. Notice that ω_u^2 is the long-run (zero frequency) variance of $\{u_t\}$, ω_φ^2 is the long-run variance of $\{\varphi'(\varepsilon_t)\}$, and $\sigma_{u\varphi}(\tau)$ is the long-run covariance of $\{u_t\}$ and $\{\varphi'(\varepsilon_t)\}$, thus λ is simply the long-run correlation coefficient between $\{u_t\}$ and $\{\varphi'(\varepsilon_t)\}$.

Remark 4 *One interesting case is obtained when* $\lambda^2 = 1$*, which implies that* $\varphi'(u_t) = \varepsilon_t$*. In this simple case, the criterion function is quadratic in* ε_t *and* $t_{\widehat{\rho}}$ *converges to the Dickey-Fuller limiting distribution. Recall that a quadratic criterion function corresponds to the Gaussian log-likelihood. Notice that when* λ^2 *increases from 0 to 1, the corresponding, say, 5% quantile of the limiting variate (8) shifts to the left, indicating that the traditional Dickey-Fuller test will be less powerful than the proposed test in the absence of Gaussianity.*

2.1 Obtaining Critical Values

Given the parameter λ, the limiting distribution of $t_{\widehat{\rho}}$ can be approximated by a direct simulation or using a polynomial approximation. The limiting distribution is the same as that of the covariate-augmented Dickey-Fuller (CADF) test of Hansen (1995). Tables of 1%, 5% and 10% critical values for the statistic $t_n(\tau)$ obtained via simulation are provided in Hansen (1995, page 1155). Thompson (2004) suggested a computationally more convenient approach which consist in approximating critical values by a third order polynomial in $1 - \lambda$. The estimated coefficients for the third order polynomial approximation to the quantiles of the asymptotic null distribution are provided in Thompson (2004, page 9). It is important to mention that Lucas (1995b) obtained critical values by computing the distribution of the test statistic from realizations of simulated data. He considered only data generating process with i.i.d. standard normal innovations. We have just shown that the parameter λ depends on the error density and, therefore, the Lucas' method will give incorrect asymptotic critical values for the robust test unless the errors are in fact normal. Likewise, the approach developed by Thompson (2004) gives incorrect critical values unless the true innovation distribution is equal to the assumed distribution. For example, if the true distribution is a student-t with 9 degrees of freedom (t-9) and one computes critical values assuming a student-t with 3 degrees of freedom (t-3), then the Thompson approach will give imprecise critical values since its polynomial approximation will be based on the value of λ for t-3 rather than for t-9 distribution. As we will show in the next section, our partially adaptive approach tends to give correct critical values because it approximates the true distribution by the data distribution and uses the latter to estimate λ and then the critical values.

3 A Unit Root Test Based on Partially Adaptive Estimation

The M estimator is asymptotically efficient when it is the maximum likelihood estimator. 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, then inference based on these method still have good sampling performance. Thus, it is important to select a criterion function that has similar characteristic as the data distribution. In this section, we consider a data-dependent approach to select an appropriate criterion function and propose a unit root test based on partially adaptive estimation.

The partially adaptive M estimation considers a parametric family of distributions. Each member of this family is indexed by some adaptation parameters. Giving the observed sample, it is possible to estimate the adaptation parameters so that the density function that best approximates the data distribution (within the parametric family) is selected. In the literature, different classes of distributions has been studied for the purpose of partially adaptive estimation (see, inter alia, Postcher and Prucha (1986), McDonald and Newey (1988), and Phillips (1994)). Taking into account of the well documented characteristic of heavy-tails in economic and financial data, we consider a partially adaptive estimator based on the student-t distributions (Postcher and Prucha 1986), although other classes of distributions may be analyzed similarly. The student-t distribution is an important class of distributions (see more discussion in, say, Hall and Joiner 1982) that contains the Cauchy distribution as a special case and the normal distribution as a limit case, and has wide applications in economic analysis. Its adaptation parameter depends on the scale and thickness parameters, which can be easily estimated from the data using the approach proposed by Potscher and Prucha (1986). Partially adaptive estimator based on this class of distribution is reasonably robust.

Giving the ADF model (4), in the presence of t-distributed innovations, the log-likelihood is given by

$$\begin{aligned} L &= \text{constant} + \frac{n}{2}\ln\Theta \\ &\quad -\frac{\nu+1}{2}\sum_{t=j+2}^{n}\ln\left\{1+\frac{\Theta}{\nu}\left[\Delta y_t-\gamma' x_t-\rho y_{t-1}-\sum_{j=1}^{k}\psi_j\Delta y_{t-j}\right]^2\right\} \end{aligned}$$

where the parameter Θ measures the spread of the disturbance distribution and ν is the degree of freedom that measures the tail thickness. Large values of ν corresponds to thin tails in distribution. For given parameters ν and Θ, denoting Θ/ν as θ, the MLE of $(\gamma,\rho,\{\psi_j\}_{j=1}^{k})$ is the solution of the following optimization problem

$$\min\sum_t \ln\left\{1+\theta\left[\Delta y_t-\gamma' x_t-\rho y_{t-1}-\sum_{j=1}^{k}\psi_j\Delta y_{t-j}\right]^2\right\}.$$

Following Potscher and Prucha (1986), let $\widetilde{\Pi}$ be the least squares estimator of Π and $\theta=\frac{\Theta}{\nu}$ be the adaptation parameter of the t-distribution, we have the following one-step partially adaptive M estimator for the ADF model:

$$\begin{aligned}
\widehat{\Pi} &= \widetilde{\Pi} + \left[\frac{1}{n}\sum_t Z_t'(w_t - 2\theta w_t^2 \widetilde{\varepsilon}_t^2) Z_t\right]^{-1} \frac{1}{n}\sum_t Z_t' w_t \widetilde{\varepsilon}_t \qquad (9)\\
& \text{where}\\
w_t &= (1+\theta\widetilde{\varepsilon}_t^2)^{-1} \text{ and } \widetilde{\varepsilon}_t = \Delta y_t - Z_t\widetilde{\Pi}.
\end{aligned}$$

In practical analysis, the parameters ν and Θ are not known and has to be estimated. We consider a two-step partially adaptive estimator of $(\gamma, \rho, \{\psi_j\}_{j=1}^k)$ in which the first step involves a preliminary estimation of the parameters ν and Θ (and thus θ). We then replace θ in (9) by its estimator and perform a second step estimation for $(\gamma, \rho, \{\psi_j\}_{j=1}^k)$. In the presence of general disturbance distributions, ν and Θ lose their original meaning. However, in the cases where $\hat{\nu} \geq 0$ and $\hat{\Theta} \geq 0$, $\hat{\nu}$ and $\hat{\Theta}$ still can be interpreted as estimators of measures of the tailthickness and the spread of the disturbance distribution, and partially adaptive estimator (9) still have good sampling properties.

Potscher and Prucha (1986) discussed the estimation of the adaptation parameters ν and Θ. In particular, if we denote $E(|u_t|^k)$ as σ_k, then for $\nu > 2$, we have

$$\frac{\sigma_2}{\sigma_1^2} = \frac{\pi}{\nu-2}\frac{\Gamma[\nu/2]^2}{\Gamma[(\nu-1)/2]^2} = \rho(\nu) \qquad (10)$$

and

$$\Theta = \frac{1}{\pi}\frac{\nu\Gamma[(\nu-1)/2]^2}{\sigma_1^2\Gamma[\nu/2]^2} = q(\nu, \sigma_1). \qquad (11)$$

Potscher and Prucha show that $\rho(.)$ is analytic and monotonically decreasing on $(2, \infty)$ with $\rho(2+) = \infty$ and $\rho(\infty) = \pi/2$. Thus, given estimator of σ_1 and σ_2, ν can be estimated by inverting $\rho(\nu)$ in 10 and thus an estimator of θ can be obtained from

$$\hat{\theta} = \frac{q(\hat{\nu}, \hat{\sigma}_1)}{\hat{\nu}} = \frac{1}{\pi}\frac{\Gamma[(\hat{\nu}-1)/2]^2}{\hat{\sigma}_1^2\Gamma[\hat{\nu}/2]^2}. \qquad (12)$$

For the estimation of σ_1 and σ_2, we may use the sample moments

$$\hat{\sigma}_k = \frac{1}{n}\sum_t |\hat{u}_t|^k.$$

Notice that $\rho(.)$ is monotonically decreasing, ν and thus θ can be estimated numerically.

We incorporate the partially adaptive estimation into the testing procedure in Section 2 and propose the following unit root test based on partially adaptive estimation:

1. We estimate the residuals $\widetilde{\varepsilon}_t$ from a preliminary ADF regression:

$$\Delta y_t = \widetilde{\gamma}' x_t + \widetilde{\rho} y_{t-1} + \sum_{j=1}^{k} \widetilde{\psi}_j \Delta y_{t-j} + \widetilde{\varepsilon}_t$$

2. Estimating the adaptation parameters. We consider the class of student-t distributions and estimate the parameters ν and Θ as described above using the residuals obtained from step 1. Denote the estimators as $\hat{\nu}$ and $\hat{\Theta}$, we obtain $\hat{\theta} = \hat{\Theta}/\hat{\nu}$.

3. Selecting the criterion function. Giving the estimated adaptation parameter, we choose the following criterion function

$$\varphi(\varepsilon) = \ln \left\{ 1 + \widehat{\theta} \, [\varepsilon]^2 \right\},$$

and calculate the corresponding M-estimator for $(\gamma, \rho, \{\psi_j\}_{j=1}^{k})$ in model (4) based on

$$\max \sum_{t=2}^{n} \varphi \left(\Delta y_t - \gamma' x_t - \rho y_{t-1} - \sum_{j=1}^{k} \psi_j \Delta y_{t-j} \right)$$

Denote the corresponding t-statistic as $t_{\widehat{\rho}}$.

4. Calculate estimate of λ^2. First we estimate the variance estimator of ε and $\varphi'(\varepsilon_t)$ by

$$\widehat{\sigma}^2 = \frac{1}{n-k-1} \sum_t \widehat{\varepsilon}_t^2, \text{ and } \widehat{\omega}_\varphi^2 = \frac{1}{n-k-1} \sum_{t=k+1}^{n} \varphi'(\widehat{\varepsilon}_t)^2$$

respectively, and let

$$\widehat{cov(\varepsilon_t, \varphi'(\varepsilon_t))} = \frac{1}{n-k-1} \sum_{t=k+1}^{n} \widehat{\varepsilon}_t \varphi'(\widehat{\varepsilon}_t),$$

we then estimate ω_u^2 and $\sigma_{u\varphi}^2$ by $\widehat{\omega}_u^2 = \widehat{\sigma}^2 / \widehat{A}(1)^2$, and $\widehat{\sigma}_{u\varphi} = \widehat{cov(\varepsilon_t, \varphi'(\varepsilon_t))} / \widehat{A}(1)$, where $\widehat{A}(1) = 1 - \sum \widehat{\psi}_j$, λ^2 can then be estimated by

$$\widehat{\lambda}^2 = \frac{\widehat{\sigma}_{u\varphi}^2}{\widehat{\omega}_\varphi^2 \widehat{\omega}_u^2}.$$

Using the estimate of λ^2, we compute the critical values using the polynomial approximation proposed by Thompson (2004).

4 Monte Carlo Experiments

In this section, we compare our partially adaptive unit root test (PADF test) with the OLS based test of Dickey and Fuller (1979) (ADF test) and the M-estimator-based tests with critical values obtained using the Lucas (1995b) and

Thompson (2004) approaches, respectively. We selected the following four innovation distributions:

D1: N(0,1),

D2: t-3,

D3: t-5,

D4: t-9,

where N(a,b) denotes the normal distribution with mean a and variance b, $t(k)$ denotes the t-distribution with k degrees of freedom.

>From the construction of the tests, it is apparent that its finite sample performance will be affected by the sample size, the distribution of the innovations ε_t, the autoregressive coefficient α, and the $I(0)$ dependence in u_t. Thus, special attention is paid here to the effects of these elements on the performance of the unit root tests.

The data generating process (DGP) in our Monte Carlo is given by the following model

$$y_t = \alpha y_{t-1} + u_t, \tag{13}$$

$$y_0 = 0. \tag{14}$$

where we assume the following serial correlation structures for the error term: (i) $u_t = 0.0 \cdot u_{t-1} + \varepsilon_t$, (ii) $u_t = 0.5 \cdot u_{t-1} + \varepsilon_t$, and (iii) $u_t = \varepsilon_t - 0.5\varepsilon_{t-1}$ with $\varepsilon_t \sim N(0,1)$. In (i) and (ii), we assume $u_0 = 0$ and $\{\varepsilon_t\}$ is a sequence of i.i.d. observations drawn from the distribution D_i, $i = 1, 2, 3, 4$. Therefore, the first structure corresponds to i.i.d innovations, the second to a AR(1) process and the third one represents a MA(1) process. The asymptotic size of each test is 5%.

The power of the test was evaluated by considering $\alpha = 0.90; 0.95$ and 0.99. The size of the test is obtained by setting $\alpha = 1$. We estimate the ADF regression (4) including intercept and intercept plus trend. We generated 2000 time series of size 200 and 500, but we only report results for $n = 200$ because they are close to the results obtained with $n = 500$.

We compute five test statistics. The first one is the "Augmented Dickey-Fuller" (ADF) t-ratio which is obtained estimating equation (4) by least squares and computing the t-ratio by (7) using the traditional covariance matrix $\widehat{\Omega} = \widetilde{\sigma}^2 (x_t x_t')^{-1}$ where $\widetilde{\sigma}^2 = \widetilde{\varepsilon}'\widetilde{\varepsilon}/(n-k-1)$. The 5% critical value for the ADF test with intercept is -2.86 for the regression with intercept and -3.41 for the intercept and trend specification. We also computed four robust tests which are described below:

1) the L_3 test:

This test consists in estimating equation (4) by maximum likelihood (ML) assuming a t-3 distribution and then computing the t-ratio by using (7) with the covariance matrix being estimated by (6). Critical values are coming from the polynomial approximation suggested in Thompson (2004) with $\lambda = 0.98$[3]. This robust test was first suggested by Lucas (1995b).

[3] When the innovations are Gaussian (as assumed by Lucas,1995.b) the estimated value of the parameter λ is equal to 0.98. This same result was obtained by Thompson (2004).

2) the T_3 test:

Likewise the L_3 test, T_3 estimates the ADF equation (4) by ML assuming a t-3 distribution and compute the t-ratio by using (7) with the covariance matrix being estimated by (6). Unlike the L_3 test where critical values are computed assuming Gaussianity, we compute critical values assuming a t-3 distribution. Thus, we are basically following the idea developed in Thompson (2004) for computation of correct asymptotic critical values.

3) the T_9 test:

This is just like the T_3 test with the t-3 distribution replaced by the t-9 distribution.

4) the $PADF$ test

We compute our PADF test estimating the ADF regression (4) using expressions given in (9), (10), (11) and (12). For the PADF t-test, we use covariance matrix (6) and its corresponding element in constructing the t-ratio (7). The critical values are also coming from the polynomial approximation suggested in Thompson (2004). Note, however, that we are not assuming any degree of freedom for the t-distribution since it is being estimated from the data (equations 10, 11 and 12).

For all tests aforementioned, we employed the Schwartz criterion to choose the number of lags k.

4.1 Results

Table 1 shows the results for the case where $u_t = \varepsilon_t$. When the innovations are Gaussian, results in Table 1 suggest that: (i) all tests have similar empirical size with T_3 being slightly oversized; (ii) the L_3 and T_3 tests have less power than the other tests for distant alternatives, say $\alpha = 0.95$ and $\alpha = 0.90$; (iii) the overal performance of the PADF test under Gaussianity is as good as the ADF test and better than the others robust test, L_3, T_3 and T_9.

When the innovation distribution is a student-t with 3 degrees of freedom, t-3, the robust test suggested by Lucas (1995b) is undersized. This happens because the critical values computed in Lucas (1995b) are incorrect unless the errors are in fact normal. Since the t-3 distribution characterizes a large departures from Gaussianity, this result does not come as a surprise. Indeed, this same finding has also been reported in Thompson (2004). We notice, on the other hand, that both the T_3 and $PADF$ tests have correct size and are much more powerful than the other tests. Finally, when the error distribution is a student-t with 9 degrees of freedom, t-9, we notice that the T_3 test seems to be slightly oversized. The PADF test has good size and, along with T_9, is more powerful than the other tests.

The above results tell us that when the likelihood function is correctly specified, the T_3 and T_9 tests performs as well as the PADF test, but if they are misspecified (and thus quasi-MLE), the PADF test performs better. To emphasized this, we now consider the t-5 innovation distribution. With t-5 distribution, the criterion function used in the computation of T_3 and T_9 and their respective critical values is misspecified. Thus, in this case of misspecification, we should not expect that quasi-maximum-likelihood tests present the best relative performance. Indeed, results in Table 1 show that when we consider the t-5 distribution, the PADF test has good size and more power than the quasi-MLE based tests.

In sum, the above results suggest that the partially adaptive test has relatively pretty good finite-sample performance: there is little loss in using the proposed test when the innovations are Gaussian, and the power gains from using the partially adaptive test is substantial when there are non-Gaussian innovations. The PADF test also peforms better than other robust tests under misspecification of the criterion function and performs as good as when the criterion function coincides with the true likelihood function. This happens because the PADF test is based on a data-dependent procedure that select an appropriate criterion function for estimation and computation of critical values.[4]

Table 1: Size and Power of 5% Tests - intercept only

Distribution (D_i)	*ADF*	L_3	T_3	T_9	*PADF*
$\alpha = 1$					
$N(0,1)$	0.051	0.058	0.066	0.062	0.055
t-3	0.047	0.029	0.046	0.049	0.050
t-5	0.052	0.043	0.057	0.051	0.054
t-9	0.051	0.055	0.066	0.056	0.058
$\alpha = 0.99$					
$N(0,1)$	0.065	0.073	0.085	0.076	0.071
t-3	0.057	0.086	0.129	0.106	0.126
t-5	0.062	0.076	0.084	0.076	0.092
t-9	0.067	0.081	0.082	0.088	0.088
$\alpha = 0.95$					
$N(0,1)$	0.320	0.272	0.299	0.317	0.320
t-3	0.306	0.597	0.679	0.596	0.676
t-5	0.321	0.421	0.480	0.460	0.490
t-9	0.312	0.341	0.381	0.390	0.400
$\alpha = 0.90$					
$N(0,1)$	0.820	0.636	0.666	0.786	0.793
t-3	0.822	0.947	0.969	0.969	0.967
t-5	0.823	0.833	0.860	0.910	0.910
t-9	0.817	0.737	0.776	0.848	0.849

The calculations in Table 1 assume that the errors are independent. To investigate the impact of serially correlated errors, Table 2 provides evidence about the small sample behavior of the test statistics decribed above. Table 2 shows results for the case where (i) $u_t = 0.5u_{t-1} + \varepsilon_t$ and $\{\varepsilon_t\}$ is a sequence

[4]However, the power of all tests is low when α is close to one, which indicates that long-memory processes may be more appropriate to model high persistence in economic time series. We thank an anonymous referee for pointing this out.

of i.i.d. observations drawn from a distribution D_i, $i = 1,2,3,4$ and (ii) $u_t = \varepsilon_t - 0.5\varepsilon_{t-1}$ and ε_t has N(0,1) distribution.

Table 2: Size and Power of 5% Tests - intercept only

$Distribution\ (D_i)$	ADF	L_3	T_3	T_9	$PADF$
$\alpha = 1$					
$N(0,1)$	0.051	0.057	0.065	0.059	0.056
t-3	0.049	0.027	0.045	0.046	0.048
t-5	0.052	0.038	0.055	0.047	0.051
t-9	0.055	0.055	0.068	0.057	0.057
$MA(1)$	0.143	0.130	0.145	0.151	0.145
$\alpha = 0.99$					
$N(0,1)$	0.064	0.069	0.078	0.074	0.073
t-3	0.062	0.084	0.124	0.104	0.121
t-5	0.061	0.073	0.095	0.081	0.095
t-9	0.071	0.080	0.091	0.090	0.090
$MA(1)$	0.228	0.196	0.214	0.234	0.225
$\alpha = 0.95$					
$N(0,1)$	0.285	0.257	0.279	0.300	0.291
t-3	0.270	0.567	0.660	0.610	0.657
t-5	0.288	0.395	0.450	0.430	0.460
t-9	0.290	0.318	0.356	0.365	0.370
$MA(1)$	0.687	0.572	0.593	0.680	0.653
$\alpha = 0.90$					
$N(0,1)$	0.712	0.555	0.592	0.703	0.697
t-3	0.742	0.908	0.943	0.950	0.951
t-5	0.748	0.770	0.810	0.850	0.860
t-9	0.743	0.669	0.707	0.780	0.780
$MA(1)$	0.971	0.881	0.892	0.955	0.940

Size is generally accurate for AR errors, but is less accurate for MA errors. As in the case with independent errors, most of the time the "robust" L_3 test has less power than the T_3, T_9 and $PADF$ tests when the distribution of the fundamental innovation in the AR(1) error is fat-tailed. The L_3 test is again undersized under t-3 distribution and has the lowest power for MA(1) errors. The PADF test again performs relatively well in terms of power and size for both AR(1) and MA(1) errors although, as occur in the other tests, it does not have accurate size for MA(1) errors. The T_3 and T_9 have a good relative performance only if the true distribution of the fundamental innovation is t-3 or t-9, respectively.

Table 3 and 4 display the results for the model with intercept and trend. As expected, the inclusion of time trend causes a reduction of power in all tests. Apart this difference, the results below are qualitatively similar to what we showed in Table 1 and 2, which were obtained estimating a model with intercept only.

Table 3: Size and Power of 5% Tests - intercept and trend.

Distribution (D_i)	ADF	L_3	T_3	T_9	$PADF$
$\alpha = 1$					
$N(0,1)$	0.056	0.064	0.080	0.066	0.062
t-3	0.060	0.021	0.043	0.048	0.047
t-5	0.060	0.037	0.055	0.054	0.055
t-9	0.061	0.055	0.070	0.063	0.064
$\alpha = 0.99$					
$N(0,1)$	0.058	0.072	0.079	0.076	0.077
t-3	0.060	0.048	0.098	0.090	0.100
t-5	0.063	0.049	0.080	0.073	0.085
t-9	0.064	0.065	0.078	0.081	0.086
$\alpha = 0.95$					
$N(0,1)$	0.198	0.193	0.222	0.237	0.233
t-3	0.191	0.400	0.528	0.482	0.530
t-5	0.191	0.270	0.350	0.321	0.360
t-9	0.199	0.218	0.266	0.271	0.273
$\alpha = 0.90$					
$N(0,1)$	0.610	0.461	0.508	0.594	0.583
t-2	0.594	0.950	0.965	0.970	0.953
t-3	0.603	0.840	0.910	0.886	0.910
t-5	0.608	0.671	0.740	0.771	0.780
t-9	0.609	0.554	0.620	0.676	0.673

Table 4: Size and Power of 5% Tests - intercept and trend included

Distribution (D_i)	*ADF*	L_3	T_3	T_9	*PADF*
$\alpha = 1$					
$N(0,1)$	0.052	0.063	0.078	0.066	0.062
t-3	0.057	0.026	0.044	0.049	0.047
t-5	0.058	0.035	0.058	0.053	0.057
t-9	0.058	0.051	0.070	0.057	0.061
$MA(1)$	0.175	0.164	0.190	0.197	0.192
$\alpha = 0.99$					
$N(0,1)$	0.058	0.072	0.085	0.078	0.077
t-3	0.061	0.048	0.094	0.085	0.092
t-5	0.066	0.052	0.080	0.067	0.080
t-9	0.065	0.067	0.081	0.078	0.079
$MA(1)$	0.220	0.195	0.221	0.245	0.229
$\alpha = 0.95$					
$N(0,1)$	0.181	0.194	0.213	0.216	0.210
t-3	0.175	0.368	0.514	0.452	0.506
t-5	0.178	0.252	0.331	0.290	0.342
t-9	0.189	0.210	0.238	0.245	0.248
$MA(1)$	0.593	0.475	0.510	0.595	0.570
$\alpha = 0.90$					
$N(0,1)$	0.528	0.392	0.445	0.524	0.501
t-3	0.498	0.770	0.871	0.851	0.870
t-5	0.514	0.595	0.672	0.696	0.711
t-9	0.524	0.486	0.543	0.598	0.599
$MA(1)$	0.917	0.802	0.826	0.898	0.881

As discussed in Section 2, The M estimator is asymptotically efficient when it is equal to the maximum likelihood estimator. In practice, however, 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, then inference based on this partial adaptation method still have good sampling performance.

In order to illustrate the above point, we added a new monte-carlo experiment to assess the performance of the PADF test when the true distribution of the innovations is not a t-distribution, but instead, a mixture of normal distributions. In this simulation, we consider the following model

$$y_t = \alpha y_{t-1} + \widetilde{\varepsilon}_t \tag{15}$$

$$y_0 = 0. \tag{16}$$

where the variable $\widetilde{\varepsilon}_t$ equals $\varepsilon_t \sim iid\ N(0,1)$ when $\upsilon_t < 0.05$,where $\upsilon_t \sim iid$ $Uniform(0,1)$, and $\widetilde{\varepsilon}_t = \varepsilon_t + w_t$, otherwise.; Here, w_t is a contaminating random variable that is being drawn from a N(0,30).

The results are displayed in Table 5. In particular, we notice that the ADF test has lower power and higher size distortion than the P-ADF test. There seems to be no doubt that the P-ADF test performs much better than the ADF test in cases where the distribution of the innovations is a mixture of normals rather than a t-distribution.

Table 5. Power and Size of 5% test.

α	0.85	0.90	0.95	0.975	0.99	1
$n = 200$						
ADF	0.987	0.827	0.295	0.113	0.064	0.057
P-ADF	0.996	0.971	0.747	0.392	0.159	0.052

5 Empirical Analysis

5.1 The uncertain unit root in real GNP.

Existence of heavy-tailed distributions in real GNP has been largely documented in economics and econometrics. In effect, Blanchard and Watson (1986) concluded that fluctuations in economic activity are characterized by a mixture of large and small shocks. Other references includes Bidarkota (2000), Balke and Fomby (1994) and Scheinkman and LeBaron (1989) who advocates that real GNP is mostly contaminated by outliers. In parallel, since the seminal work of Nelson and Plosser (1982), there has been an intense debate about the presence of stochastic trend in real GNP . Whether trend is better described as deterministic or stochastic is an important issue for point forecasting, because the two models imply very different long-run dynamics and hence different long-run forecasts. Cochrance (1988) finds little evidence of stochastic trend in GNP whereas Campbell and Mankiw (1987) claims that output fluctuations are permanent. There also be the "we don´t know" literature (Rudeebush, 1993, Christiano and Eichenbaum, 1990,) which correctly concludes that traditional ADF unit root test is unlikely to be capable of discriminating between deterministic and stochastic trend because its well known low power against distant alternatives.

5.1.1 Empirical Results Based on the PADF test

We used two series of real GNP (RGN)[5]. The first one (RGNP_{NP}) was collected from the Nelson and Plosser database and it has 81 annual observations (1909-1980). The second database (RGNP_2) were collected from the U.S. Department of Commerce, Bureau of Economic Analysis. RGNP_2 are measured in billions of fixed 1996 Dollars and are seasonally adjusted annual values and quarterly observed. Its first observation corresponds to the first quarter of 1967, totalizing 141 observations. The table below presents some descriptive information about our dataset.

[5] Both series are expressed in logarithmic terms.

Table 6. Descriptive Statistics

Series	sample size	thickness parameter	Kurtosis†	Jarque-Bera†
$RGNP_{NP}$	81	5.31	5.00	24.75**
$RGNP_2$	141	6.91	4.10	11.40**

†The symbol (**) represents rejection of the null hypothesis at 1% level of significance

Table 6 shows two measures of tails. The standard one is the kurtosis. It is well known that whenever this quantity exceeds 3, we say that the data feature excess kurtosis, or that their distribution is leptokurtic, that is, it has heavy tails. One can see that, after prewhitening the process[6], both $RGNP_{NP}$ and $RGNP_2$ have excess kurtosis. Another measure of heavy tails is the thickness parameter of the student-t distribution, ν. Small ν corresponds to heavy tails and the limiting case, $\nu \to \infty$, corresponds to the normal distribution. Again, we notice that $RGNP_{NP}$ and $RGNP_2$ have very small thickness parameters, suggesting the existence of heavy-tailed distribution for those series. Thus, our data suggest that post-war US real GNP behavior is inconsistent with linear Gaussian models.

We now turn to the unit root analysis. We employed the non-robust ADF test and the robust P-ADF test. The number of lags was chosen according to the Schwartz criterion and $\widehat{\lambda}^2$, as usual, was estimated parametrically. We also included a linear trend in the ADF regression. The results are displayed in Table 7. If one conducts unit root inference by using the non-robust ADF test, then the null of unit root could not be rejected at 5% level of significance, suggesting the presence of a stochastic trend in real GNP. This results support the literature of permanent shocks in output. As showed by our Monte Carlo simulations, the ADF test do not perform well (it has low power) when innovations are drawn from fat-tailed distributions. Results in Table 6 reveal the presence of heavy-tailed distributions and, therefore, we had better conduct unit root inference using the robust version of the ADF test, that is, the PADF test. In doing so, we reject the null of unit root for both $RGNP_{NP}$ and $RGNP_2$.This finding gives support to the literature of transitory shocks in output and suggest that the failure of rejecting the null of unit root in U.S real GNP series may be due to the use of estimation and hypothesis testing procedures that do not consider the presence of fat-tail distributions in the data. We believe that this result may be useful to investigate convergence of international (or regional) output, among other hypotheses involving real GNP.

Table 7. Unit Root Analysis

Series	Lags	Deterministic Component	ADF	PADF
$RGNP_{NP}$	1	linear trend	-3.44	-4.62**
$RGNP_2$	2	linear trend	-2.81	-3.47*

The symbol (**) represents rejection of the null hypothesis at 1% level of significance
The symbol (*) represents rejection of the null hypothesis at 5% level of significance

[6] The data were pre-whitened using a linear trend and the number of lags shown in Table 7.

5.2 Nominal interest rate and real exchange rate

In this section, we investigate the presence of unit root in other financial time series. In particular, we consider nominal interest rate and real exchange rate. We used nominal interest rate with 12-month and 3-month maturity[7], with first observation corresponding to April of 1953 and ending observation to May of 2000. As for the data on real exchange rate (RER), we used monthly data of US-dollar and UK-pound sterling based bilateral real exchange rates, that is : United kingdom-USA (UK-US), Japan-USA (JPN-US), France-US (FRA-US), Germany-US (GER-US), Japan-UK (JPN-UK), France-UK (FRA-UK), and Germany-UK (GER-UK). To construct the real exchange rate, the data on the nominal exchange rate and the price level (Consumer Price Index) are collected from the International Financial Statistics CD-Rom, which is made by the International Monetary Fund (IMF). The sample covers the Post-Bretton Woods period that runs from April 1973 to March 2001.

Table 8 presents the descriptive statistics[8]. All series seems to show evidence of deviations from Gaussianity, with the series of nominal interest rate presenting high excess kurtosis as compared to real exchange rate time series. Despite the presence of nonnormal innovations, the unit root analysis in Table 9, carried out by using the robust P-ADF test, does not suggest that the null hypothesis of unit root is rejected. This result brings out very practical consequences. For example, the presence of unit root in RER implies that PPP hypothesis does not hold in the long run even if we account for heavy tails in real exchange rates. In a recent paper, Falk and Wang (2003) reached the same conclusion by considering the effects of fat tails on critical values of cointegrating tests. In particular, they find that the Johansen´s likelihood-ratio based test are less supportive of PPP when Gaussian-based critical values are replaced by heavy-tailed-based critical values. Using a different approach, our results provide additional support to the findings of Falk and Wang.

The presence of unit root in the US nominal interest rate has puzzling the economic theory for long. In effect, Rose (1988) showed that the presence of unit root in nominal interest rate is inconsistent with the results predicted by the consumption-based capital asset pricing model (CCAPM). Furthermore, unit root in nominal interest rate is incompatible with the results predicted by optimal monetary policy models, as in Friedman (1969). These models suggest the existence of stable (constant) nominal interest rate in the long run as the result of a monetary authority that maximizes steady-state welfare. Our results indicate the presence of unit root in US nominal interest rate even when heavy tails are accounted for. Hence, we provide support for the findings in Rose (1988), which contradict the theoretical results predicted by the CCAPM and optimal monetary policy models.

[7] Three-month and twelve-month Treasury Bill Rate: Board of Governors of the Federal Reserve System, http://www.stls.frb.org/fred/

[8] The data were pre-whitened using the deterministic specification and number of lags shown in table 9.

Table 8. Descriptive Statistics

Series	sample size	thickness parameter	kurtosis	Jarque-Bera†
nominal interest rate (12M)	566	3.31	9.40	950.40**
nominal interest rate (3M)	566	2.71	15.10	3429.94**
RER (UK-US)	336	6.31	6.13	145.13**
RER (JPN-US)	336	6.91	5.09	80.64**
RER (GER-US)	336	9.11	3.90	13.07**
RER (FRA-US)	336	6.51	4.64	37.68**
RER (FRA-UK)	336	6.51	4.35	25.71**
RER (GER-UK)	336	5.91	5.94	147.02**
RER (JPN-UK)	336	7.51	4.43	28.78**

†The symbol (**) represents rejection of the null hypothesis at 1% level of significance.

Table 9. Unit root Analysis

Series	Lags	Deterministic Component	ADF	P-ADF
nominal interest rate (12M)	6	constant	-2.05	-1.16
nominal interest rate (3M)	6	constant	-2.07	-0.11
RER (UK-US)	1	constant	-2.59	-1.01
RER (JPN-US)*	1	linear trend	-2.09	-1.27
RER (GER-US)	1	constant	-1.66	-0.65
RER (FRA-US)	1	constant	-1.52	-0.21
RER (FRA-UK)	1	constant	-1.79	-1.20
RER (GER-UK)	1	constant	-1.88	-2.15
RER (JPN-UK)*	1	linear trend	-2.48	-1.69

*In order to control the possible forces that move the real exchange rate to a direction in the long run (such as Balassa-Samuleson effect), and to be consistent with past studies such as Cheung and Lai (2001), we decided to include a deterministic trend in the specification of Japanese-yen based real exchange rates.

6 Conclusion

This paper proposes a unit root test based on partially adaptive estimation. ADF type of regression is considered without assuming Gaussian innovations. Monte Carlo results indicate that the partially adaptive test has relatively pretty good finite-sample performance: there is little loss in using the proposed test when the innovations are Gaussian, and the power gains from using our partially adaptive test is substantial when there are non-Gaussian innovations. The PADF test also performs better than other robust tests under misspecification of the criterion function and performs as good as when the criterion function coincides with the true likelihood function. This happens because the PADF test is based on a data-dependent procedure that select an appropriate criterion function for estimation and computation of critical values.

As an empirical example, we apply the proposed test to some macroeconomic time series with heavy-tailed distributions. It is shown that US real GNP are featured with heavy-tailed distribution and that the traditional ADF test does not reject the null of unit root. However, this hypothesis is rejected when we use the PADF test, supporting the literature of transitory shocks in output. We also reported evidence for unit root in real exchange rate and nominal interest rate even when tail heaviness is accounted for.

7 APPENDIX

Proof of Theorems 1 and 2. The first order condition correspond to the M-estimator in section 2 is given by:

$$\text{FOC: } \sum_t \varphi' \left(\Delta y_t - \widehat{\Pi}' Z_t \right) Z_t = 0$$

or let $\psi = \varphi'$,

$$\sum_{t=1}^{n} \psi \left(\Delta y_t - \widehat{\Pi}' Z_t \right) Z_t = 0$$

Taking a Taylor expansion with respect to $\widehat{\varepsilon}_t = \Delta y_t - \widehat{\Pi}' Z_t$ around $\varepsilon_t = \Delta y_t - \Pi' Z_t$ we have

$$\sum_{t=1}^{n} \psi \left(\varepsilon_t \right) Z_t - \sum_{t=1}^{n} \psi' \left(\varepsilon_t \right) Z_t Z_t' (\widehat{\Pi} - \Pi) + R_T = 0,$$

where R_T is the remainder term.

We now introduce the standardization matrix:

$$D_n = diag\{\sqrt{n} F_n, n, \sqrt{n}, \cdots, \sqrt{n}\}$$

From the above Taylor expansion and the definition of the standardization matrix D_n, we derived the expression below, where $o_p(1)$ term refers to the standardized R_T which is $o_p(1)$ under Assumption A2. Thus,.under our regularity conditions (i.e., assumptions of Theorem 1) we have

$$D_n(\widehat{\Pi} - \Pi) = \left[\sum_{t=1}^{n} \psi' \left(\varepsilon_t \right) D_n^{-1} Z_t Z_t' D_n^{-1} + o_p(1) \right]^{-1} \sum_{t=1}^{n} \psi \left(\varepsilon_t \right) D_n^{-1} Z_t$$

the following asymptotics hold:

$$\begin{aligned}
& \sum_{t=1}^{n} \psi' \left(\varepsilon_t \right) D_n^{-1} Z_t Z_t' D_n^{-1} \\
= \quad & \sum_{t=1}^{n} \psi' \left(\varepsilon_t \right) \begin{pmatrix} n^{-1/2} F_n^{-1} x_t \\ n^{-1} y_{t-1} \\ \frac{1}{\sqrt{n}} \Delta y_{t-1} \\ \cdots \\ \frac{1}{\sqrt{n}} \Delta y_{t-k} \end{pmatrix} \left(n^{-1/2} x_t' F_n^{-1}, n^{-1} y_{t-1}, \frac{1}{\sqrt{n}} \Delta y_{t-1}, \cdots, \frac{1}{\sqrt{n}} \Delta y_{t-1} \right) \\
= \quad & \sum_{t=1}^{n} \psi' \left(\varepsilon_t \right) \begin{pmatrix} \frac{1}{n} F_n^{-1} x_t x_t' F_n^{-1} & & & \\ n^{-3/2} y_{t-1} x_t' F_n^{-1} & n^{-2} y_{t-1}^2 & & \\ \frac{1}{n} \Delta y_{t-1} x_t' F_n^{-1} & \frac{1}{n^{3/2}} \Delta y_{t-1} y_{t-1} & \frac{1}{n} \left(\Delta y_{t-1} \right)^2 & \\ \cdots.. & & & \ddots \\ \frac{1}{n} \Delta y_{t-k} x_t' F_n^{-1} & \frac{1}{n^{3/2}} \Delta y_{t-k} y_{t-1} & \frac{1}{n} \Delta y_{t-k} \Delta y_{t-1} & \frac{1}{n} \left(\Delta y_{t-k} \right)^2 \end{pmatrix} \\
\Rightarrow \quad & \delta \begin{pmatrix} \int \overline{B}_y(r) \overline{B}_y(r)' dr & 0 \\ 0 & \Gamma_y \end{pmatrix}
\end{aligned}$$

where

$$\Gamma_y = \begin{bmatrix} \gamma_y(0) & & \\ & \ddots & \\ \gamma_y(k-1) & & \gamma_y(0) \end{bmatrix}, \overline{B}_y(r)' = (X(r)', B_y(r))'.$$

Under Assumption A1, the partial sums of the vector process $(u_t, \psi(\varepsilon_t), \Delta y_{t-1}\psi(\varepsilon_t), \cdots, \Delta y_{t-k}\psi(\varepsilon_t))^\top$ follow a multivariate invariance principle (see, e.g., Phillips and Durlauf (1986, Theorem 2.1, 474-476, and 486-489):

$$n^{-1/2}\sum_{t=1}^{[nr]}(u_t, \psi(\varepsilon_t), \Delta y_{t-1}\psi(\varepsilon_t), \cdots, \Delta y_{t-k}\psi(\varepsilon_t))^\top \Rightarrow (B_u(r), B_\varphi(r), B_{\varphi 1}(r), \cdots, B_{\varphi k}(r))^\top.$$

In particular, the bivariate Brownian motion $(B_u(r), B_\varphi(r))$ ahs Covariance matrix Σ, and is independent with $(B_{\varphi 1}(r), \cdots, B_{\varphi k}(r))$. Thus, by continuous mapping theorem,

$$\sum_{t=1}^{n}\psi\left(\varepsilon_t\right) D_n^{-1} Z_t = \sum_{t=1}^{n}\psi\left(\varepsilon_t\right)\begin{pmatrix} n^{-1/2}F_n^{-1}x_t \\ n^{-1}y_{t-1} \\ \frac{1}{\sqrt{n}}\Delta y_{t-1} \\ \cdots \\ \frac{1}{\sqrt{n}}\Delta y_{t-k} \end{pmatrix} \Rightarrow \begin{pmatrix} \int \overline{B}_y(r)dB_\psi(r) \\ \Phi \end{pmatrix}$$

where $\Phi = (\int_0^1 B_{\varphi 1}(r)dr, \cdots, \int_0^1 B_{\varphi k}(r)dr)^\top$ is an k-vector of normal variates.

Thus,

$$D_n(\widehat{\Pi} - \Pi) \Rightarrow \delta^{-1}\begin{pmatrix} \int \overline{B}_y(r)\overline{B}_y(r)'dr & 0 \\ 0 & \Gamma_y \end{pmatrix}^{-1}\begin{pmatrix} \int \overline{B}_y(r)dB_\psi(r) \\ \Phi \end{pmatrix}.$$

In particular,

$$\begin{bmatrix} (n^{-1/2}F_n^{-1}\left(\widehat{\gamma} - \gamma\right) \\ n\widehat{\rho} \end{bmatrix} \Rightarrow \delta^{-1}\left(\int \overline{B}_y(r)\overline{B}_y(r)'dr\right)^{-1}\int \overline{B}_y(r)dB_\psi(r)$$

To construct a t-statistic, we estimate the covariance matrix by

$$\widehat{\Omega} = \left[\sum_{t=1}^{n}\psi'\left(\widehat{\varepsilon}_t\right)Z_tZ_t'\right]^{-1}\left[\sum_{t=1}^{n}\psi\left(\widehat{\varepsilon}_t\right)^2 Z_tZ_t'\right]\left[\sum_{t=1}^{n}\psi'\left(\widehat{\varepsilon}_t\right)Z_tZ_t'\right]^{-1}.$$

This is a heteroskedasticity consistent type covariance matrix estimator as in White (1980). If we consider the t-ratio statistic of $\widehat{\rho}$

$$t_{\widehat{\rho}} = \frac{\widehat{\rho}}{se(\widehat{\rho})}$$

$$\left[\sum_{t=1}^{n}\psi'\left(\widehat{\varepsilon}_t\right)D_n^{-1}Z_tZ_t'D_n^{-1}\right]^{-1}\left[\sum_{t=1}^{n}\psi\left(\widehat{\varepsilon}_t\right)^2D_n^{-1}Z_tZ_t'D_n^{-1}\right]\left[\sum_{t=1}^{n}\psi'\left(\widehat{\varepsilon}_t\right)D_n^{-1}Z_tZ_t'D_n^{-1}\right]^{-1}$$

$$\rightarrow\ \frac{\omega_\varphi^2}{\delta^2}\left[\begin{array}{ll}\int\overline{B}_u(r)\overline{B}_u(r)'dr & 0_{2\times k}\\ 0_{k\times 2} & \Gamma\end{array}\right]^{-1}$$

$$\begin{aligned}\widehat{\Omega}^{-1/2}D_n(\widehat{\Pi}-\Pi) &\Rightarrow \frac{\delta}{\omega_\varphi}\left[\begin{array}{ll}\int\overline{B}_u(r)\overline{B}_u(r)'dr & 0_{2\times k}\\ 0_{k\times 2} & \Gamma\end{array}\right]^{1/2}\frac{1}{\delta}\left(\begin{array}{ll}\int\overline{B}_u(r)\overline{B}_u(r)'dr & 0_{2\times k}\\ 0_{k\times 2} & \Gamma\end{array}\right)^{-1}\cdot\\ &\quad\cdot\left(\begin{array}{c}\int\overline{B}_u(r)dB_\varphi(r)\\ \Phi\end{array}\right)\\ &= \frac{1}{\omega_\varphi}\left[\begin{array}{ll}\int\overline{B}_u(r)\overline{B}_u(r)'dr & 0_{2\times k}\\ 0_{k\times 2} & \Gamma\end{array}\right]^{-1/2}\left(\begin{array}{c}\int\overline{B}_u(r)dB_\varphi(r)\\ \Phi\end{array}\right)\end{aligned}$$

$$\widehat{\Omega}_{**}^{-1/2}\left[\begin{array}{c}(n^{-1/2}F_n^{-1}(\widehat{\gamma}-\gamma)\\ n\widehat{\rho}\end{array}\right]\Rightarrow\left(\int\overline{W}_1(r)\overline{W}_1(r)'dr\right)^{-1/2}\int\overline{W}_1(r)dW_\varphi(r).$$

Thus the t-ratio

$$\begin{aligned}t_{\widehat{\rho}} &= \frac{\widehat{\rho}}{se(\widehat{\rho})}\\ &\Rightarrow \frac{\omega_\varphi}{\omega_\varphi}\left(e'\int\overline{W}_1(r)\overline{W}_1(r)'dre\right)^{-1/2}e'\int\overline{W}_1(r)dW_\varphi(r)\\ &= \left(\int W_X(r)^2dr\right)^{-1/2}\int W_X(r)dW_\varphi(r)\end{aligned}$$

where $W_X(r)=W_1(r)-\int_0^1W_1(s)X'(s)ds\left(\int_0^1X(s)X(s)'ds\right)^{-1}X(r)$ is the Hilbert projection in $L_2[0,1]$ of $W_1(r)$ onto the space orthogonal to X.

Notice that $t_{\widehat{\rho}}$ is simply the M regression counterpart of the well-known ADF t-ratio test for a unit root.

The limiting distribution of $t_{\widehat{\rho}}$ is not standard and depend on nuisance parameters since W_1 and W_φ are correlated Brownian motions. However, the limiting distribution of the t-statistic $t_{\widehat{\rho}}$ can be decomposed as a simple combination of two (independent) well-known distributions. In addition, related critical values are tabulated in the literature and thus are ready for us to use in applications. Notice that we can decompose

$$\int B_u(r)dB_\varphi(r)$$

(see, e.g. Hansen and Phillips (1990) and Phillips (1995)) as

$$\int B_udB_{\varphi.u}+\lambda_{u\psi}\int B_udB_u,$$

where $\lambda_{u\varphi} = \sigma_{u\varphi}/\omega_u^2$ and $B_{\varphi.u}$ is a Brownian motion with variance

$$\sigma_{\varphi.u}^2 = \omega_\varphi^2 - \sigma_{u\varphi}^2/\omega_u^2$$

and is independent with B_u.

$$\begin{aligned}
\widehat{\Omega}_{**}^{-1/2} D_n(\widehat{\Pi} - \Pi) &\Rightarrow \frac{1}{\omega_\varphi}\left[\int \overline{B}_u(r)\overline{B}_u(r)'dr\right]^{-1/2} \int \overline{B}_u(r) dB_\varphi(r) \\
&= \frac{1}{\omega_\varphi}\left[\int \overline{B}_u(r)\overline{B}_u(r)'dr\right]^{-1/2} \left(\int B_u dB_{\varphi.u} + \lambda_{u\psi} \int B_u dB_u,\right) \\
&= \frac{\sigma_{\varphi.u}}{\omega_\varphi}\left[\int \overline{W}_1(r)\overline{W}_1(r)'dr\right]^{-1/2} \int W_1 dW_{\varphi.1} \\
&\quad + \frac{\lambda_{u\psi}\omega_u}{\omega_\varphi}\left[\int \overline{W}_1(r)\overline{W}_1(r)'dr\right]^{-1/2} \int W_1 dW_1 \\
&= \frac{\sigma_{\varphi.u}}{\omega_\varphi}\left[\int \overline{W}_1(r)\overline{W}_1(r)'dr\right]^{-1/2} \int W_1 dW_{\varphi.1} \\
&\quad + \frac{\sigma_{u\psi}}{\omega_\varphi\omega_u}\left[\int \overline{W}_1(r)\overline{W}_1(r)'dr\right]^{-1/2} \int W_1 dW_1
\end{aligned}$$

Notice that

$$\left(\frac{\sigma_{\varphi.u}}{\omega_\varphi}\right)^2 = \frac{\omega_\varphi^2 - \sigma_{u\varphi}^2/\omega_u^2}{\omega_\varphi^2} = \frac{\omega_\varphi^2\omega_u^2 - \sigma_{u\varphi}^2}{\omega_\varphi^2\omega_u^2} = 1 - \frac{\sigma_{u\varphi}^2}{\omega_\varphi^2\omega_u^2}$$

The limiting distribution of $t_{\widehat{\rho}}$ can then be decomposed into

$$\begin{aligned}
t_{\widehat{\rho}} &= \frac{\widehat{\rho}}{se(\widehat{\rho})} \Rightarrow \sqrt{1-\lambda^2}\left(\int W_X(r)^2 dr\right)^{-1/2} \int W_X(r) dW_{\varphi.1}(r) \\
&\quad + \lambda\left(\int W_X(r)^2 dr\right)^{-1/2} \int W_X dW_1 \\
&= \sqrt{1-\lambda^2} N(0,1) + \lambda\left(\int W_X(r)^2 dr\right)^{-1/2} \int W_X dW_1
\end{aligned}$$

■

References

[1] Balke, N.S., and T.B. Fomby (1994). Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time seires. *Journal of Applied Econometrics* 9, 181-200.

[2] Beelders, O. (1998) Adaptive estimation and unit root tests. *Working Paper.* Emory University.

[3] Bidarkota, P. V. (2000). Asymmetries in the conditional mean dynamics of real GNP: robust evidence. *The Review of Economics and Statistics*, 82, 153-157.

[4] Bickel, P. (1982). On adaptive estimation. *Annals of Statistics*, 10, 647-671.

[5] Blanchard, O.J., and M. W. Watson (1986). Are business cycles all alike? in R.J. Gordon (Ed.), *The American Business Cycle: Continuity and Change* (University of Chicago Press, 123-79).

[6] Campbell, John Y. and N. Gregory Mankiw (1987). Are Output Fluctuations Transitory? *The Quarterly Journal of Economics.*

[7] Cheung, Y.W. and K.S. Lai (1997). Bandwidth selection, prewhitening, and the power of the Phillips-Perron test. *Econometric Theory* 13, 679-691

[8] Cochrane, John H. (1988) How big is the random walk in RGNP? *Journal of Political Economy,* 96,893-920.

[9] Cox, D. and Llatas, I. (1991) Maximum likelihood type estimation for nearly nonstationary autoregressive time series.*Annals of Statistics* 19, 1109-1128.

[10] Falk, B., and C-H. Wang (2003). Testing long run PPP with infinite-variance returns. Journal of Applied Econometrics, 18, 471-484.

[11] Friedman, M. (1969). The Optimum quantity of money. *In the Optimum Quantity of Money, and other Essays,* Aldine publishing Company: Chicago.

[12] Grenander, U., and M. Rosenblat (1957) *Statistical Analysis of Stationary Time Series*. New York: John Willey.

[13] Hall, D. and B. Joiner. (1982) Representations of the space of distributions useful in robust estimation of location. *Biometrika*, 69, 55-59.

[14] Hansen, B. E. (1995) Rethinking the univariate approach to unit root testing (using covariates to increase power). *Econometric Theory*, 11, 1148-1171.

[15] Hansen, B. and S. Lee. (1994) Asymptotic Theory for Garch(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, 10, 29-52.

[16] Huber, P.J. (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35, 73-101.

[17] Huber, P.J. (1973) Robust regression: asymptotics, conjectures, and monte carlo. *Annals of Statistics*, 1, 799-821.

[18] Knight, K. (1991) Limiting Theory for M estimators in an integrated finite variance process. *Econometric Theory*, 7, 200-212.

[19] Lucas, A. (1995). An outlier robust unit root test with an application to the extended Nelson-Plosser data. *Journal of Econometrics* 66, 153-174.

[20] Lucas, A. (1995). Unit root tests based on M estimators. Econometric Theory 11, 331-346.

[21] Nelson, Charles R. and Plosser, Charles I. (1982) Trends and random walks in macroeconomic time series: some evidence and implications, *Journal of Monetary Economics*, 10, 139-162

[22] Phillips, P.C.B. (1995). Robust nonstationary regression. *Econometric Theory* 12, 912-951.

[23] Phillips, P.C.B. and Z. Xiao (1998). A primer on unit root testing. *Journal of Economic Survey* 12, 423-469.

[24] Potscher, B., I. Prucha. (1986) A class of partially adaptive one-step M estimators for the nonlinear regression model with dependent observations. *Journal of Econometrics* 32, 219-251

[25] Premaratne, G., A. Bera (2005). A test for symmetry with leptokurtic financial data. *Journal of Financial Econometrics*, 3, 169-187.

[26] Rose, A. K. (1998). "Is the real interest rate stable?" The Journal of Finance, 43: 1095-1112.

[27] Rothenberg, T., and J. Stock (1997) Inference in a nearly integrated autoregressive model with nonormal innovations. *Journal of Econometrics* 80, 269-286.

[28] Seo, B. Essays on Time Series Econometrics. *PhD Dissertation*, University of Rochester.

[29] Schmidt, P. and P.C.B. Phillips (1992) Testing for a unit root in the presence of deterministic trend. *Oxford Bulletin of Economics and Statistics*, 54, 257-288.

[30] Scheickman, J. A., and B. LeBaron (1989). Non-linear dynamics and GNP data. *In W.A. Barnett et al. (Eds.), Economic Complexity: chaos, sunspots, bubles, and non-linearity* (Cambridge University Press), 213-227.

[31] Stock, J.H. (1995) Unit roots, structural breaks, and trends. In R.F. Engle and Mcfadden (eds.), *Handbook of Econometrics*, vol.4, pp. 2739-2841. Amsterdam: North-Holland.

[32] Thompson, S. B. (2004). Robust Tests for Unit Root Hypothesis Should not be "Modified". forthcoming in Econometric Theory.

[33] White, H. and I. Domowitz (1984). Nonlinear regression with dependent observations. *Econometrica*, 52, 143-161.

[34] Xiao, Zhijie (2001). Likelihood Based Inference for Trending Time Series with a Root Near Unity, *Econometric Theory*, 17, 1082-1112.