Essays on Time Series Econometrics

Emre Aylar

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Eric Zivot, Chair

Yu-chin Chen

Thomas Gilbert

Program Authorized to Offer Degree:
Department of Economics

University of Washington

Abstract

Essays on Time Series Econometrics

Emre Aylar

Chair of the Supervisory Committee:
Professor Eric Zivot
Department of Economics

This dissertation focuses on the construction of statistical tests to differentiate stationary and non-stationary time series. Chapter 1 deals with non-stationarity induced by a broken trend function and considers testing for the presence of a structural break in the trend of a univariate time-series where the date of the break is unknown. The proposed tests are robust as to whether the shocks are generated by a stationary or an integrated process. The simulation results suggest that the robust tests perform well in small samples, showing good size control and displaying very decent power regardless of the degree of persistence of the data. Chapter 2 proposes a bootstrap stationarity test that has good size control and also retains power. The test utilizes a parametric bootstrap re-sampling scheme that can generate independent re-samples and impose the null constraint on the bootstrap samples. The empirical size and power performance of the proposed test is compared with the existing bootstrap and conventional stationarity tests through Monte-Carlo studies. Simulations demonstrate that the proposed bootstrap test controls size better and has higher power than the competing methods. Finally, chapter 3 considers the initial condition problem in unit root testing and develops a powerful unit root test robust to initial condition. The proposed method estimates the trend parameters using indirect inference and results show that the proposed test statistic is robust to initial condition.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to thank my advisor Eric Zivot for his support and guidance during the completion of this thesis. I would also like to thank the rest of my thesis committee: Yu-chin Chen, Thomas Gilbert and Adrian Dobra, for their encouragement and insightful comments. I am also thankful for helpful comments and suggestions from Chang-Jin Kim, Ji-Hyung Lee, Yanqin Fan, Xuyang Ma, David Kuenzel and Selim Tuncel.

I would especially like to thank my parents for their support and encouragement throughout this journey.

Generous support from the Buechel Memorial Fellowship and the Grover and Creta Ensley Fellowship is gratefully acknowledged.

DEDICATION

To my parents, Hacer and Selim

Chapter 1

TESTS FOR A BROKEN TREND WITH STATIONARY OR INTEGRATED SHOCKS

1.1 Introduction

Many macroeconomic and financial time series can be characterized by shocks fluctuating around a broken trend function. It is very important to account for changes in the trend function and failure to model the trend function properly will lead to inconsistent parameter estimates and poor forecasts. Perron (1989) has shown that an unaccounted break in the trend function can bias unit root tests toward the nonrejection of the unit root hypothesis. Similarly, unmodeled trend breaks also lead to spurious rejections in stationarity tests as documented by Lee et al. (1997). In fact, there are many interesting economic applications for which the existence of structural change in the trend function can be of interest itself. For example, the empirical debate that convergence in incomes per capita among U.S. regions has leveled off in the mid-1970s can be explored by modeling the trend function in time series of incomes per capita in each of the U.S. regions as having a slope shift in the mid-1970s (see Sayginsoy and Vogelsang (2004)).

Tests of a break in level and/or trend of a time series exist in the literature. However, for a reliable inference on the existence of a break these tests assume a priori that the data is generated either by a stationary or an integrated process (see Bai (1994), Bai and Perron (1998) and Hansen (1992)). As it is well known, many macroeconomic and financial variables are highly persistent and it is very difficult to tell a priori that these series are stationary or integrated. Hence the structural break tests mentioned above have limitations in practice. One way to overcome this problem is to pretest the series with unit root or stationarity tests to determine if the series has a unit root. However, unit root and stationarity tests are known to have poor size and power properties in the presence of a structural break (Perron (1989), Lee et al. (1997)). We thus have a circular testing problem between tests of structural breaks and unit root/stationarity tests.

In this paper, we provide a solution to this dilemma by proposing new tests for the presence of a structural break in the trend function at an unknown date that are valid in the presence of stationary (I(0)) or integrated (I(1)) shocks. We consider tests based on two statistics, one appropriate for the case of I(0) shocks and the other appropriate for I(1) shocks. These statistics are the supremum of the trend break t-statistics, calculated for all permissible break dates, from a regression in levels and a regression in growth rates. We derive the asymptotic distributions of these statistics in both I(0) and (local to) I(1) environments. The asymptotic properties demonstrate that the test statistic appropriate for the case of I(1) shocks converges in probability to zero under I(0) shocks, while the test appropriate for I(0) shocks is always under-sized when the shocks are (local to) I(1). These properties make it possible the construction of a size-controlled union of rejections testing approach whereby we reject the null hypothesis of no trend break if either of the two tests rejects.

Recently, a few other structural break tests that are valid regardless of whether the shocks are stationary or integrated have been proposed in the econometrics literature. Sayginsoy and Vogelsang (2011)[SV] propose a Mean Wald and a Sup Wald statistic using the fixed-b asymptotic framework of Kiefer and Vogelsang (2005) in conjunction with the scaling factor approach of Vogelsang (1998) to smooth the discontinuities in the asymptotic distributions of the test statistics as the shocks go from I(0) to I(1). Harvey et al. (2009b) [HLT] employ tests that are formed as a weighted average of the supremum of the regression t-statistics, calculated for all permissible break dates, for a broken trend appropriate for the case of I(0) and I(1) shocks. The weighting function they utilize is based on the KPSS stationarity test applied to the levels and first-differenced data. Perron and Yabu (2009) [PY] consider an alternative approach based on a Feasible Generalized Least Squares procedure that uses a super-efficient estimate of the sum of the autoregressive parameters α when $\alpha = 1$. This allows tests of basically the same size with stationary or integrated shocks regardless of whether the break is known or unknown, provided that the Exp functional of Andrews and Ploberger (1994) is used in the latter case.

Our robust tests compare favorably to the other robust tests mentioned above. The advantage of our method over those of SV and HLT is that it does not involve any random scaling so that the test used is more prone to have higher power and less size distortions, as documented under the finite sample results in Section 5. Our method does well also in comparison with the method of PY; first PY assume the normality of shocks to derive pivotal asymptotic distributions for certain broken trend models, and second their structural break tests don't allow the estimation of the break date simultaneously due to the use of Exp functional. On the other hand, the test statistics we propose have pivotal asymptotic distributions without any stringent assumptions on the shocks and our method allows us to simultaneously test for a break and estimate the break date.

In our empirical application, we shed some light on the empirical debate that convergence in per-capita incomes among U.S. regions has leveled off in the mid-1970s. We employ our robust trend break tests to check the validity of these claims and find significant statistical evidence of a slowdown of convergence in per-capita incomes among U.S. regions in the late 1970s and early 1980s. The slowdown is more pronounced for relatively low income regions that are trying to catch up. These findings are important in the sense that they are robust, hence reliable regardless of the properties of the data and more importantly they have important policy implications.

It is also straightforward to extend our method to test for multiple trend breaks or to sequentially test for l versus l+1 breaks in the trend function. However, the main focus of this paper is to test the null hypothesis that there is no structural break in the trend against the alternative that there exists a break in the trend. This type of hypothesis is more relevant if one wants to evaluate the impact of a one time policy change or a new regulation on a trending variable. For example, Sidneva and Zivot (2007) consider the question of whether there was a break in the trends of two air pollutants, nitrogen oxides and volatile organic compounds, emissions around the time the Clean Air Act Amendments of 1970 were passed. The tests we propose in this paper can easily be employed to answer this question without the need to know whether the shocks are stationary or integrated.

The rest of the paper is organized as follows. Section 2 reviews the two trend break models considered and the main assumptions used in the paper. In sections 3 and 4, we investigate each trend break model separately. Section 3 outlines the testing procedure we propose for the joint broken trend model. We introduce our tests based on two statistics with their asymptotic properties under the null and the alternative and explain how the union of rejections approach works in principle. In section 4, the same steps are carried out for the disjoint broken trend model. Section 5 provides some Monte Carlo studies that demonstrate the finite sample properties of our test. In Section 6, we present our empirical study. Section 7 concludes.

1.2 Trend Break Models

We consider two trend break models: "Model 1" is a joint broken trend model where there is a break in the trend holding the level fixed, while "Model 2" allows for a simultaneous break in level and trend which we call the disjoint broken trend model. These are the same models considered by HLT, corresponds to models (3) and (4) in SV and to models II and III in PY respectively. We consider the following trend break data generating processes (DGP) to model broken trends:

$$y_t = \alpha + \beta t + \gamma DT_t(\tau^*) + u_t, \qquad t = 1, ..., T \tag{1.1}$$

$$y_t = \alpha + \beta t + \delta DU_t(\tau^*) + \gamma DT_t(\tau^*) + u_t, \qquad t = 1, ..., T \tag{1.2}$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad t = 2, ..., T, \quad u_1 = \varepsilon_1, \tag{1.3}$$

where $DU_t(\tau^*) = 1(t > T^*)$ and $DT_t(\tau^*) = 1(t > T^*)(t - T^*)$, with $T^* = \lfloor \tau^* T \rfloor$ the (potential) trend break date with associated break fraction $\tau^* \in (0, 1)$. Equation (1) refers to the joint broken trend model, while equation (2) defines the disjoint broken trend model.

We assume that ε_t in (3) satisfies Assumption 1 of SV:

Assumption 1. The stochastic process $\{\varepsilon_t\}$ is such that

$$\varepsilon_t = c(L)\eta_t, \qquad c(L) = \sum_{i=0}^{\infty} c_i L^i,$$

with $c(1)^2 > 0$ and $\sum_{i=0}^{\infty} i|c_i| < \infty$, and where $\{\eta_t\}$ is a martingale difference sequence with unit conditional variance and $sup_t E(\eta_t^4) < \infty$.

The error process $\{u_t\}$ is $I(0)$ when $|\rho| < 1$ in (3). Alternatively, $\{u_t\}$ can be modeled as a nearly $I(1)$ process by defining $\rho = \rho_T = 1 - c/T$, where $c = 0$ corresponds to the pure $I(1)$ case. We are interested in testing if there is a trend break in y_t. Our interest in this paper therefore centers on testing the null hypothesis $H_0 : \gamma = 0$ against the two-sided alternative hypothesis $H_1 : \gamma \neq 0$, independently of whether u_t is $I(0)$ or $I(1)$[1].

Remark 1. Under the conditions of Assumption 1, the long-run variance of ε_t is given by $\omega_\varepsilon^2 = lim_{T\to\infty} T^{-1} E(\sum_{t=1}^{T} \varepsilon_t)^2 = c(1)^2$. Moreover, in the $I(0)$ case the long-run variance of u_t is given by $\omega_u^2 = lim_{T\to\infty} T^{-1} E(\sum_{t=1}^{T} u_t)^2 = \omega_\varepsilon^2/(1-\rho)^2$. Both these long-run variances play important roles in our subsequent analysis.

1.3 Joint Broken Trend Model

We start by considering a time-series process y_t with a first-order linear trend and one possible change in the slope such that the trend function is always joined at the

[1]Hereafter $I(1)$ implies ρ=$1 - c/T$, with c=0 corresponding to the pure $I(1)$ case

time of the break, which we call the joint broken trend model:

$$y_t = \alpha + \beta t + \gamma DT_t(\tau^*) + u_t, \qquad t = 1, ..., T \tag{1.4}$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad t = 2, ..., T, \quad u_1 = \varepsilon_1, \tag{1.5}$$

where $DT_t(\tau^*) = 1(t > T^*)(t - T^*)$ captures the eventual break in the slope occurring at date $T^* = \lfloor \tau^* T \rfloor$ with associated break fraction $\tau^* \in (0, 1)$. The slope coefficient changes from β to $\beta + \gamma$ at time T^*. However, notice that the trend function is continuous in every period including the date at which the slope change occurs. The discontinuous case is considered in section 4.

1.3.1 Known Break Fraction

We start considering the case where the true break fraction,τ^*, is known. The unknown break fraction case will be subsequently discussed in Section 3.3. We also assume that the long-run variances, ω_u^2 and ω_ε^2, are known in the following analysis. We will relax this assumption in Section 3.3.

Suppose one knows that u_t is $I(0)$, with $\rho = 0$ in (5) and ε_t is Gaussian white noise. Then the uniformly most powerful unbiased (optimal) test of H_0 against H_1 is the standard t-test associated with γ when (4) is estimated using OLS. The t-ratio $t_0(\tau^*)$, corrected for serial correlation in errors, can be expressed as[2]

$$t_0(\tau^*) = \frac{\hat{\gamma}(\tau^*)}{\sqrt{\omega_u^2 \left[\left\{ \sum_{t=1}^T x_{DT,t}(\tau^*) x_{DT,t}(\tau^*)' \right\}^{-1} \right]_{33}}} \tag{1.6}$$

$$\hat{\gamma}(\tau^*) = \left[\left\{ \sum_{t=1}^T x_{DT,t}(\tau^*) x_{DT,t}(\tau^*)' \right\}^{-1} \sum_{t=1}^T x_{DT,t}(\tau^*) y_t \right]_3,$$

with $x_{DT,t}(\tau^*) = \{1, t, DT_t(\tau^*)\}'$.

[2]The notation $[.]_{jj}([.]_j)$ is used to denote the jj'th (j'th) element of the matrix (vector) within the square brackets.

On the other hand, if u_t is known to be pure $I(1)$, so that $\rho = 1$ in (5), and Δu_t is a Gaussian white noise process, then the optimal test is based on the t-statistic associated with γ when (4) is estimated with OLS in first differences. That is, writing

$$\Delta y_t = \beta + \gamma DU_t(\tau^*) + \Delta u_t, \qquad t = 2, ..., T, \tag{1.7}$$

where $DU_t(\tau^*) = 1(t > T^*)$, the optimal test rejects for large values of $|t_1(\tau^*)|$, where

$$t_1(\tau^*) = \frac{\hat{\gamma}(\tau^*)}{\sqrt{\omega_\varepsilon^2 \left[\left\{ \sum_{t=2}^T x_{DU,t}(\tau^*) x_{DU,t}(\tau^*)' \right\}^{-1} \right]_{22}}} \tag{1.8}$$

$$\hat{\gamma}(\tau^*) = \left[\left\{ \sum_{t=2}^T x_{DU,t}(\tau^*) x_{DU,t}(\tau^*)' \right\}^{-1} \sum_{t=2}^T x_{DU,t}(\tau^*) \Delta y_t \right]_2,$$

with $x_{DU,t}(\tau^*) = \{1, DU_t(\tau^*)\}'$.

Theorem 1 establishes the asymptotic properties of the $|t_0(\tau^*)|$ and $|t_1(\tau^*)|$ statistics under both I(0) and I(1) errors.

Theorem 1. Let the time series process $\{y_t\}$ be generated according to (4) and (5) under $H_0 : \gamma = 0$, and let Assumption 1 hold.

(i) If u_t in (5) is $I(0)$ (*i.e.*, $|\rho| < 1$), then (*a*) $|t_0(\tau^*)| \xrightarrow{d} |J_0(\tau^*)|$, *where*
$$J_0(\tau^*) = \frac{\int_0^1 RT(r, \tau^*) dW(r)}{\{\int_0^1 RT(r, \tau^*)^2 dr\}^{1/2}},$$
and (*b*) $|t_1(\tau^*)| = O_p(T^{-1/2})$.

(ii) If u_t in (5) is $I(1)$, then (*a*) $|t_0(\tau^*)| = O_p(T)$, *and* (*b*) $|t_1(\tau^*)| \xrightarrow{d} |J_1(\tau^*, c)|$, *where*
$$J_1(\tau^*, c) = \frac{\int_0^1 RU(r, \tau^*) dW_c(r)}{\{\int_0^1 RU(r, \tau^*)^2 dr\}^{1/2}},$$
where $W(r)$ *is a standard Brownian motion, and* $W_c(r) = \int_0^r e^{-(r-s)c} dW(s)$ *denotes a standard Ornstein-Uhlenbeck (OU) process on [0,1],* $RT(r, \tau^*)$ *is the continuous-time residual from the projection of* $(r - \tau^*)1(r > \tau^*)$ *onto the space*

spanned by $\{1, r\}$*, and* $RU(r, \tau^*)$ *is the residual from the projection of* $1(r > \tau^*)$ *onto* $\{1\}$.

1.3.2 Unknown Break Fraction

In this section, we consider tests of a structural break in the trend function at an unknown break fraction. Following Andrews (1993) we consider statistics based on the maximum of the sequences of statistics $\{|t_0(\tau)|, \tau \in \Lambda\}$ and $\{|t_1(\tau)|, \tau \in \Lambda\}$, where $\Lambda = [\tau_L, \tau_U]$, with $0 < \tau_L < \tau_U < 1$, and the quantities τ_L and τ_U will be referred to as the trimming parameters, and it is assumed throughout that $\tau^* \in \Lambda$. The supremum functional allows us to test for a break and determine the break date simultaneously. These statistics are given by

$$t_0^* = \sup_{\tau \in \Lambda} |t_0(\tau)| \tag{1.9}$$

and

$$t_1^* = \sup_{\tau \in \Lambda} |t_1(\tau)|, \tag{1.10}$$

with associated break fraction estimators of τ^* given by $\hat{\tau} = arg \sup_{\tau \in \Lambda} |t_0(\tau)|$ and $\tilde{\tau} = arg \sup_{\tau \in \Lambda} |t_1(\tau)|$, respectively, such that $t_0^* \equiv |t_0(\hat{\tau})|$ and $t_1^* \equiv |t_1(\tilde{\tau})|$. The break fraction estimators,$\hat{\tau}$ and $\tilde{\tau}$, are asymptotically equivalent to the corresponding minimum sum of squares break fraction estimator of Perron and Zhu (2005) and hence consistent under the alternative of fixed break magnitude.

In the following theorem we establish the asymptotic properties of t_0^* and t_1^* under both $I(0)$ and $I(1)$ environments.

Theorem 2. Let the time series process $\{y_t\}$ be generated according to (4) and (5) under $H_0 : \gamma = 0$, and let Assumption 1 hold.

(i) If u_t is $I(0)$, then $(a) t_0^* \xrightarrow{d} \sup_{\tau \in \Lambda} |J_0(\tau)|$, and $(b) t_1^* = O_p(T^{-1/2})$.

(ii) If u_t is $I(1)$, then $(a) t_0^* = O_p(T)$, and $(b) t_1^* \xrightarrow{d} \sup_{\tau \in \Lambda} |J_1(\tau, c)|$.

Remark 2. From the results of part (i) of Theorem 2 it can be seen that when u_t is $I(0)$, t_1^* converges in probability to zero. Similarly, from the results in part (ii) of Theorem 2 it can be seen that when u_t is $I(1)$, t_0^* diverges.

In view of the large sample results in Theorem 2, we would encounter two problems in practice. First, under H_0, the appropriate test statistic (either t_0^* or t_1^*) to obtain a non-degenerate limiting distribution and second the choice of long run variance standardization (either ω_ε^2 or ω_u^2) to establish a pivotal limiting null distribution both depend on whether the errors are $I(1)$ or $I(0)$, which is not known in practice. Moreover, in practice we would also need to estimate either ω_ε^2 or ω_u^2 in order to yield a feasible testing procedure. In the next section we will explore solutions to these issues.

1.3.3 Feasible Robust Tests for a Broken Trend

In this section we address the practical issues that exist in developing a feasible test for a broken trend outlined at the end of the previous section. In Section 3.3.1, we first consider the issue of long run variance estimation, and examine the behavior of the estimators under both $I(1)$ and $I(0)$ errors. In Section 3.3.2, we then use the results from section 3.3.1 to develop an operational test against a broken trend in model (4)-(5) for the situation where the order of integration is unknown. Section 3.3.2 also presents an analysis of the asymptotic size properties of the proposed tests.

Long Run Variance Estimation

We now consider estimation of the long run variances ω_ε^2 (relevant under $I(1)$ errors) and ω_u^2 (relevant under $I(0)$ errors). Estimation of ω_ε^2 is standard and is explained first. However, in order to develop a feasible test we estimate ω_u^2 using the Berk(1974)-type autoregressive spectral density estimator. The large sample properties of these estimators are provided under both $I(0)$ and $I(1)$ environments.

Estimation of ω_ε^2

First we consider estimating ω_ε^2 when the errors are known to be $I(1)$. Let $\hat{\tau}$ be a consistent estimator of the true break fraction τ^*. We first estimate the following equation with OLS to get the residuals $\hat{\varepsilon}_t(\hat{\tau})$

$$\Delta y_t = \beta + \gamma DU_t(\hat{\tau}) + \varepsilon_t, \qquad t = 2, ..., T, \tag{1.11}$$

Nonparametric long-run variance estimator $\hat{\omega}_\varepsilon^2(\hat{\tau})$ is then given by

$$\hat{\omega}_\varepsilon^2(\hat{\tau}) = \hat{\gamma}_0(\hat{\tau}) + 2\sum_{j=1}^{T-2} k(j/l)\hat{\gamma}_j(\hat{\tau}), \tag{1.12}$$

$$\hat{\gamma}_j(\hat{\tau}) = (T-1)^{-1} \sum_{t=j+2}^{T} \hat{\varepsilon}_t(\hat{\tau})\hat{\varepsilon}_{t-j}(\hat{\tau})$$

where $k(.)$ is a kernel function with associated bandwidth parameter l. In what follows we shall make use of the Bartlett kernel for $k(.)$, such that $k(j/l) = 1 - j/(l+1)$, with bandwidth parameter $l = O(T^{1/4})$.

In Theorem 3, we now establish the large sample properties of $\hat{\omega}_\varepsilon^2(\hat{\tau})$; since in practice, the order of integration is unknown we detail the asymptotic behavior of the estimator under both $I(1)$ and $I(0)$ errors.

Theorem 3. Let the conditions of Theorem 1 hold. Then

(i) If u_t is $I(1)$, then $\hat{\omega}_\varepsilon^2(\hat{\tau}) \xrightarrow{p} \omega_\varepsilon^2$

(ii) If u_t is $I(0)$, then $\hat{\omega}_\varepsilon^2(\hat{\tau}) = O_p(l^{-1})$.

Estimation of ω_u^2

We now consider estimating ω_u^2 in the case where the errors are known to be $I(0)$. Here we focus on Berk-type autoregressive spectral density estimators in the estimation of ω_u^2. The Berk-type estimator is the key in developing a feasible test that works under both $I(0)$ and $I(1)$ errors. As will be seen in the next section, estimation of ω_u^2 using an autoregressive framework allows the t_0^* statistic to be stochastically bounded under both $I(0)$ and $I(1)$ errors and enables us to utilize the *union of rejections* principle to develop a feasible test.

We estimate ω_u^2 under the null $H_0 : \gamma = 0$ and hence no trend break dummy is included in the following regression

$$y_t = \alpha + \beta t + u_t, \qquad t = 1, ..., T \tag{1.13}$$

We estimate equation (13) via OLS and get the residuals $\hat{u}_t$.The Berk-type estimator $\hat{\omega}_u^2$ is then given by

$$\hat{\omega}_u^2 = \frac{\hat{\sigma}^2}{\hat{\pi}^2}$$

where $\hat{\pi}$ and $\hat{\sigma}$ are obtained from the OLS regression

$$\Delta\hat{u}_t = \hat{\pi}\hat{u}_{t-1} + \sum_{j=1}^{k-1} \hat{\psi}_j \Delta\hat{u}_{t-j} + \hat{e}_t, \quad t = k+2, ..., T \tag{1.14}$$

with $\hat{\sigma}^2 = (T-2k-1)^{-1} \sum_{t=k+2}^{T} \hat{e}_t^2$. As is standard, we require that the lag truncation parameter, k, in (14) satisfies the condition that, as $T \to \infty$, $1/k + k^3/T \to 0$.

In Theorem 4, we now establish the asymptotic properties of $\hat{\omega}_u^2$ under both $I(0)$ and $I(1)$ environments.

Theorem 4. Let the conditions of Theorem 1 hold. Then under the null $H_0 : \gamma = 0$

(i) If u_t is $I(1)$, then $T^{-2}\hat{\omega}_u^2 \xrightarrow{d} \omega_\varepsilon^2 \Phi(c)$

(ii) If u_t is $I(0)$, then $\hat{\omega}_u^2 \xrightarrow{p} \omega_u^2$

where

$$\Phi(c) = \frac{\left\{\int_0^1 Q(r,c)^2 dr\right\}^2}{\left\{\int_0^1 Q(r,c)^2 dW_c(r)\right\}^2}$$

and $Q(r,c)$ is a continuous-time residual from the projection of $W_c(r)$ onto the space spanned by $\{1, r\}$.

Remark 3. Observe from part (ii) of Theorem 4 that $\hat{\omega}_u^2$ is a consistent estimator of ω_u^2 under $I(0)$ errors. It is also seen from part (i) of Theorem 4 that the Berk-type estimator diverges at a rate of T^2 which is crucial in our construction of a feasible test, as explained in the next section.

Feasible Tests and Asymptotic Size

Having proposed suitable long run variance estimators and having established their asymptotic properties, we are now in a position to define feasible statistics for detecting a break in the trend function. The results of Theorem 2, along with the properties of the long run variance estimators described in Theorems 3 and 4, suggest the following statistics, appropriate under $I(1)$ and $I(0)$ errors, respectively:

$$S_1 = \sup_{\tau \in \Lambda} |\hat{t}_1(\tau)| \tag{1.15}$$

$$S_0 = \sup_{\tau \in \Lambda} |\hat{t}_0(\tau)| \tag{1.16}$$

where

$$\hat{t}_1(\tau) = \frac{\hat{\gamma}(\tau)}{\sqrt{\hat{\omega}_\varepsilon^2(\tau)\left[\left\{\sum_{t=2}^T x_{DU,t}(\tau)x_{DU,t}(\tau)'\right\}^{-1}\right]_{22}}}$$

$$\hat{\gamma}(\tau) = \left[\left\{\sum_{t=2}^T x_{DU,t}(\tau)x_{DU,t}(\tau)'\right\}^{-1}\sum_{t=2}^T x_{DU,t}(\tau)\Delta y_t\right]_2,$$

and

$$\hat{t}_0(\tau) = \frac{\hat{\gamma}(\tau)}{\sqrt{\hat{\omega}_u^2\left[\left\{\sum_{t=1}^T x_{DT,t}(\tau)x_{DT,t}(\tau)'\right\}^{-1}\right]_{33}}}$$

$$\hat{\gamma}(\tau) = \left[\left\{\sum_{t=1}^T x_{DT,t}(\tau)x_{DT,t}(\tau)'\right\}^{-1}\sum_{t=1}^T x_{DT,t}(\tau)y_t\right]_3.$$

In the following lemma we now establish the asymptotic properties of the S_1 and S_0 statistics of (15) and (16), respectively, in both $I(1)$ and $I(0)$ environments.

Lemma 1. Let the time series process $\{y_t\}$ be generated according to (4) and (5) under $H_0 : \gamma = 0$, and let Assumption 1 hold.

(i) If u_t is $I(0)$, then

(*a*) $S_0 \xrightarrow{d} \sup_{\tau\in\Lambda} |J_0(\tau)|$,

(*b*) $S_1 = O_p\left\{(l/T)^{1/2}\right\}$.

(ii) If u_t is $I(1)$, then

(*a*) $S_0 \xrightarrow{d} \dfrac{\sup_{\tau\in\Lambda} |K_0(\tau, c)|}{\Phi^{1/2}(c)}$, *where*

$$K_0(\tau, c) = \frac{\int_0^1 RT(r, \tau)W_c(r)dr}{\{\int_0^1 RT(r, \tau)^2 dr\}^{1/2}}$$

(*b*) $S_1 \xrightarrow{d} \sup_{\tau\in\Lambda} |J_1(\tau, c)|$.

Asymptotic null critical values for S_1 under $I(1)$ errors with $c = 0$, and S_0 under $I(0)$ errors, are reported in Table 1, for the settings τ_L=0.1 and τ_U=0.9, and for the significance levels $\xi = 0.10, 0.05$ and 0.01. The numerical results were obtained by simulation of the appropriate limiting distributions using discrete approximations for $T = 1,000$ and 10,000 replications using normal $IID(0, 1)$ random .

Critical Values

	S_1	S_0	κ_ξ
$\xi = 0.10$	2.741	2.268	1.0238
$\xi = 0.05$	3.024	2.570	1.0065
$\xi = 0.01$	3.565	3.139	1.0048

Note: The critical values for S_0 and S_1 are for the $I(0)$ and $I(1)(c = 0)$ cases, respectively.

Table 1 Asymptotic critical values for nominal ξ-level S_0 and S_1 tests, and asymptotic κ_ξ values for U.

It is also of interest, given lack of knowledge concerning the order of integration, to examine the size properties of S_1 when $c > 0$, and also S_0 under both $c = 0$ and $c > 0$. These results are presented in Table 2, again obtained via direct simulation of the limit distributions in Lemma 1. The S_1 test becomes increasingly under-sized as c increases. Therefore, employing critical values which are appropriate for $c = 0$ will result in an under-sized test when $c > 0$. Also, notice from part (i)-(b) of Lemma 1 that S_1 converges in probability to zero in $I(0)$ case, and thus it is automatically under-sized under $I(0)$ errors.

Of particular interest is the behavior of S_0 in the (local to) $I(1)$ case. For all significance levels, we see that the asymptotic size of S_0 remains well below the nominal size across all c. Therefore, employing critical values which are appropriate for $I(0)$ case will result in an under-sized test in the (local to) $I(1)$ case.

We now turn to consideration of a feasible test that can be applied in the absence of knowledge concerning the order of integration. Our approach deliberately exploits the under-sizing phenomenon seen in the S_0 test in the (local to) $I(1)$ world, and is based on the *union of rejections* approach advocated by Harvey et al. (2009a) in a unit root testing context. Specifically, we consider the union of rejections decision rule

U : Reject H_0 if $\{S_1 > \kappa_\xi cv_\xi^1 \text{ or } S_0 > \kappa_\xi cv_\xi^0\}$

where cv_ξ^1 and cv_ξ^0 denote the ξ significance level asymptotic critical values of S_1 under $I(1)$ $(c = 0)$ errors and S_0 under $I(0)$ errors, respectively, and κ_ξ is a positive scaling constant whose role is made precise below.

If the U decision rule was to be applied with $\kappa_\xi = 1$ (i.e. without any adjustment to the asymptotic critical values used for the constituent tests in U), then the testing strategy would be asymptotically correctly sized under $I(0)$ errors, as $S_1 \xrightarrow{p} 0$. In the $I(1)$ case, the Bonferroni inequality along with the size results for S_1 and S_0 reported in Table 2, show that such a strategy could only ever be (modestly) asymptotically over-sized when c=0; indeed, the maximum possible asymptotic sizes at the $0.10, 0.05$

and 0.01 nominal significance levels are, respectively, 0.133, 0.059 and 0.0107 ,such that the size distortions will be small. However, to ensure that U is an asymptotically conservative testing strategy (i.e. asymptotically exactly correctly sized in the case of $I(0)$ errors and $I(1)$ errors when $c = 0$, and always asymptotically under-sized elsewhere), we can avoid any size distortions by suitably choosing κ_ξ.

	S_1	S_0	U
Panel A,$\xi = 0.10$			
$I(1), c = 0$	0.1000	0.0331	0.100
$I(1), c = 10$	0.0054	0.0215	0.025
$I(1), c = 20$	0.0000	0.0174	0.017
$I(1), c = 40$	0.0000	0.0141	0.014
$I(0)$	0.0000	0.1000	0.088
Panel B,$\xi = 0.05$			
$I(1), c = 0$	0.050	0.009	0.050
$I(1), c = 10$	0.001	0.006	0.006
$I(1), c = 20$	0.000	0.004	0.004
$I(1), c = 40$	0.000	0.003	0.003
$I(0)$	0.000	0.050	0.047
Panel C,$\xi = 0.01$			
$I(1), c = 0$	0.01	0.0007	0.0100
$I(1), c = 10$	0.00	0.0004	0.0004
$I(1), c = 20$	0.00	0.0001	0.0001
$I(1), c = 40$	0.00	0.0001	0.0001
$I(0)$	0.00	0.0100	0.0096

Note: The rejections for S_1 are computed using critical values for S_1 under $I(1), c = 0$ errors; the rejections for S_0 are computed using critical values for S_0 under $I(0)$ errors.

Table 2 Asymptotic sizes of nominal ξ-level tests under $I(1)$ and $I(0)$ errors.

Noting that the maximum size of U is realized when $c = 0$, choosing κ_ξ such that U has an asymptotic size of ξ in this case ensures that the procedure will be conservative. We therefore obtain κ_ξ by simulating the limit distribution of $\max\{S_1, (cv_\xi^1/cv_\xi^0)S_0\}$, calculating the ξ-level critical value for this distribution, say cv_ξ^{max}, and then computing $\kappa_\xi = cv_\xi^{max}/cv_\xi^1$. Values of κ_ξ for different ξ are shown in Table 1. Hereafter, reference to the decision rule U assumes the κ_ξ adjustment values from Table 1 are used.

Table 2 also provides asymptotic size results for U. As expected, the testing strategy is correctly sized for $I(1)$ errors when $c = 0$. When the errors are $I(1)$ with $c > 0$, U is conservative, in line with the size properties of the constituent tests S_1 and S_0 discussed above. It is only slightly conservative when the errors are $I(0)$.

Remark 4. It is important to note that the union of rejections procedure is only rendered viable due to specific behavior of the Berk-type estimator $\hat{\omega}_u^2$ under $I(1)$ errors, in that it diverges at a rate T^2; see Theorem 4(i). This ensures that S_0 is $O_p(1)$. If a typical kernel-based (e.g. Bartlett) long run variance estimator with bandwidth l, say, growing at rate smaller than T was used, then under $I(1)$ errors, it is easy to show that $\hat{\omega}_u^2$ diverges at a rate less than T^2, so that S_0 diverges to ∞. In such a case, a union of rejections approach is clearly precluded, because, regardless of the choice of κ_ξ, its size would approach one in the limit under $I(1)$ errors.

Asymptotic Power

In this section, we first investigate the local asymptotic power properties of our feasible test. In this case, we model the break magnitude as the appropriate Pitman drifts under $I(0)$ and $I(1)$ errors respectively. The local asymptotic power results demonstrate that the conservativeness of our test under $I(1)$ errors with $c > 0$ does not necessarily imply a loss of power. Next, we show that our feasible test is consistent under the fixed break magnitude by establishing the divergence rates of the constituent tests.

Consider first the case where the break magnitude, γ, is modeled as a local alternative. In this case, we may partition H_1 into two scaled components: $H_{1,0} : \gamma = \omega_u T^{-3/2}\vartheta$ when u_t is $I(0)$, and $H_{1,1} : \gamma = \omega_\varepsilon T^{-1/2}\vartheta$ when u_t is $I(1)$, where in each case ϑ is a finite non-negative constant. As we shall see below, these provide the appropriate Pitman drifts on γ under $I(0)$ and $I(1)$ errors, respectively, to obtain nondegenerate and pivotal (except ϑ) asymptotic distributions under the alternative hypothesis.

In the following theorem we establish the large sample behavior of the S_1 and S_0 statistics of (15) and (16), respectively, under the local alternative in both $I(1)$ and $I(0)$ environments.

Theorem 5. Let the time series process $\{y_t\}$ be generated according to (4) and (5), and let Assumption 1 hold.

(i) If u_t is $I(0)$, then under $H_{1,0} : \gamma = \omega_u T^{-3/2}\vartheta$

(a) $S_0 \xrightarrow{d} \sup_{\tau\in\Lambda} |L_{00}(\tau, \vartheta)|$,

where

$$L_{00}(\tau, \vartheta) = \vartheta\{\textstyle\int_0^1 RT(r, \tau)^2 dr\}^{1/2} + \dfrac{\int_0^1 RT(r, \tau)dW(r)}{\{\int_0^1 RT(r, \tau)^2 dr\}^{1/2}}$$

(b) $S_1 = O_p\left\{(l/T)^{(1/2)}\right\}$.

(ii) If u_t is $I(1)$, then under $H_{1,1} : \gamma = \omega_\varepsilon T^{-1/2}\vartheta$

(a) $S_0 \xrightarrow{d} \dfrac{\sup_{\tau\in\Lambda} |K_{01}(\tau, \vartheta, c)|}{\Phi^{1/2}(c)}$,

where

$$K_{01}(\tau, \vartheta, c) = \vartheta\{\textstyle\int_0^1 RT(r, \tau)^2 dr\}^{1/2} + \dfrac{\int_0^1 RT(r, \tau)W_c(r)dr}{\{\int_0^1 RT(r, \tau)^2 dr\}^{1/2}}$$

(b) $S_1 \xrightarrow{d} \sup_{\tau\in\Lambda} |L_{11}(\tau, \vartheta, c)|$,

where

$$L_{11}(\tau, \vartheta, c) = \vartheta\{\textstyle\int_0^1 RU(r, \tau)^2 dr\}^{1/2} + \dfrac{\int_0^1 RU(r, \tau)dW_c(r)}{\{\int_0^1 RU(r, \tau)^2 dr\}^{1/2}}$$

Table 3 shows asymptotic local powers of S_1,S_0 and U, conducted at the nominal 0.05-level. We consider the same settings as in Table 2, and the same error specifications (i.e. $I(1)$ errors with $c \geq 0$ and $I(0)$ errors). The break magnitudes are benchmarked so that the powers of S_1 for $c = 0$ in the $I(1)$ case, and S_0 in the $I(0)$ case, are equal to 0.50.

Consider first the behavior of S_1. For $I(1)$ errors, power is monotonically decreasing as c increases. Note also that for $I(0)$ errors, the power of S_1 is zero, in line with the results of Theorem 5(i)(b). The behavior of S_0 is more interesting: the power of S_0 increases significantly as the errors move from $I(0)$ to $I(1)$. The S_0 test is quite powerful, indeed power is close to 1, under the $I(1)$ environment regardless of the value of c.

	S_1	S_0	U
$I(1), c = 0$	0.500	0.987	0.997
$I(1), c = 10$	0.342	0.897	0.941
$I(1), c = 20$	0.140	0.898	0.933
$I(1), c = 40$	0.011	0.921	0.922
$I(0)$	0	0.500	0.500

Table 3 Asymptotic powers of nominal 0.05-level tests under $I(1)$ and $I(0)$ errors.

Inspection of the power performance of U shows that the conservative nature of U for $I(1)$ errors with $c > 0$, as seen in Table 2, does not translate into poor power performance for these cases. The U test essentially capitalizes on the relatively high power of the constituent tests. U generally displays power very close to the better power of the two individual tests S_1 and S_0. On the other hand, there are instances where the power of U exceeds that of either of the constituent tests S_1 and S_0, resulting from the fact that the rejections from S_1 and S_0 need not be perfectly correlated. The robust power performance of U, therefore makes a strong case for using the modified union of rejections approach in practice.

We now consider the asymptotic behavior of our feasible test under fixed alternatives in order to establish the consistency of our tests. The following theorem establishes the consistency of our tests under a fixed alternative of the form $H_1 : \gamma \neq 0$.

Theorem 6. Let the time series process $\{y_t\}$ be generated according to (4) and (5) under $H_1 : \gamma \neq 0$, and let Assumption 1 hold.

(i) If u_t is $I(0)$, then (a) $S_0 = O_p((T/k)^{1/2})$, and (b) $S_1 = O_p((lT)^{1/2})$, thus (c) $U = O_p((lT)^{1/2})$.

(ii) If u_t is $I(1)$, then (a) $S_0 = O_p(T^{1/2})$, and (b) $S_1 = O_p(T^{1/2})$, thus (c) $U = O_p(T^{1/2})$.

Under $H_1 : \gamma \neq 0$, it is seen from Theorem 6 that our feasible test, U, is consistent at rate $O_p((lT)^{1/2})$ when u_t is $I(0)$ and at rate $O_p(T^{1/2})$ when u_t is $I(1)$.

1.4 Disjoint Broken Trend Model

Although trend breaks are the central concern of this paper, we might also consider extending our analysis to allow (but not test for) a break in level occurring at the same time as the break in trend. Therefore, we consider the following model:

$$y_t = \alpha + \beta t + \delta DU_t(\tau^*) + \gamma DT_t(\tau^*) + u_t, \qquad t = 1, ..., T \tag{1.17}$$

and

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad t = 2, ..., T, \quad u_1 = \varepsilon_1. \tag{1.18}$$

In what follows we will refer to (17) and (18) together as the disjoint broken trend model. Notice that δ and γ capture the change, respectively, in the level and slope coefficients of the series at time $T^* = \lfloor \tau^* T \rfloor$. The slope coefficient changes from β to $\beta + \gamma$ and the level shifts from α to $\alpha + \delta$ at time T^*. The trend function is discontinuous at the break date T^* if $\delta \neq 0$.

The first-differenced form of the model is given by:

$$\Delta y_t = \beta + \delta D_t(\tau^*) + \gamma DU_t(\tau^*) + \Delta u_t, \qquad t = 2, ..., T, \tag{1.19}$$

where $D_t(\tau^*) = 1(t = T^* + 1)$. The null hypothesis of interest continues to be $H_0 : \gamma = 0$ against the two sided alternative: $H_1 : \gamma \neq 0$. The interest lies only on the break in the trend slope.

We propose our feasible test following the same steps as in Section 3. We first consider the appropriate test statistics for $I(0)$ and $I(1)$ cases, respectively, assuming that the true break fraction, τ^*, is known. We consequently redefine $\hat{t}_0(\tau^*)$ as follows:

$$\hat{t}_0(\tau^*) = \frac{\hat{\gamma}(\tau^*)}{\sqrt{\hat{\omega}_u^2(\tau^*)\left[\left\{\sum_{t=1}^T x_{DT,t}(\tau^*)x_{DT,t}(\tau^*)'\right\}^{-1}\right]_{44}}}, \tag{1.20}$$

$$\hat{\gamma}(\tau^*) = \left[\left\{\sum_{t=1}^T x_{DT,t}(\tau^*)x_{DT,t}(\tau^*)'\right\}^{-1}\sum_{t=1}^T x_{DT,t}(\tau^*)y_t\right]_4,$$

with $x_{DT,t}(\tau^*) = \{1, t, DU_t(\tau^*), DT_t(\tau^*)\}'$ and $\hat{\omega}_u^2(\tau^*)$ calculated as in (14) under the null $H_0 : \gamma = 0$ using the Berk-type autoregressive spectral density estimator. The residuals used in the estimation of the Berk-type long run variance are the OLS residuals $\hat{u}_t(\tau^*) = y_t - \hat{\alpha} - \hat{\beta}t - \hat{\delta}DU_t(\tau^*)$. Similarly, $\hat{t}_1(\tau^*)$ is redefined to be

$$\hat{t}_1(\tau^*) = \frac{\tilde{\gamma}(\tau^*)}{\sqrt{\hat{\omega}_\varepsilon^2(\tau^*)\left[\left\{\sum_{t=2}^T x_{DU,t}(\tau^*)x_{DU,t}(\tau^*)'\right\}^{-1}\right]_{33}}}, \tag{1.21}$$

$$\tilde{\gamma}(\tau^*) = \left[\left\{\sum_{t=2}^T x_{DU,t}(\tau^*)x_{DU,t}(\tau^*)'\right\}^{-1}\sum_{t=2}^T x_{DU,t}(\tau^*)\Delta y_t\right]_3,$$

with $x_{DU,t}(\tau^*) = \{1, D_t(\tau^*), DU_t(\tau^*)\}'$, and $\hat{\omega}_\varepsilon^2(\tau^*)$ calculated nonparametrically as in (12) with bandwidth parameter $l = O(T^{1/4})$ but using the OLS residuals $\hat{\varepsilon}_t(\tau^*) = \Delta y_t - \tilde{\beta} - \tilde{\delta}D_t(\tau^*) - \tilde{\gamma}(\tau^*)DU_t(\tau^*)$.

In order to accommodate unknown break point, we again consider statistics based on the maximum of the sequences of statistics $\{|\hat{t}_0(\tau)|, \tau \in \Lambda\}$ and $\{|\hat{t}_1(\tau)|, \tau \in \Lambda\}$. These statistics are given by:

$$S_0 = \sup_{\tau \in \Lambda} |\hat{t}_0(\tau)| \tag{1.22}$$

$$S_1 = \sup_{\tau \in \Lambda} |\hat{t}_1(\tau)| \tag{1.23}$$

As in SV, the null hypothesis H_0 must be restated as $H_0 : \gamma = \delta = 0$ in the case of unknown break fraction in order to obtain a pivotal limiting null distribution for our test statistic. The following theorem, whose proof is a straightforward generalization of those in Section 3 and is therefore omitted, details the large sample behavior of the redefined S_0 and S_1 statistics under H_0.

Theorem 7. Let the time series process $\{y_t\}$ be generated according to (17) and (18) under $H_0 : \gamma = \delta = 0$, and let Assumption 1 hold.

(i) If u_t is $I(0)$, then (a) $S_0 \xrightarrow{d} \sup_{\tau \in \Lambda} |J_{U,0}(\tau)|$, (b) $S_1 = O_p\left\{(l/T)^{1/2}\right\}$.

where

$$J_{U,0}(\tau) = \frac{\int_0^1 RT_U(r,\tau^*)dW(r)}{\{\int_0^1 RT_U(r,\tau^*)^2 dr\}^{1/2}}.$$

(ii) If u_t is $I(1)$, then (a) $S_0 \xrightarrow{d} \sup_{\tau \in \Lambda} \left|\frac{K_{U,0}(\tau,c)}{\Phi^{1/2}(c,\tau)}\right|$, (b) $S_1 \xrightarrow{d} \sup_{\tau \in \Lambda} |J_1(\tau,c)|$,

where

$$K_{U,0}(\tau,c) = \frac{\int_0^1 RT_U(r,\tau)W_c(r)dr}{\{\int_0^1 RT_U(r,\tau)^2 dr\}^{1/2}}, \quad \textit{and}$$

$$\Phi(c,\tau) = \frac{\left\{\int_0^1 Q(r,c,\tau)^2 dr\right\}^2}{\left\{\int_0^1 Q(r,c,\tau)^2 dW_c(r)\right\}^2}$$

and $Q(r,c,\tau)$ is a continuous-time residual from the projection of $W_c(r)$ onto the space spanned by $\{1, r, 1(r > \tau)\}$, and $RT_U(r,\tau)$ is a continuous-time residual from the projection of $(r-\tau)1(r > \tau)$ onto the space spanned by $\{1, r, 1(r > \tau)\}$.

Remark 5. Observe from the result given in part $(ii)(b)$ of Theorem 7 that the limiting distribution of S_1 from the disjoint broken trend model is identical to that for joint broken trend model given in Lemma 1 $(ii)(b)$. This is because the regressor $D_t(\tau)$ has an asymptotically negligible effect on S_1.

Critical Values

	S_1	S_0	κ_ξ
$\xi = 0.10$	2.741	2.901	1.0288
$\xi = 0.05$	3.024	3.163	1.0114
$\xi = 0.01$	3.565	3.655	1.00

Note: The critical values for S_0 and S_1 are for the $I(0)$ and $I(1)(c = 0)$ cases, respectively.

Table 4 Asymptotic critical values for nominal ξ-level S_0 and S_1 tests, and asymptotic κ_ξ values for U.

Asymptotic null critical values for S_1 under $I(1)$ errors with $c = 0$, and S_0 under $I(0)$ errors, are reported in Table 4, for the settings $\tau_L = 0.1$ and $\tau_U = 0.9$, and for the significance levels $\xi = 0.10, 0.05$ and 0.01. Table 5 presents, similar to Table 2, the asymptotic size properties of S_1 when $c > 0$, and also S_0 under both $c = 0$ and $c > 0$. From Table 5 we see that the S_1 test becomes increasingly under-sized as c increases. Similarly, the asymptotic size of S_0 remains well below the nominal size across all c. This under-sizing phenomenon seen in the S_0 test under the (local to)$I(1)$ errors and in the S_1 test when $c > 0$ or errors are $I(0)$ renders the *union of rejections* principle viable. Specifically, our feasible test which is based on the union of rejections decision rule can be stated as follows:

$$U : \text{Reject } H_0 \quad \text{if } \{S_1 > \kappa_\xi cv_\xi^1 \text{ or } S_0 > \kappa_\xi cv_\xi^0\}$$

where cv_ξ^1 and cv_ξ^0 denote the ξ significance level asymptotic critical values of S_1 under $I(1)$ $(c = 0)$ errors and S_0 under $I(0)$ errors, respectively, and κ_ξ is as explained in Section 3.3.2.

	S_1	S_0	U
Panel A, $\xi = 0.10$			
$I(1), c = 0$	0.1000	0.0281	0.100
$I(1), c = 10$	0.0054	0.0037	0.008
$I(1), c = 20$	0.0000	0.0033	0.003
$I(1), c = 40$	0.0000	0.0022	0.002
$I(0)$	0.0000	0.1000	0.082
Panel B, $\xi = 0.05$			
$I(1), c = 0$	0.050	0.0083	0.050
$I(1), c = 10$	0.001	0.0014	0.002
$I(1), c = 20$	0.000	0.0013	0.001
$I(1), c = 40$	0.000	0.0011	0.001
$I(0)$	0.000	0.0500	0.044
Panel C, $\xi = 0.01$			
$I(1), c = 0$	0.01	0.0005	0.010
$I(1), c = 10$	0.00	0.0001	0.000
$I(1), c = 20$	0.00	0.0001	0.000
$I(1), c = 40$	0.00	0.0001	0.000
$I(0)$	0.00	0.0100	0.010

Note: The rejections for S_1 are computed using critical values for S_1 under $I(1), c = 0$ errors; the rejections for S_0 are computed using critical values for S_0 under $I(0)$ errors.

Table 5 Asymptotic sizes of nominal ξ-level tests under $I(1)$ and $I(0)$ errors.

The asymptotic size results for our feasible test, U, are also presented in Table 5. As expected, the testing strategy is correctly sized for $I(1)$ errors when $c = 0$. When the errors are $I(1)$ with $c > 0$, U is conservative, in line with the size properties of the constituent tests S_1 and S_0 discussed above. It is also almost correctly sized when the errors are $I(0)$.

We now turn to consistency properties of our feasible test. We consider fixed alternatives of the form $H_1 : \gamma \neq 0$, with δ now unrestricted. In the following theorem, we show that our feasible test is consistent under a fixed alternative by establishing the divergence rates of our feasible test under both $I(0)$ and $I(1)$ errors.

Theorem 8. Let the time series process $\{y_t\}$ be generated according to (17) and (18) under $H_1 : \gamma \neq 0$, and let Assumption 1 hold.

(i) If u_t is $I(0)$, then (a) $S_0 = O_p((T/k)^{1/2})$, and (b) $S_1 = O_p((lT)^{1/2})$, thus (c) $U = O_p((lT)^{1/2})$.

(ii) If u_t is $I(1)$, then (a) $S_0 = O_p(T^{1/2})$, and (b) $S_1 = O_p(T^{1/2})$, thus (c) $U = O_p(T^{1/2})$.

The proof of theorem is very similar to that of Theorem 6 and, hence, is omitted. Under H_1, when u_t is $I(0)$, we obtain that U is consistent at rate $O_p((lT)^{1/2})$, while if u_t is $I(1)$, U is consistent at rate $O_p(T^{1/2})$.

1.5 Finite Sample Results

In this section we report the finite sample size and power performances of our structural break tests. We employ 10% trimming ($\tau_L = 0.1, \tau_U = 0.9$) throughout our Monte Carlo simulations. All the results reported in this section are for two-sided tests conducted at the 0.05 nominal asymptotic significance level, and were computed over 5,000 replications.

We employ the following data generating process (DGP), which is a simplified version of (17) and (18), to carry out our simulations

$$y_t = \delta DU_t(\tau^*) + \gamma DT_t(\tau^*) + u_t, \qquad t = 1, ..., T \tag{1.24}$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad t = 2, ..., T, \quad u_1 = \varepsilon_1. \tag{1.25}$$

where $\varepsilon_t \sim NIID(0, 1)$. In Section 5.1, we provide the finite sample size properties of our proposed tests and compare the results to those of the tests proposed in HLT.

In Section 5.2, we subsequently investigate the finite sample power properties of our test, again relative to the tests of HLT. We form our union of rejections decision rule as detailed in Section 3.3 for the joint broken trend model, and as detailed in Section 4 for the disjoint case.

Here we briefly outline the tests proposed by HLT before we report the finite sample results. As in this paper, HLT also utilize the two test statistics S_0 and S_1, the supremum of the t-ratios for the levels and the differenced data, respectively. The first difference is that HLT normalize the t-statistics for the levels data using a nonparametric estimator of the long run variance ω_u^2, with bandwidth parameter $l = O(T^{1/4})$. The S_0 statistic, thus, diverges to infinity under $I(1)$ errors with this choice of the long run variance estimator. Second, instead of using a union of rejections principle, which is infeasible due to their choice of the long run variance estimator, they take a weighted average of the S_0 and S_1 statistics.They adopt the stationarity test statistics of KPSS, Q_0 and Q_1, calculated from the residuals of the levels data $\{\hat{u}_t(\tau^*)\}_{t=1}^T$ and the differenced data $\{\hat{\varepsilon}_t(\tau^*)\}_{t=2}^T$ respectively to form their weight function, $\lambda(.,.)$, of the form

$$\lambda(Q_0(\tau^*), Q_1(\tau^*)) = exp[-\{gQ_0(\tau^*)Q_1(\tau^*)\}^2], \tag{1.26}$$

where

$$Q_0(\tau^*) = \frac{\sum_{t=1}^T \left(\sum_{i=1}^t \hat{u}_i(\tau^*)\right)^2}{T^2\hat{\omega}_u^2(\tau^*)}, \qquad Q_1(\tau^*) = \frac{\sum_{t=2}^T \left(\sum_{i=2}^t \hat{\varepsilon}_i(\tau^*)\right)^2}{(T-1)^2\hat{\omega}_\varepsilon^2(\tau^*)}$$

and g is a positive constant. Finally, the constant m_ξ is chosen such that for a given significance level, ξ, the critical values for the test coincide under $I(0)$ and $I(1)$ errors. The robust structural break test statistic of HLT, t_λ, can then be written as

$$t_\lambda = \{\lambda(Q_0(\tau^*), Q_1(\tau^*)) \times S_0\} + m_\xi \{[1 - \lambda(Q_0(\tau^*), Q_1(\tau^*))] \times S_1\}. \tag{1.27}$$

HLT show that under $I(0)$ errors $t_\lambda = S_0 + o_p(1)$ and under $I(1)$ errors $t_\lambda = m_\xi S_1 + o_p(1)$.

1.5.1 *Size Properties*

Table 6 reports the empirical size performance of our U test and t_λ test of HLT for each of joint and disjoint broken trend models. These we obtained by setting $\delta = \gamma = 0$ in (24). The AR parameter in (25) were varied over $\rho = 1 - (c/T)$ for $c \in \{0, 10, 20, 50, T\}$. We consider three different sample sizes; $T = 100$, $T = 150$, and $T = 300$. In the computation of our U test statistic, the long run variance estimator $\hat{\omega}_u^2$ uses values of k (in (14)) determined according to the BIC criterion with $k_{max} = \lfloor 4(T/100)^{1/4} \rfloor$. We set the bandwidth parameter $l = \lfloor 4(T/100)^{1/4} \rfloor$ in the computation of all other long variance estimators. Finally, we set $g = 250$ in (25), since it gives rise to the most accurate weights for both $I(1)$ and $I(0)$ cases.

Joint	T=100		T=150		T=300	
	U	t_λ	U	t_λ	U	t_λ
c=0	9.0%	21.3%	8.6%	16.2%	6.7%	10.2%
c=10	2.8%	15.4%	2.6%	13.0%	2.0%	8.8%
c=20	2.6%	14.9%	2.5%	16.3%	1.7%	14.5%
c=50	3.6%	10.7%	3.4%	16.7%	3.0%	20.8%
c=T	5.2%	5.4%	5.0%	5.7%	5.1%	6.1%
Disjoint	T=100		T=150		T=300	
	U	t_λ	U	t_λ	U	t_λ
c=0	8.4%	28.1%	7.8%	22.9%	5.7%	15.4%
c=10	2.2%	19.9%	2.1%	18.0%	1.2%	13.9%
c=20	2.3%	17.6%	2.0%	20.9%	1.2%	22.1%
c=50	3.2%	11.7%	2.4%	18.2%	1.3%	24.8%
c=T	4.6%	6.6%	3.8%	6.2%	4.4%	5.3%

Table 6 Empirical sizes of nominal 0.05 level tests

In the case of $I(1)$ errors ($c = 0$), Table 6 shows that our U test is slightly oversized in finite samples. However, we see that the size distortions get smaller as the sample size increases. On the other hand, when the errors are $I(0)$ $(c > 0)$ the U test seems to be undersized, except for the white noise case $(c = T)$. The undersizing effect is more pronounced as the persistence and the sample size increases. Also, the undersizing phenomenon is more obvious in the disjoint broken trend model. However, keep in mind that all these finite sample results are consistent with the asymptotic properties of the U test given in Tables 2 and 5.

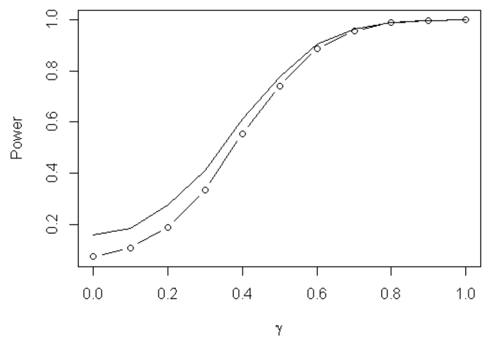
The finite sample size properties of the t_λ test can also be seen in Table 6. The t_λ test is oversized for persistent data regardless of the sample size and the model. The size distortion can be as high as 25% at a 5% nominal level. The main source of the size distortions is the way the t_λ test is constructed. In finite samples, the weight function does not pick the appropriate statistic quite accurately[3] and since the S_0 test statistic is large (asymptotically unbounded) for persistent data, the t_λ test rejects too often.

A simple comparison of the results from Table 6 favors our U test over the t_λ test of HLT in terms of the finite sample size. In the next section, we compare the finite sample power properties of these two tests.

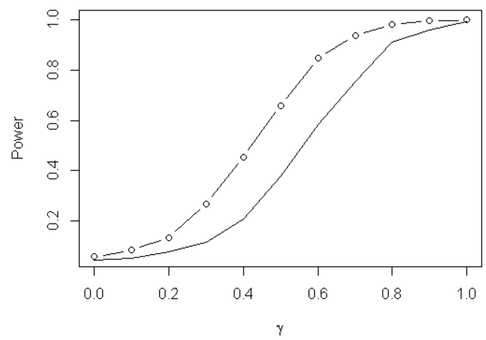
1.5.2 Power Properties

Figures 1.1-1.2 and 1.3 present the empirical power performance of the tests. The data were generated according to (24) and (25) for a grid of γ values, covering the range $[0, 1]$ in steps of 0.1. In order to save some space, we only present results for the joint broken trend model, where we set $\delta = 0$. The results for the disjoint broken trend model are qualitatively similar and are available upon request. We consider two break fractions $\tau^* \in \{0.25, 0.5\}$ and again let the AR parameter vary over $\rho = 1 - (c/T)$ for $c \in \{0, 10, 20, 50, T\}$ in (25). The parameters values used in the estimation of the long run variances and weight function are the same as in the size simulations.

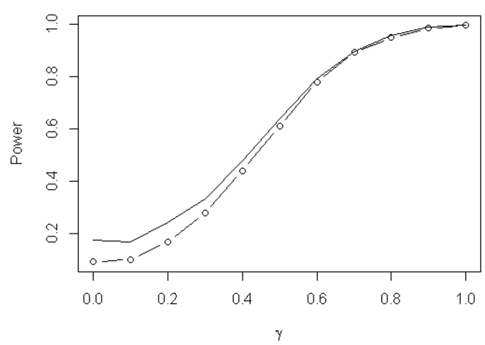
[3] For highly persistent data, the appropriate test statistic is S_1



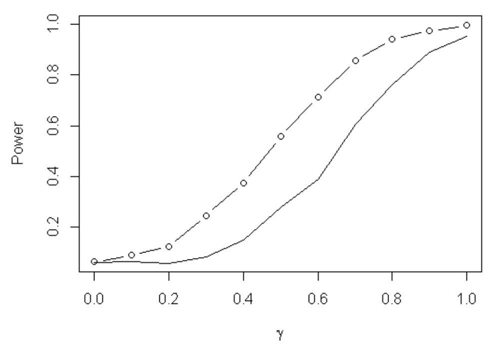
(a) $c = 0,\tau^* = 0.5$



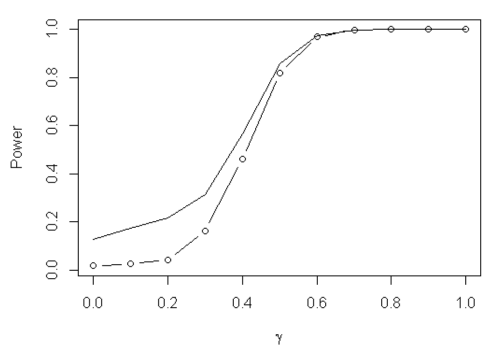
(b) $c = 0,\tau^* = 0.5$, size-adjusted



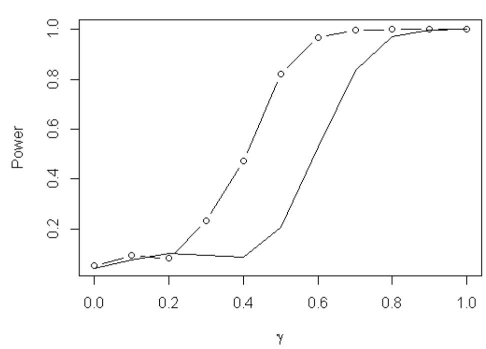
(c) $c = 0,\tau^* = 0.25$



(d) $c = 0,\tau^* = 0.25$, size adjusted

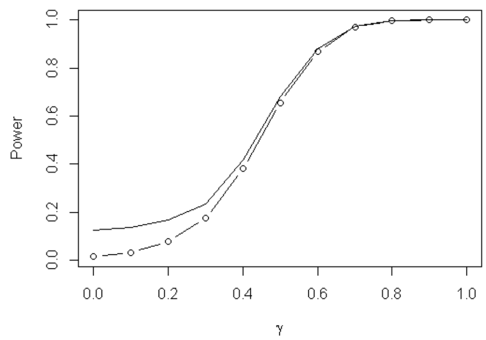


(e) $c = 10,\tau^* = 0.5$

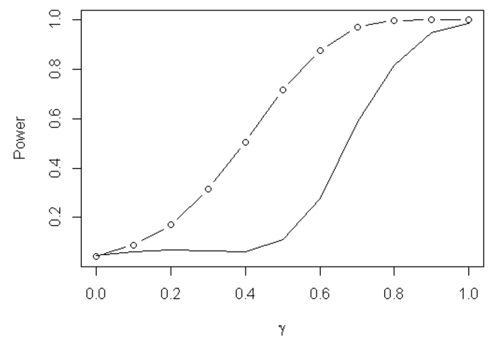


(f) $c = 10,\tau^* = 0.5$, size-adjusted

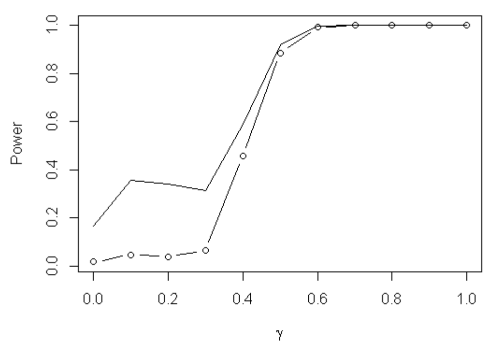
Figure 1.1: Power: Joint Broken Trend Model, $T = 150$, U:—o—, t_λ:—



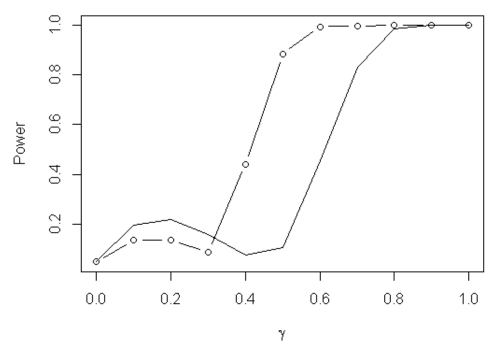
(a) $c = 10, \tau^* = 0.25$



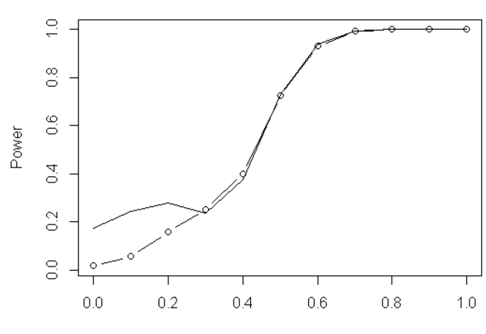
(b) $c = 10, \tau^* = 0.25$, size-adjusted



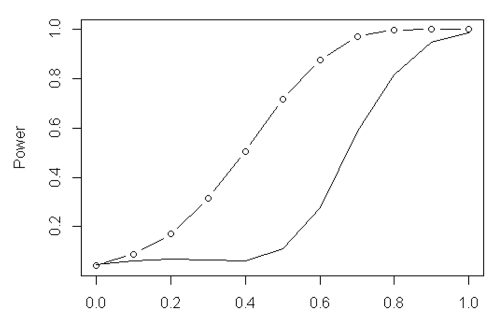
(c) $c = 20, \tau^* = 0.5$



(d) $c = 20, \tau^* = 0.5$, size adjusted

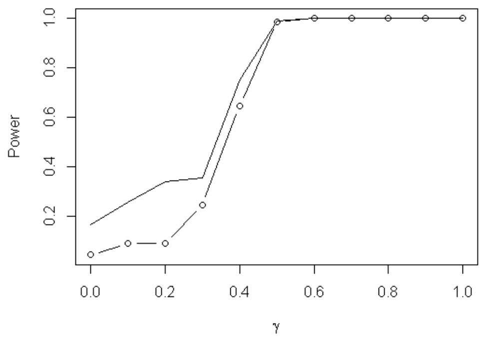


(e) $c = 20, \tau^* = 0.25$

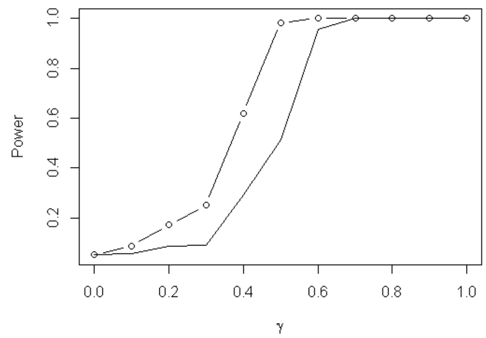


(f) $c = 20, \tau^* = 0.25$, size-adjusted

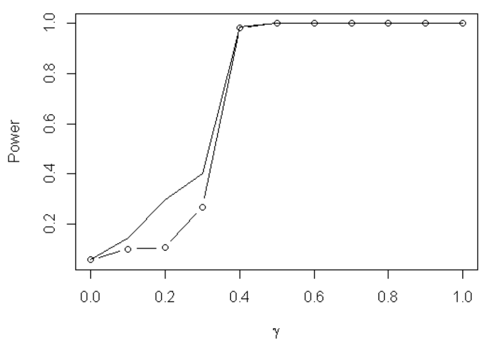
Figure 1.2: Power: Joint Broken Trend Model, $T = 150$, U:—o—, t_λ:—



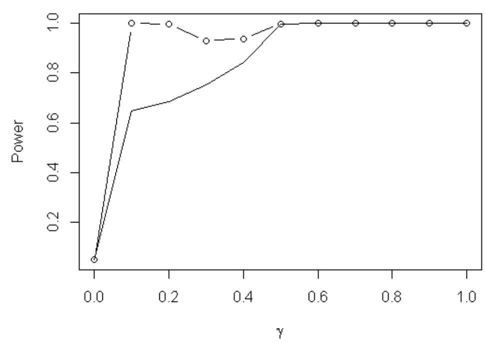
(a) $c = 50, \tau^* = 0.5$



(b) $c = 50, \tau^* = 0.5$, size-adjusted



(c) $c = T, \tau^* = 0.5$



(d) $c = T, \tau^* = 0.25$

Figure 1.3: Power: Joint Broken Trend Model, $T = 150$, U:—o—, t_λ:—

Consider first the left panels of Figures 1.1-1.2 and 1.3. The power of t_λ is slightly higher than U for small values of γ and the two tests are equivalent in terms of power for medium to large values of γ. However, these power advantages of t_λ over U for small γ can be attributed to the size distortions of t_λ. Because the t_λ tests are oversized (see Table 6), we also report size-adjusted powers. The size-adjusted power curves are given on the right panels of Figures 1.1-1.2 and 1.3. The main conclusion drawn from these results is that any power gain of t_λ over U is only due to the oversizing property of t_λ. This is not surprising; since both U and t_λ utilize the same statistics (S_0 and S_1) in their construction, we don't expect one test to be superior in extracting more information from the data regarding a potential break.

The power functions for U do not appear to depend to any noticeable degree on the location of the break. Comparing results between $\tau^* = 0.5$ and $\tau^* = 0.25$ we see that there is no power loss as the break is located away from the middle of the sample. This is also supported by an unreported simulation where we set $\tau^* = 0.75$. Consequently, our U test is as powerful for early and late breaks as for the mid-point breaks.

1.6 Empirical Application

This section considers an empirical application related to convergence of regional per-capita incomes in the U.S. We apply our trend break tests to relative per-capita income of the eight U.S. regions, to investigate the validity of the empirical debate that convergence in per- capita incomes among U.S. regions leveled off in the mid-1970s,which can be explored by modeling the trend function in each region as having a slope shift in the mid-1970s.

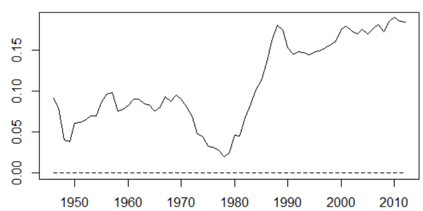
Solow's (1956) seminal neoclassical model of growth predicts that once the determinants of steady-state per-capita income have been controlled for, economies exhibit convergence. Empirical tests of convergence have used both cross-sectional and time series techniques. The cross sectional tests generally find evidence for convergence across countries (Baumol (1986) and Barro (1991)) and across regions within the U.S. and within Western Europe (Barro and Sala-i Martin (1992)).

In time series context, there are two conditions for convergence. Shocks to relative regional per-capita incomes should be temporary (stochastic convergence), and regions having per-capita incomes initially above the average should exhibit slower growth than those regions having per-capita incomes initially below the average (β-convergence). Loewy and Papell (1996) find evidence of stochastic convergence for U.S. regional per-capita incomes. Similarly, Carlino and Mills (1993), and Tomljanovich and Vogelsang (2002) show that U.S. regions have also achieved β-convergence. Moreover, all of these studies suggest there is a trend break in per-capita regional incomes with break date of 1946.

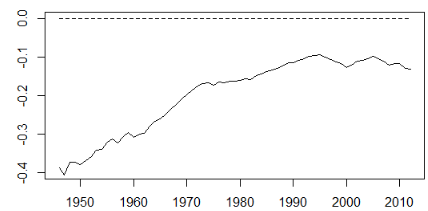
In this paper, we take a slightly different approach and rather focus on the rate of convergence of regional per-capita income in the U.S. We are interested in the empirical question that given U.S. regions converge, is there any change in the convergence pattern over time? This empirical question is motivated by two observations; first is the empirical debate that convergence in per capita incomes among U.S. regions has leveled off in the mid-1970s (Sayginsoy (2004)) and second is a simple visual inspection of the relative per-capita income data of eight U.S. regions. Figure 1.4 plots the log of the relative per-capita income for eight U.S. regions defined by the Bureau of Economic Analysis. From these figures, we can clearly see that the convergence has leveled off for certain regions (Southeast, Southwest, New England, Mideast). Hence, in this empirical exercise we are looking for statistical evidence to support these arguments.

The empirical question we pose above can be explored by modeling the trend function in each region as having a slope shift. Hence, we fit the following disjoint broken trend model to the data

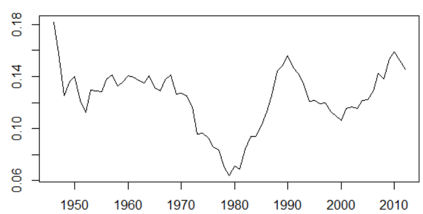
$$y_t = \alpha + \beta t + \delta DU_t(\tau^*) + \gamma DT_t(\tau^*) + u_t, \qquad t = 1, ..., T.$$



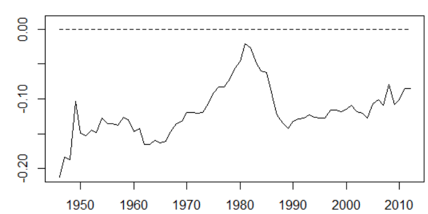
(a) New England



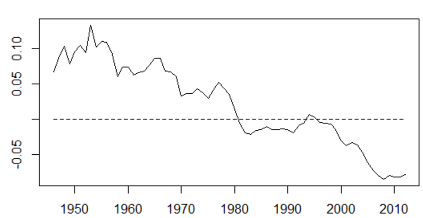
(b) Souteast



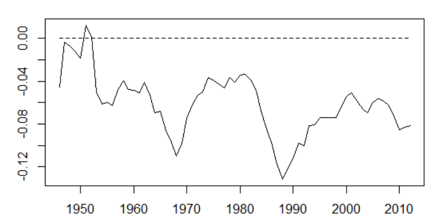
(c) Mideast



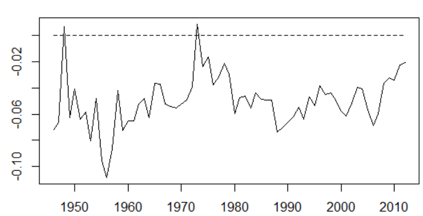
(d) Southwest



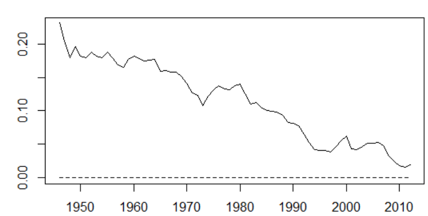
(e) Great Lakes



(f) Rocky Mountains



(g) Plains



(h) Far West

Figure 1.4: Natural logarithm of regional relative per-capita income

where y_t is the log of relative regional per-capita income. We prefer this model over the joint broken trend model to avoid any misspecification issue. We use our robust trend break test to check the significance of the parameter γ. If γ is statistically significant and the sign of γ is different from that of β, then we will have statistical evidence of a slowdown in convergence.

Annual BEA data on per-capita income for the U.S. regions during the period 1946-2012 are employed in this study. We consider the time series properties of per-capita income in the U.S. regions relative to per-capita income in the nation. Since previous studies found 1946 as a break date, we let the data start at 1946 to avoid multiple breaks in the data.

	γ	β	Estimated Break Date
Far West	0.0011	-0.0026	1992
Great Lakes	**-0.0028***	-0.0029	1994
Mideast	**0.0032***	-0.0018	**1980**
New England	**0.0043****	-0.0007	**1979**
Plains	0.0045	-0.0045	1958
Rocky Mountains	0.0026	-0.0008	1986
Southeast	**-0.0071*****	0.0073	**1986**
Southwest	**-0.0034***	0.0029	**1982**

Note: ***, **, and * denote statistical significance at the 1%, 5%, 10% levels, respectively.

Table 7 Empirical results using the disjoint broken trend model and U test

Table 7 presents the empirical results. In Table 7, we report the estimates of slope coefficients β and γ, and the break date estimates for each region. We determine statistical significance of γ using our robust break test. As explained above, for a given region if γ is statistically different from zero and at the same time γ and β have opposite signs, then we can conclude that there is a slowdown in per-capita income convergence for that region. Thus, the results in Table 7 provides statistical evidence of a slowdown in per-capita income convergence for Mideast, New England, Southeast, and Southwest regions at around late 1970s and early 1980s . These findings are also significant in the sense that Southeast and Southwest regions have the lowest per-capita income in U.S. historically.

Having found some evidence of a slowdown in regional per-capita income convergence, a relevant question is then "How can we explain these findings?". A potential answer is the following; the late 1970s and 1980s were characterized by shifts in relative income and productivity in many regions of the United States due to the oil price shocks, sharp recession, and subsequent economics recovery that was led by New England, the Mideast and California (Tomljanovich and Vogelsang (2002)).

1.7 Conclusion

In this paper we present tests for the presence of a structural break in the trend slope of a univariate time series which do not require knowledge of the form of serial correlation in the data and are valid regardless of the errors being $I(0)$ or $I(1)$. We consider two models: a joint and a disjoint broken trend model. We propose a union of rejections based procedure using two statistics; one appropriate for stationary errors and the other for integrated errors. We provide representations for and critical values from the asymptotic distributions of our proposed statistics (also for the union of rejections approach) under the null hypothesis of no trend break, together with representations for and numerical evaluation of their asymptotic local power functions under both $I(1)$ and $I(0)$ environments. We also provide finite sample results using Monte Carlo simulations and these simulations demonstrate that our proposed method performs well in small samples, regardless of the (unknown) order of integration of the data.

As an empirical exercise we consider the convergence issue in U.S. regional incomes and using our robust tests we find that there is a slowdown in per-capita income convergence among the U.S. regions after late 1970s or early 1980s, especially for the relatively low income regions that are trying to catch up. These findings are important in the sense that they are robust, hence reliable regardless of the properties of the data and they have important policy implications.

It is also straightforward to extend our method to test for multiple trend breaks or to sequentially test for l versus $l + 1$ breaks in the trend function. However, the main focus of this paper is on single trend break cases. Single trend break tests can be quite useful in policy analysis. For example, our robust tests can be employed to evaluate the impact of a one time policy change or a new regulation on a trending variable, without the need to know whether the data is generated by a stationary or integrated process. Overall, we believe that the robust tests we propose in this paper should prove useful in practical applications, particularly when a possible trend break in a macroeconomic or financial data is an important consideration itself, and where there is uncertainty regarding the order of integration of the data.

Appendix

In what follows, due to invariance of the statistics concerned, we can set $\alpha = \beta = 0$ without loss of generality

Proof of Theorem 1.

(i)(a) Using the Frisch-Waugh-Lovell Theorem (FWLT) we can write $t_0(\tau^*)$ in the form

$$t_0(\tau^*) = \left\{ \frac{T^{-3/2} \sum RT_t(\tau^*) u_t}{T^{-3} \sum RT_t(\tau^*)^2} \right\} \times \frac{1}{\sqrt{\omega_u^2 / T^{-3} \sum RT_t(\tau^*)^2}},$$

where $RT_t(\tau^*)$, $t = 1, ..., T$, are the OLS residuals from the regression of $DT_t(\tau^*)$ onto 1 and t. We establish the following weak convergence results using the standard time-series properties,

$$t_0(\tau^*) \xrightarrow{d} \left\{ \frac{\omega_u \int_0^1 RT(r, \tau^*) dW(r)}{\int_0^1 RT(r, \tau^*)^2 dr} \right\} \times \frac{1}{\sqrt{\omega_u^2 / \int_0^1 RT(r, \tau^*)^2 dr}},$$

where $W(r)$ is a standard Brownian motion, defined via, $\omega_u^{-1} T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t \xrightarrow{d} W(r)$.

(b) Again appealing to the FWLT, $t_1(\tau^*)$ can be expressed as

$$T^{1/2} t_1(\tau^*) = \left\{ \frac{\sum RU_t(\tau^*) \Delta u_t}{T^{-1} \sum RU_t(\tau^*)^2} \right\} \times \frac{1}{\sqrt{\omega_\varepsilon^2 / T^{-1} \sum RU_t(\tau^*)^2}}$$

where $RU_t(\tau^*)$,$t = 1, ..., T$, are the OLS residuals from the regression of $DU_t(\tau^*)$ onto 1. $RU_t(\tau^*)$ can also be written as

$$RU_t(\tau^*) = \begin{cases} \tau^* - 1, & t \leq T^* \\ \tau^* & t > T^* \end{cases}$$

which implies that $\sum RU_t(\tau^*) \Delta u_t = \tau^* u_T + u_{T^*} + (1 - \tau^*) u_1 = O_p(1)$, since u_t is $I(0)$. Entirely standard results, all other terms on the right hand side are also $O_p(1)$. As a result,$T^{1/2} t_1(\tau^*) = O_p(1)$, which establishes the result in (b).

(ii) (a) We have that

$$T^{-1} t_0(\tau^*) = \left\{ \frac{T^{-5/2} \sum RT_t(\tau^*) u_t}{T^{-3} \sum RT_t(\tau^*)^2} \right\} \times \frac{1}{\sqrt{\omega_u^2 / T^{-3} \sum RT_t(\tau^*)^2}},$$

Consequently, since all the stochastic terms appearing above are of $O_p(1)$ with non-degenerate limiting distributions, it follows that $T^{-1} t_0(\tau^*) = O_p(1)$.

Turning to the result in (b), observe that

$$\begin{aligned} t_1(\tau^*) &= \left\{ \frac{T^{-1/2} \sum RU_t(\tau^*)\Delta u_t}{T^{-1} \sum RU_t(\tau^*)^2} \right\} \times \frac{1}{\sqrt{\omega_\varepsilon^2 / T^{-1} \sum RU_t(\tau^*)^2}} \\ &\xrightarrow{d} \left\{ \frac{\omega_\varepsilon \int_0^1 RU(r,\tau^*)dW_c(r)}{\int_0^1 RU(r,\tau^*)^2 dr} \right\} \times \frac{1}{\sqrt{\omega_\varepsilon^2 / \int_0^1 RU(r,\tau^*)^2 dr}}, \end{aligned}$$

using standard results.Rearranging delivers the stated result in (b).

Proof of Theorem 2. The proof of (i)(a) and (ii)(b) follows by applying the Continuous Mapping Theorem as in Zivot and Andrews (1992) to show the convergence in distribution from the space $D[0,1]$ to the space $C(0,1)$. t_0^* and t_1^* statistics can be written as continuous functionals of the processes $\left(T^{-3} \sum RT_t(.)^2, T^{-3/2} \sum RT_t(.)u_t\right)'$ and $\left(T^{-1} \sum RU_t(.)^2, T^{-1/2} \sum RU_t(.)\Delta u_t\right)'$, respectively. Using similar arguments from Zivot and Andrews (1992) we can show the joint weak convergence of these processes:

$$\left(T^{-3} \sum RT_t(.)^2, T^{-3/2} \sum RT_t(.)u_t\right)' \xrightarrow{d} \left(\int_0^1 RT(r,.)^2 dr, \int_0^1 RT(r,.)dW(r)\right)'$$

$$\left(T^{-1} \sum RU_t(.)^2, T^{-1/2} \sum RU_t(.)\Delta u_t\right)' \xrightarrow{d} \left(\int_0^1 RU(r,.)^2 dr, \int_0^1 RU(r,.)dW_c(r)\right)'$$

The stated results in Theorem 2 (i)(a) and (ii)(b) then follow directly from Theorem 1 (i)(a) and (ii)(b),respectively, using applications of the CMT, noting the continuity of the sup function. The result in Theorem 2 (i)(b) follows from Theorem 1 (i)(b) and the result that $t_1(\tau) = O_p(T^{-1/2})$ uniformly in τ. Finally, for the result in Theorem 2 (ii)(a) we appeal to Theorem 1 (ii)(a) and the fact that $t_0(\tau) = O_p(T)$ uniformly in τ.

Proof of Theorem 3. For the proofs of part (i) and (ii) we use the fact that $\hat{\varepsilon}_t$ is asymptotically identical to Δu_t and that $\hat{\tau}$ is consistent for τ^*.

(i) $\hat{\omega}_\varepsilon^2(\hat{\tau}) \xrightarrow{p} lim_{T\to\infty} T^{-1} E\left(\sum_{t=2}^T \Delta u_t\right)^2 = \omega_\varepsilon^2$, by the standard results for the nonparametric kernel estimators.

(ii) $\hat{\omega}_\varepsilon^2(\hat{\tau}) \xrightarrow{p} lim_{T\to\infty}T^{-1}E\left(\sum_{t=2}^{T}\Delta u_t\right)^2 = 0$, since Δu_t is over-differenced when u_t is $I(0)$. However, it follows from Leybourne, Taylor, and Kim (2007) that $l\hat{\omega}_\varepsilon^2 \xrightarrow{p} -2\sum_{s=1}^{\infty} s\psi_s'$ where $\psi_s' = E(\Delta u_t \Delta u_{t-s})$. Consequently, $l\hat{\omega}_\varepsilon^2 = O_p(1)$, which establishes the result in (ii).

Proof of Theorem 4.

(i) For $r \in \Lambda$, straightforward extensions of results in Perron and Vogelsang (1992) yield the result that $T^{-1/2}\hat{u}_{\lfloor rT \rfloor} \xrightarrow{d} \omega_\varepsilon Q(r,c)$ where $Q(r,c)$ is a continuous time residual from the projection of $W_c(r)$ onto the space spanned by $\{1, r\}$. Then, from (14) we obtain, using the CMT, that

$$T\hat{\pi} \xrightarrow{d} \frac{\sigma_\eta}{\omega_\varepsilon}\frac{\int_0^1 Q(r,c)dW_c(r)}{\int_0^1 Q(r,c)^2 dr}$$

and, since $\hat{\sigma}^2 \xrightarrow{p} \sigma_\eta^2 = 1$, we therefore obtain that

$$T^{-2}\hat{\omega}_u^2 = \frac{\hat{\sigma}^2}{(T\hat{\pi})^2} \xrightarrow{d} \omega_\varepsilon^2 \frac{\left\{\int_0^1 Q(r,c)^2 dr\right\}^2}{\left\{\int_0^1 Q(r,c)^2 dW_c(r)\right\}^2} = \omega_\varepsilon^2 \Phi(c)$$

as required.

(ii)Under the null, $\hat{u}_t$ is asymptotically identical to u_t, hence $\hat{\omega}_u^2$ behaves asymptotically as if calculated directly from u_t, and therefore $\hat{\omega}_u^2 \xrightarrow{p} \omega_u^2$.

Proof of Lemma 1.

(i)(a) The proof of Lemma 1 (i)(a) follows from the results in Theorem 1 (i)(a), Theorem 2 (i)(a) and Theorem 4 (ii) .

(b) The proof of Lemma 1 (i)(b) is a trivial extension to the results in Theorem 1 (i)(b), Theorem 2 (i)(b) and Theorem 3 (ii) combined with the result that $l\hat{\omega}_\varepsilon^2(\tau) \xrightarrow{p} -2\sum_{s=1}^{\infty} s\psi_s'$ uniformly in τ.

(ii)In order to establish the result in (a), notice first that

$$\hat{t}_0(\tau) = \left\{\frac{T^{-5/2}\sum RT_t(\tau)u_t}{T^{-3}\sum RT_t(\tau)^2}\right\} \times \frac{1}{\sqrt{T^{-2}\hat{\omega}_u^2/T^{-3}\sum RT_t(\tau)^2}},$$

From standard weak convergence results, namely, the CMT and the Functional Limit

Theorem (FCLT), we can establish that:
$\hat{t}_0(\tau) \xrightarrow{d} \left\{ \frac{\omega_\varepsilon \int_0^1 RT(r,\tau)W_c(r)dr}{\int_0^1 RT(r,\tau)^2 dr} \right\} \times \frac{1}{\sqrt{\omega_\varepsilon^2 \Phi(c) / \int_0^1 RT(r,\tau)^2 dr}}$
Then the rest follows from the arguments given in Theorem 2.

(b)The proof of Lemma 1 (ii)(b) follows from the results in Theorem 1 (ii)(b), Theorem 2 (ii)(b) and Theorem 3 (i).

Proof of Theorem 5. In order to establish the results in Theorem 5, notice first that under $H_{1,0}$ and $H_{1,1}$ the level breaks are of order $o_p(1)$, and hence they have no asymptotic effect on $\hat{\omega}_u^2$. Then, the proof of Theorem 5 follows from trivial extensions to the results in Theorem 1, Theorem 2 and Lemma 1. The proof is therefore omitted in the interest of brevity.

Proof of Theorem 6. We omit the constant and trend regressors from (4) and the constant regressor from (7), for technical expediency, since our focus is only on establishing the orders in probability of the statistics under H_1. These particular regressors have no effect on any of the orders involved, but just introduce algebraic complexities.

(i) To establish the result in part (a), we follow the same steps outlined in Harvey et al.(2009b). We first derive the order of $\hat{t}_0(\tau^*)$ under the fixed alternative $H_1 : \gamma \neq 0$. Notice first that

$$\begin{aligned}(T/k)^{-1/2}\hat{t}_0(\tau^*) &= \left\{ \gamma + \frac{\sum DT_t(\tau^*)u_t}{\sum DT_t(\tau^*)^2} \right\} \times \frac{1}{\sqrt{T^{-2}k^{-1}\hat{\omega}_u^2 / T^{-3}\sum DT_t(\tau^*)^2}}, \\ &= \{\gamma + o_p(1)\}O_p(1) = O_p(1)\end{aligned}$$

which follows from Lee et.al (1997) that $\hat{\omega}_u^2 = O_p(T^2k)$ (since the long run variance is estimated ignoring the existing break) and that $T^{-3}\sum DT_t(\tau^*)^2 \to (1-\tau^*)^3/3$. Now, it is straightforward to establish that for any $\tau \in \Lambda$, we may write
$(T/k)^{-1/2}|\hat{t}_0(\tau)| = \sqrt{\left(\frac{T^{-3}\sum y_t^2}{\hat{\sigma}^2(\tau)} - T^{-2} \right) \times \frac{\hat{\sigma}^2(\tau)}{T^{-2}k^{-1}\hat{\omega}_u^2}},$

where $\hat{\sigma}^2(\tau) = T^{-1}\sum_{t=1}^{T} \hat{u}_t(\tau)^2$ and $\hat{u}_t(\tau) = y_t - \hat{\alpha} - \hat{\beta}t - \hat{\gamma}(\tau)DT_t(\tau)$. From above we see that the stated result will hold if $\hat{\sigma}^2(\hat{\tau}) - \hat{\sigma}^2(\tau^*)$ is asymptotically negligible. Following the results in Harvey et al. (2009b) we now demonstrate that $\hat{\sigma}^2(\hat{\tau}) - \hat{\sigma}^2(\tau^*) = O_p(T^{-1})$. It is straightforward to show that

$$\hat{\sigma}^2(\hat{\tau}) - \hat{\sigma}^2(\tau^*) = -T^{-1}\left[\frac{(\sum DT_t(\hat{\tau})y_t)^2}{\sum DT_t(\hat{\tau})^2} - \frac{(\sum DT_t(\tau^*)y_t)^2}{\sum DT_t(\tau^*)^2}\right],$$

from which it is easy to demonstrate that the dominant term of the right hand side of above equality is of the form

$$-\gamma^2 T^{-1}\left[\frac{(\sum DT_t(\hat{\tau})DT_t(\tau^*))^2}{\sum DT_t(\hat{\tau})^2} - \sum DT_t(\tau^*)^2\right].$$

After some lengthy manipulations, the dominant term of above expression can be shown to be given by

$$-\gamma^2\frac{(\hat{d}T)^2}{36}(\hat{\tau}-1)^3(4\tau^* - \hat{\tau} - 3),$$

where $\hat{d} = \tau^* - \hat{\tau}$. From Theorem 3 of Perron and Zhu (2005, p.75) we have that $\hat{d} = O_p(T^{-3/2})$ since our break fraction estimator $\hat{\tau}$ can be shown to have the same rate of consistency as the minimum sum of squares break fraction estimator of Perron and Zhu(2005). As a result, the dominant term above is $O_p(T^{-1})$, thus $\hat{\sigma}^2(\hat{\tau}) - \hat{\sigma}^2(\tau^*) = O_p(T^{-1})$. Consequently, $T^{-3/2}(\left|\hat{t}_0(\hat{\tau})\right| - \left|\hat{t}_0(\tau^*)\right|) \xrightarrow{p} 0$, which establishes the result in (a).

The proof of the result in (b) directly follows from the Proof of Theorem 3 (i)(b) in Harvey et al. (2009b).

(ii)(a) We again establish first the behavior of $\hat{t}_0(\tau^*)$ under H_1. Observe first that

$$\begin{aligned} T^{-1/2}\hat{t}_0(\tau^*) &= \left\{\gamma + \frac{\sum DT_t(\tau^*)u_t}{\sum DT_t(\tau^*)^2}\right\} \times \frac{1}{\sqrt{T^{-2}\hat{\omega}_u^2 / T^{-3}\sum DT_t(\tau^*)^2}}, \\ &= \{\gamma + o_p(1)\}O_p(1) = O_p(1) \end{aligned}$$

which again follows from Lee et.al (1997) that $T^{-2}\hat{\omega}_u^2 = O_p(1)$. Again, for any $\tau \in \Lambda$, we may write

$$T^{-1/2}|\hat{t}_0(\tau)| = \sqrt{\left(\frac{T^{-3}\sum y_t^2}{T^{-1}\hat{\sigma}^2(\tau)} - T^{-1}\right) \times \frac{T^{-1}\hat{\sigma}^2(\tau)}{T^{-2}\hat{\omega}_u^2}},$$

so again we need to establish the behavior of the difference between the OLS variance estimators evaluated at τ^* and $\hat{\tau}$. Using the results from part (i)(a), we obtain that the dominant term of $T^{-1}(\hat{\sigma}^2(\hat{\tau}) - \hat{\sigma}^2(\tau^*))$ is given by
$$-\gamma^2 \frac{\hat{d}^2 T}{36}(\hat{\tau} - 1)^3(4\tau^* - \hat{\tau} - 3),$$
Next, again utilizing the results from Theorem 3 of Perron and Zhu (2005), we may show that $\hat{d} = O_p(T^{-1/2})$ and, hence, we obtain that $T^{-1}(\hat{\sigma}^2(\hat{\tau}) - \hat{\sigma}^2(\tau^*)) = O_p(1)$. So it follows that $T^{-1/2}(\left|\hat{t}_0(\hat{\tau})\right| - \left|\hat{t}_0(\tau^*)\right|) = O_p(1)$, establishing (a).

The proof of the result in (b) again, directly follows from the Proof of Theorem 3 (ii)(b) in Harvey et al. (2009b).

Chapter 2

BOOTSTRAP STATIONARITY TEST WITH AN APPLICATION TO PURCHASING POWER PARITY

2.1 Introduction

Whether a time series is stationary or it has a unit root has been one of the most important issues in time series econometrics. Unit root tests have been the primary statistical method to differentiate these two alternatives. A testing procedure proposed by Kwiatkowski et al. (1992, KPSS hereafter) takes the stationarity as the null hypothesis instead and complements the unit root tests. Stationarity tests can be useful in cases where unit root tests have size distortion or low power. Moreover, the stationarity null may be more appropriate when testing for some economic theories where the hypothesis to be tested under the null is the one that economic theories suggest. For instance, purchasing power parity predicts that the relative values of currencies fluctuates temporarily around the relative price levels between countries and that in the long run nominal exchange rates converge to these relative prices. Hence, it is more natural to test stationarity of real exchange rates and to look for evidence against mean-reversion to test for purchasing power parity.

This useful stationarity test, however, is known to have significant size distortions in the presence of highly persistent yet stationary data. These size distortions were first documented by KPSS and later Caner and Killian (2001) reported that the size distortions are common for all conventional stationarity tests. The main source of the size distortion of the KPSS test is underestimation of long run variances. Muller (2005), for instance, provides a theoretical explanation of size distortions of the KPSS test in relation to long run variance estimation under local to unity framework.

In this paper, we propose a bootstrap stationarity test that has a good size control over the null parameter space. First we utilize a parametric re-sampling scheme that can generate independent bootstrap samples. This parametric re-sampling scheme, proposed by Kuo and Tsong (2005), allows us to replicate the dependence structure of the original data. In addition, the null constraint can easily be imposed on the generated data with this scheme. The bootstrap tests have correct size given that the dependence structure is replicated successfully and the null constraint is imposed. Second we select our bootstrap test statistic to increase the power of our bootstrap test. This choice of the test statistic is what separates our procedure from other bootstrap stationary tests in the literature. Kuo and Tsong (2005) and Lee and Lee (2012) choose the Lagrange multiplier test statistic (KPSS test statistic) as their bootstrap test statistic which is slightly different from our choice. Our test statistic has a well defined limit under the null and diverges under the alternative. However, our test statistic diverges at a faster rate than the LM test statistic hence we expect our bootstrap procedure to have better power properties than competing bootstrap tests. Our simulation studies also confirm these findings and the effectiveness of our bootstrap procedure.

The rest of the paper is organized as follows. Section 2 reviews conventional stationarity tests and their size distortions. We introduce our bootstrap procedure in Section 3. Section 4 provides some Monte Carlo studies where we compare the finite sample properties of our bootstrap test with existing bootstrap and conventional stationarity tests. We utilize our bootstrap procedure to test purchasing power parity as an empirical application, and the results are given in Section 5. Section 6 concludes and mentions future research topics.

2.2 Stationarity Tests

In this section we present the idea behind the stationarity tests and their size distortions through KPSS test. There is no loss of generality since other conventional stationarity tests are either parametric version of KPSS (Leybourne and McCabe (1994)) or employing an equivalent model which only transforms the parameter to be tested (Saikkonen and Luukkonen (1993)). Moreover, these tests also share the size problem of the KPSS test. KPSS considers the following unobserved component model as the data generating process;

$$y_t = \mu + \beta t + r_t + \epsilon_t$$

where $r_t = r_{t-1} + u_t$ is the random walk component with u_t are iid $(0, \sigma_u^2)$ and ϵ_t is a stationary error. Then the stationarity (null) hypothesis can be formulated as;

$$H_0 : \sigma_u^2 = 0 \; versus \; H_1 : \sigma_u^2 > 0$$

The KPSS test statistic is both the one-sided Lagrange multiplier statistic and locally best invariant test statistic under the normality assumption. However, they modify the test statistic in a nonparametric way to allow for general forms of temporal dependence in stationary error process ϵ_t. The KPSS test statistic is then; $KPSS_i = \hat{\sigma}^{-2} T^{-2} \sum_{t=1}^{T} S_t^2 \;\; i = \mu, \tau$ and the partial sum process is defined as $S_t = \sum_{i=1}^{t} e_i$ where e_i is the residual from the regression of y on an intercept for $KPSS_\mu$ and on an intercept and time trend for $KPSS_\tau$. The sum of the squared partial sums is normalized by an estimate of the long run variance $\sigma^2 = \lim_{T\to\infty} T^{-1} E(S_t^2)$.

KPSS show that under the null and some regularity conditions, the limiting distributions of the Lagrange multiplier test statistic can be represented as:

$$KPSS_i \xrightarrow{d} \int_0^1 V_i^2(r) dr \; for \; i = \mu, \tau$$

where $V_\mu(r)$ is a standard Brownian bridge: $V_\mu(r) = W(r) - rW(1)$, and $V_\tau(r) = W(r) + (2r - 3r^2)W(1) + (-6r + 6r^2) \int_0^1 W(s)ds$ is a second-level Brownian bridge and $W(r)$ is a Wiener process.

KPSS construct a consistent estimator of the long-run variance σ^2 , say $\hat{\sigma}^2$, from the residuals e_t ; specifically they use an estimator of the form;

$$\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} e_t^2 + 2T^{-1}\sum_{s=1}^{l} w(s,l) \sum_{t=s+1}^{T} e_t e_{t-s}$$

Here $w(s,l)$ is an optimal weighting function that corresponds to the choice of a spectral window. KPSS use the Bartlett window $w(s,l) = 1 - s/(l+1)$ as in Newey and West (1987), which guarantees the nonnegativity of $\hat{\sigma}^2$. For consistency of $\hat{\sigma}^2$, it is necessary that the lag truncation parameter $l \to \infty$ as $T \to \infty$ but at a slower rate; $l = o(T)$. In practice, the lag truncation parameter is chosen as $l = [k(T/100)^{1/4}]$, where k is a constant, and [.] is the largest integer function, following from Schwert (1989).

It is now widely known that the KPSS and other conventional stationarity tests have significant size distortions in the presence of highly persistent but stationary data. It is KPSS that first documented the size problems and later Caner and Killian (2001) provided an extensive simulation study which demonstrates that all conventional stationarity tests share this problem. In order to address this size problem, we provide a simple simulation study in Table 1. Table 1 provides empirical rejection frequencies of the KPSS test at a 5% nominal level with varying degrees of persistence of a simple AR(1) model. We present our results with three different bandwidth numbers to highlight the impact of the choice of lag truncation parameter on the size of the test. To save some space, we will not report the simulations for the tests with an intercept, as they share very similar qualitative outcomes as reported here.

As seen from Table 1, the KPSS test has significant size distortions, ranging from 15% to 99% as opposed to 5% nominal level, when the data is highly persistent yet stationary. Table 1 is also quite informative about the sources of the size distortions of the KPSS test. First, for a given sample size, size distortions get smaller as the bandwidth/lag truncation parameter increases. Thus, one source of size distortions is the underestimation of the long run variance. However, a more accurate estimate of the long run variance is not enough to circumvent the size problem. The asymptotic results of KPSS do not provide a good approximation to the finite sample distribution of the test statistic when the data is highly persistent. This is the second source of the size distortions and it can clearly be seen from Table 1: given a fixed , an increase in sample size does not help alleviate but exacerbate the degree of size distortions. Muller (2005) and Kuo and Tsong (2005) give a theoretical explanation about this surprising finding.

T	ρ	$KPSS_\tau(4)$	$KPSS_\tau(12)$	$KPSS_\tau(18)$
100	0.98	85.72%	41.16%	23.96%
	0.94	80.62%	29.5%	13.96%
	0.9	68.46%	20.6%	9.96%
	0.5	13.68%	5.14%	3.94%
300	0.98	97.76%	71.66%	50.26%
	0.94	89.52%	42.48%	24.82%
	0.9	73.78%	24.48%	14.42%
	0.5	11.64%	5.06%	4.56%
600	0.98	99.22%	79.16%	57.46%
	0.94	89.72%	44%	25.5%
	0.9	69.98%	21.62%	14.48%
	0.5	10.7%	6.02%	5.46%

Notes: 1. The DGP is $y_t = \rho y_{t-1} + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$

2. The rejection frequencies are calculated with 10000 replications

3. $KPSS_\tau(k)$ is calculated with a bandwidth number set to $l = [k(T/100)^{1/4}]$.

Table 1: Size performance of $KPSS_\tau$ at 5% nominal level

Kuo and Tsong (2005) demonstrate, under the local-to-unity framework, that the KPSS test statistic is $O_p(T/l)$. Local to unity framework provides an alternative asymptotics to the distribution of the test statistic for highly persistent data. This alternative explains clearly why the rejection frequencies in Table 1 increase with the sample size; as the sample size increase the KPSS statistic diverges at rate T/l.

2.3 Bootstrap Procedure

In time series literature, when conventional asymptotics fail to deliver good approximations to the finite sample distributions bootstrap turns out to be quite useful to improve small sample properties. Ferretti and Romo (1996) used bootstrap in unit root testing and provided an alternative to augmented Dickey-Fuller and Phillips-Perron tests in unit root tests with correlated errors. Nankervis and Savin (1996) introduced bootstrap t test in the AR(1) model to fix the size distortions of the conventional t test of the AR parameter. Hansen (1999) provided the grid bootstrap method to construct confidence interval for the largest autoregressive root and achieved uniform coverage probability where the conventional methods failed to do so.

In this paper, we propose a bootstrap stationarity test to deal with the size distortions of the conventional stationarity tests addressed in Section 2. There exits bootstrap stationarity tests by Kuo and Tsang (2005) and Lee and Lee (2012). Our bootstrap test utilizes Kuo and Tsangs resampling scheme, however differs from their test in the choice of the test statistic as will be explained later in this section.

There are two major issues one needs to deal with for bootstrap tests to be successful ; i) generating independent bootstrap resamples, ii) imposing the null constraint on the bootstrap resamples. We overcome these issues by utilizing an ARIMA model which is equivalent (up to second order moments) to the unobserved component model of section 2. Specifically, the unobserved component model $y_t = \beta t + r_t + \epsilon_t$ with $r_t = r_{t-1} + u_t$ is second order equivalent in moments to ARIMA(p,1,1) model;

$$\Phi(L)(1-L)y_t = \beta + (1-\theta L)\eta_t, \quad \eta_t \overset{iid}{\sim} (0, \sigma_\eta^2)$$

where $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$ with roots outside the unit circle. This equivalence is the basis of the stationarity test of Leybourne and McCabe (1994). They utilized above ARIMA(p,1,1) model to take care of the serial correlation and used the residuals to construct the Lagrange multiplier test statistic of KPSS. In our bootstrap resampling scheme, we make use of the parametric ARIMA(p,1,1) model to capture the autocorrelation in the series from which resamples can thus be independently drawn.

We also exploit one-to-one correspondence of model parameters to impose the null constraint on bootstrap resamples. To show this explicitly, let $\lambda = \sigma_u^2/\sigma_\epsilon^2$ then the moving average parameter can be written as $\theta = \frac{1}{2}(\lambda + 2 - (\lambda^2 + 4\lambda)^{1/2})$. Hence, the stationarity null $\sigma_u^2 = 0$ implies $\theta = 1$; a moving average unit root and this idea was exploited in constructing the stationarity test proposed by Saikkonen and Luukkonen (1993). We impose the null constraint by imposing a moving average unit root on generated bootstrap resamples.

Finally, it is the choice of the test statistic that separates our bootstrap test from other bootstrap stationarity tests. Kuo and Tsong (2005) and Lee and Lee (2012) pick the KPSS test statistic, $KPSS = (T^{-2}\sum_{t=1}^{T} S_t^2)/\hat{\sigma}^2$, to bootstrap. On the other hand, we focus only on the numerator; the normalized sum of squared partial sums to gain some power advantages and lets denote it as $NSSPS = T^{-2}\sum_{t=1}^{T} S_t^2$. Under the alternative hypothesis the numerator of the KPSS test statistic is $O_p(T^2)$ and the denominator is $O_p(lT)$. Hence, under the alternative KPSS test statistic diverges at rate T/l whereas the NSSPS test statistic diverges at a faster rate T^2. We then expect this faster divergence rate to translate into a higher power for our bootstrap test than the bootstrap tests using the KPSS test statistic. Moreover, with our choice of the test statistic we also avoid estimating the long run variance in each bootstrap resample. Our simulations indicate that estimation of long run variance in each bootstrap resample deteriorates the size of the bootstrap tests.

Our bootstrap resampling scheme is described in the following.

1. Given a sample $\{y_t\}_{t=1}^T$ generated from the unobserved component model, fit an ARIMA(p,1,1) to the series y_t using the maximum likelihood principle. Specifically, for intercept only, the model to be estimated is $\Delta y_t = \sum_{i=1}^p \rho_i \Delta y_{t-1} + \eta_t - \theta\eta_{t-1}$, while for intercept and trend $\Delta y_t = \beta + \sum_{i=1}^p \rho_i \Delta y_{t-1} + \eta_t - \theta\eta_{t-1}$. The resulting estimated parameters and residuals are denoted by $\hat{\rho}_i, \hat{\beta}, \hat{\theta}$ and $\hat{\eta}_t$.

2. Center the residuals $\hat{\eta}_t$ by $\bar{\eta}_t = \hat{\eta}_t - \frac{1}{T-1}\sum_{t=2}^T \hat{\eta}_t$.

3. Draw a bootstrap sample of size T with replacement from the empirical distribution function of the centered residuals $\{\bar{\eta}_t\}$, and denote it by η_t^*.

4. Set the initials that $y_1^8 = y_1, ..., y_p^* = y_p$, and generate the bootstrap samples $\{y_t^*\}$ based on the recursive relation using the estimated parameters from step 1 and setting $\theta = 1$ such that $\Delta y_t^* = \sum_{i=1}^p \hat{\rho}_i \Delta y_{t-1}^* + \eta_t^* - \eta_{t-1}^*$ (for intercept only) and $\Delta y_t^* = \hat{\beta} + \sum_{i=1}^p \hat{\rho}_i \Delta y_{t-1}^* + \eta_t^* - \eta_{t-1}^*$ (for intercept and trend).

5. Calculate $NSSPS_\mu$ and $NSSPS_\tau$ using $\{y_t^*\}_{t=1}^T$, denoted by $NSSPS_\mu^*$ and $NSSPS_\tau^*$, respectively.

6. Repeat step 3 to step 5 B times

7. Compute the empirical distribution function for $NSSPS_\mu^*$ or $NSSPS_\tau^*$, and use the empirical distribution function as an approximation to the cumulative distribution function of the bootstrap null distribution for the test statistics.

8. Compute the intended bootstrap critical values, based on the bootstrap null distribution in the preceding step.

Some remarks are in order here. In step 1, ARIMA(p,1,1) is fit to data to capture the dependence structure of the underlying data generating process. As a result, the residuals are independent and this enables us to generate independent bootstrap samples in steps 3 and 4. The lag order p in general needs to increase with sample size. In practice, the appropriate lag order is usually chosen by some information criteria. We follow this practice in the subsequent Monte Carlo study.

In step 4 we generate our bootstrap samples $\{y_t^*\}$ and two points are worth mentioning. First, in step 4 we replicate the dependence structure of the underlying data generating process on the bootstrap samples $\{y_t^*\}$. This is accomplished by using the parameter estimates from step 1. Second, in step 4 we impose the null (stationarity) constraint on the bootstrap samples by setting $\theta = 1$ regardless of the estimate $\hat{\theta}$ from step 1.

By the bootstrap invariance principle and continuous mapping theorem, the asymptotic limit of the bootstrap version of $NSSPS$ test statistic is equivalent to that of the original $NSPSS$ test statistic and it has the following form under the null

$$NSSPS_i^* \xrightarrow{d} \sigma^2 \int_0^1 V_i^2(r)dr \quad for\ i = \mu, \tau$$

where σ^2 and $V(r)$ are as defined in section 2. The null distribution of the original test statistic is not pivotal and depends on the nuisance parameter σ^2. However, since our bootstrap resampling scheme is able to replicate the autocorrelation structure, the bootstrap test statistic adapts automatically and directly approaches the same limiting distribution. Another example of bootstrap test statistic with non-pivotal limit distribution can be found in Ferretti and Romo (1996).

2.4 Simulation Studies

In this section we investigate the finite sample performance of our bootstrap NSSPS test and compare it to asymptotic and bootstrap KPSS test. First, we look the empirical size performance of competing tests. The simulation setup under the null is the same as that in Table 1 where the data generating process is an AR(1): $y_t = \rho y_{t-1} + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0, 1)$. We consider two different sample sizes: T=100 and T=300 are approximately the sample sizes encountered in practice. To observe the impact of persistence of the data on the size performance we vary parameter ρ; $\rho = \{0, 0.3, 0.5, 0.8, 0.9, 0.92, 0.94, 0.96, 0.98\}$. As it is obvious from this set of values, we are mainly interested in the size performances for highly persistent yet stationary data. In estimating the long run variance for the KPSS test, sample size dependent lag truncation parameters are considered such that $l = [k(T/100)^{1/4}]$.

In our bootstrap resampling scheme, we apply the AIC rule to choose the autoregressive lag order. Other lag order selections, as long as they satisfy the growth condition, may be considered in practice. The number of replications for the asymptotic tests are 10,000, while 1,000 replications are performed for the bootstrap tests with 100 bootstrap resamples in each replication. The rejection rates for all tests are computed and reported at nominal 5% level.

Table 2 presents the size performances of the tests for models with intercept. To save some space, we will not report the simulation results of the tests for models with intercept and time trend, as they share very similar qualitative outcomes as reported here. The main conclusion from Table 2 is that our bootstrap test, controls size better than both the asymptotic and bootstrap KPSS tests. The size distortions of the asymptotic KPSS test is similar to the size distortions in Table 1. The size distortions increase with sample size and persistence of the data, and decreases with the bandwidth. The bootstrap KPSS test reduces size distortions and has empirical size close to nominal level. Contrary to the asymptotic counterpart, size distortions of the bootstrap KPSS decrease with sample size. Increasing bandwidth improves the empirical size performance of the bootstrap KPSS test as well, however a larger bandwidth also leads to a reduction in power, as in Table 3, so there is a size-power tradeoff in the choice of bandwidth for the bootstrap KPSS test.

Our bootstrap NSSPS test displays size control close to that of the bootstrap KPSS with the larger bandwidth when the sample size is 100 but it clearly outperforms all other alternatives when the sample size increases to 300. The same bootstrap resampling scheme is used for both KPSS and our NSSPS test, then what accounts

T	ρ	$KPSS_\mu(4)$		$KPSS_\mu(12)$		$NSSPS_\mu$
		asym.	*boot.*	*asym.*	*boot.*	*boot.*
100	0.98	81.9%	19.6%	47.3%	15.3%	12.0%
	0.96	73.4%	15.1%	36.1%	10.2%	10.4%
	0.94	66.1%	11.8%	29.3%	9.3%	9.0%
	0.92	58.7%	11.3%	23.3%	8.1%	8.1%
	0.9	53.3%	11.1%	20.1%	7.4%	7.9%
	0.8	32.7%	8.9%	9.5%	5.1%	6.8%
	0.5	12.1%	5.5%	4.7%	5.0%	7.3%
	0.3	7.2%	6.0%	4.2%	5.8%	6.1%
	0	4.1%	4.6%	3.3%	5.7%	6.3%
300	0.98	92.1%	21.7%	57.6%	17.6%	10.9%
	0.96	81.8%	12.9%	40.7%	9.7%	5.3%
	0.94	70.9%	10.3%	29.9%	8.4%	4.3%
	0.92	59.7%	9.2%	23.9%	7.2%	4.7%
	0.9	52.1%	8.1%	18.7%	6.5%	5.0%
	0.8	28.5%	6.0%	10%	5.7%	5.7%
	0.5	9.3%	5.9%	5.7%	5.6%	6.0%
	0.3	7.4%	5.8%	4.8%	5.5%	5.7%
	0	4.7%	5.2%	3.8%	5.2%	5.5%

Notes: 1. The DGP is $y_t = \rho y_{t-1} + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$

2. $KPSS_\mu(k)$ is calculated with a bandwidth number set to $l = [k(T/100)^{1/4}]$

3. *boot.* and *asym.* denote the bootstrap test and asymptotic test respectively.

Table 2: Size performance of stationarity tests at 5% nominal level

for the better performance of NSSPS over KPSS? We believe that NSSPS has better size control than KPSS because estimation of long run variance in each bootstrap resample deteriorates the size of the KPSS tests. Especially for highly persistent data, the estimate of the long run variance is quite noisy and this noisy estimate is the source of the different size performances of bootstrap NSPSS and KPSS. Finally, we explain the size distortions of our bootstrap test when $\rho \geq 0.95$ for $T = 100$ and $\rho \geq 0.975$ for $T = 300$ in Table 2. These are due to imprecise estimates of the ARIMA model in step 1 of our resampling scheme. The estimates of the autoregressive and moving average parameters are not accurate due to near cancellation effect when ρ is very close to unity and this is the source of aforementioned size distortions of NSSPS in Table 2.

T	$\lambda = \sigma_u^2/\sigma_\epsilon^2$	$KPSS_\mu(4)$		$KPSS_\mu(12)$		$NSSPS_\mu$
		asym.	*boot.*	*asym.*	*boot.*	*boot.*
100	1	82.1%	82.0%	61.2%	60.7%	84.1%
	0.1	75.9%	73.2%	58.6%	59.3%	76.3%
	0.01	50.6%	56.3%	42.6%	46.9%	63.1%
	0.001	14.6%	17.6%	13.4%	13.2%	20.5%
	0.0001	5.7%	8.5%	5.8%	6.8%	8.0%
300	1	96.2%	97.2%	81.6%	83.3%	99.8%
	0.1	95.7%	94.7%	80.9%	81.2%	94.9%
	0.01	88.0%	88.4%	75.3%	77.1%	94.5%
	0.001	55.4%	57.2%	50.8%	52.0%	61.3%
	0.0001	15.2%	17.2%	14.8%	15.3%	18.1%

Notes: 1. The DGP is $y_t = r_t + \epsilon_t$, and $r_t = r_{t-1} + u_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$ and $u_t \overset{iid}{\sim} N(0, \sigma_u^2)$

2. $KPSS_\mu(k)$ is calculated with a bandwidth number set to $l = [k(T/100)^{1/4}]$

3. *boot.* and *asym.* denote the bootstrap test and asymptotic test respectively

4. The figures reported for asymptotic tests are the size-adjusted empirical power.

Table 3: Power performance of stationarity tests at 5% nominal level

Table 3 presents the empirical power performances of the tests for models with intercept only. The data generating process used in this simulation is the unobserved component model of section 2 with $\lambda = \sigma_u^2/\sigma_\epsilon^2$ is defined as the signal-to-noise ratio. The autoregressive lag order in our resampling scheme is again selected by AIC. The number of replications for the asymptotic tests are 10,000, while 1,000 replications are performed for the bootstrap tests with 100 bootstrap resamples in each replication. The rejection rates for all tests are computed and reported at nominal 5% level.

As we argued in Section 3, our choice of the test statistic results in superior power of our NSSPS test to the bootstrap KPSS test. When we compare our bootstrap NSSPS test to the bootstrap KPSS, we see significantly higher power for the NSSPS relative to the KPSS with a large bandwidth number. The difference is less pronounced for the KPSS with a smaller bandwidth. However, it is important to keep in mind that the bootstrap NSSPS controls size much better than the bootstrap KPSS with small bandwidth. Thus, simulation results from Table 2 and Table 3 combined clearly favors our bootstrap NSSPS test over the bootstrap KPSS test. We also report the size-adjusted powers of the asymptotic KPSS test and see that the size improvement of our bootstrap test does not come at the expense of power loss.

2.5 Empirical Results

We apply our proposed bootstrap test to a set of real exchange rate data to test purchasing power parity. We seek stationarity, or mean-reverting property in the real exchange rates that corresponds to the notion of long-run purchasing power parity. Even though stationarity in the real exchange rates is not enough to prove purchasing power parity (PPP), rejection of stationarity is strong evidence against it. The stationarity test maintains what economic theory predicts under the null and rejects the theory only when there is strong evidence against it. However, asymptotic stationarity tests may result in spurious rejections of PPP due to size distortions underlined in section 2. As Table 1 suggests, the asymptotic KPSS test rejects the stationarity null too often when the data is indeed stationary. Moreover, real exchange rate data is characterized by a highly persistent autoregressive process and this is exactly where the size distortions are most severe. In order to have a more reliable result we test PPP with the bootstrap NSSPS test developed in section 3.

The real exchange rates are constructed from the consumer price index series and the exchange rate series for the price of U.S. dollars in respective currency. Data is obtained from the IMF publication, International Financial Statistics. Quarterly data is available for the following 9 developed countries over the period 1973.I-2012.IV (post-Bretton Woods period): Australia, Canada, Denmark, Japan, Norway, New Zealand, Sweden, Switzerland, and United Kingdom.

Country	$KPSS_{\mu}(12)$	p	$NSSPS_{\mu}$
	asym.		*boot.*
Australia	0.3008	1	0.075
Canada	**0.3625***	2	**0.047**††
Denmark	0.2353	4	0.048
Japan	**0.609****	5	**0.194**†
Norway	0.0968	1	0.011
New Zealand	0.2907	1	0.065
Sweden	**0.6196****	4	0.179
Switzerland	**0.4641****	1	0.091
United Kingdom	0.2218	1	0.007

Notes: 1. Quarterly data is avilable over 1973.I-2012.IV

2. p is the lag order of the ARIMA(p,1,1) fitted to the data and is chosen by AIC

3. $KPSS_{\mu}(12)$ is calculated with a bandwidth number set to $l = [12(T/100)^{1/4}]$

4. **(*) represents a rejection at 5%(10%) level using the asymptotic test, and $^{\dagger\dagger}(^{\dagger})$ using the bootstrap test

5. The figures below $NSSPS_{\mu}$ is calculated with the data, and critical values (not reported here) are computed using 5000 replications with 1000 resamples

Table 4: The results for purchasing power parity

Table 4 reports the stationarity test results for the real exchange rates of 9 countries. Since the real exchange rates do not exhibit trending behavior, models with an intercept only are considered in this study. First, lets look at the results of the asymptotic KPSS test. The critical values for the asymptotic KPSS test can be found in KPSS (1992) and these are 0.463 and 0.347 at the 5% and 10% significance levels respectively. Based on these critical values, the asymptotic KPSS test rejects purchasing power parity for Japan, Sweden and Switzerland at 5% significance level and for Canada at 10%. Thus, the asymptotic KPSS test finds evidence against purchasing power parity for almost half of the countries studied.

We finally report the results of our bootstrap test. Due to the size distortions of the asymptotic KPSS test, some of the rejections mentioned in previous paragraph may be spurious. This is also suggested by the results of our bootstrap stationarity test. The rejections for Sweden and Switzerland are overturned by the bootstrap test, whereas for Canada and Japan purchasing power parity is still rejected by the bootstrap method. Hence, our bootstrap test provides more evidence for the existence of PPP than the asymptotic test. Since the size and power performance of the bootstrap test surpasses that of the asymptotic test, the bootstrap test offers a more reliable conclusion.

2.6 Conclusion

We propose a bootstrap stationarity test to reduce possible size distortions of the asymptotic stationarity test. Our bootstrap test utilizes the resampling scheme propped by Kuo and Tsong (2005). We show how the resampling scheme generates independent bootstrap resamples and imposes the stationarity constraint on the generated resamples. Next, we select our bootstrap test statistic to improve upon the existing bootstrap methods and we demonstrate how the faster divergence rate of our test statistic under the alternative translates into higher power.

Monte Carlo studies show that our bootstrap test not only offers higher power but also controls the size better than other bootstrap tests. Finally, we apply our test to real exchange rates to test purchasing power parity. We find less evidence against PPP than what the asymptotic tests provide.

Chapter 3

A POWERFUL ROBUST UNIT ROOT TEST

3.1 Introduction

This paper considers the initial condition problem in unit root testing. There are three major problems in unit root tests: i) serial correlation in residuals, ii) uncertainty over trend specification, iii) uncertainty over initial condition. A significant amount of effort has been put in to deal with i and ii. As a result, optimal and simple to use tests have been proposed that can handle these two issues. However, recent studies have shown that these class of optimal tests perform very poorly if the initial displacement is large. In this paper, we robustify these tests to initial condition by utilizing indirect inference method.

The power of any unit root test depends on the deviation of the initial observation y_0 from its deterministic part, the so-called initial condition. The class of optimal tests proposed by Elliott et al. (1996) (ERS hereafter) have been shown to be even more sensitive to the initial condition. Specifically, when the initial deviation is large, these optimal tests suffer very significant power losses, certainly enough that these tests have trivial power. As discussed in Elliott and Muller (2006), although there are many situations where one would not expect the initial deviation to be large relative to other data points, equally the initial condition may be relatively large in other situations. The former case happens, for example when the first observation of the sample is dated quite some time after the initiation of a mean-reverting process, whereas the latter can occur when the sample data happen to be chosen to begin after a break in the series or where the beginning of the sample happen to coincide with the start of the process. Consequently, in practice it is difficult to rule out small or large initial conditions a priori. This is problematic because, as discussed in Elliott and Muller (2006), we observe only the initial observation rather than the initial condition.

The optimal unit root tests of ERS attain their power curves by detrending (demeaning) the data efficiently before applying the unit root tests. However, detrending technique used in optimal tests leads to power distortions due to initial condition. This is due to the inconsistent estimation of the trend parameters due to initial displacement. In this paper, I first show that indirect inference method can estimate trend parameters consistently regardless of the initial displacement. I use the one-to-one relationship between the structural and reduced form parameters of the model to robustify the estimation of the trend parameters. Moreover, indirect inference also provides unbiased estimates of the persistence parameters which further improves the power of the unit root tests. Once the data is detrended consistently independent of the initial displacement, existing methods such as augmented Dickey-Fuller procedure can be employed to test for a unit root. This two step procedure results in a powerful unit root test which is robust to initial condition.

My simulation results indicate that the power of my robust unit root test is not sensitive to initial condition. Also, over the reasonable range of initial displacement levels ($\pm$ 4 standard errors) the power of my test improves upon Dickey-Fuller t-test in which no detrending is used. Compared to the optimal tests the power gain of my test is relatively low. However, for large initial displacement levels optimal tests have almost no power while my test still retains its high power. This result underlines the robustness-power tradeoff of the unit root tests.

The rest of the paper is organized as follows. Section 2 reviews the class of optimal unit root tests proposed by ERS. In section 3, I introduce the initial condition problem in unit root testing and demonstrate the poor power properties of the optimal tests in the presence of large initial displacements. Section 4 details the construction of my robust unit root tests, and section 5 provides preliminary simulation results . Finally section 5 concludes.

3.2 Optimal Unit Root Tests

Let the time series y_t be the stochastic process generated by the linear model

$$y_t = \beta' z_t + x_t \tag{3.1}$$

and the first-order autoregressive (AR) process

$$x_t = \rho x_{t-1} + \epsilon_t \tag{3.2}$$

where z_t represents a vector of deterministic terms, e.g. $z_t = 1$ or $z_t = (1, t)'$, and ϵ_t is a mean-zero stationary process with long-run variance ω^2. We consider testing the null hypothesis $H_0 : \rho = 1$ versus $H_1 : |\rho| < 1$.

If ϵ_t were known to be Gaussian then the optimal test (test with highest power) can be constructed against any point alternative $\bar{\rho}$, using the Neyman-Pearson Lemma. The power of this optimal test plotted as a function of each alternative results in an upper bound for the power of unit root tests. This upper bound constitutes the power envelope of the Gaussian model. However, there is no uniformly most powerful test of the unit root hypothesis against all stationary alternatives. Nevertheless, ERS derive a class of test statistics, using asymptotic approximations based on the local-to-unity alternatives $\rho = 1 + c/T$ with $c < 0$, that come very close to the power envelope for a wide range of alternatives.

3.2.1 Point Optimal Invariant Tests

In this subsection, I first demonstrate the construction of the feasible test statistic that is optimal against the point alternative $\bar{\rho} = 1 + c/T, c < 0$. Define the T-dimensional column vector y_ρ and the matrix Z_ρ ($T \times 1$ for $z_t = 1$ and $T \times 2$ for $z_t = (1, t)'$) by

$$y_\rho = (y_1, y_2 - \rho y_1, \ldots.., y_T - \rho y_{T-1})'$$

$$Z_\rho = (z_1, z_2 - \rho z_1, ..., z_T - \rho z_{T-2})'$$

Notice that all of the elements in y_ρ and Z_ρ, except the first, are quasi-differenced using the operator $1 - \rho L$. Next, define $S(\rho)$ as the sum of squared residuals from a least square regression of y_ρ on Z_ρ. That is,

$$S(\rho) = (y_\rho - Z_\rho \hat{\beta}_\rho)'(y_\rho - Z_\rho \hat{\beta}_\rho)$$

where $\hat{\beta}_\rho = (Z'_\rho Z_\rho)^{-1} Z'_\rho y_\rho$. ERS show that the feasible point optimal unit root test against the alternative $\bar{\rho} = 1 + \bar{c}/T$ has the form

$$P_T = [S(\bar{\rho}) - \bar{\rho} S(1)]/\hat{\omega}^2 \tag{3.3}$$

where $\hat{\omega}^2$ is a consistent estimate of ω^2.

A disadvantage of point optimal tests is that it is only efficient for the given alternative, and hence there are practical limitations. However, through some simulation studies, ERS discover that if $\bar{\rho} = 1 + \bar{c}/T$ is chosen such that the power of P_T is tangent to the power envelope at 50% power, then the overall power of P_T is quite close to the power envelope for a wide range of $\bar{\rho}$ values less than unity. For a given sample size T, the value of $\bar{\rho}$ that results in P_T having 50% power depends on $\bar{c}$ and the type of the deterministic terms in z_t. ERS show that if $z_t = 1$ then $\bar{c} = -7$ and if $z_t = (1, t)'$ then $\bar{c} = -13.5$.

3.2.2 DF-GLS Tests

The point optimal unit root tests explained above (3) attains their optimal power properties due to the fact that the unknown parameters β of the deterministic terms(trend function) are efficiently estimated under the alternative with $\bar{\rho} = 1 + \bar{c}/T$. That is, $\hat{\beta}_{\bar{\rho}} = (Z'_{\bar{\rho}} Z_{\bar{\rho}})^{-1} Z'_{\bar{\rho}} y_{\bar{\rho}}$. ERS use this insight to derive an efficient version of the augmented Dickey-Fuller(ADF) t-statistic,which they call the DF-GLS test. They construct this t-statistic as follows. First, using the trend parameters $\hat{\beta}_{\bar{\rho}}$ estimated under the alternative, define the detrended data

$$y_t^d = y_t - \hat{\beta}_{\bar{\rho}}' z_t$$

ERS call this detrending procedure GLS detrending. Next, using the GLS detrended data, estimate by least squares the ADF test regression which omits the deterministic terms

$$\Delta y_t^d = \pi y_{t-1}^d + \sum_{j=1}^{p} \psi_j \Delta y_{t-j}^d + \varepsilon_t \tag{3.4}$$

and compute the t-statistic for testing $\pi = 0$. When $z_t = 1$, ERS show that the asymptotic distribution of the DF-GLS test is the same as the ADF t-test, but has higher asymptotic power (against local alternatives) than the DF t-test. Furthermore, ERS show that the DF-GLS test has essentially the same asymptotic power as the ERS point optimal test when $\bar{c} = -7$. When $z_t = (1, t)'$ the asymptotic distribution of the DF-GLS test, however, is different from the ADF t-test. ERS and Ng and Perron (2001) provide critical values for the DF-GLS test in this case. ERS show that the DF-GLS test has the same asymptotic power as the ERS point optimal test with $\bar{c} = -13.5$, and has higher power than the DF t-test against local alternatives.

3.3 Initial Condition Problem

Consider,again, the following model:

$$y_t = \beta' z_t + x_t$$

$$x_t = \rho x_{t-1} + \epsilon_t$$

$$x_0 = \xi$$

The initial value ξ can be treated as a fixed but unknown nuisance parameter. Let $y = (y_0, y_1, ..., y_T)'$, $x = (x_0, x_1, ..., x_T)'$, and $Z = (z_0, z_1, ..., z_T)'$ then we can represent the model above compactly as

$$y = Z\beta + x$$

Note that $E[x] = \xi R(\rho)$, where $R(r)$ is the $(T+1) \times 1$ vector $R(r) = (1, r, r^2, ..., r^T)'$. Let $A(r)$ be the $(T+1) \times (T+1)$ matrix with ones on its diagonal, $-r$ on the lower diagonal, and zeros elsewhere, and let $\epsilon = (0, \epsilon_1, \epsilon_2, ..., \epsilon_T)'$. Then $A(\rho)(x - \xi R(\rho)) = \epsilon$ and the model can be written as

$$y = Z\beta + \xi R(\rho) + A(\rho)^{-1}\epsilon$$

The initial condition ξ is also a nuisance parameter like β and the covariance matrix of ϵ. We are not interested in its value, however its impact on the data generating process must be considered to build useful tests and assess their performance.

Under the null hypothesis ($\rho = 1$) different values of ξ induce mean shifts in the data. Hence when $\rho = 1$, ξ can not be individually identified independent of the mean of the data. This implies that tests that are invariant to the mean will also be automatically invariant to initial condition, so that ξ does not affect their size properties. Under the alternative hypothesis ($\rho < 1$), however, different ξ implies adding a different geometrically decaying series $\xi\rho^t$ to the data. This leads to an extra difference between the null and alternative models, so in turn will affect the power properties of the tests.

It is a well documented fact that the power of unit root tests depend on the initial condition in finite samples; see Evans and Savin (1981,1984) and Stock (1994) for instance. Conventional asymptotic analysis assume ξ to be either fixed or random but stochastically bounded. An application of a Functional Central Limit Theorem (FCLT) to

$$T^{-1/2} x_{Ts} = T^{-1/2} \rho^{[Ts]} \xi + T^{-1/2} \sum_{l=1}^{[Ts]} \rho^{[Ts]-l} \epsilon_l \tag{3.5}$$

with [.] denoting the largest smaller integer function then suggests that ξ has no impact on the asymptotic distributions either under the null or alternative hypotheses as the first term is $o_p(1)$.

More appropriate asymptotic approximations for small sample inference, when ξ is of similar magnitude to variation in the data after deterministic terms are removed, appears when the first term is $O(1)$. Relevant asymptotics for the unit root testing problem require ρ to become ever closer to unity as the sample size T increases. The appropriate rate of convergence of ρ to one is achieved by setting $\rho = 1 - \gamma T^{-1}$ for a fixed γ. A stationary series in this setting will have an unconditional variance that is proportional to $(1-\rho^2)^{-1} = T(2\gamma)^{-1} + o(T)$. Taking the root of the unconditional variance as the natural scale for the initial condition, which implies that ξ is an $O(T^{1/2})$ variable. But with $\xi = O(T^{1/2})$, the first term in (5) has the same order of magnitude as the second, so that the initial condition does not vanish asymptotically.

With ξ being a relevant nuisance parameter, some method is required to account for it in the testing procedure. A 'plug-in' approach, substituting ξ with an estimator $\hat{\xi}$, fails because ξ can not be estimated accurately to leave power unaffected by the substitution. Alternatively one might consider constructing a test invariant to ξ. But with ρ unknown, the relevant group of transformations

$$y \to y + zR(r) \qquad \forall z \; \forall r < 1$$

is so large that a requirement of invariance with respect to this group yields tests with trivial power. Moreover, the application of invariance is inappropriate for ξ because the form of the induced deterministic $R(\rho)$ depends on ρ and hence on the supported hypothesis.

3.3.1 Optimal Unit Root Tests in the Presence of Large Initial Deviations

In this section, the power properties of the optimal unit root tests is investigated with varying degrees of initial deviations. More specifically, Monte-Carlo simulations are employed to demonstrate the deteriorating power performance of the optimal unit root tests as the initial deviation increases. The following model with a deterministic linear trend function is used as the data generating process:

$$y_t = \mu + \beta t + x_t \tag{3.6}$$

where

$$x_t = \rho x_{t-1} + \epsilon_t, \quad with\, \epsilon_t \sim NIID(0, \sigma^2) \tag{3.7}$$

We can also rewrite equation (6) after substituting the recursive relationship in (5) as:

$$y_t = \mu + \beta t + \rho[\rho^{t-1}x_0 + \sum_{j=0}^{t-1} \rho^{t-j+1}\epsilon_j] + \epsilon_t \tag{3.8}$$

Define the standardized initial deviation as $x_0^* = \dfrac{x_0}{\sigma} = \dfrac{(y_0 - \mu)}{\sigma}$. As explained in the previous section, the power of unit root tests depend on the initial deviation in small samples. The following Monte-Carlo simulation is set up to investigate the power properties of the DF-GLS tests under different x_0^*. The parameters are set as, $\mu = 1$ and $\beta = 0.1$, however since DF-GLS tests are invariant to both μ and β, these choices are not critical for power performance and this finding is also justified by unreported simulations where μ and β are assigned different values. In order to draw the finite sample power curves, ρ parameter is varied according to $\rho = 1 - c/T$ with $T \in \{100, 300\}$

Figure-3.1 plots the finite sample power curves for $T = 100$ (the results for $T = 300$ is qualitatively very similar, thus not reported here). From Figure-3.1, it is seen that the power of DF-GLS test decrease as x_0^* increases, and DF-GLS has very low power for large x_0^*. The power simulations are also carried out for the point optimal invariant tests and very similar power properties is observed with respect to initial condition.

The optimal unit root tests of Section 2 improves the power by detrending the data using the GLS technique. The poor power performance of these tests for large x_0^* also stems from the very same reason. Unreported simulations indicate that these optimal tests fail to estimate the trend parameters, μ and β, consistently when x_0* is large. Thus, deterministic terms are not properly removed with GLS detrending. As a result, $\hat{\rho}$ is estimated spuriously high for all ρ, and this explains the poor power performance.

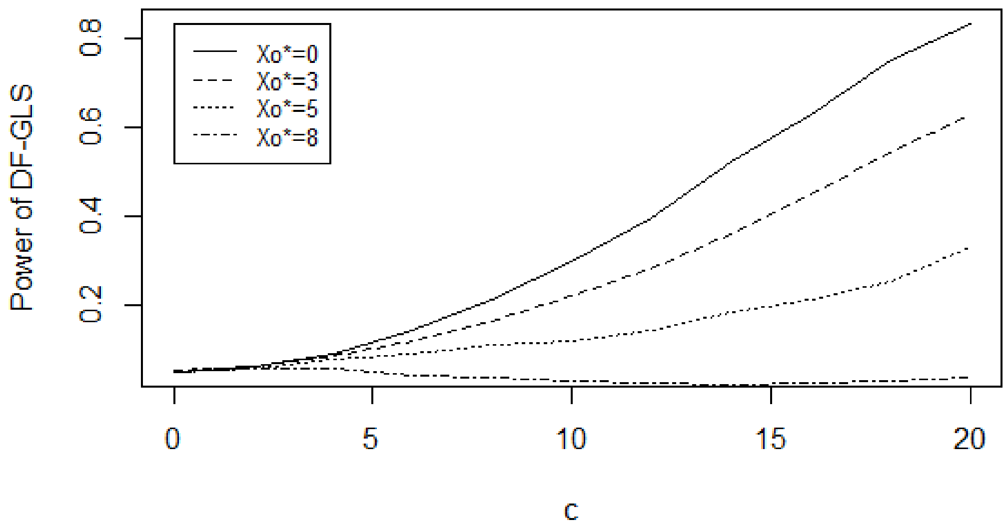


Figure 3.1: Finite Sample Powercurves of DF-GLS

3.4 Robust Unit Root Test

In this section, I construct a powerful and robust (to initial condition) unit root test using the insights from sections 2 and 3. The power of the unit root tests is improved by detrending the data , as in the optimal unit root tests, before testing for a unit root. Moreover, the tests constructed are robustified to initial condition by utilizing a detrending technique that is not sensitive to the initial deviation. I show that indirect inference method can be used to estimate the trend parameters accurately regardless of the initial deviation. In order to understand the robust trend estimation more clearly, next subsection briefly summarizes the indirect inference method. Next, robust trend estimation using indirect inference is demonstrated, and finally the construction of the robust unit root test is explained in detail.

3.4.1 Indirect Inference

Let the true data generating process $y_t = f(x_t, \epsilon_t, \theta_0)$ be easy to simulate but difficult for the estimation of θ_0. The true data generating process is called the structural model, and θ_0 the true structural parameter. The objective is to estimate the true structural parameter θ_0, however in some cases direct estimation, i.e. standard methods, is either infeasible or very difficult. Thus, we take an indirect approach and use a simulation-based method to estimate θ_0.

We replace the true model, $y_t = f(x_t, \epsilon_t, \theta_0)$, with an auxiliary model and use this auxiliary model to estimate auxiliary parameters λ using the original data. Define the data vectors $\vec{y}$ and $\vec{x}$, and let $\psi(\vec{y}, \vec{x}, \lambda)$ be an appropriate objective function for the auxiliary model;

$$\hat{\lambda} = argmax_\lambda \, \psi_T(\vec{y}, \vec{x}, \lambda)$$

where $\hat{\lambda}$ is an estimate of the auxiliary parameters λ using the original data.

Next, simulate data using the true model defined above for some values of θ,

$$y_t^s(\theta) = f(x_t, \epsilon_t^s, \theta)$$

and estimate the auxiliary parameters, λ, with the simulated data using the same objective function.

$$\hat{\lambda}^s(\theta) = argmax_\lambda \, \psi_T^s(\vec{y^s}, \vec{x}, \lambda)$$

In order to eliminate the simulation bias, we repeat the same simulation process, with the same set of values for θ, S times, and define

$$\hat{\lambda}_S(\theta) = \frac{1}{S} \sum_{s=1}^{S} \hat{\lambda}^s(\theta)$$

Intuitively, the ideal set of values for θ lead $\hat{\lambda}_S(\theta)$ to be as close as possible to $\hat{\lambda}$, and using this insight define the indirect inference estimator as:

$$\hat{\theta}(\Omega) = argmin_\theta \, (\hat{\lambda} - \hat{\lambda}_S(\theta))'\Omega(\hat{\lambda} - \hat{\lambda}_S(\theta))$$

where Ω is a positive definite weight matrix. For the identification of the structural parameters we require the dimension of the auxiliary parameters to be as large as the dimension of the structural parameters; $dim(\theta) \leq dim(\lambda)$.

The following simple example illustrates the use of indirect inference. Let's say, the true data generating process is a moving average process with order one, MA(1), $y_t = \epsilon_t + \theta\epsilon_{t-1}$ and we would like to estimate the moving average parameter θ. Since ϵ_t is not observable, estimation of θ is not straightforward. Instead, if we take the autoregressive process with order p, AR(p), $y_t = \sum_{i=1}^{p} \lambda_i y_{t-i} + u_t$ as the auxiliary model and follow the steps listed above we can easily estimate θ.

3.4.2 Robust Trend Estimation

In this section, we demonstrate the estimation of the trend parameters robustly regardless of the initial condition. As explained in Section 2, detrending the data before testing for a unit root improves the power. Since our trend parameter estimates are accurate regardless of the initial deviation, detrending the data using these trend parameter estimates will improve power irrespective of the initial condition. We assume that the following unobserved component model is the true data generating process;

$$y_t = \mu + \beta t + x_t \tag{3.9}$$

$$x_t = \rho x_{t-1} + \epsilon_t \tag{3.10}$$

where (μ, β) are the trend parameters, and ρ is the largest autoregressive root. In the following subsection we first present trend estimation in the presence of *i.i.d* ϵ_t's, and next we look at the more general case where ϵ_t's are serially correlated.

AR(1) Case

If ϵ_t in (10) is $i.i.d$, then y_t is an AR(1) process. The structural model in (9) and (10) implies the following reduced form model

$$y_t = \alpha_0 + \alpha_1 t + \rho y_{t-1} + \epsilon_t \tag{3.11}$$

where $\alpha_0 = [\mu(1-\rho) + \beta\rho]$ and $\alpha_1 = \beta(1-\rho)$. There is a one-to-one correspondence between the reduced and structural parameters, and we exploit this correspondence to estimate the trend parameters, μ and β, using indirect inference.

Let the structural model in (9) and (10) be the true model, and the reduced form model in (11) be the auxiliary model. Then we can define $\theta = (\mu, \beta, \rho)'$ and $\lambda = (\alpha_0, \alpha_1, \rho)'$, as the structural parameters and auxiliary parameters respectively. The relationship between the structural and auxiliary parameters is given above, and notice that this one-to-one correspondence does not depend on x_0. Indirect inference utilizes this relationship to estimate trend parameters and hence estimates the trend parameters consistently regardless of the initial condition..

We first estimate the auxiliary model with the original data and get $\hat{\lambda}$. Next, for a given θ we simulate data as:

$$y_t^s = \mu + \beta t + x_t^s,\; x_t^s = \rho x_{t-1}^s + \epsilon_t^s \; with \, \epsilon_t^s \sim NIID(0,1)$$

Using the simulated data we re-estimate the auxiliary model to find $\hat{\lambda}^s(\theta)$. We repeat the simulation $s = 1, 2, ..., S$ times, and get $\hat{\lambda}_S(\theta) = \frac{1}{S}\sum_{s=1}^{S} \hat{\lambda}^s(\theta)$. We finally find the indirect inference estimator with:

$$\hat{\theta}^{ii} = argmin_\theta \, (\hat{\lambda} - \hat{\lambda}_S(\theta))'(\hat{\lambda} - \hat{\lambda}_S(\theta))$$

where $\hat{\theta}^{ii} = (\hat{\mu}^{ii}, \hat{\beta}^{ii}, \hat{\rho}^{ii})'$. We do not need to use a weight matrix here, since the dimension of the structural parameters is the same as that of auxiliary parameters.

AR(p) Case

If ϵ_t in (10) is serially correlated, which is a more realistic assumption, then we need to make some adjustments to take care of serial correlation. Let, $(1 - \tau_1 L - \tau_2 L^2 - .... - \tau_{p-1} L^{p-1})\epsilon_t = u_t$, with u_t is *i.i.d*, then y_t is an AR(p) process and the reduced form can be written as

$$y_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^{p} \delta_i y_{t-i} + u_t \tag{3.12}$$

Defining $\theta = (\mu, \beta, \rho, \tau_1, ..., \tau_{p-1})'$ and $\lambda = (\alpha_0, \alpha_1, \delta_1, ..., \delta_p)'$, we can show that there is a one-to-one correspondence between the true parameters θ and the auxiliary parameters λ that does not depend on x_0. Hence, we can use indirect inference to estimate the trend parameters, μ and β, consistently regardless of the initial condition.

We first estimate the auxiliary model (12) with the original data and get $\hat{\lambda}$. Next, for a given θ we simulate data as:

$$y_t^s = \mu + \beta t + x_t^s,\ x_t^s = \rho x_{t-1}^s + \epsilon_t^s$$

where

$$\epsilon_t^s = \sum_{j=1}^{p-1} \tau_j \epsilon_{t-j} + u_t^s \; with \, u_t^s \sim NIID(0,1)$$

Using the simulated data we re-estimate the auxiliary model to find $\hat{\lambda}^s(\theta)$. We repeat the simulation $s = 1, 2, ..., S$ times, and get $\hat{\lambda}_S(\theta) = \dfrac{1}{S}\sum_{s=1}^{S} \hat{\lambda}^s(\theta)$. We finally find the indirect inference estimator with:

$$\hat{\theta}^{ii} = argmin_\theta \, (\hat{\lambda} - \hat{\lambda}_S(\theta))'(\hat{\lambda} - \hat{\lambda}_S(\theta))$$

where $\hat{\theta}^{ii} = (\hat{\mu}^{ii}, \hat{\beta}^{ii}, \hat{\rho}^{ii}, \hat{\tau}_1^{ii}, ..., \hat{\tau}_{p-1}^{ii})'$. We do not need to use a weight matrix here, since the dimension of the structural parameters is the same as that of auxiliary parameters.

3.4.3 Robust Unit Root Test

The test that we propose follows the same steps with DF-GLS and differ only in the estimation of the trend parameters. The following three steps summarize our testing procedure:

1. Estimate the trend parameters using indirect inference as explained in the previous section

2. Detrend the data using these trend parameter estimates

3. Use the detrended data to test for a unit root

The use of indirect inference in step 1 robustifies our test to initial condition and detrending data, as in step 2, will improve the power of the test.

The steps 2 and 3 can be illustrated as follows; we detrend the data using the indirect inference trend parameter estimates;

$$y_t^d = y_t - \hat{\mu}^{ii} - \hat{\beta}^{ii} t$$

We finally estimate the ADF regression without deterministic terms;

$$\Delta y_t^d = \pi y_{t-1}^d + \sum_{j=1}^{p} \psi_j \Delta y_{t-j}^d + \varepsilon_t$$

and our robust unit root test is simply the t-test for testing $\pi = 0$.

3.5 Preliminary Simulation Results

In this section we present preliminary simulation results that demonstrates the finite sample power performance of our robust unit root test. First we show that our test is not sensitive to initial condition, and preserves its high power in the presence of large initial deviations. Next, we compare the finite sample power performance of our test to that of the optimal tests (DF-GLS) and the conventional tests (DF t-test) through Monte-Carlo simulations. We use the same simulation set-up used in Section 3.1, and Figure 3.2 replicates Figure 3.1 with the results for our robust test.

Figure 3.2 clearly shows that our test is robust to initial condition. The robustness is due to the accurate estimation of the trend parameters regardless of the initial condition. Thus, the data is detrended successfully before testing for a unit root, for all levels of initial deviation.

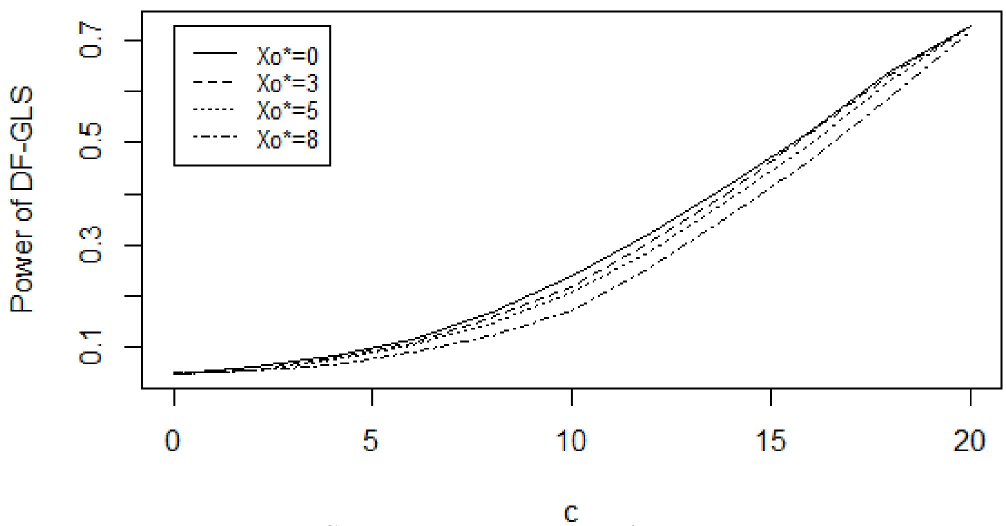


Figure 3.2: Finite Sample Powercurves of the Robust Unit Root Test

The next set of figures are provided to compare the finite sample power performance of our test to other popular methods. From Figures 3.3 and 3.4, we see that our robust test improves the power upon the DF t-test for small initial deviations, however compared to the optimal test (DF-GLS) the power gain is relatively small. However, for large initial deviations our test preserves its power while the optimal test has trivial power as Figures 3.5 and 3.6 suggest. Also notice that for large initial deviations, DF t-test has the highest power and this is also pointed out by Elliott and Muller (2006). Elliott and Muller (2006) proves that for very large initial deviations, DF t-test has optimal power properties.

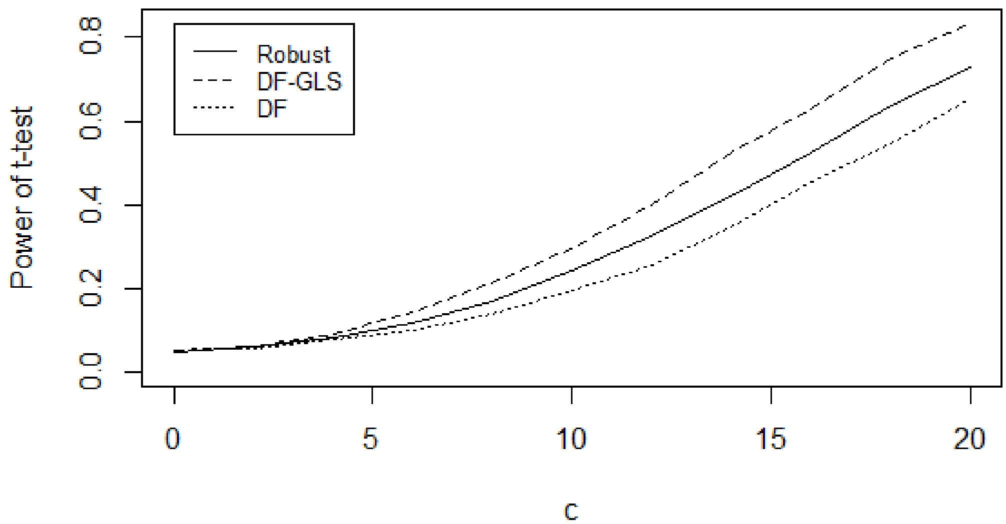


Figure 3.3: Finite Sample Powercurves when $X_0^* = 0$

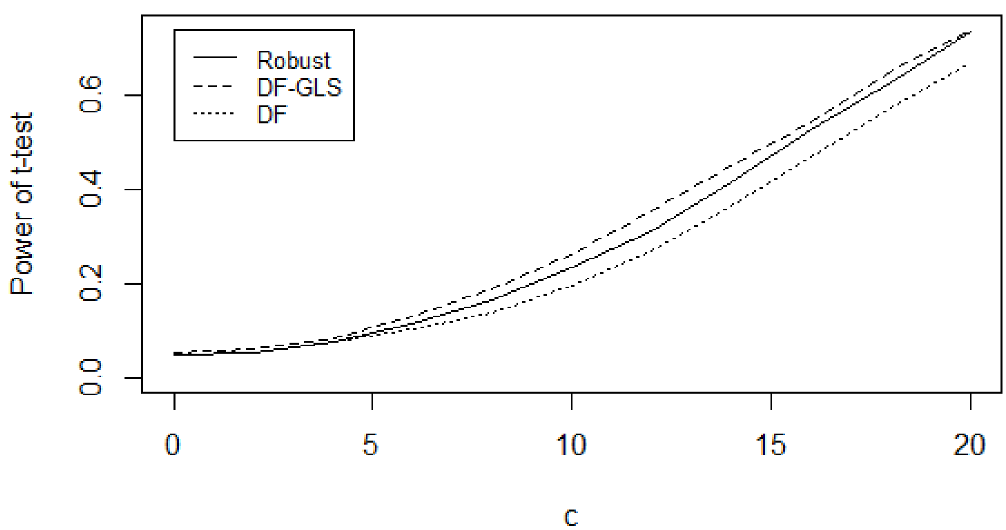


Figure 3.4: Finite Sample Powercurves when $X_0^* = 2$

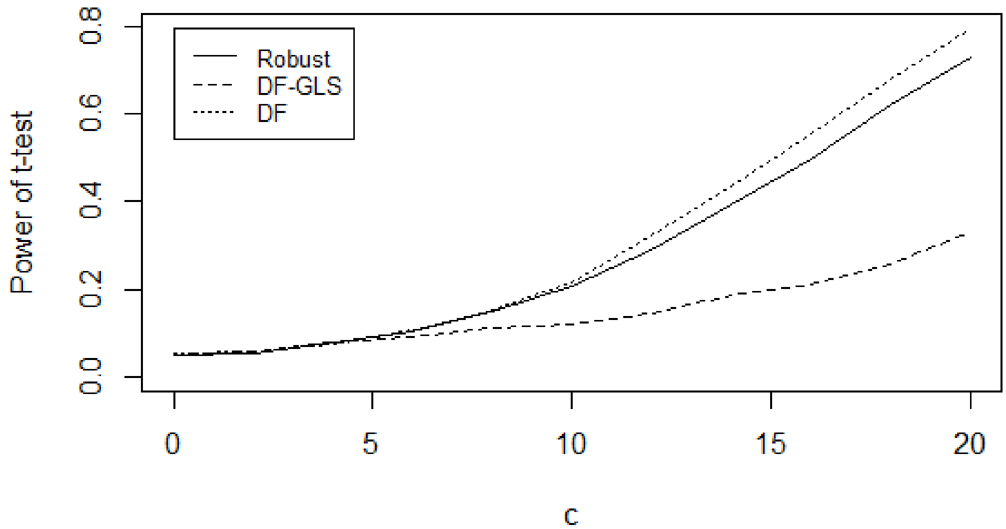


Figure 3.5: Finite Sample Powercurves when $X_0^* = 5$

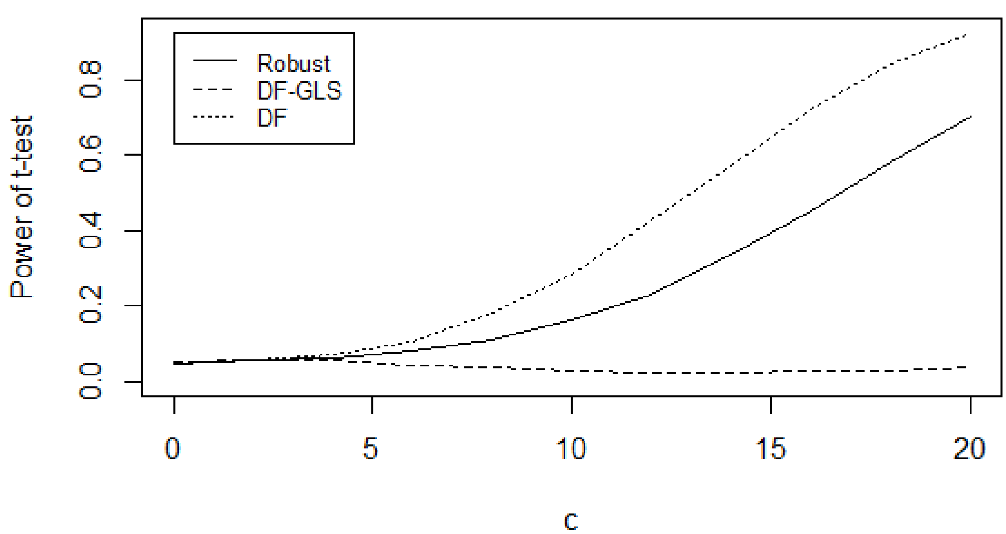


Figure 3.6: Finite Sample Powercurves when $X_0^* = 8$

3.6 Conclusion

In this paper, we consider the initial condition problem in unit root tests and trend estimation. Majority of the unit root tests are not robust to initial condition and may perform very poorly due to initial displacement. In this paper, we robustify the unit root tests to initial condition by utilizing indirect inference method. Our contribution in this paper is twofold; first we prove that indirect inference method can estimate trend function parameters consistently regardless of the initial displacement. Second, we improve the power of the unit root tests by using the robust indirect inference trend parameter estimates to detrend the data. The simulation results indicate that the power of our robust unit root test is not sensitive to initial condition. However, through some Monte Carlo studies, we also demonstrate the robustness-power tradeoff of the unit root tests.

BIBLIOGRAPHY

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. Econometrica: Journal of the Econometric Society, 821-856.

Andrews, D. W., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. Econometrica: Journal of the Econometric Society, 1383-1414.

Bai, J. (1994). Least squares estimation of a shift in linear processes. Journal of Time Series Analysis, 15(5), 453-472.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. Econometrica, 47-78.

Barro, R. J. (1991). Economic growth in a cross section of countries. The Quarterly Journal of Economics, 106(2), 407-443.

Barro, R. J., & Sala-i-Martin, X. (1992). Convergence. Journal of Political Economy, 223-251.

Baumol, W. J. (1986). Productivity growth, convergence, and welfare: what the long-run data show. The American Economic Review, 1072-1085.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. The Annals of Statistics, 2(3), 489-502.

Caner, M., & Killian L. (2001). Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate. Journal of International Money and Finance 20, 639-657

Carlino, G. A., & Mills, L. O. (1993). Are US regional incomes converging?: A time series analysis. Journal of Monetary Economics, 32(2), 335-346.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient Tests for an Autoregressive Unit Root. Econometrica, 64(4), 813-836.

Elliott, G., & Muller, U. K. (2006). Minimizing the impact of the initial condition on testing for unit roots. Journal of Econometrics, 135(1), 285-310.

Evans, G. B. A., & Savin, N. E. (1981). Testing for unit roots: 1. Econometrica: Journal of the Econometric Society, 753-779.

Evans, G. B. A., & Savin, N. E. (1984). Testing for unit roots: 2. Econometrica: Journal of the Econometric Society, 1241-1269.

Ferretti, N., & Romo, J. (1996). Unit Root Bootstrap Tests for AR(1) Models. Biometrika 83, 849-860.

Hansen, B. E. (1992). Testing for parameter instability in linear models. Journal of policy Modeling, 14(4), 517-533.

Hansen, B. E. (1999). The Grid Bootstrap and the Autoregressive Model. The Review of Economics and Statistics 81, 594-607.

Harvey, D. I., Leybourne, S. J., & Taylor, A. R. (2009a). Unit root testing in practice:

dealing with uncertainty over the trend and initial condition. Econometric Theory, 25(3), 587.

Harvey, D. I., Leybourne, S. J., & Taylor, A. R. (2009b). Simple, robust, and powerful tests of the breaking trend hypothesis. Econometric Theory, 25(4), 995.

Harvey, D. I., Leybourne, S. J., & Taylor, A. R. (2010). Robust methods for detecting multiple level breaks in autocorrelated time series. Journal of Econometrics, 157(2), 342-358.

Kiefer, N. M., & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory, 21(06), 1130-1164.

Kuo, B.-S., & Tsong, C.-C. (2005). Bootstrap Inference for Stationarity. HECER Discussin Papers 50, 1-31.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., & Shin, Y. (1992). Testing the Null 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.

Lee, J. & Lee, Y.I. (2012). Size Improvement of the KPSS Test Using Sieve Bootstraps. Economics Letters 116, 483-486.

Lee, J., Huang, C. J., & Shin, Y. (1997). On stationary tests in the presence of structural breaks. Economics Letters, 55(2), 165-172.

Leybourne, S. & McCabe, B. (1994). A Consistent Test for a Unit Root. Journal of Business and Economic Statistics 12, 157-166.

Leybourne, S., Taylor, R., & Kim, T. H. (2007). CUSUM of SquaresBased Tests for a Change in Persistence. Journal of Time Series Analysis, 28(3), 408-433.

Loewy, M. B., & Papell, D. H. (1996). Are US regional incomes converging? Some further evidence. Journal of Monetary Economics, 38(3), 587-598.

Muller, U. (2005). Size and Power of Tests of Stationarity in Highly Autocorrelated Time Series. Journal of Econometrics 128, 195-213.

Nankervis, J. & Savin, N. (1996). The Level and Power of the Bootstrap t Test in the AR(1) Model with Trend. Journal of Business and Economic Statistics 14, 161-168.

Newey, W.K., & West, K.D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703-708.

Ng, S., & Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. Econometrica, 69(6), 1519-1554.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. Econometrica: Journal of the Econometric Society, 1361-1401.

Perron, P., & Vogelsang, T. J. (1992). Testing for a unit root in a time series with a changing mean: corrections and extensions. Journal of Business & Economic Statistics, 10(4), 467-470.

Perron, P., & Yabu, T. (2009). Testing for shifts in trend with an integrated or stationary noise component. Journal of Business & Economic Statistics, 27(3).

Perron, P., & Zhu, X. (2005). Structural breaks with deterministic and stochastic trends. Journal of Econometrics, 129(1), 65-119.

Saikkonen, P. & Luukkonen, R. (1993). Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models. Journal of the American Statistical Association 188, 596-601.

Sayginsoy, O. (2004). Powerful and Serial Correlation Robust Tests of the Economic Convergence Hypothesis. Econometrics 0503014, EconWPA.

Sayginsoy, O., & Vogelsang, T. (2004). Powerful tests of structural change that are robust to strong serial correlation. Working paper 04/08, SUNY at Albany.

Sayginsoy, O., & Vogelsang, T. J. (2011). Testing for a shift in trend at an unknown date: a fixed-b analysis of heteroskedasticity autocorrelation robust OLS-based tests. Econometric Theory, 27(5), 992.

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

Sidneva, N., & Zivot, E. (2007). Trends of U.S. emissions of nitrogen oxides and volatile organic compounds. Working Paper 07/11, University of Washington

Solow, R. M. (1956). A contribution to the theory of economic growth. The Quarterly Journal of Economics, 70(1), 65-94.

Stock, J. H. (1994). Unit roots, structural breaks and trends. Handbook of econometrics, 4, 2739-2841.

Tomljanovich, M., & Vogelsang, T. J. (2002). Are US regions converging? Using new econometric methods to examine old issues. Empirical Economics, 27(1), 49-62.

Vogelsang, T. J. (1998). Trend function hypothesis testing in the presence of serial correlation. Econometrica, 123-148.

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. Journal of Business & Economic Statistics, 20(1), 25-44.

VITA

Emre Aylar was born on July 20, 1984, in Corum, Turkey. He received his Bachelor's Degree in Economics from Bogazici University in 2007. Next, he completed a Masters Degree in Economics at Iowa State University in May 2009. In September 2009, he enrolled as a graduate student in Economics at the University of Washington and in June 2014 he graduated with a Doctor of Philosophy in Economics. He has accepted an assistant professor position in the Department of Economics at Lund University, Sweden to begin in September 2014.