Robust inference in nonstationary time series models

Author: Zhijie Xiao

Persistent link: http://hdl.handle.net/2345/2683

This work is posted on eScholarship@BC, Boston College University Libraries.

Post-print version of an article published in Journal of Econometrics 169(2): 211-223. doi:10.1016/j.jeconom.2012.01.027.

These materials are made available for use in research, teaching and private study, pursuant to U.S. Copyright Law. The user must assume full responsibility for any use of the materials, including but not limited to, infringement of copyright and publication rights of reproduced materials. Any materials used for academic research or otherwise should be fully credited with the source. The publisher or original authors may retain copyright to the materials.

Robust Inference in Nonstationary Time Series Models

Zhijie Xiao*

Boston College

November 16, 2010

Abstract

This paper studies robust inference in unit root and cointegration models. The analysis covers a range of important inference problems including: testing stationarity against unit roots; testing for structure change in nonstationary regressions; and testing for cointegration. We analyze these inference problems in a unified regression framework, although seperate analysis is given for each specific case when it is needed. The proposed inference procedures are constructed based on residuals of robust M-estimations. The limiting behavior of the proposed tests is investigated, and a monte carlo experiment is conducted. The proposed tests are easy to use and have advantages in the presence of non-Gaussian data.

JEL Code: C12, C22. Key Words: Cointegration, M-estimation, Robust Inference, Structural Change, Unit Root.

^{*}I thank Bobby Mariano, Jun Yu, Peter Phillips, Roger Koenker, Ted Juhl, two referees, and participants at the 2008 Singapore conference in honor of Peter Phillips for their helpful comments. Address correspondence: Department of Economics, Boston College, Chestnut Hill, MA 02467. Tel: 617-552-1709. Email: xiaoz@bc.edu. And Tsinghua University, Beijing, China.

1 Introduction

Nonstationarity is an important empirical feature in many economic and financial time series. Since the influential article by Nelson and Plosser (1982), hundreds of economic time series have been examined by unit root tests (against a stationary alternative) or stationarity tests (against a unit root alternative). The problem is particularly delicate and interesting in the multivariate case, where several time series may have nonstationary characteristics and the (cointegration) interrelationships of these variables are the main object of study.

In the past two decades, econometricians have focused a great deal of attention on the development of inference procedures in time series models with nonstationarity under the null or the alternative. The majority of these procedures are constructed based on least square methods in linear regression models and have likelihood interpretations when the data are iid Gaussian. In the absence of Gaussianity, asymptotic results of these procedures generally still hold but these methods are usually less efficient than inference methods that exploit the distributional information. Monte Carlo evidence indicates that the least squares estimator can be very sensitive to certain outliers, and inference procedures based on the least square estimation may have poor performance in these cases. In empirical analysis, many applications in nonstationary time series involve financial data like exchange rates whose distributions are heavy-tailed and thus not normally distributed. It is therefore important to consider estimation and inference procedures which are robust to departures from Gaussianity and can be applied to nonstationary time series.

The current paper investigates robust inference in nonstationary time series models. The study in this paper covers several important inference problems in unit roots and cointegration that have attracted a great deal of research attention in the recent 20 years, including: testing for trend stationarity against the unit root alternative; testing for cointegration against the alternative of no-cointegration; and testing for structural change in time series regressions with nonstationary regressors, as well as many other similar inference problems. We study these inference problems in a unified regression framework and construct testing procedures based on residuals from robust M-estimations. Asymptotic results are developed, and separate discussions on each special case are provided whenever necessary. A Monte Carlo experiment is conducted to compare the M-estimation based inference with the OLS regression based inference.

Related to this topic, Phillips (1995) studied robust cointegrating regression estimation and developed Fully Modified M-estimators for cointegrating regressions. Juhl (1999) proposed a test for cointegration using M estimators based on singular value decomposition. deJong, Amsler and Schimidt (2007) developed an indicator based KPSS test for stationarity around mean. Xiao (2009) studied quantile regression of cointegrated time series. Another related inference problem (not covered in the current paper) is testing for the null of a unit root in autoregression processes with nonnormal innovations. Cox and Llatas (1991) considered M-estimation and inference for nearly integrated autoregressive models without deterministic trends. Xiao (2001(b)) studied M-estimation and unit root testing for trending processes with nonnormal innovations. Lucas (1995) considered unit root tests based on M-estimators. Herce (1996) studied a unit root test based on LAD estimation and Koenker and Xiao (2004) proposed a quantile regression based unit root test.

The rest of this paper is organized as follows: The inference problems and a unified regression model are introduced in Section 2; Section 3 proposes the testing procedures, asymptotic behavior of these tests are analyzed. Further discussions related to the implementation of the proposed tests are given in Section 4. Section 5 reports monte carlo results and concludes. A sketch of proofs is provided in the Appendix.

2 The Inference Problems and A Unified Regression Model

2.1 The Inference Problems

In this paper, we study robust inference in nonstationary time series based on a unified regression framework. For convenience, we first introduce three important inference problems in the unit root and cointegration literature, and our later analysis in this paper will be focused on these three problems. We emphasize that although our discussion only focuses on these three inference problems, there is no doubt that these models can be modified in various ways and the analysis in this paper can be extended to other inference problems - for example, similar testing procedures can be constructed in testing for fractional cointegration, or testing for stationarity against long memory alternatives.

2.1.1 Testing Trend Stationarity v.s. Unit Root

The first inference problem is testing for trend stationarity against the unit root alternative. In this problem, the observed time series y_t is modelled as

$$y_t = \theta' z_t + u_t, \ t = 1, ..., n,$$

where the regressor is a deterministic component which can be expressed as $\theta' z_t$, θ is a vector of coefficient and z_t is a deterministic function of known form, say, $z_t = (1, t, ..., t^{q-1})'$. The leading cases of the deterministic component are $z_t = 1$ and $z_t = (1, t)$. The stochastic component of y_t is given by u_t . We are interested in testing $H_{01} : u_t$ is a stationary (I(0)) process, against the alternative H_{A1} : u_t is an unit root (I(1)) process.

2.1.2 Testing for Structure Change in Nonstationary Regression Models

The second inference problem is testing for parameter instability in regression models with I(1) regressors. In particular, we consider the following regression

$$y_t = \theta'_t X_t + u_t,$$

where $X_t = (z'_t, x'_t)'$ is a vector of nonstationary regressors and u_t is an I(0) residual. More specifically, x_t is an *p*-dimensional vector of I(1) regressors, and z_t is a *q*-dimensional deterministic function of known form. We want to test the null of constant regression parameters θ , i.e. H_{02} : $\theta_t = \theta_0$, against alternatives that θ_t is not constant over *t*. To study asymptotic power properties of the tests, we consider sequences of local alternatives H_{A2} : $\theta_t = \theta_0 + n^{-1/2} D_n^{-1} g(t/n)$, where $g(\cdot)$ is an arbitrary k-dimensional bounded function defined on the [0,1] interval, k = p + q is the dimension of X and D_n is a scaling matrix (that standardize the regressors) defined later in this Section. In the simple special case $g(r) = \Delta \theta 1 (r \leq r^*)$, this corresponds to the leading case of the alternative with a one time shift.

2.1.3 Testing for Cointegration

In the third model we are interested in testing the null hypothesis of cointegration against the alternative of no cointegration. In particular, we consider I(1) processes y_t and x_t (1 and p-dimensional respectively) and test H_{03} : y_t and x_t are cointegrated, against the alternative H_{A3} : y_t and x_t are not cointegrated.

The above inference problem can be tested based on the residuals of a cointegrating regression. If we consider a regression of y_t on x_t ,

$$y_t = \beta' x_t + v_t, \tag{1}$$

we may construct tests for cointegration based on stationarity of the residuals v_t from the above regression. However, in many economic applications of cointegration, the residual term v_t in the cointegrating regression (1) is correlated with regressors x_t . To deal with this endogeneity problem, we need to modify either the original regression (1) or the estimator $\hat{\beta}$. Several approaches have been suggested in the previous literature to deal with endogeneity, including nonparametric fullymodification on the estimator from the original regression; augmented regression using leads and lags of Δx_t ; and the canonical regression method. In this paper, for the convenience of putting these inference problems in a unified regression framework (but notice that other methods can be applied to M-estimation as well), we introduce leads and lags of Δx_t into the regression model to absorb the correlation. In particular, we assume that v_t has the following representation

$$v_t = \sum_{j=-K}^{K} \pi'_j \Delta x_{t-j} + u_t,$$

where u_t is a stationary process under the null and $E(\Delta x_{t-j}u_t) = 0$, for any j, and consider the following cointegrating regression with leads and lags:

$$y_t = \beta' x_t + \sum_{j=-K}^{K} \pi'_j \Delta x_{t-j} + u_t.$$

More generally, we may allow for a general deterministic component z_t in the cointegration model and consider:

$$y_t = \alpha' z_t + \beta' x_t + \sum_{j=-K}^{K} \pi'_j \Delta x_{t-j} + u_t = \theta' X_t + u_t.$$
(2)

The null hypothesis of cointegration then corresponds to "the residual term u_t in the cointegrating regression (2) is I(0)", and the alternative of no-cointegration corresponds to an integrated residual process u_t .

2.2 A Unified Regression Model

To analyze the inference problems that we list in Section 2.1, we consider the following linear regression model

$$y_t = \theta' X_t + u_t,\tag{3}$$

where y_t is the observed time series, X_t is the k-dimension vector of regressors, and u_t is the residual process. The general form of the regressors can be written as $X_t = (z'_t, x'_t, s'_t)'$, where z_t is a q-dimensional deterministic function of known form; x_t is an p dimensional vector of I(1) process; and s_t is a p_2 -dimensional vector of I(0) process, $k = p + q + p_2$. The three inference problems discussed in Section 2.1 correspond to three special cases where X_t takes different forms:

- 1. In the first inference problem of decision between I(0) v.s. I(1), p = 0 and $p_2 = 0$, $X_t = z_t$ is a q-dimensional deterministic regressor. The residual process u_t is I(0) (stationary) under the null H_{01} and I(1) (unit root process) under the alternative H_{A1} .
- 2. In inference problem 2 of testing for structure changes in nonstationary regression models, $p_2 = 0$, thus k = p + q, and $X_t = (z'_t, x'_t)'$. In this model, the regressors z_t and x_t are deterministic and stochastic trends. We want to test for parameter instability of θ . The residual u_t is I(0) under both the null and the alternative.
- 3. In the third problem of testing the null of cointegration, $X_t = (z'_t, x'_t, s'_t)'$, where x_t is I(1) and s_t is an I(0) vector of leads and lags of Δx_t , i.e. Δx_{t-j} . The residual process u_t is I(0) under the null of cointegration between y_t and x_t ; and u_t is I(1) under the alternative of no cointegration.

We construct testing procedures based on the residuals from regression (3). A common feature of these inference problems is: The fluctuation of the regression residual processes from model (3) under the alternative hypotheses is larger than the fluctuation under the null. Consequently, we may formulate these inference problems based on regression (3) and test them by looking at the fluctuation in the residuals. If the residuals display too much fluctuation, we should reject the null hypotheses. This general principle applies to all three inference problems that we considered in Section 2.1, and can be easily extended to other similar inference problems, although the detailed behavior of the tests in these models are different.

If u_t were known, we could look at its fluctuation via the partial sum process:

$$U_n(r) = \frac{1}{\omega_u \sqrt{n}} \sum_{t=1}^{[nr]} u_t,\tag{4}$$

where ω_u^2 is the long run variance of u_t . Under the null hypotheses and appropriate regularity assumptions, $U_n(r)$ converges weakly to a standard Brownian motion W(r). Consider a continuous functional $h(\cdot)$ that measures the fluctuation of the partial sum process, we can use $h(U_n(r))$ as a test statistic for the null. In practice, u_t is unobservable. In order to test these hypotheses, we need to estimate θ from regression model (3) first and then look at the fluctuation in the estimated residuals. If we estimate regression (3) by OLS method, denote the OLS regression estimator of θ as $\check{\theta}$, and denote the residuals as $\check{u}_t = y_t - \check{\theta}' X_t$, then testing procedures can be constructed based the partial sum process of the OLS estimated residuals \check{u}_t .

Residual based tests are important inference methods in unit root and cointegration models. Based on OLS regression residuals, tests for trend stationarity against unit roots were proposed by Kwiatkowski, Phillips, Schmidt and Shin (hereafter KPSS, 1992), Xiao (2001(a)); tests for parameter instability in nonstationary regression models were studied by Hansen (1992), Hao and Inder (1996); and tests for cointegration were studied by Shin (1999), Xiao and Phillips (2002) among many other researchers.

Testing procedures based on least square residuals usually have relatively good performance when the data are Gaussian, but are less efficient than more robust methods in the presence of non-Gaussianity. Giving the cumulated evidence of nonnormality in many financial and economic time series, it is useful to consider inference procedures based on more robust methods. Following the idea of Huber (1964) for the location problem, Relles (1968) and Huber (1973) introduced a class of the so-called M estimators which generally have good properties over a wide range of distributions.

If we consider a general criterion function ρ , the so-called M estimator of θ based on model (3) can be obtained from the following optimization problem:

$$\widehat{\theta} = \arg\min_{\theta} \sum_{t=1}^{n} \rho \left(y_t - \theta' X_t \right).$$
(5)

In the simple case that u_t are unobserved i.i.d. errors with log density $-\rho(u)$, $\sum_t \rho(y_t - \theta' X_t)$ is the log likelihood function of the random sample and the above M estimator corresponds to the conditional MLE estimator of θ .

Examples

Example 1: If $-\rho(\cdot)$ is chosen as the log density of normal distribution, $\rho(u) = \frac{1}{2}u^2$, the M estimation reduces to the conventional OLS estimation.

Example 2: If we take $-\rho(\cdot)$ to be the log density of a logistic distribution, then

$$\rho\left(x\right) = -\log f(x)$$

where

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2}, -\infty < x < \infty.$$

Example 3: If we take $-\rho(\cdot)$ to be the log density of a double exponential distribution, then $\rho(x) = |x|, -\infty < x < \infty$.

In this paper, we study inference procedures for nonstationary time series based on residuals from the above M estimation regression.

2.3 Asymptotic Behavior of The M Estimators

For convenience of the asymptotic analysis, we first introduce some assumptions. We should stress that we are not seeking to achieve the weakest possible regularity conditions in asymptotic analysis. Some conditions are assumed for the simplicity of proofs, and may be replaced by weaker conditions without affecting the asymptotic results.

We first introduce some assumptions for the regression error u_t and the regressors X_t . ASSUMPTION 1: u_t is strictly stationary and strong mixing with mixing numbers $\alpha(m)$ that satisfy the summability condition: $\sum_{1}^{\infty} \alpha(m)^{(b-a)/ba} < \infty$ for some b > a > 2.

ASSUMPTION 2: (i) z_t is a q-dimensional deterministic function of known form, and there exists a scaling matrix G_n such that $G_n^{-1}z_{[nr]} \to Z(r)$, as $n \to \infty$, uniformly in $r \in [0, 1]$. (ii) x_t is an p dimensional vector of I(1) regressors such that $x_t = x_{t-1} + \xi_t$, $t = 1, \dots, n$, where the initial observation of x_t is taken to be any random variable with finite variance. $\{\xi_t\}$ satisfies the same mixing assumption in Assumption 1 and has b-th moments. (iii) s_t is an p_2 dimensional vector of random variables satisfy the same assumptions as ξ_t .

Assumptions 1 and 2 assume that u_t , Δx_t , and s_t are weakly dependent and satisfy the specified mixing condition. The above assumptions ensure some invariance principles that are convenient for our asymptotic analysis and there is no doubt that they could be replaced by a variety of similar conditions. Under Assumption 2, the partial sum process constructed from ξ_t satisfies a multivariate invariance principle:

$$n^{-1/2} \sum_{t=1}^{[nr]} \xi_t \Rightarrow B(r), 0 \le r \le 1,$$

where B(r) is a p dimensional Brownian motion with long-run variance Ω . We assume that Ω is nonsingular so that elements within x_t are not cointegrated themselve. Since the regressors X_t are different in the three models, the convergence rate of $\hat{\theta}$ is dependent on the choice of model. We introduce a standardization matrix D_n that take different forms in different models. Corresponding to the general form of $X_t = (z'_t, x'_t, s'_t)'$, $D_n = diag[G_n, G_{2n}, I_{p_2}]$, where $G_{2n} = diag[\sqrt{n}, \dots, \sqrt{n}]$ is a p-dimensional diagonal matrix and I_{p_2} is a p_2 -dimensional identity matrix. In the first inference problem, $D_n = G_n$. For the leading case of a linear trend, $G_n = diag[1, n]$ and Z(r) = (1, r)'. If x_t is a general p-th order polynomial trend, $G_n = diag[1, n, \dots, n^p]$ and $Z(r) = (1, r, \dots, r^p)$. In the second inference problem, $D_n = diag[G_n, G_{2n}]$. For example, if the deterministic component is only an intercept term, then $D_n = diag[1, \sqrt{n}, \dots, \sqrt{n}]$).

We also need some assumptions regarding the criterion function. The following assumption is a standard condition in M-estimation asymptotic analysis. ASSUMPTION 3: $\rho(\cdot)$ possesses derivatives ψ and ψ' . $\psi(u_t)$ has b-th moments for some b > a > 2,

ASSUMPTION 3: $\rho(\cdot)$ possesses derivatives ψ and ψ . $\psi(u_t)$ has b-th moments for some b > a > 2, $E[\psi(u_t)] = 0, \ \delta = E[\psi'(u_t)], \ \text{and} \ \psi'$ is Lipschitz continuous.

The differentiability of ρ enables us to conduct a Taylor expansion for the criterion function. Many M-estimation procedures satisfy this assumption, but some do not. This assumption also rules out example 3 in section 2.2. In example 3, $-\rho(u)$ is the log density of a double exponential distribution and is not everywhere smooth. However, the derivative of $\rho(u) = |u|$ exists except at the point u = 0, and is given by $\psi(u) = 1 - 2I(u < 0)$. Residual based test may still be constructed based on $\psi(u)$. To facilitate asymptotic analysis of nonstationary M estimation with nonsmooth criterion function like the case of example 3, we make the following assumption as an alternative of Assumption 3.

ASSUMPTION 3': $\rho(\cdot)$ possesses derivatives $\psi(\cdot)$ and $\psi'(\cdot)$ everywhere except a finite number of points. $\psi(\cdot)$ and $\psi'(\cdot)$ are bounded measurable functions and can be treated as generalized functions. u_t has a continuous density with f(u) > 0 on $\mathcal{U} = \{u : 0 < F(u) < 1\}, E[\psi(u_t)] = 0, \delta = E[\psi'(u_t)],$ and $\psi(u_t)$ has b-th moments for some b > a > 2.

In Assumption 3', we allow for the criterion function to be non-differentiability at a finite number of points, and assume that the criterion functions can be treated as generalized functions as Gel'fand and Vilenkin (1964), Phillips (1995). The moment conditions on $\psi(u)$ is needed to establish the weak convergence results. The asymptotic behavior of the residuals \hat{u}_t will be dependent on the limiting bahavior of $\hat{\theta}$, which in turn depend on the weak limit of the partial sums of $\psi(u_t)$. Under our mixing condition and moment conditions, as n goes to ∞ , $n^{-1/2} \sum_{1}^{[nr]} \psi(u_t) \Rightarrow B_{\psi}(r)$, where $B_{\psi}(r)$ is a Brownian motion with variances ω_{ψ}^2 .

The following assumption is simply an analog of the first order condition. ASSUMPTION 4: The estimator $\hat{\theta}$ satisfy $n^{-1/2} \sum_{t=1}^{n} D_n^{-1} X_t \psi \left(y_t - \hat{\theta}' X_t \right) = o_p(1)$.

The results of this paper can be obtained under different types of identification conditions. In this paper, to cover a wide range of models and to focus on the discussion of testing procedures, we follow a similar approach as Phillips (1995) and Xiao(2001b), and assume that the following high level condition holds under the null.

ASSUMPTION 5: $n^{1/2}D_n(\widehat{\theta} - \theta) = o_p(n^{1/4}).$

Assumption 5 is standard in the development of M-estimator asymptotics. It is similar to Assumption (b) in Theorem 5.1 of Phillips (1995) and the assumption on $\tilde{\varepsilon}_t - \varepsilon_t$ in Theorem 1 of Lucas (1995). Notice that alternative regularity assumptions can be made in place of the above assumptions. For example, we may derive the asymptotics based on, say, convexity of the criterion function. For other types of regularity assumptions, see, e.g., Knight (1989) and Pollard (1991) for asymptotic theory with convex criterion functions.

We summarize the asymptotic behavior of the M-estimator in Theorem 1.

Theorem 1: Under Assumptions 1 - 5, as $n \to \infty$,

$$\sqrt{n}D_n(\widehat{\theta} - \theta) \Rightarrow \frac{1}{\delta} \left[\begin{array}{c} \left[\int \overline{B}(r)\overline{B}(r)'dr \right]^{-1} \int \overline{B}(r)dB_{\psi}(r) \\ \Phi \end{array} \right],$$

where the partition is conformable with dividing the regressors as $(z'_t, x'_t)'$ and s_t , $B_{\psi}(r) = \omega_{\psi} W_1(r)$, and $\overline{B}(r) = (Z(r)', B(r)')'$, $B(r) = \Omega^{1/2} W_2(r)$, $W_1(r)$ and $W_2(r)$ are 1 and p dimensional standardized Brownian motions and are independent with each other, and Φ is a p_2 dimensional multivariate normal variate with covariance matrix equals to $\Gamma_s(0)^{-1}\Omega_{s\psi}\Gamma_s(0)^{-1}$,

where $\Omega_{s\psi} = \sum_{h=-\infty}^{\infty} \gamma_{\psi}(h) \Gamma_s(h)$, $\gamma_{\psi}(h) = \mathbb{E}[\psi(u_t) \psi(u_{t+h})]$, $\Gamma_s(h) = \mathbb{E}[s_t s'_{t+h}]$. For the special case corresponding to the first inference problem:

$$\sqrt{n}D_n(\hat{\theta}-\theta) \Rightarrow \frac{1}{\delta} \left[\int Z(r)Z(r)'dr\right]^{-1} \int Z(r)dB_{\psi}(r).$$

3 Inference Based on M-Estimation Residuals

3.1 The Tests and Their Limiting Behavior Under The Null

We want to construct testing procedures based on residuals from the above M-estimation regression. In particular, we test the aforementioned inference problems by looking at the fluctuation in the residual processes $\hat{u}_t = y_t - \hat{\theta}' X_t$. To look at the fluctuation in \hat{u}_t , it might be natural to consider constructing testing statistics based on the partial sum process

$$\widehat{U}_n(r) = \frac{1}{\widehat{\omega}_u \sqrt{n}} \sum_{t=1}^{[nr]} \widehat{u}_t,$$

where $\hat{\omega}_u^2$ is a consistent estimator of ω_u^2 . Under the null hypotheses and additional regularity conditions, the partial sum process $\hat{U}_n(r)$ converges to functionals of Brownian motions. For Inference Problem 1,

$$n^{-1/2} \sum_{t=1}^{[nr]} \widehat{u}_t \Rightarrow B_u(r) - \frac{1}{\delta} \int_0^1 dB_\psi(s) Z(s)' \left[\int_0^1 Z(s) Z(s)' ds \right]^{-1} \int_0^r Z(s) ds,$$

and for Inference Problems 2 and 3,

$$n^{-1/2} \sum_{t=1}^{[nr]} \widehat{u}_t \Rightarrow B_u(r) - \frac{1}{\delta} \int_0^1 dB_\psi(s)\overline{B}(s)' \left[\int_0^1 \overline{B}(s)\overline{B}(s)' ds \right]^{-1} \int_0^r \overline{B}(s) ds$$

where $B_u(r)$ is the weak limit of $n^{-1/2} \sum_{t=1}^{[nr]} u_t$, and $B(\cdot)$ is uncorrectated with $B_u(r)$ and $B_{\psi}(r)$.

The limiting variate of the M-estimator is generally a functional of the limiting trend function Z(r) and B(s) and, more importantly, the Brownian motion $B_{\psi}(r)$. Notice that $B_u(r)$ and $B_{\psi}(r)$ are correlated Brownian motions and have different variances (unless OLS estimation is used). Thus, the limiting processes of the partial sums of the residuals are dependent on nuisance parameters that reflect the correlation between $B_u(r)$ and $B_{\psi}(r)$. Consequently, a simple functional of $\hat{U}_n(r)$ can not be used as test statistics for these inference problems.

For this reason, we consider the partial sum process based on the *score transformation* of \hat{u}_t , i.e. the process $\psi(\hat{u}_t)$, and construct residual based tests using the score process. The asymptotic behavior of the partial sum of the score process $\psi(\hat{u}_t)$ is summarized in the following Theorem.

Theorem 2: Under Assumptions 1 - 5, as $n \to \infty$,

$$n^{-1/2} \sum_{t=1}^{[nr]} \psi\left(\widehat{u}_t\right) \Rightarrow B_{\psi}(r) - \int_0^1 dB_{\psi}(s)\overline{B}(s)' \left[\int_0^1 \overline{B}(s)\overline{B}(s)'ds\right]^{-1} \int_0^r \overline{B}(s)ds.$$

For the special case corresponding to the first inference problem:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\psi\left(\widehat{u}_{t}\right) \Rightarrow B_{\psi}(r) - \int_{0}^{1} dB_{\psi}(r)Z(r) \left[\int_{0}^{1} Z(s)Z(s)'ds\right]^{-1} \int_{0}^{r} Z(s)ds.$$

- **Remark 1:** Notice that, the stationary covariates s_t do not affect the limit of the partial sum process, and thus the limits corresponding to the second and third models are the same. However, although the limiting null process of the partial sum of the score process $\psi(\hat{u}_t)$ are the same for these models, their asymptotic behavior under the alternatives are different - see additional results in Section 3.2.
- **Remark 2:** An alternative approach to deal with the nuisance parameters in the residual process is to use bootstrap or simulation-based inference.

After the transformation, the leading term and the drift term coming from the preliminary estimation in the limiting partial sum process contain the same component $B_{\psi}(r)$ which can be removed by a simple standardization. Thus, we may test our hypotheses based on the following standardized partial sums of the score process

$$\widetilde{W}_{n}(r) = \frac{1}{\widehat{\omega}_{\psi}\sqrt{n}} \sum_{t=1}^{[nr]} \psi\left(\widehat{u}_{t}\right), \qquad (6)$$

where $\hat{\omega}_{\psi}^2$ is a consistent nonparametric estimate of ω_{ψ}^2 ,

$$\widehat{\omega}_{\psi}^{2} = \sum_{j=-\ell}^{\ell} k(\frac{j}{\ell}) C_{\widehat{\psi}\widehat{\psi}}(j).$$
(7)

In this formula, $C_{\widehat{\psi}\widehat{\psi}}(j)$ is the sample covariance defined as $n^{-1}\sum'\psi(\widehat{u}_t)\psi(\widehat{u}_{t+j})$, where \sum' signifies summation over $1 \leq t, t+j \leq n, k(\cdot)$ is the lag window defined on [-1,1] with k(0) = 1, and ℓ is the bandwidth parameter satisfying the property that $\ell \to \infty$ and $\ell/n \to 0$ as the sample size $n \to \infty$, (see, *inter alia*, Andrews, 1991). Inference procedures may be constructed based on a continuous functional $h(\cdot)$ that measures the fluctuation of $\widetilde{W}_n(r)$. It will be convenient in what follows to make the following assumptions about the functional $h(\cdot)$. In practice, the leading choices of $h(\cdot)$ are the classical Kolmogoroff-Smirnoff or Cramer-von Mises type measures.

ASSUMPTION 6: h is continuous and $h(\theta\lambda(r)) = \theta^{\tau}h(\lambda(r))$ for some $\tau > 0$.

Limiting null distribution of the tests can be immediately obtained by the continuous mapping theorem. We summarize the results of the proposed tests in Theorem 3.

Theorem 3: Under Assumptions 1 -6, as $n \to \infty$,

$$h(\widetilde{W}_n(r)) \Rightarrow h(\widetilde{W}(r)),$$

where

$$\widetilde{W}(r) = W_1(r) - \int_0^1 dW_1(r)\overline{W}_2(r) \left[\int_0^1 \overline{W}_2(s)\overline{W}_2(s)'ds\right]^{-1} \int_0^r \overline{W}_2(s)ds$$

and $\overline{W}_2(r)' = (Z(r)', W_2(r)')'$. For the special case corresponding to model 1:

$$\widetilde{W}(r) = W(r) - \int_0^1 dW(r)Z(r) \left[\int_0^1 Z(s)Z(s)'ds\right]^{-1} \int_0^r Z(s)ds$$

Critical values of these tests depend on the choice of Z(r) and metric $h(\cdot)$, we provide sources of critical values for the classical Kolmogoroff-Smirnoff or Cramer-von Mises measures in section 4. Also see section 4 for more discussions on the choices of h.

Remark 3: We introduced the tests by looking at the fluctuation in the residuals. Corresponding to certain choice of the metric $h(\cdot)$, the testing procedures can also be derived based on alternative ways (say, based on a LM principle under appropriate additional assumptions).

Examples (Continued)

- **Example 1:** If $-\rho(\cdot)$ is chosen as the log normal, the corresponding score function is simply $\psi(u) = u$.
- **Example 2:** If we take $-\rho(\cdot)$ to be the log density of a logistic distribution, then the corresponding score function is the well-known Wilcoxon (or Hodges-Lehmann) score,

$$\psi(u) = -\frac{f'(u)}{f(u)} = 1 - \frac{2e^{-u}}{1 + e^{-u}} = \frac{e^u - 1}{e^u + 1}.$$

Example 3: If we take $-\rho(\cdot)$ to be the log density of a double exponential distribution, the corresponding score function is given by $\psi(u) = 1 - 2I(u < 0)$.

3.2 Asymptotic Behavior Under The Alternatives

The asymptotic behavior of the proposed test under the alternatives are different across the three models in Section 2.1. In particular, the power properties of the first and third inference problems are similar, and are different from the second inference problem - testing for structure changes. We first analyze the power property of the second inference problem, and then investigate the asymptotic behavior of the tests corresponding to the first and the third models.

For the second model, the asymptotic power property under the local alternatives H_{A2} : $\theta_t = \theta_0 + n^{-1/2} D_n^{-1} g(t/n)$, is summarized in Theorem 4 below.

Theorem 4: Under the local alternative H_{A2} and the other regularity assumptions, as $n \to \infty$,

$$\widetilde{W}_n(r) \Rightarrow \widetilde{W}(r) + \frac{\delta}{\omega_{\psi}} \left[\int_0^r g^\top \overline{B} - \int g^\top \overline{BB}^\top \left[\int \overline{BB}^\top \right]^{-1} \int_0^r \overline{B} \right]$$

where $\widetilde{W}(r)$ is the same as the limiting process in Theorem 3 and

$$\int_{0}^{r} g^{\top}\overline{B} = \int_{0}^{r} g(s)^{\top} \overline{B}(s) ds, \quad \int g^{\top}\overline{B}\overline{B}^{\top} = \int_{0}^{1} g(r)^{\top}\overline{B}(r)\overline{B}(r)^{\top} dr,$$
$$\int \overline{B}\overline{B}^{\top} = \int_{0}^{1} \overline{B}(s)\overline{B}(s)' ds, \quad \int_{0}^{r} \overline{B} = \int_{0}^{r} \overline{B}(s) ds.$$

For the first and the third inference problems, under the alternatives H_{A1} or H_{A3} , the residual process is an I(1) process. We first make the following assumption on the model behavior under the alternatives H_{A1} or H_{A3} .

ASSUMPTION 7: Under H_{A1} or H_{A3} , the regression residual process \hat{u}_t is I(1), and $n^{-1/2}\hat{u}_{[nr]} \Rightarrow \eta(r)$, where $\eta(r)$ is a functional of Brownian motions.

For the purpose of analyzing asymptotic behavior of nonlinear transformation (ψ) of an integrated process, we introduce the following concept of asymptotically homogeneous function studied by Park and Phillips (1999).

Definition: A transformation G is said to be *asymptotically homogeneous* iff

$$G(\lambda x) = \nu(\lambda)H(x) + R(x,\lambda)$$

where H is locally integrable, and R has the following property:

- 1. $|R(x,\lambda)| \leq a(\lambda)P(x)$, where $\limsup_{\lambda\to\infty} a(\lambda)/\nu(\lambda) = 0$, and P is locally integrable, or
- 2. $|R(x,\lambda)| \leq b(\lambda)Q(\lambda x)$, where $\limsup_{\lambda\to\infty} b(\lambda)/\nu(\lambda) < \infty$, and Q is locally integrable and vanishes at infinity in the sense $Q(x) \to 0$ as $|x| \to \infty$.

For convenience of our asymptotic analysis, we make the following assumption about the score function.

ASSUMPTION 8: The score function $\psi(\cdot)$ is asymptotically homogeneous.

Many score functions are asymptotically homogeneous with $\nu(\lambda) = 1$.

Examples (Continued)

- **Example 1:** Corresponding to the case of normal, the score function is simply $\psi(u) = u$, and thus $\widetilde{W}_n(r)$ reduces to $\widetilde{U}_n(r)$.
- **Example 2:** If we take $-\rho(\cdot)$ to be the log density of a logistic distribution, then the corresponding score function is the well-known Wilcoxon (or Hodges-Lehmann) score. We re-write the score as

$$\psi(x) = 2\frac{e^x}{e^x + 1} - 1 = 2g(x) - 1,$$

where g(x) is a "distribution function"-like transformation satisfying the property

$$g(x) \to \begin{cases} 1, & x \to \infty \\ 0, & x \to -\infty \end{cases}$$

Thus, using the results of Park and Phillips (1999), $\psi(x)$ is asymptotically homogeneous with $\nu(\cdot) = 1$.

Example 3: If we take $-\rho(\cdot)$ to be the log density of a double exponential distribution, the corresponding score function is given by $\psi(x) = 1 - 2I(x < 0)$. Again, by results of Park and Phillips (1999), $\psi(x)$ is asymptotically homogeneous with $\nu(\cdot) = 1$.

Under the alternative hypotheses H_{A1} or H_{A3} , both the numeritor and the denominator in $\widetilde{W}_n(r)$ diverge as $n \to \infty$. However, the numeritor $n^{-1/2} \sum_{t=1}^{[nr]} \psi(\widehat{u}_t)$ diverges faster than the denominator $\widehat{\omega}_{\psi}$. In particular, under the alternatives and asymptotic homogeneity of the score function, $n^{-1/2} \sum_{t=1}^{[nr]} \psi(\widehat{u}_t) = O_p(\sqrt{n\nu}(\sqrt{n}))$ and $\widehat{\omega}_{\psi} = o_p(\sqrt{n\nu}(\sqrt{n}))$. Thus, the process $\widetilde{W}_n(r)$ diverges to infinity as $n \to \infty$. The asymptotic properties of the tests for the inference problems 1 and 3 under the alternatives H_{A1} and H_{A3} are summarized below.

- **Theorem 5:** Under H_{A1} or H_{A3} and Assumptions 6 8, as $n \to \infty$, $\Pr\left[h(\widetilde{W}_n(r)) > B_n\right] \to 1$ for any nonstochastic sequence $B_n = o(n^{1/2}\ell^{-1/2})$.
- **Remark 4:** The behavior of $h(\widetilde{W}_n(r))$ under the alternative hypotheses H_{A1} or H_{A3} is similar to that of other existing tests (say, the KPSS or CUSUM tests) in the sense that the divergence rate of $h(\widetilde{W}_n(r))$ under the alternative is dependent on the bandwidth expansion rate ℓ .
- **Remark 5:** The stationarity and cointegration tests (against the unit root alternatives) are consistent tests under regularity conditions that ensure invariance principles to hold and long-run variances be consistently estimated. Regularity assumptions such as the existence of long-run variance estimators are typically used in the literature and are sufficient, but not necessary, for, say, the invariance principles. These regularity assumptions restrict the processes into subclasses of (the general sense) I(0) or I(1) processes. Without such type sufficient restrictions on the model, it is impossible to consistently discriminate between I(0) and I(1) processes. For example, Pötscher (2002) show that the minimax risk for estimating the value of the long-run variance is infinite; Faust (1996) studied this issue and concludes that the two classes of processes: I(1) sequences and I(0) sequences, are nearly observationally equivalent if no further restrictions are imposed. See Müller (2008) for a recent study on this issue, and additional discussions in Section 5 on its relation with data-dependent bandwidth selection.

4 Implementation of the Tests

The previous section provides a general residual-based inference method. In principle, any metric h that measures the fluctuation in $\widetilde{W}_n(r)$ can be used in constructing the tests. The classical Kolmogoroff-Smirnoff or Cramer-von Mises type measures are of particular interest. In this section, we discuss several important implementations of this general method.

4.1 The Kolmogoroff-Smirnoff Test

The CUSUM type statistics based on the classical Kolmogoroff-Smirnoff measure is a natural choice of h. We may consider the following test based on the cumulated sum process $\widetilde{W}_n(r)$:

$$h(\widetilde{W}_n(r)) = \sup_{r \in R} \left| \widetilde{W}_n(r) \right|,\tag{8}$$

where $R \subseteq [0, 1]$. Usually we take R = [0, 1], in this case, the testing statistic is

$$\sup_{0 \le r \le 1} \left| \widetilde{W}_n(r) \right| = \max_{1 \le k \le n} \left| \frac{1}{\widehat{\omega}_{\psi} \sqrt{n}} \sum_{t=1}^k \psi\left(\widehat{u}_t\right) \right|.$$

Under Assumptions 1 - 5, as $n \to \infty$,

$$\sup_{r \in R} \left| \widetilde{\Psi}_n(r) \right| \Rightarrow \sup_{r \in R} \left| \widetilde{W}(r) \right|.$$

where $\widetilde{W}(r)$ is given by Theorem 3. For inference problem 1, the limiting process given by

$$\widetilde{W}(r) = W(r) - \int_0^1 dW(r)Z(r) \left[\int_0^1 Z(s)Z(s)'ds\right]^{-1} \int_0^r Z(s)ds$$

is a generalized Brownian bridge process. Critical values of these tests can be found in Xiao (2001a). For inference problems 2 and 3, the limiting distribution of the test depends on both the trend and the known dimension number p, and may be found from Hao and Inder (1996) and Xiao and Phillips (2002).

4.2 The Cramer-von Mises Type Test

Another functional that is frequently used in measuring the fluctuation in a process is the Cramervon Mises metric. For a suitably chosen weight function w(r), we can construct the following Cramer-von Mises type test:

$$h(\widetilde{W}_n(r)) = \int_{r \in R} w(r) [\widetilde{W}_n(r)]^2 dr,$$

where w(r) is a weighting function. Under the null

$$\int_{r \in R} w(r) [\widetilde{W}_n(r)]^2 dr \Rightarrow \int_{r \in R} w(r) \widetilde{W}(r)^2 dr.$$

In particular, choosing w(r) = 1, and R = [0, 1], we obtain the following test with the conventional limiting distribution:

$$\int_0^1 \widetilde{W}_n(r)^2 dr = \frac{1}{\widehat{\omega}_{\psi}^2 n^2} \sum_{k=1}^n \left[\sum_{t=1}^k \psi\left(\widehat{u}_t\right) \right]^2 \Rightarrow \int_0^1 \widetilde{W}(r)^2 dr.$$
(9)

For inference problem 1, the limiting distribution is the same as the KPSS test, and critical values for the leading cases can be found in KPSS (1992). For inference problems 2 and 3, the limiting distributions of the tests are the same as that of the test of Shin (1999), where we can also find the critical values for the leading cases.

4.3 Other Choices of h

In additional to the classical Kolmogoroff-Smirnoff or Cramer-von Mises measures, other choices of h may be used. For example, the MOSUM or Range measures of fluctuation can be used. We may consider the following MOSUM statistic:

$$h(\widetilde{W}_n(r)) = \sup_{r \in R(\theta)} \left| \widetilde{W}_n(r+\theta) - \widetilde{W}_n(r) \right|.$$
(10)

where $0 < \theta < 1$ is a prespecified bandwidth parameter of moving windows, indicating the proportion of \hat{u}_t used to construct the moving sum. $R(\theta) \subseteq [0, 1]$ is an interval such that both r and $r + \theta \in [0, 1]$ when $r \in R(\theta)$. Usually we choose $R(\theta) = [0, 1 - \theta]$ or $R(\theta) = [\delta, 1 - \theta - \delta]$. Under H_0 ,

$$\sup_{r \in R(\theta)} \left| \widetilde{W}_n(r+\theta) - \widetilde{W}_n(r) \right| \Rightarrow \sup_{r \in R(\theta)} \left| \widetilde{W}(r+\theta) - \widetilde{W}(r) \right|.$$

We may also use the range functional based on the difference between the maximum and the minimum values of the empirical process. Testing statistics based the range functional are constructed as follows:

$$h(\widetilde{W}_n(r)) = \sup_{r \in R} \left| \widetilde{W}_n(r) \right| - \inf_{r \in R} \left| \widetilde{W}_n(r) \right|,$$

where $R \subseteq [0, 1]$. Under our conditions and the null,

$$\sup_{r \in R} \left| \widetilde{W}_n(r) \right| - \inf_{r \in R} \left| \widetilde{W}_n(r) \right| \Rightarrow \sup_{r \in R} \left| \widetilde{W}(r) \right| - \inf_{r \in R} \left| \widetilde{W}(r) \right|.$$

5 Monte Carlo

We conduct a Monte Carlo experiment to examine the finite sample performance of the proposed inference procedures. There are many important factors (for example, the short run dynamics and related bandwidth selection issue) that affects the size and power properties of these tests, and there has been a large amount of monte carlo study in the previous literature investigating these issues for the OLS regression based tests. Since the main purpose of this paper is to develop robust inference procedures, we focus our attention on the robustness issue in the monte carlo study. For this reason, we consider mainly models with iid innovations with different tail thickness, and compare the M-estimation based tests with the OLS regression based tests in the presence of various tail behavior. We only consider a few representative cases with correlated errors terms, instead of looking at a wide range short term dynamics.

For the M-estimation in our monte carlo, we consider the LAD estimation, corresponding to example (3) in the previous sections where $\rho(x) = |x|$, and $\psi(u) = 1 - 2I(u < 0)$. For choices of the metric *h* that measures the fluctuation in $\widetilde{W}_n(r)$, we consider both the Kolmogoroff-Smirnoff measure and the Cramer-von Mises measure. Corresponding to the Kolmogoroff-Smirnoff measure, the testing statistic is given by

$$\max_{1 \le k \le n} \left| \frac{1}{\widehat{\omega}_{\psi} \sqrt{n}} \sum_{t=1}^{k} \psi\left(\widehat{u}_{t}\right) \right|,$$

and corresponding to the Cramer-von Mises type measure, the testing statistic is given by

$$\frac{1}{\widehat{\omega}_{\psi}^2 n^2} \sum_{k=1}^n \left[\sum_{t=1}^k \psi\left(\widehat{u}_t\right) \right]^2.$$

We use the Bartlett kernel k(x) = 1 - |x| in estimating $\widehat{\omega}_{\psi}^2$. For the bandwidth parameter ℓ , we consider two choices, denoted as ℓ_1 and ℓ_2 . The first choice is a fixed bandwidth, we choose $\ell_1 = 0$, corresponding to the best choice with iid errors. For the second bandwidth choice, we use the following partially data-dependent bandwidth suggested in Xiao (1998)

$$\ell_2 = \min\{\mu_k \hat{\delta}_k n^{1/(2\nu+1)}, B(n)\},\tag{11}$$

where $\mu_k \hat{\delta}_k n^{1/(2\nu+1)}$ is a data-dependent plug-in bandwidth that minimizes the mean squared error in variance estimation in stationarity time series. In particular, μ_k is a constant associated with the kernel function, δ_k is a function of the kernel and the error distribution, and ν is the characteristic exponent of the kernel. This bandwidth formula $\mu_k \hat{\delta}_k n^{1/(2\nu+1)}$ has been studied by Andrews (1991) in the context of estimation of a covariance matrix for stationary time series. However, this datadependent bandwidth can not be directly used in distinguishing between I(0) and I(1) processes because it diverges to ∞ too fast under the alternatives (also see Xiao (2003), Xiao and Phillips (2002) for related discussions on this issue). For this reason, we introduce an upper bound function B(n) that prevents ℓ_2 from being too big under the alternatives¹. We use $B(n) = [2n^{1/3}]$ in our monte carlo. Corresponding to the Bartlett estimator that we use in our experiment, $\nu = 1$, $\mu_k =$ 1.1447, and, if we use AR(1) plug-in, $\hat{\delta}_k = 4\hat{\rho}^2/(1-\hat{\rho}^2)^4$, where $\hat{\rho}$ is the first order autoregression coefficient. Thus

$$\ell_2 = \min\left\{1.1447 \times \left(\frac{4\hat{\rho}^2 n}{(1-\hat{\rho}^2)^4}\right)^{1/3}, \, [2n^{1/3}]\right\}.$$

We consider the three inference problems discussed in Section 2 based on model (3). For the first problem of testing trend stationarity, we consider the model with $X_t = (1, t)'$. Under the null,

¹Although it would be desirable to consider a fully data dependent bandwidth, it is infeasible in our inference problems. Such a problem is a natural consequence that the unit root ($\alpha = 1$) is on the boundary of the stationary range $|\alpha| < 1$, and is again related to the observational equivalence issue mentioned in Remark 5.

we consider the following two cases of errors: (1) Null1: $u_t = \varepsilon_t$; and (2) Null2: $u_t = \alpha u_{t-1} + \varepsilon_t$, where $\alpha = 0.5$. For the model under the alternative hypothesis, we follow KPSS(1992) and consider a component model $u_t = u_{1t} + \varepsilon_{2t}$, where $u_{1t} = u_{1,t-1} + \varepsilon_{1t}$, and ε_{1t} and ε_{2t} are iid random variables with the same tail thickness and with variance σ_1^2 and σ_2^2 , and are independent with each other. The variance ratio $\lambda = \sigma_1/\sigma_2$ measures the relative importance of the random walk component. We consider two cases with different variance ratios: (1) Alter1: $\lambda = 100$; and (2) Alter2: $\lambda = 0.1$.

Special attention is paid here to the performance of these tests in the presence of various tail behavior of ε_t . In particular, we consider the following distributions for ε_t (or ε_{jt} , j = 1, 2): ε_t are iid (i) normal, (ii) t_3 , (iii) t_2 , and (iv) t_1 . The true value of θ is (0,0). The sample size that we use in the experiments is n = 200. The testing statistics are constructed based on OLS and M regression of y_t on $X_t = (1, t)'$. The nominal size is 5% in our experiments.

Empirical size and power are reported in Table 1. We first look at the empirical sizes under "Null1" for different error distributions. In general, the M-estimation based tests have more robust size property than the OLS-based tests. As the tail thickness increases, the OLS-based tests tend to under-reject the null. The under-rejection is particularly serious for the OLS-based Kolmogoroff-Smirnoff test in both Case 3 (t_2) and Case 4 (t_1), and for the Cramer von-Mises test in Case 4 (t_1). For the data generating process that we consider in this experiment, the Cramer von Mises tests in general has better size properties than the Kolmogoroff-Smirnoff tests.

For the results corresponding to "Null2", the tests using bandwidth $\ell_1 = 0$ overrejects because short run dynamics is not taken into account. The tests using the partially data-dependent bandwidth ℓ_2 has reasonable performance in general. Again, results are qualitatively similar to the size properties under "Null1": the M-estimation based tests have more robust size property than the OLS-based tests; as the tail thickness increases, the OLS-based tests tend to under-reject the null.

We next look at the power property of these tests. The power results are mixed. For a large λ (Alter1), the M-estimation based tests are not more powerful than the OLS-based tests even in the presence of thicktailness. For small λ (Alter2), the M-estimation based tests have lower power when the errors are normally distributed, but have significantly higher power than the OLS-based tests in the presence of non-Gaussian errors. This is because that efficiency gain of using M-estimation is obtained in the presence of stationary non-Gaussian errors and is translated to more robust size properties. When the regression errors are nonstationary non-Gaussian processes, the efficiency issue is more complicate and the benefit from using M estimation may depend on the structure of the process.

Table 1: Testing for $I(0)$ vs $I(1)$: Empirical Size and Power										
	Koln	nogoroff-S	Smirnoff	Tests	Cra	Cramer von-Mises Tests				
	OLS-Based		M-Estimation		OLS-I	OLS-Based		M-Estimation		
Bandwidth:	ℓ_1	ℓ_2	ℓ_1	ℓ_2	ℓ_1	ℓ_2	ℓ_1	ℓ_2		
	Case 1: $\varepsilon_t = \text{Normal}$									
Null1	0.0354	0.0322	0.0464	0.0406	0.0500	0.0466	0.0521	0.0506		
Null2	0.5762	0.0220	0.4378	0.0378	0.5626	0.0716	0.3972	0.0688		
Alter1	1	0.3730	0.9996	0.4700	1	0.7434	0.9998	0.7434		
Alter2	0.6750	0.4198	0.5594	0.3492	0.7160	0.5698	0.5904	0.4736		
Case 2: $\varepsilon_t = t_3$										
Null1	0.0310	0.0304	0.0464	0.0415	0.0486	0.0476	0.0530	0.0516		
Null2	0.5792	0.0278	0.4970	0.0395	0.5544	0.0592	0.4388	0.0604		
Alter1	1	0.4280	0.9996	0.4798	1	0.7488	0.9996	0.7078		
Alter2	0.6696	0.4108	0.7630	0.5044	0.7206	0.5678	0.7816	0.6242		
			Case	e 3: $\varepsilon_t = t$	2					
Null1	0.0280	0.0260	0.0450	0.0434	0.0420	0.0392	0.0534	0.0524		
Null2	0.5962	0.0248	0.550	0.0456	0.5770	0.0616	0.4806	0.0608		
Alter1	1	0.5138	0.9998	0.4968	1	0.7592	1	0.7093		
Alter2	0.6642	0.4274	0.8850	0.6444	0.7030	0.5624	0.8910	0.7416		
Case 4: $\varepsilon_t = t_1$										
Null1	0.0248	0.0234	0.0438	0.0430	0.0244	0.0214	0.0548	0.0492		
Null2	0.6142	0.0430	0.6090	0.0520	0.6398	0.045	0.6096	0.066		
Alter1	1	0.7548	1	0.7014	0.9986	0.8168	1	0.7714		
Alter2	0.6144	0.4826	0.9942	0.8492	0.6356	0.5324	0.9946	0.9066		

The second experiment considers testing for the null of cointegration. The data were generated from model (3) with $X_t = (1, t, x_t)'$, where $x_t = x_{t-1} + v_t$, and v_t are iid random variables. We consider the same error processes as the first experiment, i.e. "Null1" and "Null2" under the null of cointegration, and "Alter1" and "Alter2" under the alternative hypothesis. ε_t (or ε_{jt} , j = 1, 2) and v_t are iid distributed random variables and are independent with each other. Again, our attention focuses on the tail behavior of error distribution. In particular, we consider the following combinations of error distributions: Case 1: ε_t (or ε_{jt}) = Normal, v_t = Normal; Case 2: ε_t (or ε_{jt}) $= t_3, v_t = t_3$; Case 3: ε_t (or ε_{jt}) = t_3, v_t = Normal; Case 4: ε_t (or ε_{jt}) = t_2, v_t = Normal; Case 5: ε_t (or ε_{jt}) = t_1, v_t = Normal. The true value of θ is (0, 0, 1).

Results for the cointegration tests are reported in Table 2, and are similar to those results in the first experiment. The OLS based display better performance in the first case where both ε_t and v_t are normally distributed. However, the M-estimation based tests have more robust size property than the OLS-based tests, although the size distortion of the OLS based cointegration tests is relatively smaller than the distortion of the OLS based stationarity tests. The power property of cointegration tests is also similar to that of the stationarity tests.

Table 2: Testing for Cointegration: Empirical Size and Power									
	Koln	nogoroff-S	Smirnoff	Tests	Cra	Cramer von-Mises Tests			
	OLS-Based		M-Estimation		OLS-1	OLS-Based		M-Estimation	
	ℓ_1	ℓ_2	ℓ_1	ℓ_2	ℓ_1	ℓ_2	ℓ_1	ℓ_2	
Case 1: $\varepsilon_t = \text{Normal}, v_t = \text{Normal}$									
Null1	0.0374	0.0342	0.0454	0.0416	0.0446	0.0438	0.0490	0.0484	
Null2	0.5862	0.0354	0.4520	0.0478	0.5600	0.0666	0.3886	0.0672	
Alter1	0.9998	0.3950	0.9994	0.4456	1	0.7062	0.9998	0.6304	
Alter2	0.6016	0.3800	0.4882	0.3202	0.6438	0.5124	0.5212	0.4148	
Case 2: $\varepsilon_t = t_3, v_t = t_3$									
Null1	0.0330	0.0360	0.0458	0.0428	0.0442	0.0450	0.0482	0.0526	
Null2	0.5828	0.0396	0.5120	0.0506	0.5564	0.0720	0.4362	0.0724	
Alter1	1	0.4338	0.9996	0.4428	1	0.7072	0.9998	0.6368	
Alter2	0.5906	0.3772	0.7054	0.4644	0.6460	0.4970	0.7248	0.5598	
			Case 3: a	$\varepsilon_t = t_3, v_t$	= Normal				
Null1	0.0376	0.0354	0.0494	0.0468	0.0498	0.0508	0.0526	0.0544	
Null2	0.5942	0.0428	0.5190	0.0485	0.5678	0.0714	0.4498	0.0728	
Alter1	1	0.4506	0.9998	0.4594	1	0.7196	0.9998	0.6444	
Alter2	0.6090	0.3858	0.7000	0.4664	0.6528	0.5086	0.7180	0.5650	
Case 4: $\varepsilon_t = t_2, v_t = \text{Normal}$									
Null1	0.0344	0.0324	0.0500	0.0502	0.0406	0.0428	0.0530	0.0536	
Null2	0.6002	0.0464	0.5488	0.0488	0.5632	0.0682	0.4800	0.0702	
Alter1	1	0.5224	0.9996	0.4796	1	0.7214	0.9996	0.6538	
Alter2	0.6028	0.3882	0.8420	0.5884	0.6358	0.4780	0.8492	0.6608	
Case 5: $\varepsilon_t = t_1, v_t = \text{Normal}$									
Null1	0.0364	0.0380	0.0504	0.0486	0.0312	0.0332	0.0558	0.0540	
Null2	0.7080	0.0778	0.7042	0.0526	0.6114	0.0636	0.6284	0.0750	
Alter1	1	0.7534	0.9998	0.5726	0.9996	0.7752	0.9998	0.7312	
Alter2	0.5700	0.4712	0.9908	0.7940	0.5842	0.4710	0.9906	0.8582	

The last part of our experiment considers testing for structural instability in regressions with nonstationary regressors. Notice that the behavior of the second inference problem and the third inference problem are the same under the null, thus we focus our attention on the power analysis. In particular, we consider, under the alternative, the following model:

$$y_t = \theta'_t X_t + u_t$$

where $X_t = (1, t, x_t)'$, and $\theta_t = (0, 0, 1)$, for $t = 1, \dots, [n/2]$, and $\theta_t = (0, 0, 1.1)$, for $t = [n/2]+1, \dots, n$. We consider the same error distributions, i.e. Case 1 - Case 5, as in the second experiment, and consider iid errors in the regression. The empirical powers are reported in Table 3. In the case when ε_t is normal, the OLS based tests are more powerful than the M estimation based tests although the difference is not huge. The M estimation based tests have much better power than OLS based tests in the presence of non-Gaussian ε_t . The power of OLS based tests decreases as the tail of error distribution gets thicker.

Table 3: Testing for Structural Break - Empirical Power										
Koln	nogoroff-S	Smirnoff	Tests	Cra	Cramer von-Mises Tests					
OLS-	OLS-Based		M-Estimation		OLS-Based		M-Estimation			
ℓ_1	ℓ_2	ℓ_1	ℓ_2	ℓ_1	ℓ_2	ℓ_1	ℓ_2			
Case 1: $\varepsilon_t = \text{Normal}, v_t = \text{Normal}$										
0.4844	0.4572	0.4026	0.3606	0.5166	0.4912	0.4062	0.3890			
Case 2: $\varepsilon_t = t_3, v_t = t_3$										
0.4650	0.4582	0.5666	0.5296	0.5014	0.4950	0.5750	0.5560			
Case 3: $\varepsilon_t = t_3, v_t = \text{Normal}$										
0.2610	0.2394	0.3528	0.3142	0.2920	0.2732	0.3558	0.3392			
Case 4: $\varepsilon_t = t_2, v_t = \text{Normal}$										
0.1344	0.1322	0.3350	0.3080	0.1634	0.1634	0.3284	0.3262			
Case 5: $\varepsilon_t = t_1, v_t = \text{Normal}$										
0.0406	0.0448	0.2730	0.2566	0.0334	0.0362	0.2778	0.2716			

In summary, we conclude that, for testing the null of trend stationarity or the null of cointegration, the M-estimation method-based tests have more robust size property than the OLS-based tests. In particular, the M-estimation method-based tests have much better size in the presence of non-Gaussian errors. Power properties of these tests are more complicate and depend on the structure of the error process. If we consider a component model like the KPSS (1992) paper, the M estimation-based tests have better power in the presence of non-Gaussian errors for a small variance ratio. If the error process is dominated by a simple random walk, the M-estimation based tests no longer has power advantage over the OLS method. In this case, appropriate method that takes into account of the nonstationarity should be combined with M estimation to improve the performance of the estimation and inference procedure - of course, in such a case, the limiting distributions need to be modified correspondingly. For the second inference problem, the M-estimation based tests have both better size and power properties in the presence of non-Gaussian errors.

6 Appendix: A Sketch of Proofs

THEOREM 1. We first consider the case with smooth criterion functions. Notice that under the null $y_t - \hat{\theta}' X_t = u_t - (\hat{\theta} - \theta_0) X_t$, and, under Assumption 3 - 5, by a Taylor expansion of $\psi \left(y_t - \hat{\theta}' X_t \right)$ around u_t we obtain

$$n^{-1/2} \sum_{t=1}^{n} D_n^{-1} X_t \psi(u_t) - \left[\frac{1}{n} \sum_{t=1}^{n} \psi'(u_t) D_n^{-1} X_t X_t' D_n^{-1} \right] n^{1/2} D_n(\widehat{\theta} - \theta) = o_p(1).$$
(12)

Consequently,

$$n^{1/2}D_n(\widehat{\theta} - \theta) = \left[\frac{1}{n}\sum_{t=1}^n \psi'(u_t) D_n^{-1}X_t X_t' D_n^{-1}\right]^{-1} n^{-1/2} \sum_{t=1}^n D_n^{-1}X_t \psi(u_t) + o_p(1).$$

Notice that, under our assumptions, as $n \to \infty$,

$$\frac{1}{n} \sum_{t=1}^{n} \psi'(u_t) G_n^{-1} z_t z_t' G_n^{-1}
= \frac{1}{n} \sum_{t=1}^{n} \mathbf{E} \left[\psi'(u_t) \right] G_n^{-1} z_t z_t' G_n^{-1} + \frac{1}{n} \sum_{t=1}^{n} \left[\psi'(u_t) - \mathbf{E} \left[\psi'(u_t) \right] \right] G_n^{-1} z_t z_t' G_n^{-1}
\rightarrow \delta \int Z(r) Z(r)' dr$$

and, similarly,

$$n^{-3/2} \sum_{t=1}^{n} G_n^{-1} z_t x_t' \psi'(u_t) \rightarrow \delta \int Z(r) B(r)^{\top} dr,$$

$$n^{-2} \sum_{t=1}^{n} x_t x_t' \psi'(u_t) = n^{-1} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} E\psi'(u_t) + o_p(1) \Rightarrow \delta \int B(r) B(r)' dr,$$

$$n^{-1} \sum_{t=1}^{n} s_t s_t' \psi'(u_t) \rightarrow \delta \Gamma_s(0),$$

$$n^{-1/2} \sum_{t=1}^{n} G_n^{-1} z_t \psi(u_t) = \sum_{t=1}^{n} G_n^{-1} z_t \frac{\psi(u_t)}{\sqrt{n}} \Rightarrow \int_0^1 Z(r) dB_{\psi}(r),$$

$$n^{-1} \sum_{t=1}^{n} x_t \psi(u_t) = \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{\psi(u_t)}{\sqrt{n}} \Rightarrow \int_0^1 B(r) dB_{\psi}(r),$$

$$n^{-1/2} \sum_{t=1}^{n} s_t \psi(u_t) \Rightarrow \int_0^1 dB_{s\psi}(r) = N(0, \Omega_{s\psi})$$

$$n^{-1}\sum_{t=1}^{n} G_{n}^{-1} z_{t} s_{t}' \psi'(u_{t}) = \frac{1}{\sqrt{n}} \left[\sum_{t=1}^{n} G_{n}^{-1} z_{t} \frac{s_{t}'}{\sqrt{n}} \psi'(u_{t}) \right] = O_{p} \left(\frac{1}{\sqrt{n}} \right)$$
$$n^{-3/2} \sum_{t=1}^{n} x_{t} s_{t}' \psi'(u_{t}) = \frac{1}{\sqrt{n}} \left[\sum_{t=1}^{n} \frac{x_{t}}{\sqrt{n}} \frac{s_{t}'}{\sqrt{n}} \psi'(u_{t}) \right] = O_{p} \left(\frac{1}{\sqrt{n}} \right),$$

notice that $\int_0^1 dB_{s\psi}(r) = B_{s\psi}(1)$ is a p_2 dimensional normal variate with covariance matrix equals to $\Omega_{s\psi}$. If we partition θ into $(\theta'_1, \theta'_2)'$, where θ_1 is the sub-vector of coefficient corresponding to $(z'_t, x'_t)'$, and θ_2 is the sub-vector of coefficient corresponding to s_t , and let $D_{1n} = diag[G_n, G_{2n}]$, then $\sqrt{n}D_{1n}(\hat{\theta}_1 - \theta_1) \Rightarrow \frac{1}{\delta} \left[\int \overline{B}(r)\overline{B}(r)'dr\right]^{-1} \int \overline{B}(r)dB_{\psi}(r)$, and $\sqrt{n}(\hat{\theta}_2 - \theta_2) \Rightarrow \frac{1}{\delta}\Phi$.

In the case where $\psi(\cdot)$ is not everywhere differentiable, define

$$H_n(g) = n^{-1/2} \sum_{t=1}^n D_n^{-1} X_t \psi \left(u_t - n^{-1/2} g' D_n^{-1} X_t \right).$$

Because $\psi(\cdot)$ is not everywhere differentiable and we can not directly take a Taylor expansion with $\psi(\cdot)$ to obtain (12), we proceed by treating the function $\psi(\cdot)$ as a generalized function with a

smooth regular sequence $\psi_m(\cdot)$ defined based on an appropriate set of test functions (see Phillips (1995) for more discussions and related literature), then $\psi_m \to \psi$, and $\psi'_m \to \psi'$, as $m \to \infty$. $H_n(g)$ is then a generalized process defined by the following regular sequence of processes:

$$H_{nm}(g) = n^{-1/2} \sum_{t=1}^{n} D_n^{-1} X_t \psi_m \left(u_t - n^{-1/2} g' D_n^{-1} X_t \right)$$

Expanding $\psi_m(\cdot)$ around u_t gives

$$H_{nm}(g) = n^{-1/2} \sum_{t=1}^{n} D_n^{-1} X_t \psi_m(u_t) - \left[\frac{1}{n} \sum_{t=1}^{n} \psi'_m(u_t) D_n^{-1} X_t X'_t D_n^{-1}\right] g + \left[\frac{1}{n} \sum_{t=1}^{n} \left[\psi'_m(u_t) - \psi'_m(u_t^*)\right] D_n^{-1} X_t X'_t D_n^{-1}\right] g$$

where u_t^* is between u_t and $u_t - n^{-1/2}g' D_n^{-1} X_t$. Notice that ψ'_m is a regular sequence which is differentiable with a bounded derivative for each m, thus

$$\left|\psi'_{m}\left(u_{t}^{*}\right)-\psi'_{m}\left(u_{t}\right)\right|\leq K_{m}n^{-1/2}\left|g'D_{n}^{-1}X_{t}\right|$$

for some $K_m > 0$, and thus, for any $g = o_p(n^{1/4})$,

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\left[\psi_{m}'\left(u_{t}^{*}\right)-\psi_{m}'\left(u_{t}\right)\right]D_{n}^{-1}X_{t}X_{t}'D_{n}^{-1}\right\| \leq K_{m}n^{-3/2}\left\|g\right\|\sum_{t=1}^{n}\left\|D_{n}^{-1}X_{t}\right\|^{3} \to 0$$

 thus

$$H_{nm}(g) = n^{-1/2} \sum_{t=1}^{n} D_n^{-1} X_t \psi_m(u_t) - \left[\frac{1}{n} \sum_{t=1}^{n} \psi'_m(u_t) D_n^{-1} X_t X'_t D_n^{-1}\right] g + g \times o_p(1).$$

Let $g = n^{1/2} D_n\left(\widehat{\theta} - \theta_0\right)$ and denote $H_{nm}(n^{1/2} D_n\left(\widehat{\theta} - \theta_0\right))$, $H_n(n^{1/2} D_n\left(\widehat{\theta} - \theta_0\right))$ as $H_{nm}(\widehat{\theta})$ and $H_n(\widehat{\theta})$ for simplicity, then, under Assumption 5,

$$H_{nm}(\widehat{\theta}) = n^{-1/2} \sum_{t=1}^{n} D_n^{-1} X_t \psi_m(u_t) - \left[\frac{1}{n} \sum_{t=1}^{n} \psi'_m(u_t) D_n^{-1} X_t X'_t D_n^{-1} \right] \left[n^{1/2} D_n\left(\widehat{\theta} - \theta_0\right) \right] \\ + \left[n^{1/2} D_n\left(\widehat{\theta} - \theta_0\right) \right] \times o_p(1).$$

Thus,

$$n^{1/2}D_n\left(\widehat{\theta} - \theta_0\right) = \left[\frac{1}{n}\sum_{t=1}^n \psi'_m\left(u_t\right)D_n^{-1}X_tX_t'D_n^{-1} + o_p(1)\right]^{-1} \left[n^{-1/2}\sum_{t=1}^n D_n^{-1}X_t\psi_m\left(u_t\right) - H_{nm}(\widehat{\theta})\right].$$

We now examine, as $n \to \infty$, the limit behavior of each component in the above expression. First,

$$n^{-1/2} \sum_{t=1}^{n} D_n^{-1} X_t \psi_m(u_t) = \begin{bmatrix} n^{-1/2} \sum_{t=1}^{n} G_n^{-1} z_t \psi_m(u_t) \\ n^{-1} \sum_{t=1}^{n} x_t \psi_m(u_t) \\ n^{-1/2} \sum_{t=1}^{n} s_t \psi_m(u_t) \end{bmatrix} \Rightarrow \begin{bmatrix} \int_0^1 Z(r) dB_{\psi_m}(r) \\ \int_0^1 B(r) dB_{\psi_m}(r) \\ \int_0^1 dB_{s\psi_m}(r) \end{bmatrix}, \quad (13)$$

where $B_{\psi_m}(r)$ is a Brownian motion with variance $\omega_{\psi_m}^2 = \sum_{h=-\infty}^{\infty} \mathbb{E}[\psi_m(u_t)\psi_m(u_{t+h})]$ and, as $m \to \infty$,

$$\lim_{m \to \infty} \begin{bmatrix} \int_0^1 Z(r) dB_{\psi_m}(r) \\ \int_0^1 B(r) dB_{\psi_m}(r) \\ \int_0^1 dB_{s\psi_m}(r) \end{bmatrix} = \begin{bmatrix} \int_0^1 Z(r) dB_{\psi}(r) \\ \int_0^1 B(r) dB_{\psi}(r) \\ \int_0^1 dB_{s\psi}(r) \end{bmatrix}.$$

Next, notice that u_t is uncorrelated with s_t , as $n \to \infty$,

$$n^{-1} \sum_{t=1}^{n} \psi'_{m}(u_{t}) G_{n}^{-1} z_{t} s'_{t} \to 0, \ n^{-3/2} \sum_{t=1}^{n} \psi'_{m}(u_{t}) x_{t} s'_{t} \to 0, \ \forall m$$

and

$$n^{-1}\sum_{t=1}^{n}\psi'_{m}\left(u_{t}\right)s_{t}s'_{t} \to \mathrm{E}\left(\psi'_{m}\left(u_{t}\right)s_{t}s'_{t}\right) = \delta_{m}\Gamma_{s}(0),$$

where $\delta_m = E\left(\psi'_m\left(u_t\right)\right)$, thus

$$\begin{bmatrix} \frac{1}{n} \sum_{t=1}^{n} \psi'_m(u_t) D_n^{-1} X_t X'_t D_n^{-1} \end{bmatrix} \Rightarrow \begin{bmatrix} \delta_m \int Z(r) Z(r)' dr & \delta_m \int Z(r) B(r)' dr & 0 \\ \delta_m \int B(r) Z(r)' dr & \delta_m \int B(r) B(r)' dr & 0 \\ 0 & 0 & \delta_m \Gamma_s(0) \end{bmatrix}$$

Notice that $\lim_{m\to\infty} \delta_m = \delta = E(\psi'(u_t))$, the right hand side quantity of the above expression converges to

$$\delta \begin{bmatrix} \int Z(r)Z(r)'dr & \int Z(r)B(r)'dr & 0\\ \int B(r)Z(r)'dr & \int B(r)B(r)'dr & 0\\ 0 & 0 & \Gamma_s(0) \end{bmatrix}, as \ m \to \infty.$$

Finally notice that $\lim_{m\to\infty} H_{nm}(\widehat{\theta}) = H_n(\widehat{\theta}) = o_p(1)$ under Assumption 4, we have

$$n^{1/2}D_n\left(\widehat{\theta}-\theta_0\right) \Rightarrow \left[\begin{array}{c} \delta^{-1}\left[\int \overline{B}(r)\overline{B}(r)'dr\right]^{-1}\int_0^1 \overline{B}(r)dB_{\psi}\left(r\right)\\ \delta^{-1}\Gamma_s(0)^{-1}\int_0^1 dB_{s\psi}\left(r\right)\end{array}\right].$$

THEOREM 2. If ρ is smooth, notice that $\hat{u}_t = y_t - \hat{\theta}' X_t$, by a Taylor expansion of $\psi(\hat{u}_t)$ around u_t we obtain

$$n^{-1/2} \sum_{t=1}^{[nr]} \psi(\widehat{u}_t)$$

$$= n^{-1/2} \sum_{t=1}^{[nr]} \psi(u_t) - n^{-1/2} \sum_{t=1}^{[nr]} \psi'(u_t) (\widehat{\theta} - \theta)' X_t + o_p(1)$$

$$= n^{-1/2} \sum_{t=1}^{[nr]} \psi(u_t) - n^{-1/2} \sum_{t=1}^{[nr]} \delta(\widehat{\theta} - \theta)' X_t - n^{-1/2} \sum_{t=1}^{[nr]} [\psi'(u_t) - \delta] (\widehat{\theta} - \theta)' X_t + o_p(1)$$

$$\Rightarrow B_{\psi}(r) - \delta \left[\lim \sqrt{n} D_{1n}(\widehat{\theta}_1 - \theta_1)' \right] \left(\int_0^r \overline{B}(s) ds \right)$$

since $\frac{1}{n} \sum_{t=1}^{[nr]} s_t \to 0.$

If $\psi(\cdot)$ is not everywhere differentiable, we again treat the function $\psi(\cdot)$ as a generalized function as in the proof of Theorem 1. If we denote $\Psi_n(r) = n^{-1/2} \sum_{t=1}^{[nr]} \psi(\widehat{u}_t)$, then $\Psi_n(r)$ is a generalized process defined by the following regular sequence of processes

$$\Psi_{n,m}(r) = n^{-1/2} \sum_{t=1}^{[nr]} \psi_m(\hat{u}_t) \,.$$

Expanding $\psi_m(\cdot)$ around u_t gives

$$\Psi_{n,m}(r) = n^{-1/2} \sum_{t=1}^{[nr]} \psi_m(u_t) - n^{-1/2} \sum_{t=1}^{[nr]} \psi'_m(u_t) (\widehat{\theta} - \theta)' X_t$$

$$+ n^{-1/2} \sum_{t=1}^{[nr]} \left[\psi'_m(u_t) - \psi'_m(u_t^*) \right] (\widehat{\theta} - \theta)' X_t,$$
(14)

where u_t^* is a middle value. We next show that $n^{-1/2} \sum_{t=1}^{[nr]} \left[\psi'_m(u_t) - \psi'_m(u_t^*) \right] (\hat{\theta} - \theta)' X_t \xrightarrow{P} 0$, and derive the limiting variates of the first two terms in the above expansion. First, notice that ψ'_m is a regular sequence, thus $\left| \psi'_m(u_t^*) - \psi'_m(u_t) \right| \le K_m \left| (\hat{\theta} - \theta)' X_t \right|$ for some $K_m > 0$, and thus under Assumption 3',

$$\sup_{r} \left| n^{-1/2} \sum_{t=1}^{[nr]} \left[\psi'_{m} \left(u_{t} \right) - \psi'_{m} \left(u_{t}^{*} \right) \right] \left(\widehat{\theta} - \theta \right)' X_{t} \right|$$

$$\leq n^{-1/2} \sum_{t=1}^{n} \left| \psi'_{m} \left(u_{t} \right) - \psi'_{m} \left(u_{t}^{*} \right) \right| \left| \left(\widehat{\theta} - \theta \right)' X_{t} \right|$$

$$\leq K_{m} n^{-3/2} \sum_{t=1}^{n} \left\| n^{1/2} \left[D_{n} \left(\widehat{\theta} - \theta_{0} \right) \right] \right\|^{2} \left\| D_{n}^{-1} X_{t} \right\|^{2}$$

$$\xrightarrow{P} 0, \text{ under Assumption 5.}$$

For the first term on the right hand side of (14), as $n \to \infty$, $n^{-1/2} \sum_{t=1}^{[nr]} \psi_m(u_t) \Rightarrow B_{\psi_m}(r)$, where $B_{\psi_m}(r)$ is a Brownian motion with variance $\omega_{\psi_m}^2 = \sum_{h=-\infty}^{\infty} \mathbb{E}[\psi_m(u_t) \psi_m(u_{t+h})]$ and, as $m \to \infty$, the limiting process of $B_{\psi_m}(r)$ is $B_{\psi}(r)$. For the second term in (14), as $n \to \infty$,

$$n^{-1/2} \sum_{t=1}^{[nr]} \psi'_m(u_t) \left(\widehat{\theta} - \theta\right)' X_t \Rightarrow \left[\lim \sqrt{n} D_n(\widehat{\theta} - \theta)'\right] \delta_m \left[\begin{array}{c} \int_0^r \overline{B}(s) ds \\ 0 \end{array} \right]$$

Thus, $\lim_{n\to\infty} \Psi_{n,m}(r) = \Psi_{\infty,m}(r)$, where

$$\Psi_{\infty,m}(r) = B_{\psi_m}(r) - \int_0^1 dB_{\psi}(s)\overline{B}(s) \left[\int_0^1 \overline{B}(s)\overline{B}(s)'ds\right]^{-1} \int_0^r \overline{B}(s)ds.$$

Let
$$m \to \infty$$
, $\Psi_{\infty,m}(r) \Rightarrow \Psi(r) = B_{\psi}(r) - \int_0^1 dB_{\psi}(s)\overline{B}(s) \left[\int_0^1 \overline{B}(s)\overline{B}(s)'ds\right]^{-1} \int_0^r \overline{B}(s)ds$. Thus,
 $n^{-1/2} \sum_{t=1}^{[nr]} \psi\left(\widehat{u}_t\right) \Rightarrow \Psi(r) = \omega_{\psi} \left[W_{\psi}(r) - \int_0^1 dW_{\psi}(s)\overline{W}(s) \left[\int_0^1 \overline{W}(s)\overline{W}(s)'ds\right]^{-1} \int_0^r \overline{W}(s)ds\right].$

THEOREM 4. Under the local alternative $H_{A2}: \theta_t = \theta_0 + \frac{1}{\sqrt{n}} D_n^{-1} g(t/n)$, use a similar analysis as those in the proof of Theorem 1 and Theorem 2, we obtain that

$$n^{1/2} D_n(\widehat{\theta} - \theta_0) = \left[\frac{1}{n} \sum_{t=1}^n \psi'(u_t) D_n^{-1} X_t X_t' D_n^{-1} \right]^{-1} n^{-1/2} \sum_{t=1}^n D_n^{-1} X_t \psi(u_t) + \left[\frac{1}{n} \sum_{t=1}^n \psi'(u_t) D_n^{-1} X_t X_t' D_n^{-1} \right]^{-1} \left[\frac{1}{n} \sum_{t=1}^n \psi'(u_t) D_n^{-1} X_t X_t' D_n^{-1} g(t/n) \right] + o_p(1)$$

$$\Rightarrow \frac{1}{\delta} \left[\int \overline{B}(s) \overline{B}(s)' ds \right]^{-1} \int_0^1 \overline{B}(r) dB_{\psi}(r) + \left[\int \overline{B}(s) \overline{B}(s)' ds \right]^{-1} \int \overline{B}(r) \overline{B}(r)^\top g(r) dr.$$

Notice that under H_{A2} , $\hat{u}_t = u_t - \left(\hat{\theta} - \theta_0\right) X_t + \left(\frac{1}{\sqrt{n}}D_n^{-1}g\left(t/n\right)\right) X_t$, by a similar analysis as the proof of Theorem 2, and notice that, if we partition $g\left(t/n\right)$ as $\left(g_1\left(t/n\right)^{\top}, g_2\left(t/n\right)^{\top}\right)^{\top}$ (conformable with $(z'_t, x'_t)'$),

$$\frac{1}{n} \sum_{t=1}^{[nr]} \psi'(u_t) g_1(t/n)^\top G_n^{-1} z_t \to \delta \int g_1(r)^\top Z(r) dr,$$

$$\frac{1}{n} \sum_{t=1}^{[nr]} \psi'(u_t) g_2(t/n)^\top n^{-1/2} x_t \to \delta \int g_2(r)^\top B(r) dr,$$

we have $n^{-1/2} \sum_{t=1}^{[nr]} \psi\left(\widehat{u}_t\right) \Rightarrow$

$$B_{\psi}(r) - \int_{0}^{1} dB_{\psi}(r) \overline{B}(r)^{\top} \left[\int \overline{B}(s) \overline{B}(s)^{\top} ds \right]^{-1} \int_{0}^{r} \overline{B}(s) ds$$
$$-\delta \int g(r)^{\top} \overline{B}(r) \overline{B}(r)^{\top} dr \left[\int \overline{B}(s) \overline{B}(s)' ds \right]^{-1} \int_{0}^{r} \overline{B}(s) ds + \delta \int_{0}^{r} g(s)^{\top} \overline{B}(s) ds.$$

Thus

$$\begin{split} \widetilde{W}_{n}(r) &= \frac{1}{\sqrt{n}\widehat{\omega}_{\psi}}\sum_{t=1}^{\lfloor nr \rfloor}\psi\left(\widehat{u}_{t}\right) \\ \Rightarrow & W_{1}(r) - \int_{0}^{1}dW_{1}(r)\overline{W}_{2}(r)\left[\int_{0}^{1}\overline{W}_{2}(s)\overline{W}_{2}(s)'ds\right]^{-1}\int_{0}^{r}\overline{W}_{2}(s)ds. \\ &\quad -\frac{\delta}{\omega_{\psi}}\int g(r)^{\top}\overline{B}(r)\overline{B}(r)^{\top}dr\left[\int\overline{B}(s)\overline{B}(s)'ds\right]^{-1}\int_{0}^{r}\overline{B}(s)ds + \frac{\delta}{\omega_{\psi}}\int_{0}^{r}g\left(s\right)^{\top}\overline{B}(s)ds \\ &= \widetilde{W}(r) \\ &\quad +\frac{\delta}{\omega_{\psi}}\left[\int_{0}^{r}g\left(s\right)^{\top}\overline{B}(s)ds - \int g(r)^{\top}\overline{B}(r)\overline{B}(r)^{\top}dr\left[\int\overline{B}(s)\overline{B}(s)'ds\right]^{-1}\int_{0}^{r}\overline{B}(s)ds\right]. \end{split}$$

THEOREM 5: Notice that the residual process is I(1) under the alternative hypotheses, and the score function $\psi(\cdot)$ is asymptotically homogeneous, by a similar argument as Park and Phillips (1999, Theorem 5.3), $n^{-1/2} \sum_{t=1}^{[nr]} \psi(\hat{u}_t)$ diverges at rate $\sqrt{n\nu}(\sqrt{n})$. On the other side, the nonparametric spectral density estimate $\hat{\omega}^2_{\psi}$ diverges as well. To prove consistency of the tests, we need to show that, as $n \to \infty$, $\hat{\omega}_{\psi} = o_p(\sqrt{n\nu}(\sqrt{n}))$. Notice that

$$\frac{1}{\nu(\sqrt{n})^{2}\ell}\widehat{\omega}_{\psi}^{2} = \frac{k(0)}{\nu(\sqrt{n})^{2}\ell n}\sum_{t}\psi(\widehat{u}_{t})\psi(\widehat{u}_{t}) + \frac{2}{\nu(\sqrt{n})^{2}\ell}\sum_{j=1}^{\ell}k(\frac{j}{\ell})\frac{1}{n}\sum_{t}\psi(\widehat{u}_{t})\psi(\widehat{u}_{t+j}).$$

Under the alternative, $n^{-1/2}\hat{u}_{[nr]} \Rightarrow \eta(r)$, where $\eta(r)$ is a function of Brownian motions. In addition, under Assumption 8,

$$\psi(\widehat{u}_t) = \psi\left(\sqrt{n}\frac{\widehat{u}_t}{\sqrt{n}}\right) = \nu\left(\sqrt{n}\right)H\left(\frac{\widehat{u}_t}{\sqrt{n}}\right) + R\left(\frac{\widehat{u}_t}{\sqrt{n}},\sqrt{n}\right),$$

thus,

$$\begin{aligned} &\frac{1}{n\nu\left(\sqrt{n}\right)^2}\sum_t\psi\left(\widehat{u}_t\right)\psi(\widehat{u}_t)\\ &= \frac{1}{n\nu\left(\sqrt{n}\right)^2}\sum_t\left[\nu\left(\sqrt{n}\right)H\left(\frac{\widehat{u}_t}{\sqrt{n}}\right) + R\left(\frac{\widehat{u}_t}{\sqrt{n}},\sqrt{n}\right)\right]\left[\nu\left(\sqrt{n}\right)H\left(\frac{\widehat{u}_t}{\sqrt{n}}\right) + R\left(\frac{\widehat{u}_t}{\sqrt{n}},\sqrt{n}\right)\right]\\ &= \frac{1}{n}\sum_tH\left(\frac{\widehat{u}_t}{\sqrt{n}}\right)H\left(\frac{\widehat{u}_t}{\sqrt{n}}\right) + o_p(1)\\ &\Rightarrow \int_0^1H\left(\eta(r)\right)^2dr.\end{aligned}$$

Thus $\frac{k(0)}{\nu(\sqrt{n})^2 \ell n} \sum_t \psi(\widehat{u}_t) \psi(\widehat{u}_t) = O_p(\ell^{-1}) = o_p(1)$, consequently

$$\frac{1}{\nu\left(\sqrt{n}\right)^{2}\ell}\widehat{\omega}_{\psi}^{2} = \frac{2}{\ell}\sum_{j=1}^{\ell}k(\frac{j}{\ell})\frac{1}{n\nu\left(\sqrt{n}\right)^{2}}\sum^{\prime}\psi\left(\widehat{u}_{t}\right)\psi(\widehat{u}_{t+j}) + o_{p}(1).$$

By similar arguments as Phillips (1991) and deJong, Amsler and Schmidt (2007), as $n \to \infty$, we have

$$\frac{1}{M}\sum_{j=1}^{\ell}k(\frac{j}{\ell})\frac{1}{n\nu\left(\sqrt{n}\right)^2}\sum_t\psi\left(\widehat{u}_t\right)\psi(\widehat{u}_{t+j}) \Rightarrow \int_0^1k(u)du\int_0^1H\left(\eta(r)\right)^2dr,$$

thus $\frac{1}{\nu(\sqrt{n})^2\ell}\widehat{\omega}_{\psi}^2 = 2\int_0^1 k(u)du\int_0^1 H(\eta(r))^2 dr + o_p(1)$. Consequently $\widehat{\omega}_{\psi} = O_p(\sqrt{\ell}\nu(\sqrt{n}))$, and thus $\widetilde{W}_n(r) = O_p(\sqrt{n/\ell})$.

7 References

- Andrews, D.W.K., 1991, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," Econometrica, 59, 817–858.
- Cox, D. and Llatas, I., 1991, "Maximum likelihood type estimation for nearly non-stationary autoregressive time series," Annals of Statistics 19, 1109-1128.
- deJong, R., C. Amsler and P. Schmidt, 2007, A robust version of the KPSS test based on indicators, Journal of Econometrics, 137, 311-333.
- Faust, J., 1996. Near observational equivalence and theoretical size problems with unit root tests. Econometric Theory, 12, 724-731.
- Gel'fand and Vilenkin, 1964, Generalized functions, Vol. 4, New York, Academic Press.
- Hansen, B., 1992, Tests for parameter instability in regressions with I(1) processes, Journal of Business and Economic Statistics, 10, 321-355.
- Hao, K. and B. Inder, 1996, "Diagnostic test for structural change in cointegrated regression models," Economic Letters 50, 179-187.
- Hercé, M., 1996, Asymptotic Theory of LAD estimation in a unit root process with finite variance errors, Econometric Theory, 12, 129-153.
- Huber, P.J., 1964, Robust estimation of a location parameter, Annals of Mathematical Statistics, 35, 73-101.
- Huber, P.J., 1973, Robust regression: asymptotics, conjectures, and monte carlo, Annals of Statistics, 1, 799-821.
- Juhl, T., 1999, "Testing for cointegration using M estimators", preprint.
- Knight, K., 1991, "Limit Theory for M estimators in an Integrated Infinite Variance Process," Econometric Theory, 7, 200-212.

- Koenker, R. and Z. Xiao, 2004, Unit Root Quantile Regression Inference, Journal of the American Statistical Association, 2004, Vol. 99, No. 467, 775-787.
- Kuan, C.M. and K. Hornik, 1995, "The generalized fluctuation test: A unifying view" Econometric Reviews, 14, 135-161.
- Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin, 1992, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?", Journal of Econometrics, 54, 159–178.
- Lucas, A., 1995, "Unit Root tests based on M estimators," Econometric Theory 11, 331-346.
- Müller, U.K., 2008. The impossibility of consistent discrimination between I(0) and I(1) processes, Econometric Theory 24 (2008), 616 – 630.
- Nelson, C.R. and C. Plosser, 1982, "Trends and random walks in macro-economic time series: some evidence and implications," Journal of Monetary Economics, 10, 139–162.
- Park, J. and P. Phillips, 1999, Asymptotics for nonlinear transformations of integrated time series, Econometric Theory, 15, 269-298.
- Phillips, P.C.B., 1991, "Spectral regression for Cointegrated time series," Nonparametric and semiparametric methods in economics and statistics, eds. W. Barnett etc., Cambridge University Press.
- Phillips, P.C.B., 1995, Robust Nonstationary Regression, Econometric Theory 12, 912-951.
- Ploberger, W., W. Kramer, and K. Kontrus, 1989, "A new test for structural stability in the linear regression model," Journal of Econometrics, 40, 307–318.
- Pötscher, B. M., 2002, Lower risk bounds and properties of confidence sets for ill-posted estimation problems with applications to spectral density and persistency estimation, unit roots and estimation of long memory parameters., Econometrica, 1035 -1068.
- Relles, D., 1968, Robust regression by modified least squares, Yale University, New Haven.
- Shin, Y., 1994, "A residual based test of the null of cointegration against the alternative of no cointegration," Econometric Theory, 10, 91—115.
- Xiao, Zhijie, 1998, A Note on Bandwidth Selection in Hypothesis Tests with Integrated Time Series, mimeo, University of Illinois.
- Xiao, Zhijie, 2001(a), Testing the Null hypothesis of Stationarity against an Autoregressive Unit Root Alternative, Journal of Time Series Analysis, Vol. 22, No. 1, 87-105.
- Xiao, Zhijie, 2001(b), Likelihood Based Inference in Trending Time Series with a Root Near Unity, Econometric Theory 17, 1082-1112.
- Xiao, Zhijie, and P.C.B. Phillips, 2002, A CUSUM test for Cointegration Using Regression Residuals, Journal of Econometrics 108, 43-61.