Unit Root Testing with Stationary Covariates and a Structural Break in the Trend Function

Sebastian Fossati*
University of Alberta

This version: January 10, 2013

Abstract

The issue of testing for a unit root allowing for a structural break in the trend function is considered. The focus is on the construction of more powerful tests using the information in relevant multivariate data sets. The proposed test adopts the GLS detrending approach and uses correlated stationary covariates to improve power. As it is standard in the literature, the break date is treated as unknown. Asymptotic distributions are derived and a set of asymptotic and finite sample critical values are tabulated. Asymptotic local power functions show that power gains can be large. Finite sample results show that the test exhibits small size distortions and power that can be far beyond what is achievable by univariate tests.

Keywords: unit root test, GLS detrending, structural break.
JEL Codes: C22, C32.

*Contact: Department of Economics, University of Alberta, Edmonton, AB T6G 2H4, Canada. Email: sfossati@ualberta.ca. Web: http://www.ualberta.ca/~sfossati/.

1 Introduction

Testing for an autoregressive unit root against trend stationary alternatives has become standard in the empirical analysis of macroeconomic time series. While a substantial literature is devoted to the construction of unit root tests with good size and power, the work of Perron (1989) demonstrates the need to allow for structural breaks in the trend function of common U.S. macroeconomic time series. This paper proposes a test of the unit root hypothesis that allows for a structural break in the trend function and uses the information in relevant multivariate data sets to improve power. The test is simple to construct and exhibits power that can be far beyond what is achievable by univariate tests.

Perron (1989) shows that the power of unit root tests against trend stationary alternatives can be severely reduced when the true data generating process (DGP) involves structural breaks in the trend function. He proposes unit root tests that are valid when a break in the trend function is present given that the location of the break is known, a priori. This assumption was heavily criticized (see, e.g., Christiano, 1992) and, as a consequence, the subsequent literature has focused on methods that endogenously determine the location of the break; see, Zivot and Andrews (1992), Perron (1997), Perron and Rodríguez (2003), Rodríguez (2007), Papell and Prodan (2007), Kim and Perron (2009), Harris et al. (2009), and Carrion-i-Silvestre et al. (2009). While unit root tests with an estimated break date are robust to the presence of a structural break in the trend function, the tests can have very low power against local alternatives that are close to unity.

The recent literature has focused on improving the power of unit root tests that allow for structural breaks in the trend function. Carrion-i-Silvestre et al. (2009) develop tests that allow for multiple breaks in the level and the slope of the trend function both under the null and the alternative hypotheses. While these tests have improved power when a break is in fact present, the tests suffer from large liberal size distortions when the break does not occur.[1] To solve this problem, Carrion-i-Silvestre et al. (2009) suggest the use of a pre-test to determine the presence of a break. A similar strategy is recommended in Harris et al. (2009). They propose a break date estimator that also relies on a pre-test and yields efficient unit root tests. In finite samples, the properties of these unit root tests are directly related to the performance of the pre-test and, in particular, the tests suggested to determine the presence of a break have low power when the break is of small or moderate size. This result implies that the unit root tests proposed in Carrion-i-Silvestre et al. (2009) and Harris et al. (2009) can have very low power when large structural breaks are not present.[2]

This paper follows the approach initiated by Hansen (1995) and uses the information contained in correlated stationary covariates to construct unit root tests that can have higher power than univariate tests. Hansen (1995) proposes a unit root test with covariates which uses the Dickey-Fuller regression equation (Dickey and Fuller, 1979; Said and Dickey, 1984) augmented with leads and lags of stationary covariates to improve power. The role of the covariates is to soak up part of the variability in the variable of interest, generating tighter confidence intervals and more powerful tests.[3] Based on the results in Elliott et al. (1996), Pesavento (2006) suggests constructing Hansen's test using generalized least squares (GLS) detrended data and shows that this approach yields tests with good size and power. This paper contributes to the literature proposing a test of the unit root hypothesis that allows for a structural break in the trend function and uses correlated stationary covariates to improve power. As it is standard in the literature, the break date is treated as unknown. The test uses GLS detrended data and can be seen as an extension of the test proposed in Pesavento (2006) to the case where the trend function exhibits a structural break. Alternatively, it can be seen as an extension of a test proposed in Perron and Rodríguez (2003) that incorporates stationary covariates to improve power.[4]

In the next section, I describe the DGP and the three structural break models considered. In section 3, I discuss the construction of covariate unit root tests for the case where the trend function exhibits a structural break and discus the estimation of the break date. The asymptotic distributions of the test are derived and approximated numerically. The asymptotic local power functions are computed for the trending cases and show large power gains. In section 4, the finite sample properties of the test are analyzed using Monte Carlo simulation and results show that the test exhibits small size distortions and power that can be far beyond what is achievable by univariate tests. Section 5 concludes.

2 The Model with a Structural Break

Consider a DGP of the form

$$x_t = \psi_x' z_{xt} + u_{xt}, \tag{1}$$

$$y_t = \psi_y' z_{yt}(\delta_0) + u_{yt}, \tag{2}$$

and

$$A(L) \begin{pmatrix} u_{xt} \\ (1-\rho L)u_{yt} \end{pmatrix} = \varepsilon_t \qquad t = 1, 2, ..., T, \tag{3}$$

where y_t is univariate (and potentially non-stationary), x_t is a stationary process of dimension $(m \times 1)$, and $A(L)$ is a matrix polynomial in the lag operator with first element equal to the identity matrix. z_{xt} and $z_{yt}(\delta_0)$ contain deterministic terms to be defined later. As in Elliott and Jansson (2003), the error process $\{\varepsilon_t\}$ satisfies the following assumptions:

Assumption 1. $|A(r)| = 0$ *has all roots outside the unit circle.*

Assumption 2. $E_{t-1}(\varepsilon_t) = 0$, $E_{t-1}(\varepsilon_t \varepsilon_t') = \Sigma$, *and* $E \|\varepsilon_t\|^{4+\zeta} < \infty$ *(a.s.) for some*

$\zeta > 0$, *where* Σ *is positive definite,* $E_{t-1}(\cdot)$ *denotes the conditional expectation with respect to* $\{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$*, and* $\|\cdot\|$ *is the Euclidean norm.*

Assumption 3. u_0,u_{-1},...,u_{-k} *are* $O_p(1)$ *with* $u_t = [u'_{xt}, u_{yt}]'$.

Define $u_t(\rho) = [u'_{xt}, u_{yt}(\rho)]' = [u'_{xt}, (1 - \rho L)u_{yt}]' = A(L)^{-1}\varepsilon_t$ and let $\Gamma(k) = E[u_t(\rho)u_{t+k}(\rho)']$ be the covariance function of $u_t(\rho)$. Following Pesavento (2006), Assumption 4 requires that $\Gamma(k)$ is absolute summable.

Assumption 4. *The covariance function of* $u_t(\rho)$ *is absolute summable such that* $\sum_{j=-\infty}^{+\infty} \|\Gamma(k)\| < \infty$ *and* $\sum_{j=-\infty}^{+\infty} j\,\|\Gamma(k)\| < \infty$.

Remark 1. Under Assumptions 1-4, $T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} u_t(\rho) \Rightarrow \Omega^{1/2}W(r)$, where "$\Rightarrow$" denotes weak convergence in distribution, $\lfloor\cdot\rfloor$ is the integer part, $W(r)$ is a multivariate Wiener process defined on the space of continuous functions on the interval [0,1], and $\Omega = A(1)^{-1}\Sigma A(1)^{-1'}$ is 2π times the spectral density at frequency zero of $u_t(\rho)$.

Consider the partition of Ω

$$\Omega = \begin{pmatrix} \Omega_{xx} & \omega_{xy} \\ \omega_{yx} & \omega_{yy} \end{pmatrix}$$

and define $R^2 = \omega_{yy}^{-1}\omega_{yx}\Omega_{xx}^{-1}\omega_{xy}$, the frequency zero correlation between the shocks to x_t and the quasi-differences of y_t. Although $0 \leq R^2 \leq 1$, following Elliott and Jansson (2003) it is assumed that $0 \leq R^2 < 1$, ruling out the case where, under the null, the partial sum of x_t cointegrates with y_t.

Three structural break models are considered for the deterministic components. Let T_0 be the time of the structural break in (2), and assume $T_0 = \lfloor T\delta_0 \rfloor$ with $\delta_0 \equiv T_0/T \in (0, 1)$ the break fraction parameter. Then $z_{yt}(\delta_0) = (1, 1(t > T_0), t, 1(t > T_0)(t - T_0))'$, where $1(\cdot)$ is the indicator function, and $\psi_y = (\mu_{y1}, \mu_{y2}, \beta_{y1}, \beta_{y2})'$. In the case of covariates, the trend components are $z_{xt} = (1, t)'$ and $\psi_x = (\mu_x, \beta_x)'$, with μ_x and β_x

of dimension $(m \times 1)$. Elliott and Jansson (2003) and Pesavento (2006) consider five cases for the deterministic part of the model: (1) $\mu_{y1} = \mu_{y2} = \beta_{y1} = \beta_{y2} = 0$ and $\mu_x = \beta_x = 0$; (2) $\mu_{y2} = \beta_{y1} = \beta_{y2} = 0$ and $\mu_x = \beta_x = 0$; (3) $\mu_{y2} = \beta_{y1} = \beta_{y2} = 0$ and $\beta_x = 0$; (4) $\mu_{y2} = \beta_{y2} = 0$ and $\beta_x = 0$; (5) $\mu_{y2} = \beta_{y2} = 0$. The first case then corresponds to a model with no deterministic terms. The second has no constant or trend in x_t but a constant in y_t. The third case includes a constant in x_t and a constant in y_t. Case 4 includes a constant in x_t and a constant and trend in y_t. Case 5 allows for constants and trends in both x_t and y_t. The models analyzed here extend these cases to incorporate a structural break in the deterministic component of y_t.

Model A. Structural break in the intercept: The "crash" model in Perron (1989) allows for a one-time structural break in the intercept of y_t, a level shift. There are four relevant cases for the deterministic part of this model:

Case 2-A: $\beta_{y1} = \beta_{y2} = 0$ and $\mu_x = \beta_x = 0$.

Case 3-A: $\beta_{y1} = \beta_{y2} = 0$ and $\beta_x = 0$.

Case 4-A: $\beta_{y2} = 0$ and $\beta_x = 0$.

Case 5-A: $\beta_{y2} = 0$.

For this model, the set of deterministic components of y_t, $z_{yt}(\delta_0)$, is given by

$$z_{yt}(\delta_0) = (1, 1(t > T_0))' \tag{4}$$

for cases 2-A and 3-A, and

$$z_{yt}(\delta_0) = (1, 1(t > T_0), t)' \tag{5}$$

for cases 4-A and 5-A.

Model B. Structural break in the slope: The "changing growth" model in Perron (1989) allows for a one-time structural break in the slope of y_t. There are two relevant cases for the deterministic part of this model:

Case 4-B: $\mu_{y2} = 0$ and $\beta_x = 0$.

Case 5-B: $\mu_{y2} = 0$.

For this model, $z_{yt}(\delta_0)$ is given by

$$z_{yt}(\delta_0) = (1, t, 1(t > T_0)(t - T_0))' \tag{6}$$

for both cases.

Model C. Structural break in the slope and the intercept: This model allows for a one-time structural break in the slope and the intercept of y_t. There are two relevant cases for the deterministic part of this model:

Case 4-C: $\beta_x = 0$.

Case 5-C: No restrictions.

For this model, $z_{yt}(\delta_0)$ is given by

$$z_{yt}(\delta_0) = (1, 1(t > T_0), t, 1(t > T_0)(t - T_0))' \tag{7}$$

for both cases. Note that in the most general case (Model C), a structural break in the intercept ($\mu_{y2} \neq 0$) and the slope ($\beta_{y2} \neq 0$) of y_t is allowed while no structural break is allowed in the deterministic component of x_t.[5]

3 The Test and Asymptotic Distributions

In this section, I discuss the construction and asymptotic properties of covariate unit root tests that allow for a structural break in the trend function. Section 3.1 considers the construction of the statistic under the assumption that the break date is known. In section 3.2, the asymptotic distributions of the tests are derived for the case outlined in section 3.1. Section 3.3 considers the case where the break date is not known. I discuss the estimation of the break fraction and the asymptotic distributions of the tests in this case. Section 3.4 considers the selection of the non-centrality parameter for GLS detrending. In section 3.5, I approximate the asymptotic distributions and compute the asymptotic local power functions of the tests.

3.1 The test statistic

The objective is to test whether the univariate time series y_t is an integrated process of order one ($\rho = 1$) against the alternative of stationarity ($|\rho| < 1$). Let $x_t^d = x_t - \hat{\psi}_x' z_{xt}$ be the detrended x_t with $\hat{\psi}_x$ the least squares (OLS) estimate of ψ_x. Because the covariates are stationary (by assumption), OLS detrending is sufficient. The variable of interest (y_t), however, is GLS detrended, i.e. detrended under the local alternative as in Elliott et al. (1996). Let $y_t^d = y_t - \tilde{\psi}_y' z_{yt}(\delta_0)$ be the detrended y_t with $\tilde{\psi}_y$ the GLS estimate of ψ_y such that

$$\tilde{\psi}_y = \arg\min_{\psi_y} \sum_{t=1}^{T} \left[y_t^{\bar{\rho}} - \psi_y' z_{yt}^{\bar{\rho}}(\delta_0) \right]^2, \tag{8}$$

where $y_t^{\bar{\rho}} = (y_1, (1-\bar{\rho}L)y_t)$ and $z_{yt}^{\bar{\rho}}(\delta_0) = (z_{y1}(\delta_0), (1-\bar{\rho}L)z_{yt}(\delta_0))$ for $t = 1, ..., T$, with $\bar{\rho} = 1 + \bar{c}/T$, and $\bar{c} \leq 0$ is the non-centrality parameter. The selection of $\bar{c}$ is discussed below. The choice of z_{xt} and $z_{yt}(\delta_0)$ depends on the selected deterministic case and structural break model as defined in the previous section. The test statistic is based on the covariate augmented Dickey-Fuller regression

$$\Delta y_t^d = \phi y_{t-1}^d + \sum_{j=-k}^{k} \pi_{xj}' x_{t-j}^d + \sum_{j=1}^{k} \pi_{yj} \Delta y_{t-j}^d + e_{tk} \quad t = k+2, ..., T-k. \tag{9}$$

Note that deterministic terms are not included in (9) as the data has already been detrended and the lead and lag orders k are restricted to be equal. As it is standard in the literature, the truncation lag k in (9) satisfies the following condition:

Assumption 5. *$k \to \infty$ and $k^3/T \to 0$ as $T \to \infty$.*

Ng and Perron (1995) show that traditional model selection criteria such as the Bayesian or Akaike information criterion satisfy the upper bound condition in Assumption 5 and, hence, can be used to select k. While Pesavento (2006) recommends using the value of k selected by minimizing a modified Akaike information criterion (MAIC) of the type suggested in Ng and Perron (2001) from an univariate regression on the GLS detrended y_t, Fossati (2012) shows that forcing the lead and lag orders in (9) to be equal results in a small power reduction in the tests that can be avoided by removing this constraint.

The CADF-GLS(δ_0) test statistic is the t-statistic for testing $\phi = 0$ ($\rho = 1$) in (9) and rejects for large negative values. The critical values depend on the correlation between y_t^d and x_t^d, so a consistent estimate of R^2 is needed. Hansen (1995) suggests using a nonparametric estimator of the form

$$\hat{R}^2 = 1 - \left(\frac{\hat{\theta}_{21}^2}{\hat{\theta}_{11}\hat{\theta}_{22}} \right), \tag{10}$$

where

$$\hat{\Theta} = \begin{pmatrix} \hat{\theta}_{11} & \hat{\theta}_{12} \\ \hat{\theta}_{21} & \hat{\theta}_{22} \end{pmatrix} = \sum_{i=-M}^{M} w(i/M) \frac{1}{T} \sum_{t} \hat{\nu}_{t-i}\hat{\nu}_t', \tag{11}$$

with $\hat{\nu}_t = \left(\hat{e}_{tk} + \sum_{j=-k_1}^{k_2} \hat{\pi}_{xj}' x_{t-j}^d, \hat{e}_{tk}\right)'$, $w(\cdot)$ is a kernel weight function, e.g. the Bartlett or Parzen kernels, and M is a bandwidth.[6] In this paper, all estimations are performed using the Parzen kernel and a bandwidth determined following Andrews (1991).

3.2 Asymptotic distributions with a known break date

I start considering the limiting distribution of the tests in the case where the break date is known. The asymptotic distributions are derived using a local-to-unity framework

(Phillips, 1987a) where $\rho = 1+c/T$ with $c \leq 0$ fixed as $T \to \infty$. Theorem 1 summarizes the asymptotic distributions for the case where a one-time break in the intercept is allowed (Model A).

Theorem 1 (Model A)**.** *Let y_t be generated by equations* (1) *to* (3) *with $\rho = 1 + c/T$ and under Assumptions* 1 *to* 5*, and the CADF-GLS(δ_0) be the t-statistic for testing $\phi = 0$ in regression* (9) *with data obtained from local GLS detrending at $\bar{\rho} = 1 + \bar{c}/T$. Then, as $T \to \infty$, for Model A we have*

$$t_{\hat{\phi}}(\delta_0) \Rightarrow \left(\int_0^1 J_{xyc}^{d\ 2}\right)^{1/2} \left[\left(\int_0^1 J_{xyc}^{d\ 2}\right)^{-1} \left(\int_0^1 J_{xyc}^d dW_y + \Lambda_{c\bar{c}}\right) + c\right] \equiv H^{iA}\left(c, \bar{c}, R^2\right)$$

where i refers to the deterministic case ($i = 2, 3, 4, 5$), and

1. $\Lambda_{c\bar{c}} = 0$ *and* $J_{xyc}^d(r) = J_{xyc}(r)$ *for case 2.*
2. $\Lambda_{c\bar{c}} = Q^{1/2} W_x(1) \int_0^1 J_{xyc}^d$ *and* $J_{xyc}^d(r) = J_{xyc}(r)$ *for case 3.*
3. $\Lambda_{c\bar{c}} = Q^{1/2} W_x(1) \int_0^1 J_{xyc}^d - V_{c\bar{c}} \left[\int_0^1 J_{xyc}^d - c \int_0^1 r J_{xyc}^d\right]$ *and* $J_{xyc}^d(r) = J_{xyc}(r) - rV_{c\bar{c}}$ *for case 4.*
4. $\Lambda_{c\bar{c}} = Q^{1/2} \left[-2W_x(1) + 6 \int_0^1 W_x\right] \int_0^1 J_{xyc}^d + Q^{1/2} \left[6W_x(1) - 12 \int_0^1 W_x\right] \int_0^1 r J_{xyc}^d$ $-V_{c\bar{c}} \left[\int_0^1 J_{xyc}^d - c \int_0^1 r J_{xyc}^d\right]$ *and* $J_{xyc}^d(r) = J_{xyc}(r) - rV_{c\bar{c}}$ *for case 5.*

$J_{xyc}(r)$ is an Ornstein-Uhlenbeck process such that $J_{xyc}(r) = W_{xy}(r) + c \int_0^r e^{(r-s)c}$ $W_{xy}(s)ds$ with $W_{xy}(r) = Q^{1/2} W_x(r) + W_y(r)$ where $W_x(r)$ and $W_y(r)$ are univariate independent standard Brownian motions, and $Q = R^2/(1 - R^2)$. $V_{c\bar{c}} = b_1/a_1$, $b_1 = (1-\bar{c})J_{xyc}(1) + \bar{c}^2 \int_0^1 r J_{xyc}$, and $a_1 = 1 - \bar{c} + \bar{c}^2/3$. Unless otherwise stated, all integrals are over r with r suppressed, e.g. $\int_0^1 J_{xyc}^d = \int_0^1 J_{xyc}^d(r)dr$.

Remark 2. In the case of Model A, the tests have the same asymptotic distributions established in Pesavento (2006, Theorem 2) as a break in the intercept is a special case of the "slowly evolving trend" considered in Elliott et al. (1996).

Note that the limiting distributions in Theorem 1 are independent of the break fraction parameter (δ_0). This results does not hold when a break in the slope is present. The limiting distributions of the tests in the case where a break in the slope is allowed (Models B and C) are summarized in Theorem 2.

Theorem 2 (Models B and C)**.** *Let y_t be generated by equations* (1) *to* (3) *with $\rho = 1 + c/T$ and under Assumptions* 1 *to* 5*, and the CADF-GLS(δ_0) be the t-statistic for testing $\phi = 0$ in regression* (9) *with data obtained from local GLS detrending at $\bar{\rho} = 1 + \bar{c}/T$. Then, as $T \to \infty$, for Models B and C we have*

$$t_{\hat{\phi}}(\delta_0) \Rightarrow \left(\int_0^1 J_{xyc}^{d\,2}\right)^{1/2} \left[\left(\int_0^1 J_{xyc}^{d\,2}\right)^{-1} \left(\int_0^1 J_{xyc}^d dW_y + \Lambda_{c\bar{c}}\right) + c\right] \equiv H^{iB}\left(c, \bar{c}, R^2, \delta_0\right)$$

where i refers to the deterministic case ($i = 4, 5$), and

$$\Lambda_{c\bar{c}} = \Lambda_x - V_{c\bar{c}}^{(1)} \left[\int_0^1 J_{xyc}^d - c\int_0^1 r J_{xyc}^d\right] - V_{c\bar{c}}^{(2)} \left[\int_{\delta_0}^1 J_{xyc}^d - c\int_{\delta_0}^1 (r - \delta_0) J_{xyc}^d\right],$$

with

1. $\Lambda_x = Q^{1/2} W_x(1) \int_0^1 J_{xyc}^d$ *for case 4.*
2. $\Lambda_x = Q^{1/2}\left[-2W_x(1) + 6\int_0^1 W_x\right]\int_0^1 J_{xyc}^d + Q^{1/2}\left[6W_x(1) - 12\int_0^1 W_x\right]\int_0^1 r J_{xyc}^d$ *for case 5.*

$J_{xyc}^d(r, \delta_0) = J_{xyc}(r) - rV_{c\bar{c}}^{(1)} - (r - \delta_0)V_{c\bar{c}}^{(2)} 1(r > \delta_0)$ *with $J_{xyc}(r)$ an Ornstein-Uhlenbeck process such that $J_{xyc}(r) = W_{xy}(r) + c\int_0^r e^{(r-s)c} W_{xy}(s)ds$ with $W_{xy}(r) = Q^{1/2}W_x(r) + W_y(r)$ where $W_x(r)$ and $W_y(r)$ are univariate independent standard Brownian motions, and $Q = R^2/(1 - R^2)$. $V_{c\bar{c}}^{(1)} = (\lambda_1 b_1 + \lambda_2 b_2)$ and $V_{c\bar{c}}^{(2)} = (\lambda_2 b_1 + \lambda_3 b_2)$ where $b_1 = (1 - \bar{c})J_{xyc}(1) + \bar{c}^2 \int_0^1 r J_{xyc}$, $b_2 = (1 - \bar{c} + \bar{c}\delta_0)J_{xyc}(1) - J_{xyc}(\delta_0) + \bar{c}^2 \int_{\delta_0}^1 (r - \delta_0) J_{xyc}$, $\lambda_1 = a_3/a_4$, $\lambda_2 = -a_2/a_4$, $\lambda_3 = a_1/a_4$, $a_1 = 1 - \bar{c} + \bar{c}^2/3$, $a_2 = 1 - \delta_0 - \bar{c}(1 - \delta_0) - \bar{c}^2\delta_0(1 - \delta_0^2)/2 + \bar{c}^2(1 - \delta_0^3)/3$, $a_3 = 1 - \delta_0 - \bar{c}(1 - 2\delta_0 + \delta_0^2) - \bar{c}^2\delta_0(1 - \delta_0) + \bar{c}^2(1 - \delta_0^3)/3$,* *and* $a_4 = a_1 a_3 - a_2^2$. *Unless otherwise stated, all integrals are over* r *with* r *suppressed, e.g.* $\int_0^1 J_{xyc}^d = \int_0^1 J_{xyc}^d(r, \delta_0)dr$.

Remark 3. As in Perron and Rodríguez (2003), since a break in the intercept is a special case of the "slowly evolving trend" considered in Elliott et al. (1996), the limiting distributions for Models B and C are the same.

Similar to other tests in the literature, the distributions depend on the value of the parameter R^2. Furthermore, the limiting distributions of the tests are nonstandard and depend on the alternative chosen ($\bar{c}$). This dependence implies that no uniformly most powerful test exists for this problem and power depends on the selected alternative. The case when the break date is unknown is discussed next.

3.3 The case when the break date is unknown

While in some applications the break date may be known, a priori, oftentimes the researcher will want to treat the break date as unknown. An estimator of the break fraction considered in the literature consists in selecting the break date that maximizes the absolute value of the t-statistic on one of the break parameters (μ_{y2},β_{y2}) in (8); see, Perron (1997), Vogelsang and Perron (1998), Perron and Rodríguez (2003), and Rodríguez (2007).[7] If Model A is specified, $\hat{\delta}$ is estimated as the break date that maximizes the absolute value of the t-statistic on the coefficient of the change in the intercept. For Models B or C, $\hat{\delta}$ is chosen to maximize the absolute value of the t-statistic on the coefficient of the change in the slope. Denote the resulting estimator

$$\hat{\delta} = \arg \max_{\delta \in [\epsilon, 1-\epsilon]} \begin{cases} |t_{\hat{\mu}_{y2}}(\delta)| & \text{for Model A} \\ |t_{\hat{\beta}_{y2}}(\delta)| & \text{for Models B and C} \end{cases} \tag{12}$$

where ϵ refers to the required trimming with $\epsilon = 0.15$ being used throughout.

In the case of Model A, the distributions do not depend on the break fraction and, asymptotically, any estimate of δ_0 can be used. In finite samples however, a good estimate of the break date is important. In the case of Models B and C, the distributions depend on whether the break is present or not. Vogelsang and Perron (1998) show that, under the null hypothesis, if $\beta_{y2} \neq 0$ then $\hat{\delta}$ is a consistent estimator of the break fraction and the asymptotic distributions of the tests are the same as in the case where the break date is known (Theorem 2). In this case, critical values can be tabulated for different values of the break fraction. While the use of critical values computed under this approach (i.e., assuming $\beta_{y2} \neq 0$) yields more powerful tests when the break is in fact present, the tests are severely over-sized when a break is not present ($\beta_{y2} = 0$). Asymptotically, this problem can be avoided with a pre-test to determine the presence of a break, a strategy adopted in Carrion-i-Silvestre et al. (2009) and Harris et al. (2009). In finite samples, however, the performance of these unit root tests is directly related to the performance of the pre-test. In particular, the tests suggested to determine the presence of a break have low power when the break is of small or moderate size directly affecting the size of the unit root tests (see Carrion-i-Silvestre et al., 2009; Harris et al., 2009).

A conservative approach suggested in Vogelsang and Perron (1998) uses critical values obtained under the assumption that no break is present. Remark 4 summarizes the asymptotic results for the case when, under the null hypothesis, there is no change in the slope of the trend function ($\beta_{y2} = 0$).

Remark 4. Under the null hypothesis, when no break in the slope of the trend function is present ($\beta_{y2} = 0$) we have that $t_{\hat{\beta}_{y2}}(\delta) \Rightarrow V_{c\bar{c}}^{(2)} \lambda_3^{-1/2}$ with $V_{c\bar{c}}^{(2)}$ and λ_3 defined in Theorem 2. Then,

$$\hat{\delta} = \arg \max_{\delta \in [\epsilon, 1-\epsilon]} |t_{\hat{\beta}_{y2}}(\delta)| \Rightarrow \arg \max_{\delta \in [\epsilon, 1-\epsilon]} |V_{c\bar{c}}^{(2)} \lambda_3^{-1/2}| \equiv \delta^*. \tag{13}$$

For Models B and C, $t_{\hat{\phi}}(\hat{\delta}) \Rightarrow H^{iB}\left(c, \bar{c}, R^2, \delta^*\right)$ for $i = 4, 5$.

The asymptotic distributions of the tests are then given by

$$\text{CADF-GLS}(\hat{\delta}) \Rightarrow \begin{cases} H^{iA}\left(c, \bar{c}, R^2\right) & \text{for Model A and } i = 2, 3, 4, 5 \\ H^{iB}\left(c, \bar{c}, R^2, \delta^*\right) & \text{for Models B and C and } i = 4, 5 \end{cases} \tag{14}$$

with terms defined in Theorems 1 and 2 and Remark 4. When critical values computed under the conservative approach are used, the tests have an asymptotic size which is equal to the asymptotic level when the break is not present and an asymptotic size which is less than the asymptotic level when the break is present. Under the alternative hypothesis, however, the conservative approach yields less powerful tests. Because low power is less problematic when good covariates are available, I adopt the conservative approach suggested in Vogelsang and Perron (1998) and given by (12) and (14).[8]

3.4 Selection of the non-centrality parameter $\bar{c}$

Elliott et al. (1996) recommend using the value of the non-centrality parameter $\bar{c}$ such that the asymptotic power of the point-optimal unit root test is 50%. For covariate tests, however, the "optimal" $\bar{c}$ will depend on R^2. Nevertheless, Elliott and Jansson (2003) recommend using the values of $\bar{c}$ that correspond to the case where the covariates have no useful information ($R^2 = 0$), noting that low power is less of an issue as R^2 rises above zero.[9] Following this recommendation and the results of Elliott et al. (1996) we have $\bar{c} = -7$ for models 2-A and 3-A, and $\bar{c} = -13.5$ for models 4-A and 5-A. In the case of Models B and C, when $R^2 = 0$ the tests are asymptotically equivalent to the ADF-GLS test of Perron and Rodríguez (2003). Therefore, for models 4-B, 5-B, 4-C, and 5-C we have $\bar{c} = -22.5$. While the limiting distributions for Models B and C are also a function of δ^*, which is now random, results in Perron and Rodríguez (2003) show that this approach yields tests with power functions that lie very close to the power envelope.

3.5 Asymptotic local power functions and critical values

The asymptotic local power functions for Model A are analyzed in Pesavento (2006) and, hence, not discussed here. For Models B and C, the asymptotic local power functions are given by $\pi^*(c, \bar{c}, R^2) = Pr[H^{iB}(c, \bar{c}, R^2, \delta^*) < b(\bar{c}, R^2)]$ where $b(\bar{c}, R^2)$ is such that $Pr[H^{iB}(0, \bar{c}, R^2, \delta^*) < b(\bar{c}, R^2)] = 0.05$ and $H^{iB}(c, \bar{c}, R^2, \delta^*)$ is defined in Theorem 2 and Remark 4. The asymptotic distributions are simulated approximating the Wiener processes on [0,1] as the partial sums of *i.i.d.* $N(0, 1)$ random variables using 1,000 steps and 10,000 replications. Figure 1 displays the asymptotic local power functions for case 5. In each panel, the lowest curve corresponds to $R^2 = 0$ and the curves are strictly increasing in R^2 with $R^2 = 0$, 0.3, 0.5, 0.7, and 0.9. The increase in power due to an increase in R^2 is substantial. Consider the asymptotic power in case 5 at the local alternative $\bar{c} = -5$.[10] Similar to the univariate versions of the test, the asymptotic power when $R^2 = 0$ is about 7%, barely above the asymptotic level. Power increases to 11% when $R^2 = 0.3$, to 19% when $R^2 = 0.5$, and to 72% when $R^2 = 0.9$.

[FIGURE 1 ABOUT HERE]

Critical values are tabulated for the limiting distributions under the null hypothesis ($c = 0$). The asymptotic distributions given by (14) are simulated approximating the Wiener processes on [0,1] as the partial sums of *i.i.d.* $N(0, 1)$ random variables using 1,000 steps and 100,000 replications. Finite sample critical values are tabulated using data generated by a random walk with *i.i.d.* $N(0, 1)$ errors, with the initial observation set to zero, $T \in \{100, 250\}$, and 50,000 replications. Figure 2 displays the asymptotic distributions for case 5. In each panel, the curve with the most negative mode corresponds to $R^2 = 0$, and the modes are strictly increasing in R^2 with $R^2 =$ 0, 0.3, 0.5, 0.7, and 0.9. The distributions show that, for a given level, critical values increase (become less negative) as R^2 increases. 1, 5, and 10% critical values for the CADF-GLS($\hat{\delta}$) test statistic for selected values of R^2 are given in Tables 1-3. When

$R^2 = 0$ the asymptotic critical values are very close to the corresponding values in Rodríguez (2007) for Model A and Perron and Rodríguez (2003) for Models B and C. Critical values are determined for the estimated $\hat{R}^2$, and linear interpolation can be used for intermediate values of $\hat{R}^2$.

[FIGURE 2 ABOUT HERE]

[TABLES 1-3 ABOUT HERE]

4 Finite Sample Properties

In this section, the finite sample size and power properties of the CADF-GLS($\hat{\delta}$) test are evaluated using Monte Carlo simulation. In section 4.1, I consider the case where the covariates behave in accordance with the assumptions of the model, i.e. the covariates are stationary and do not exhibit structural breaks in the trend function. The case where the covariates are non-stationary or highly persistent is analyzed in Hansen (1995). He notes that including non-stationary covariates invalidates the asymptotic results derived in this paper and, based on this result, recommends taking first differences before including highly persistent variables in the regression. Section 4.2 considers the consequences of ignoring a structural break in the trend function of the covariates.

4.1 Well behaved covariates

To analyze the finite sample size and power properties of the test in the case where the covariates behave in accordance with the assumptions of the model, consider a process generated by a model of the form

$$x_t = u_{xt}, \tag{15}$$

$$y_t = \mu_y 1(t > T_0) + \beta_y 1(t > T_0)(t - T_0) + u_{yt}, \tag{16}$$

and

$$\begin{pmatrix} u_{xt} \\ (1-\rho L)u_{yt} \end{pmatrix} = e_t, \tag{17}$$

where $e_t \sim i.i.d.\ N(0, \Sigma)$ for $t > 0$, and $e_t = 0$ for $t \leq 0$. Σ is a symmetric matrix of dimension (2×2) with diagonal elements equal to 1 and off-diagonal elements equal to the degree of correlation R. Structural breaks of several magnitudes are considered, with the level shift $\mu_y \in \{0, 0.5, 1, 5\}$ and, for each value of μ_y, $\beta_y \in \{-2.0, -1.8, -1.6, ..., 2.0\}$. The model is specified for $R \in \{0, 0.3, 0.5, 0.7, 0.9\}$, and three values of the break fraction $\delta_0 \in \{0.3, 0.5, 0.7\}$. The sample size $T = 100$ in all cases and, as a consequence, $T_0 \in \{30, 50, 70\}$. To avoid issues related to the determination of the lead and lag orders, all tests are constructed for $k_i = 0$ for $i = 1, 2, 3$. Empirical power is evaluated at $\rho = 0.9$. The results are based on 1,000 replications, with the same set of random errors used across values of μ_y and β_y, and reported only for case 5-C. Case 4-C shows almost the same results as case 5-C, and results for Model B are very similar to the results for Model C when $\mu_y = 0$. To determine the relative performance of the CADF-GLS($\hat{\delta}$) test, I also construct the ADF-GLS($\hat{\delta}$) test of Perron and Rodríguez (2003) with the break date selected using (12).

Figure 3 presents the empirical size of the tests for $T_0 = 50$. The tests exhibit small size distortions when $|\beta_y|$ is small (near zero) for all values of μ_y. The tests are conservative, with an empirical size that decreases as $|\beta_y|$ increases and, as shown in Harris et al. (2009), the empirical size remains relatively unchanged for values of $|\beta_y|$ greater than 1. This result is due to the fact that critical values are obtained assuming $\mu_y = \beta_y = 0$. When a break in the slope of the trend function is in fact present and under the null hypothesis, the tests do not reject as often as the level of the tests suggest (hence, the conservative approach). The tests also show an asymmetric response to the true position of the break. When $T_0 = 30$ and μ_y is large, size distortions are larger (the tests are more conservative) for positive values of β_y. The opposite is true for $T_0 = 70$ and μ_y large. The results are similar for different values of R^2 and the ADF-GLS($\hat{\delta}$).

[FIGURE 3 ABOUT HERE]

Figure 4 presents the empirical power of the tests for $T_0 = 50$. When $R^2 = 0$, the power of the CADF-GLS($\hat{\delta}$) test is almost identical to the power of the ADF-GLS($\hat{\delta}$) test. This result holds even for a sample as small as 100 observations indicating that the efficiency loss resulting from the inclusion of redundant regressors when $R^2 = 0$ is not important. However, as R^2 increases, the power of the test rises considerably. For small values of μ_y, the ADF-GLS($\hat{\delta}$) has power around 10% while if $R^2 = 0.25$ the covariate test has power around 20%, about a 100% gain. For $R^2 = 0.49$ the covariate test has power close to 40%, and for $R^2 = 0.81$ the test has power that can be as large as 90%. When a large break in the intercept is present ($\mu_y = 5$), the power of the tests dips down for $\beta_y = 0$ or near zero. Although in this case power gains are not as important, power improves fast as β_y moves away from zero. Overall, large power gains are available and the CADF-GLS($\hat{\delta}$) test exhibits power that can be far beyond what is achievable when covariates are not included in the regression equation.

[FIGURE 4 ABOUT HERE]

4.2 Covariates with a structural break in the trend

The results in this paper are based on the important assumption that x_t does not exhibit structural breaks in the trend function. This section evaluates the finite sample properties of the CADF-GLS($\hat{\delta}$) test in the case where the trend function of x_t exhibits a level shift. Consider a process generated by a model of the form

$$x_t = \mu_x 1(t > T_0^x) + u_{xt}, \tag{18}$$

$$y_t = \mu_y 1(t > T_0) + \beta_y 1(t > T_0)(t - T_0) + u_{yt}, \tag{19}$$

and

$$\begin{pmatrix} u_{xt} \\ (1-\rho L)u_{yt} \end{pmatrix} = e_t, \qquad (20)$$

where $e_t \sim i.i.d.\ N(0,\Sigma)$ for $t > 0$, and $e_t = 0$ for $t \leq 0$. Σ is a symmetric matrix of dimension (2×2) with diagonal elements equal to 1 and off-diagonal elements equal to the degree of correlation R. The level shift in x_t is $\mu_x \in \{0, 0.5, 1, 5\}$ and, for each value of μ_x, $\beta_y \in \{-2.0, -1.8, -1.6, ..., 2.0\}$. $\mu_y = 0$ in all cases. The sample size $T = 100$, three positions of the break in x_t are considered with $T_0^x \in \{30, 50, 70\}$, $T_0 = 50$, and $R \in \{0, 0.3, 0.5, 0.7, 0.9\}$. Again, to avoid issues related to the determination of the lead and lag orders, all tests are constructed for $k_i = 0$ for $i = 1, 2, 3$. Empirical power is evaluated at $\rho = 0.9$. The results are based on 1,000 replications, with the same set of random errors used across values of μ_x and β_y, and reported only for case 5-C and $T_0^x = 50$. Results for other values of T_0^x are qualitatively similar and, hence, not presented.

Figure 5 presents the empirical size of the tests for the case when a break in the trend function of x_t is present. For small values of μ_x the tests exhibit modest size distortions for all values of β_y. This result is important because, in empirical applications, small level shifts may be hard to identify. When μ_x is large ($\mu_x = 5$) the tests show liberal size distortions, particularly large for small values of $|\beta_y|$. For large values of R^2, when a break in the trend function of x_t is present and under the null hypothesis, the tests reject too often.

[FIGURE 5 ABOUT HERE]

Figure 6 presents the empirical power of the tests. Results show that for small values of μ_x the tests still exhibit important power gains as R^2 increases, with a small power loss for very large values of R^2. When μ_x is large, however, the covariates behave like a non-stationary variable and the tests show a very important power loss.

[FIGURE 6 ABOUT HERE]

In sum, when level shifts in the trend function of the covariates are not accounted for, the CADF-GLS($\hat{\delta}$) test can exhibit liberal size distortions and low power. These distortions can be even more important when the covariates exhibit a break in the slope. Nevertheless, power improvements are still available as long as the break in the trend function of the covariates is of small magnitude.

5 Conclusion

This paper proposed a test of the unit root hypothesis that allows for a structural break in the trend function and uses correlated stationary covariates to improve power. Three structural break models were considered and, as it is standard in the literature, the test endogenously determines the break date. Similar to other unit root tests in the recent literature, the proposed test is based on GLS detrended data. The statistic is simple to construct and the test can be seen as an extension of the tests proposed in Pesavento (2006) and Perron and Rodríguez (2003). The asymptotic local power functions of the test were approximated numerically and show that large power gains are available. The finite sample properties were analyzed using Monte Carlo simulation and results show that the test exhibits small size distortions and power that can be far beyond what is achievable by univariate tests. With good covariates, the proposed test should dominate other tests available.

A Appendix

A.1 Proof of Theorem 2

The proof of Theorem 1 is omitted as it is basically the same as that of Theorem 2. The following lemmas, provided without proof, present some auxiliary results that will be used later.

Lemma A.1. *Consider the model generated by equations* (1) *to* (3) *and under the assumptions of the model, with* $\rho = 1 + c/T$, *then as* $T \to \infty$ *we have*

$$\omega_{y.x}^{-1/2} T^{-1/2} u_{y\lfloor Tr \rfloor} \Rightarrow J_{xyc}(r),$$

where $J_{xyc}(r)$ *is an Ornstein-Uhlenbeck process such that*

$$J_{xyc}(r) = W_{xy}(r) + c \int_0^r e^{(r-s)c} W_{xy}(s) ds,$$

with $W_{xy}(r) = Q^{1/2} W_x(r) + W_y(r)$, *where* $W_x(r)$ *and* $W_y(r)$ *are univariate independent standard Brownian motions,* $Q = R^2/(1 - R^2)$, *and* $\omega_{y.x} = \omega_{yy} - \omega_{yx}\Omega_{xx}^{-1}\omega_{xy}$. $Q^{1/2} W_x(r) = \bar{\gamma}' \tilde{W}_x(r)$, $\bar{\gamma}' = \omega_{y.x}^{-1/2} \omega_{yx} \Omega_{xx}^{-1/2}$ *so that* $\bar{\gamma}'\bar{\gamma} = R^2/(1 - R^2)$, *and* $W(r) = \left[\tilde{W}_x(r)', W_y(r)\right]'$.

The proof of Lemma A.1 follows from results in Elliott and Jansson (2003) and Pesavento (2006). Lemma A.2 presents limiting results for the GLS estimates of the coefficients in the trend function obtained from (8). The proof of Lemma A.2 follows from results in Perron and Rodríguez (2003).

Lemma A.2. *Assume* y_t *is generated by* (2) *with* $\rho = 1 + c/T$. *Let* $\tilde{\psi}_y$ *be the GLS estimates of the coefficients in the trend function given by* (8) *using* $\bar{\rho} = 1 + \bar{c}/T$.

(i) When $z_{yt}(\delta_0)$ *is given by* (4), *then as* $T \to \infty$ *we have*

$\tilde{\mu}_{y1} - \mu_{y1} \Rightarrow u_1,$

$$\tilde{\mu}_{y2} - \mu_{y2} \Rightarrow lim_{T\to\infty} u_{y(T_0+1)}(\bar{\rho}) \equiv u^*.$$

(ii) When $z_{yt}(\delta_0)$ *is given by* (5)*, then as* $T \to \infty$ *we have*

$$\tilde{\mu}_{y1} - \mu_{y1} \Rightarrow u_1,$$
$$\tilde{\mu}_{y2} - \mu_{y2} \Rightarrow lim_{T\to\infty} u_{y(T_0+1)}(\bar{\rho}) \equiv u^*,$$
$$T^{1/2}(\tilde{\beta}_{y1} - \beta_{y1}) \Rightarrow \omega_{y.x}^{1/2}\ b_1/a_1 \equiv \omega_{y.x}^{1/2}\ V_{c\bar{c}}.$$

(iii) When $z_{yt}(\delta_0)$ *is given by* (6)*, then as* $T \to \infty$ *we have*

$$\tilde{\mu}_{y1} - \mu_{y1} \Rightarrow u_1,$$
$$T^{1/2}(\tilde{\beta}_{y1} - \beta_{y1}) \Rightarrow \omega_{y.x}^{1/2}(\lambda_1 b_1 + \lambda_2 b_2) \equiv \omega_{y.x}^{1/2}\ V_{c\bar{c}}^{(1)},$$
$$T^{1/2}(\tilde{\beta}_{y2} - \beta_{y2}) \Rightarrow \omega_{y.x}^{1/2}(\lambda_2 b_1 + \lambda_3 b_2) \equiv \omega_{y.x}^{1/2}\ V_{c\bar{c}}^{(2)}.$$

(iv) When $z_{yt}(\delta_0)$ *is given by* (7)*, then as* $T \to \infty$ *we have*

$$\tilde{\mu}_{y1} - \mu_{y1} \Rightarrow u_1,$$
$$\tilde{\mu}_{y2} - \mu_{y2} \Rightarrow lim_{T\to\infty} u_{y(T_0+1)}(\bar{\rho}) \equiv u^*,$$
$$T^{1/2}(\tilde{\beta}_{y1} - \beta_{y1}) \Rightarrow \omega_{y.x}^{1/2}(\lambda_1 b_1 + \lambda_2 b_2) \equiv \omega_{y.x}^{1/2}\ V_{c\bar{c}}^{(1)},$$
$$T^{1/2}(\tilde{\beta}_{y2} - \beta_{y2}) \Rightarrow \omega_{y.x}^{1/2}(\lambda_2 b_1 + \lambda_3 b_2) \equiv \omega_{y.x}^{1/2}\ V_{c\bar{c}}^{(2)}.$$

Lemma A.3 presents standard results for OLS estimated trend functions. See, e.g., Lütkepohl (2005, Proposition C.18 on page 705).

Lemma A.3. *Assume* x_t *is generated by* (1)*. Let* $\hat{\psi}_x$ *be the OLS estimates of the coefficients in the trend function.*

(i) When $z_{xt} = \{1\}$*, then as* $T \to \infty$ *we have*

$$T^{1/2}(\hat{\mu}_x - \mu_x) \Rightarrow \Omega_{xx}^{1/2}\tilde{W}_x(1).$$

(ii) When $z_{xt} = \{1, t\}$*, then as* $T \to \infty$ *we have*

$$T^{1/2}(\hat{\mu}_x - \mu_x) \Rightarrow \Omega_{xx}^{1/2}\left\{4\tilde{W}_x(1) - 6\left[\tilde{W}_x(1) - \int_0^1 \tilde{W}_x(r)dr\right]\right\},$$
$$T^{3/2}(\hat{\beta}_x - \beta_x) \Rightarrow \Omega_{xx}^{1/2}\left\{-6\tilde{W}_x(1) + 12\left[\tilde{W}_x(1) - \int_0^1 \tilde{W}_x(r)dr\right]\right\}.$$

The derivation of the regression equation follows the results in Said and Dickey (1984), Saikkonen (1991), and in particular Pesavento (2006). As a consequence, only a sketch of the proof is presented. From Brillinger (1975, page 296) we know that under Assumptions 1-4 we can write

$$u_{y,t}(\rho) = \sum_{j=-\infty}^{+\infty} \tilde{\pi}'_{xj} u_{x,t-j} + \eta_t, \tag{21}$$

with $\sum_{j=-\infty}^{+\infty} \left\| \tilde{\pi}'_{xj} \right\| < \infty$, $E(u_{x,t}\eta'_{t+k}) = 0$ for any $k = 0, \pm 1, \pm 2, ...$, and $2\pi f_{\eta\eta}(0) = \omega_{yy} - \omega_{yx}\Omega_{xx}^{-1}\omega_{xy} = \omega_{y.x}$ where $f_{\eta\eta}(0)$ is the spectral density of η_t at frequency zero.

For case 5-C we have $y_t^d = y_t - \tilde{\mu}_{y1} - \tilde{\mu}_{y2} 1(t > T_0) - \tilde{\beta}_{y1} t - \tilde{\beta}_{y2} 1(t > T_0)(t - T_0)$ or

$$\begin{aligned} y_t^d = {} & u_{yt} - (\tilde{\mu}_{y1} - \mu_{y1}) - (\tilde{\mu}_{y2} - \mu_{y2}) 1(t > T_0) \\ & - (\tilde{\beta}_{y1} - \beta_{y1}) t - (\tilde{\beta}_{y2} - \beta_{y2}) 1(t > T_0)(t - T_0), \end{aligned} \tag{22}$$

and after some algebra we get

$$\begin{aligned} \Delta y_t^d = {} & (\rho - 1) y_{t-1}^d - (1 - \rho)(\tilde{\mu}_{y1} - \mu_{y1}) - (1 - \rho)(\tilde{\mu}_{y2} - \mu_{y2}) 1(t > T_0) \\ & - (1 - \rho L)(\tilde{\beta}_{y1} - \beta_{y1}) t - (1 - \rho L)(\tilde{\beta}_{y2} - \beta_{y2}) 1(t > T_0)(t - T_0) \\ & + u_{yt}. \end{aligned} \tag{23}$$

Plugging (21) into (23) and noting that $u_{xt}^d = x_t^d + (\hat{\mu}_x - \mu_x) + (\hat{\beta}_x - \beta_x) t$ we get

$$\Delta y_t^d = \alpha y_{t-1}^d + \sum_{j=-\infty}^{+\infty} \tilde{\pi}'_{xj} x_{t-j}^d + \bar{\eta}_t, \tag{24}$$

where $\alpha = \rho - 1$, and

$$\bar{\eta}_t = \tilde{\pi}_x(1)'(\hat{\mu}_x - \mu_x) + \sum_{j=-\infty}^{+\infty} \tilde{\pi}'_{xj}(\hat{\beta}_x - \beta_x)(t-j)$$
$$- (1-\rho)(\tilde{\mu}_{y1} - \mu_{y1}) - (1-\rho)(\tilde{\mu}_{y2} - \mu_{y2})1(t > T_0)$$
$$- (1-\rho L)(\tilde{\beta}_{y1} - \beta_{y1})t - (1-\rho L)(\tilde{\beta}_{y2} - \beta_{y2})1(t > T_0)(t - T_0)$$
$$+ \eta_t,$$

with $\tilde{\pi}_x(1)' = \sum_{j=-\infty}^{+\infty} \tilde{\pi}'_{xj} = \omega_{yx}\Omega_{xx}^{-1}$. Note that η_t is uncorrelated at all leads and lags with x_t^d but it may still be serially correlated. Assume $\Phi(L)\eta_t = \tilde{e}_t$, where $\tilde{e}_t$ is white noise. Then

$$\Delta y_t^d = \phi y_{t-1}^d + \sum_{j=-\infty}^{+\infty} \pi'_{xj} x_{t-j}^d + \sum_{j=1}^{+\infty} \pi_{yj} \Delta y_{t-j}^d + e_t, \tag{25}$$

where $\phi = \Phi(1)(\rho - 1)$, and $e_t = \Phi(L)\bar{\eta}_t$ or

$$e_t = \Phi(1)\tilde{\pi}_x(1)'(\hat{\mu}_x - \mu_x) + \Phi(L) \sum_{j=-\infty}^{+\infty} \tilde{\pi}'_{xj}(\hat{\beta}_x - \beta_x)(t-j)$$
$$- \Phi(1)(1-\rho)(\tilde{\mu}_{y1} - \mu_{y1}) - \Phi(1)(1-\rho)(\tilde{\mu}_{y2} - \mu_{y2})1(t > T_0)$$
$$- \Phi(L)(1-\rho L)(\tilde{\beta}_{y1} - \beta_{y1})t - \Phi(L)(1-\rho L)(\tilde{\beta}_{y2} - \beta_{y2})1(t > T_0)(t - T_0)$$
$$+ \tilde{e}_t. \tag{26}$$

Since the sequence $\{\tilde{\pi}_{xj}\}$ is absolute summable, we can approximate (25) with

$$\Delta y_t^d = \phi y_{t-1}^d + \sum_{j=-k_1}^{k_2} \pi'_{xj} x_{t-j}^d + \sum_{j=1}^{k_3} \pi_{yj} \Delta y_{t-j}^d + e_{tk_1k_2k_3}. \tag{27}$$

For simplicity, and without loss of generality, assume $k_1 = k_2 = k_3 = k$. Then

$$\Delta y_t^d = \phi y_{t-1}^d + \sum_{j=-k}^{k} \pi'_{xj} x_{t-j}^d + \sum_{j=1}^{k} \pi_{yj} \Delta y_{t-j}^d + e_{tk}, \tag{28}$$

where

$$e_{tk} = e_t + \sum_{|j|>k} \pi'_{xj} x^d_{t-j} + \sum_{j>k} \pi_{yj} \Delta y^d_{t-j}.$$

Under Assumption 5 and following the results in Pesavento (2006) we obtain

$$(T-2k)(\hat{\phi}-\phi) = \left[(T-2k)^{-2} \sum_{t=k+1}^{T-k} y^{d\ 2}_{t-1}\right]^{-1} \left[(T-2k)^{-1} \sum_{t=k+1}^{T-k} y^d_{t-1} e_t\right] + o_p(1).$$

The derivation of this result follows exactly from Pesavento (2006) and, hence, is omitted. Alternatively, we can assume $\pi'_{xj} = 0$ for $|j| > k$ and $\pi_{yj} = 0$ for $j > k$ which implies assuming that the regression is correctly specified ($e_{tk} = e_t$). This approach is used in Stock (1991) and Elliott et al. (1996) and yields the same results.

From (22) and using the results in Lemmas A.1 and A.2 we have that

$$T^{-1/2} y^d_{\lfloor Tr \rfloor} \Rightarrow \omega^{1/2}_{y.x} \left[J_{xyc}(r) - rV^{(1)}_{c\bar{c}} - (r-\delta_0)V^{(2)}_{c\bar{c}} 1(r > \delta_0)\right].$$

Let $J^d_{xyc}(r, \delta_0) = J_{xyc}(r) - rV^{(1)}_{c\bar{c}} - (r-\delta_0)V^{(2)}_{c\bar{c}} 1(r > \delta_0)$ so that

$$T^{-1/2} y^d_{\lfloor Tr \rfloor} \Rightarrow \omega^{1/2}_{y.x} J^d_{xyc}(r, \delta_0).$$

Then, by the continuous mapping theorem and arguments as in Phillips (1987a,b) we have

$$T^{-3/2} \sum_{t=1}^{T} y^d_t \Rightarrow \omega^{1/2}_{y.x} \int_0^1 J^d_{xyc}(r, \delta_0) dr,$$

and

$$T^{-2} \sum_{t=1}^{T} y^{d\ 2}_t \Rightarrow \omega_{y.x} \int_0^1 J^d_{xyc}(r, \delta_0)^2 dr.$$

Finally,

$$(T-2k)^{-2} \sum_{t=k+1}^{T-k} y^{d\ 2}_{t-1} \Rightarrow \omega_{y.x} \int_0^1 J^d_{xyc}(r, \delta_0)^2 dr \tag{29}$$

if $k/T \to 0$ as $T \to \infty$.

Consider now the term $(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d e_t$. e_t is given by (26) and using Lemma A.3, the continuous mapping theorem, and noting that $\bar{\gamma}' = \omega_{y.x}^{-1/2}\omega_{yx}\Omega_{xx}^{-1/2}$ and $\bar{\gamma}'\tilde{W}_x(r) = \sqrt{\frac{R^2}{1-R^2}}W_x(r) = Q^{1/2}W_x(r)$ we have

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d \Phi(1)\tilde{\pi}_x(1)'(\hat{\mu}_x - \mu_x) \Rightarrow \\ \Phi(1)\omega_{y.x}Q^{1/2}\left[-2W_x(1) + 6\int_0^1 W_x(r)\right]\int_0^1 J_{xyc}^d$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d \Phi(L)\sum_{j=-\infty}^{\infty}\tilde{\pi}_{xj}'(\hat{\beta}_x - \beta_x)(t-j) \Rightarrow \\ \Phi(1)\omega_{y.x}Q^{1/2}\left[6W_x(1) - 12\int_0^1 W_x(r)\right]\int_0^1 rJ_{xyc}^d$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d \Phi(1)(1-\rho)(\tilde{\mu}_{y1} - \mu_{y1}) \Rightarrow 0$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d \Phi(1)(1-\rho)(\tilde{\mu}_{y2} - \mu_{y2})1(t > T_0) \Rightarrow 0$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d \Phi(L)(1-\rho L)(\tilde{\beta}_{y1} - \beta_{y1})t \Rightarrow \\ \Phi(1)\omega_{y.x}V_{c\bar{c}}^{(1)}\left[\int_0^1 J_{xyc}^d - c\int_0^1 rJ_{xyc}^d\right]$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d \Phi(L)(1-\rho L)(\tilde{\beta}_{y2} - \beta_{y2})1(t > T_0)(t - T_0) \Rightarrow \\ \Phi(1)\omega_{y.x}V_{c\bar{c}}^{(2)}\left[\int_{\delta_0}^1 J_{xyc}^d - c\int_{\delta_0}^1 (r-\delta_0)J_{xyc}^d\right]$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d \tilde{e}_t \Rightarrow \Phi(1)\omega_{y.x}\int_0^1 J_{xyc}^d dW_y.$$

Collecting terms we have

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^d e_t \Rightarrow \Phi(1)\omega_{y.x}\left[\int_0^1 J_{xyc}^d dW_y + \Lambda_{c\bar{c}}\right],$$

where

$$\Lambda_{c\bar{c}} = \Lambda_x - V_{c\bar{c}}^{(1)} \left[\int_0^1 J_{xyc}^d - c \int_0^1 r J_{xyc}^d \right] - V_{c\bar{c}}^{(2)} \left[\int_{\delta_0}^1 J_{xyc}^d - c \int_{\delta_0}^1 (r - \delta_0) J_{xyc}^d \right],$$

with

$$\Lambda_x = Q^{1/2} \left[-2W_x(1) + 6 \int_0^1 W_x \right] \int_0^1 J_{xyc}^d + Q^{1/2} \left[6W_x(1) - 12 \int_0^1 W_x \right] \int_0^1 r J_{xyc}^d$$

for case 5-C, and

$$\Lambda_x = Q^{1/2} W_x(1) \int_0^1 J_{xyc}^d$$

for case 4-C. Finally,

$$(T - 2k)^{-1} \sum_{t=k+1}^{T-k} y_{t-1}^d e_t \Rightarrow \Phi(1) \omega_{y.x} \left[\int_0^1 J_{xyc}^d dW_y + \Lambda_{c\bar{c}} \right] \tag{30}$$

if $k/T \to 0$ as $T \to \infty$. Then

$$(T - 2k)(\hat{\phi} - \phi) \Rightarrow \Phi(1) \left[\int_0^1 J_{xyc}^{d\,2} \right]^{-1} \left[\int_0^1 J_{xyc}^d dW_y + \Lambda_{c\bar{c}} \right]. \tag{31}$$

Let $s^2 = (T - 2k)^{-1} \sum_{t=k+1}^{T-k} \hat{e}_{tk}^2$ where $s \xrightarrow{p} \omega_{y.x}^{1/2} \Phi(1)$. Then

$$(T - 2k) SE(\hat{\phi}) = (T - 2k)\ s \left[\sum_{t=k+1}^{T-k} y_{t-1}^{d\,2} \right]^{-1/2} \Rightarrow \Phi(1) \left[\int_0^1 J_{xyc}^{d\,2} \right]^{-1/2}. \tag{32}$$

The CADF-GLS(δ_0) test statistic is

$$t_{\hat{\phi}}(\delta_0) = \frac{\hat{\phi}}{SE(\hat{\phi})} = \frac{(T - 2k)\phi}{(T - 2k) SE(\hat{\phi})} + \frac{(T - 2k)(\hat{\phi} - \phi)}{(T - 2k) SE(\hat{\phi})}.$$

From (31) and (32), and noting that $(T - 2k)\phi = (T - 2k)(\rho - 1)\Phi(1) \xrightarrow{p} c\ \Phi(1)$ as

$T \to \infty$, we have

$$t_{\hat{\phi}}(\delta_0) \Rightarrow \left(\int_0^1 J_{xyc}^{d\ 2}\right)^{1/2} \left[\left(\int_0^1 J_{xyc}^{d\ 2}\right)^{-1} \left(\int_0^1 J_{xyc}^{d} dW_y + \Lambda_{c\bar{c}}\right) + c\right].$$

A.2 Proof of Remark 4

From Lemma A.2 it is straightforward to obtain the limiting distribution of $t_{\hat{\beta}_{y2}}(\delta)$ when $\beta_{y2} = 0$. Rewriting (8) in matrix notation, we obtain $\bar{\Delta} y = \psi_y' \bar{\Delta} z_y + \bar{\Delta} u_y$ where $\bar{\Delta} = 1 - \bar{\rho} L$. Then, from standard OLS results we know that the $VAR(\tilde{\psi}_y) = s^2 \left[\bar{\Delta} z_y' \bar{\Delta} z_y\right]^{-1}$. Define the scaling matrix $D_T = diag\{1, 1, T^{1/2}, T^{1/2}\}$, then from Lemma A.2 we have

$$D_T \left[\bar{\Delta} z_y' \bar{\Delta} z_y\right]^{-1} D_T \Rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & a_1 & a_2 \\ 0 & 0 & a_2 & a_3 \end{pmatrix}^{-1} \tag{33}$$

with terms defined in Theorem 2. From Lemma A.2, (33), and noting that $s \xrightarrow{p} \omega_{y.x}^{1/2}$, we have

$$t_{\hat{\beta}_{y2}}(\delta) \Rightarrow V_{c\bar{c}}^{(2)} \lambda_3^{-1/2} \tag{34}$$

with terms again defined in Theorem 2.

Once weak convergence for a fixed δ is established (Theorem 2), the proof of Remark 4 follows directly from (34), the results of Zivot and Andrews (1992), and the continuous mapping theorem. See Banerjee et al. (1992) and Vogelsang and Perron (1998) for a more detailed discussion.

Notes

[1] The tests also show liberal size distortions when the break is of small size, with size distortions increasing with the sample size. Similar unit root tests proposed in Kim and Perron (2009) also show large size distortions when the break is small.

[2] See figures 11 to 13 in Carrion-i-Silvestre et al. (2009) and 6 to 8 in Harris et al. (2009).

[3] See Caporale and Pittis (1999) for a complete discussion of the effects introduced by adding stationary covariates to the Dickey-Fuller regression equation, Elliott and Jansson (2003) who derive the power envelope for the case where constants and/or time trends are included in the regression and propose a point-optimal unit root test with stationary covariates that has maximal power against a local alternative, and Galvao (2009) who proposes a quantile unit root test using covariates.

[4] Independent work by Liu and Rodríguez (2007) considers a similar problem. Their analysis extends the point-optimal unit root tests with stationary covariates proposed by Elliott and Jansson (2003) to the breaking trend case. Since the tests developed here are of the Dickey-Fuller type, our papers are therefore complementary.

[5]The test considered here allows only for one break in the trend function. The results in this paper can be extended to the case where two breaks are allowed in the trend function, as Lumsdaine and Papell (1997) extend Zivot and Andrews (1992), or to the case where multiple structural breaks are allowed adopting the approach developed in Carrion-i-Silvestre et al. (2009).

[6] Note that R^2 in this paper corresponds to $(1-\rho^2)$ in Hansen (1995).

[7]Another estimator, suggested by Zivot and Andrews (1992), uses the break date that gives the least favorable result for the null hypothesis, i.e. $\hat{\delta}$ is chosen to minimize the t-statistic for testing $\phi = 0$ in (9). The resulting estimator is then $\hat{\delta} = \arg\min_{\delta\in[0,1]} t_{\hat{\phi}}(\delta)$. Preliminary results show that the use of the this estimator yields less powerful tests and, as a consequence, the results are not presented.

[8]Good covariates satisfy two conditions: (1) Need to be stationary variables. (2) Need to be correlated with the quasi-differenced y_t at the zero frequency. So any covariates that exhibit contemporaneous, leading, or lagging correlation would work. In economics, for example, any theory that involves y_t could be used to identify reasonable covariates.

[9]Juhl and Xiao (2003) show that the point-optimal unit root tests with stationary covariates of Elliott and Jansson (2003) have power functions that are tangent to the asymptotic Gaussian power envelope at approximately 0.75 instead of 0.50, the value at which the power functions are tangent in the case of the univariate tests of Elliott et al. (1996). This result suggests that other values of $\bar{c}$ could potentially yield small power improvements.

[10]When $T = 100$, the local alternative $\bar{c} = -5$ corresponds to an autoregressive root of 0.95.

Acknowledgements

I would like to thank two anonymous referees, Maurice Priestley, Eric Zivot, Charles Nelson, Chang-Jin Kim, Richard Startz, and Drew Creal for helpful comments and Bruce Hansen and Gabriel Rodríguez for providing their code. I would also like to acknowledge financial support from the Buechel Memorial Fellowship and Grover and Creta Ensley Fellowship, which funded part of this research while I was a graduate student at the University of Washington.

References

Andrews, D.W.K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59(3), 817-858.

Banerjee, A., Lumsdaine, R., and Stock, J. (1992): "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence", *Journal of Business & Economic Statistics*, 10(3), 271-287.

Blanchard, O., and Quah, D. (1989): "The dynamic effect of aggregate demand and aggregate supply disturbances", *American Economic Review*, 79, 655-673.

Brillinger, D.R. (1975): *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, Inc..

Caporale, G.M., and Pittis, N. (1999): "Unit root testing using covariates: some theory and evidence", *Oxford Bulletin of Economics and Satistics*, 61, 583-595.

Carrion-i-Silvestre, J.L., Kim, D., and Perron, P. (2009): "GLS-based unit root tests with multiple structural breaks under the null and the alternative hypotheses", *Econometric Theory*, 25, 1754-1792.

Christiano, L.J. (1992): “Searching for a break in GNP”, *Journal of Business & Economic Statistics*, 10(3), 237-250.

Dickey, D.A., and Fuller, W.A. (1979): “Distribution of the estimator for autoregressive time series with a unit root”, *Journal of the American Statistical Association*, 74, 427-431.

Elliott, G., and Jansson, M. (2003): “Testing for unit roots with stationary covariates”, *Journal of Econometrics*, 115, 75-89.

Elliott, G., Rothenberg, T.J., and Stock, J. (1996): “Efficient tests for an autoregressive unit root”, *Econometrica*, 64, 813-836.

Fossati, S. (2012): “Covariate unit root tests with good size and power”, *Computational Statistics and Data Analysis*, 56, 3070-3079.

Galvao, A., (2009): “Unit root quantile autoregression testing using covariates”, *Journal of Econometrics*, 152, 165-178.

Hansen, B.E. (1995): “Rethinking the univariate approach to unit root testing: using covariates to increase power”, *Econometric Theory*, 11, 1148-1172.

Harris, D., Harvey, D.I., Leybourne, S.J., and Taylor, A.M.R. (2009): “Testing for a unit root in the presence of a possible break in the trend”, *Econometric Theory*, 25, 1545-1588.

Juhl, T., and Xiao, Z. (2003): “Power functions and envelopes for unit root tests”, *Econometric Theory*, 19, 240-253.

Kim, D., and Perron, P. (2009): “Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses”, *Journal of Econometrics*, 148, 1-13.

Liu, H., and Rodríguez, G. (2007): "Point optimal unit root tests using GLS detrended data and covariates in structural change models", Working Paper.

Lumsdaine, R.L., and Papell, D.H. (1997): "Multiple trend breaks and the unit root hypothesis", *The Review of Economics and Statistics*, 79, 212-218.

Lütkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis.* Springer.

Ng, S., and Perron, P. (1995): "Unit root tests in arma models with data-dependent methods for selection of the truncation lag", *Journal of the American Statistical Association*, 90, 268-281.

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

Papell, D.H., and Prodan, R. (2004): "The uncertain unit root in U.S. real GDP: evidence with restricted and unrestricted structural change", *Journal of Money, Credit and Banking*, 36, 423-427.

Papell, D.H., and Prodan, R. (2007): "Restricted structural change and the unit root hypothesis", *Economic Inquiry*, 45, 834-853.

Perron, P. (1989): "The great crash, the oil price shock and the unit root hypothesis", *Econometrica*, 57, 1361-1401.

Perron, P. (1997): "Further evidence on breaking trend functions in macroeconomic variables", *Journal of Econometrics*, 80, 355-385.

Perron, P. (2006): "Dealing with structural breaks", in Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory, K. Patterson and T.C. Mills (eds.), Palgrave Macmillan, 2006, 278-352.

Perron, P., and Rodríguez, G. (2003): "GLS detrending, efficient unit root tests and structural change", *Journal of Econometrics*, 115, 1-27.

Pesavento, E. (2006): "Near-optimal unit root tests with stationary covariates with better finite sample size", Working Paper, Emory University.

Phillips, P.C.B. (1987a): "Toward and unified asymptotic theory for autoregression". *Biometrika*, 74, 535-547.

Phillips, P.C.B. (1987b): "Time series regression with a unit root", *Econometrica*, 55, 277-302.

Rodríguez, G. (2007): "Finite sample behaviour of the level shift model using quasi-differenced data", *Journal of Statistical Computation and Simulation*, 77(10), 889-905.

Rudebusch, G.D. (1993): "The uncertain unit root in real GNP", *The American Economic Review*, 83, 264-272.

Said, S., and Dickey, D.A. (1984): "Testing for unit roots in autoregressive moving-average models of unknown order", *Biometrika*, 71, 599-607.

Saikkonen, P. (1991): "Asymptotically efficient estimation of cointegration regressions", *Econometric Theory*, 7, 1-21.

Stock, J.H. (1991): "Confidence intervals for the largest autoregressive root in U.S. economic time series", *Journal of Monetary Economics*, 28(3), 435-460.

Vogelsang, T., and Perron, P. (1998): "Additional tests for a unit root allowing for a break in the trend function at an unknown time", *International Economic Review*, 39, 1073-1100.

Zivot, E., and 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*, 10(3), 251-270.

Table 1: Critical values for Model A.

Case	Size	T	R^2 0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2-A	0.01	100	-3.03	-3.01	-3.03	-3.02	-3.01	-2.95	-2.99	-2.95	-2.89	-2.84
		250	-2.83	-2.83	-2.83	-2.82	-2.82	-2.81	-2.81	-2.78	-2.74	-2.69
		∞	-2.55	-2.55	-2.57	-2.54	-2.54	-2.55	-2.54	-2.50	-2.47	-2.45
	0.05	100	-2.45	-2.44	-2.44	-2.41	-2.40	-2.35	-2.34	-2.31	-2.25	-2.20
		250	-2.26	-2.25	-2.24	-2.21	-2.20	-2.20	-2.17	-2.13	-2.10	-2.04
		∞	-1.94	-1.94	-1.93	-1.93	-1.91	-1.89	-1.89	-1.84	-1.82	-1.76
	0.10	100	-2.16	-2.15	-2.14	-2.10	-2.08	-2.04	-2.01	-1.97	-1.91	-1.86
		250	-1.96	-1.95	-1.92	-1.90	-1.88	-1.86	-1.83	-1.80	-1.76	-1.70
		∞	-1.62	-1.61	-1.60	-1.59	-1.57	-1.54	-1.53	-1.49	-1.46	-1.41
3-A	0.01	100	-3.03	-2.99	-2.99	-2.95	-2.92	-2.83	-2.83	-2.73	-2.66	-2.57
		250	-2.82	-2.82	-2.78	-2.75	-2.71	-2.66	-2.64	-2.56	-2.48	-2.42
		∞	-2.55	-2.53	-2.53	-2.47	-2.43	-2.41	-2.35	-2.26	-2.23	-2.16
	0.05	100	-2.45	-2.42	-2.38	-2.33	-2.29	-2.23	-2.16	-2.07	-1.99	-1.92
		250	-2.26	-2.22	-2.18	-2.12	-2.08	-2.05	-1.97	-1.89	-1.80	-1.69
		∞	-1.94	-1.91	-1.87	-1.82	-1.77	-1.71	-1.66	-1.57	-1.49	-1.40
	0.10	100	-2.16	-2.12	-2.08	-2.02	-1.97	-1.90	-1.83	-1.73	-1.64	-1.54
		250	-1.96	-1.92	-1.87	-1.81	-1.75	-1.69	-1.61	-1.54	-1.43	-1.30
		∞	-1.62	-1.58	-1.53	-1.48	-1.42	-1.35	-1.28	-1.19	-1.08	-0.95
4-A	0.01	100	-3.86	-3.79	-3.76	-3.71	-3.66	-3.57	-3.50	-3.43	-3.26	-3.13
		250	-3.59	-3.57	-3.54	-3.47	-3.45	-3.37	-3.31	-3.23	-3.08	-2.97
		∞	-3.42	-3.38	-3.34	-3.28	-3.24	-3.19	-3.10	-3.03	-2.94	-2.92
	0.05	100	-3.26	-3.22	-3.17	-3.10	-3.05	-2.96	-2.87	-2.77	-2.64	-2.50
		250	-3.06	-3.01	-2.96	-2.90	-2.85	-2.77	-2.70	-2.60	-2.48	-2.36
		∞	-2.85	-2.80	-2.76	-2.69	-2.64	-2.58	-2.51	-2.41	-2.35	-2.29
	0.10	100	-2.98	-2.93	-2.87	-2.80	-2.73	-2.64	-2.55	-2.43	-2.29	-2.16
		250	-2.79	-2.72	-2.66	-2.60	-2.54	-2.46	-2.37	-2.28	-2.16	-2.03
		∞	-2.56	-2.51	-2.46	-2.40	-2.33	-2.27	-2.20	-2.09	-2.03	-1.97
5-A	0.01	100	-3.86	-3.78	-3.74	-3.68	-3.63	-3.52	-3.43	-3.35	-3.17	-3.04
		250	-3.59	-3.56	-3.52	-3.43	-3.39	-3.30	-3.24	-3.14	-2.99	-2.85
		∞	-3.42	-3.37	-3.31	-3.24	-3.19	-3.11	-3.02	-2.92	-2.82	-2.74
	0.05	100	-3.26	-3.21	-3.15	-3.06	-2.99	-2.91	-2.80	-2.67	-2.52	-2.36
		250	-3.06	-3.00	-2.93	-2.87	-2.79	-2.69	-2.61	-2.48	-2.33	-2.16
		∞	-2.85	-2.78	-2.73	-2.65	-2.58	-2.50	-2.40	-2.27	-2.15	-2.01
	0.10	100	-2.98	-2.91	-2.84	-2.76	-2.68	-2.58	-2.46	-2.32	-2.17	-2.00
		250	-2.78	-2.71	-2.63	-2.55	-2.48	-2.37	-2.26	-2.14	-1.99	-1.80
		∞	-2.56	-2.49	-2.42	-2.34	-2.26	-2.17	-2.06	-1.92	-1.78	-1.62

Table 2: Critical values for Model B.

Case	Size	T	R^2 0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
4-B	0.01	100	-4.66	-4.57	-4.47	-4.39	-4.26	-4.13	-4.03	-3.88	-3.66	-3.52
		250	-4.45	-4.39	-4.33	-4.22	-4.14	-4.04	-3.89	-3.77	-3.67	-3.67
		∞	-4.32	-4.26	-4.19	-4.09	-4.01	-3.93	-3.82	-3.74	-3.69	-3.76
	0.05	100	-4.07	-3.96	-3.86	-3.76	-3.64	-3.49	-3.35	-3.17	-2.95	-2.73
		250	-3.92	-3.83	-3.74	-3.63	-3.53	-3.41	-3.26	-3.11	-2.95	-2.80
		∞	-3.80	-3.72	-3.63	-3.53	-3.43	-3.32	-3.19	-3.06	-2.93	-2.84
	0.10	100	-3.77	-3.66	-3.56	-3.44	-3.31	-3.17	-3.01	-2.82	-2.60	-2.33
		250	-3.64	-3.54	-3.44	-3.33	-3.22	-3.07	-2.92	-2.75	-2.57	-2.37
		∞	-3.53	-3.44	-3.34	-3.23	-3.13	-3.00	-2.86	-2.69	-2.54	-2.39
5-B	0.01	100	-4.65	-4.56	-4.46	-4.38	-4.25	-4.11	-4.00	-3.86	-3.62	-3.46
		250	-4.45	-4.39	-4.33	-4.20	-4.12	-4.00	-3.86	-3.71	-3.57	-3.46
		∞	-4.32	-4.26	-4.18	-4.08	-3.98	-3.90	-3.77	-3.65	-3.52	-3.47
	0.05	100	-4.07	-3.96	-3.86	-3.75	-3.63	-3.48	-3.34	-3.15	-2.93	-2.68
		250	-3.92	-3.83	-3.73	-3.62	-3.52	-3.39	-3.24	-3.08	-2.88	-2.69
		∞	-3.80	-3.72	-3.63	-3.52	-3.41	-3.29	-3.16	-3.00	-2.84	-2.70
	0.10	100	-3.77	-3.66	-3.55	-3.43	-3.31	-3.16	-3.00	-2.80	-2.57	-2.30
		250	-3.64	-3.54	-3.44	-3.32	-3.21	-3.06	-2.91	-2.73	-2.53	-2.29
		∞	-3.53	-3.44	-3.33	-3.23	-3.11	-2.98	-2.83	-2.66	-2.49	-2.30

Table 3: Critical values for Model C.

Case	Size	T	R^2									
			0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
4-C	0.01	100	-4.63	-4.52	-4.45	-4.36	-4.26	-4.14	-4.04	-3.91	-3.75	-3.54
		250	-4.38	-4.32	-4.28	-4.20	-4.11	-4.01	-3.90	-3.81	-3.71	-3.70
		∞	-4.32	-4.26	-4.19	-4.09	-4.01	-3.93	-3.82	-3.74	-3.69	-3.76
	0.05	100	-4.04	-3.94	-3.86	-3.75	-3.65	-3.51	-3.37	-3.21	-3.02	-2.79
		250	-3.85	-3.78	-3.70	-3.61	-3.51	-3.39	-3.28	-3.13	-2.99	-2.86
		∞	-3.80	-3.72	-3.63	-3.53	-3.43	-3.32	-3.19	-3.06	-2.93	-2.84
	0.10	100	-3.75	-3.66	-3.55	-3.45	-3.34	-3.20	-3.04	-2.86	-2.66	-2.41
		250	-3.59	-3.50	-3.41	-3.31	-3.21	-3.08	-2.94	-2.78	-2.61	-2.44
		∞	-3.53	-3.44	-3.34	-3.23	-3.13	-3.00	-2.86	-2.69	-2.54	-2.39
5-C	0.01	100	-4.63	-4.51	-4.44	-4.35	-4.24	-4.13	-4.00	-3.87	-3.66	-3.42
		250	-4.38	-4.33	-4.26	-4.17	-4.08	-3.99	-3.86	-3.73	-3.57	-3.45
		∞	-4.32	-4.26	-4.18	-4.08	-3.98	-3.90	-3.77	-3.65	-3.52	-3.47
	0.05	100	-4.04	-3.94	-3.85	-3.74	-3.64	-3.50	-3.36	-3.18	-2.98	-2.72
		250	-3.85	-3.78	-3.69	-3.60	-3.50	-3.37	-3.23	-3.09	-2.91	-2.72
		∞	-3.80	-3.72	-3.63	-3.52	-3.41	-3.29	-3.16	-3.00	-2.84	-2.70
	0.10	100	-3.75	-3.66	-3.55	-3.44	-3.32	-3.18	-3.02	-2.83	-2.63	-2.36
		250	-3.59	-3.50	-3.41	-3.30	-3.19	-3.05	-2.91	-2.74	-2.56	-2.34
		∞	-3.53	-3.44	-3.33	-3.23	-3.11	-2.98	-2.83	-2.66	-2.49	-2.30

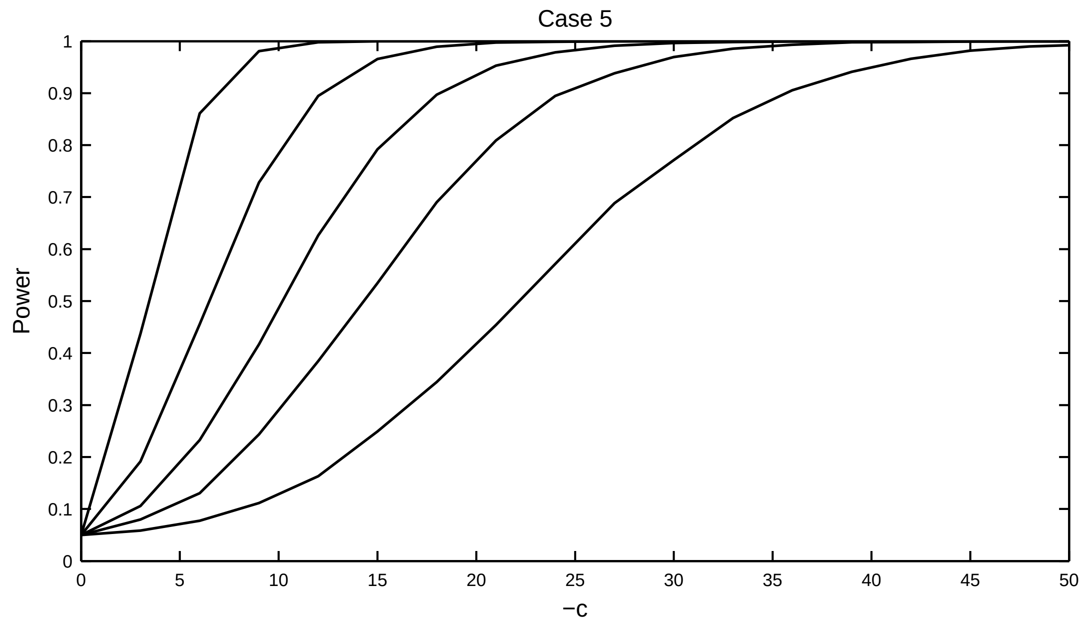

Figure 1: Asymptotic power curves of Models B and C for case 5 (Theorem 2). Curves for $R^2 = 0$, 0.3, 0.5, 0.7, and 0.9, where power is increasing in R^2.

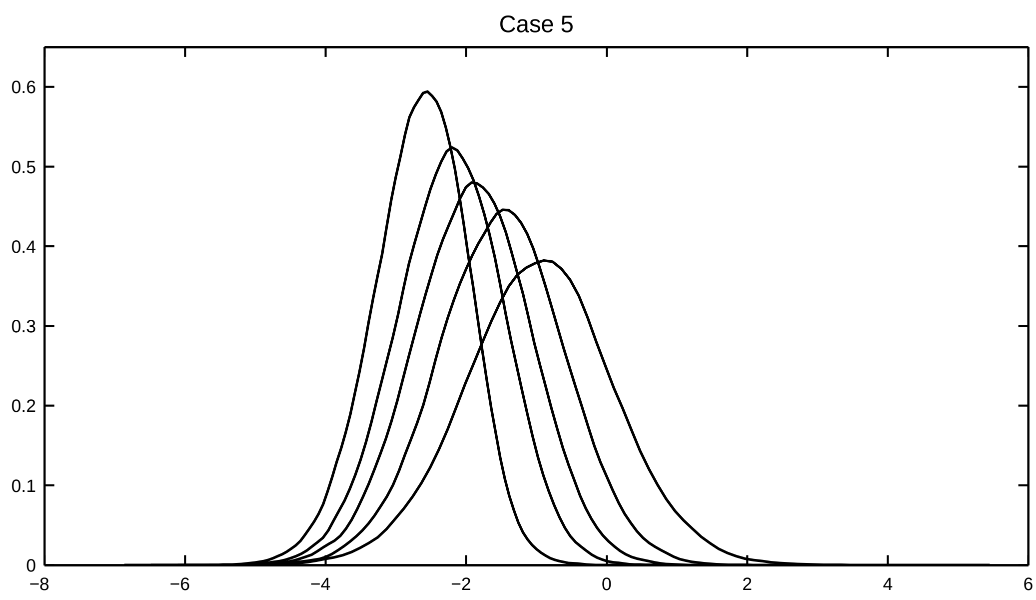

Figure 2: Asymptotic density functions of Models B and C for case 5 (Theorem 2). Densities for $R^2 = 0$, 0.3, 0.5, 0.7, and 0.9, where the distribution's mode shifts right as R^2 increases.

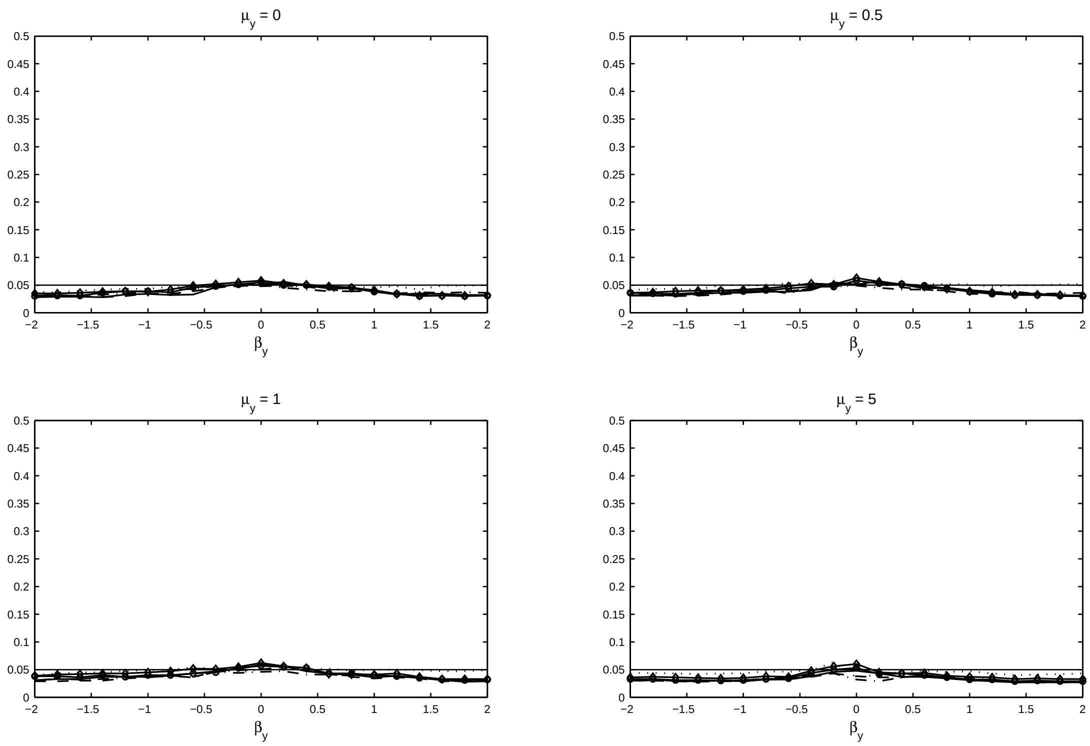

Figure 3: Empirical size for case 5-C, $T = 100$, $T_0 = 50$, and 5% nominal size. ADF-GLS($\hat{\delta}$): $-\diamond-$; and CADF-GLS($\hat{\delta}$) with $R^2 = 0$: $-\circ-$, $R^2 = 0.09$: —— , $R^2 = 0.25$: $--$, $R^2 = 0.49$: $-\cdot-$, $R^2 = 0.81$: $\cdots$.

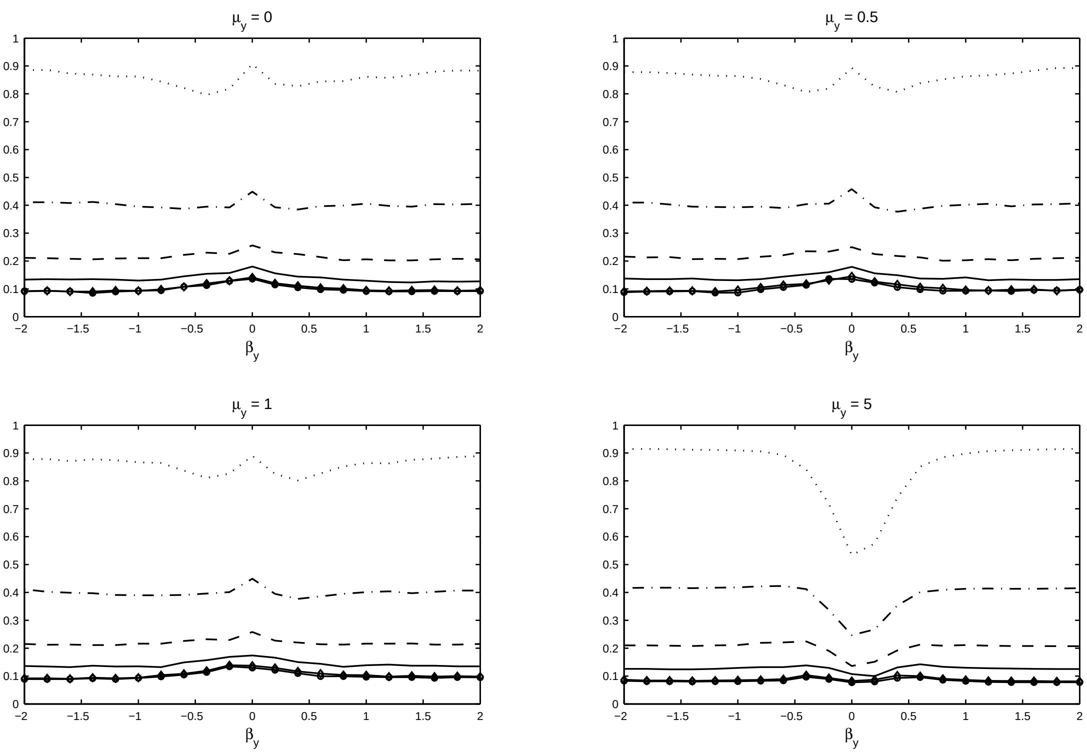

Figure 4: Empirical power for case 5-C, $T = 100$, $T_0 = 50$, and $\rho = 0.90$. ADF-GLS($\hat{\delta}$): $-\diamond-$; and CADF-GLS($\hat{\delta}$) with $R^2 = 0$: $-\circ-$, $R^2 = 0.09$: —— , $R^2 = 0.25$: $--$, $R^2 = 0.49$: $- \cdot -$, $R^2 = 0.81$: $\cdots$.

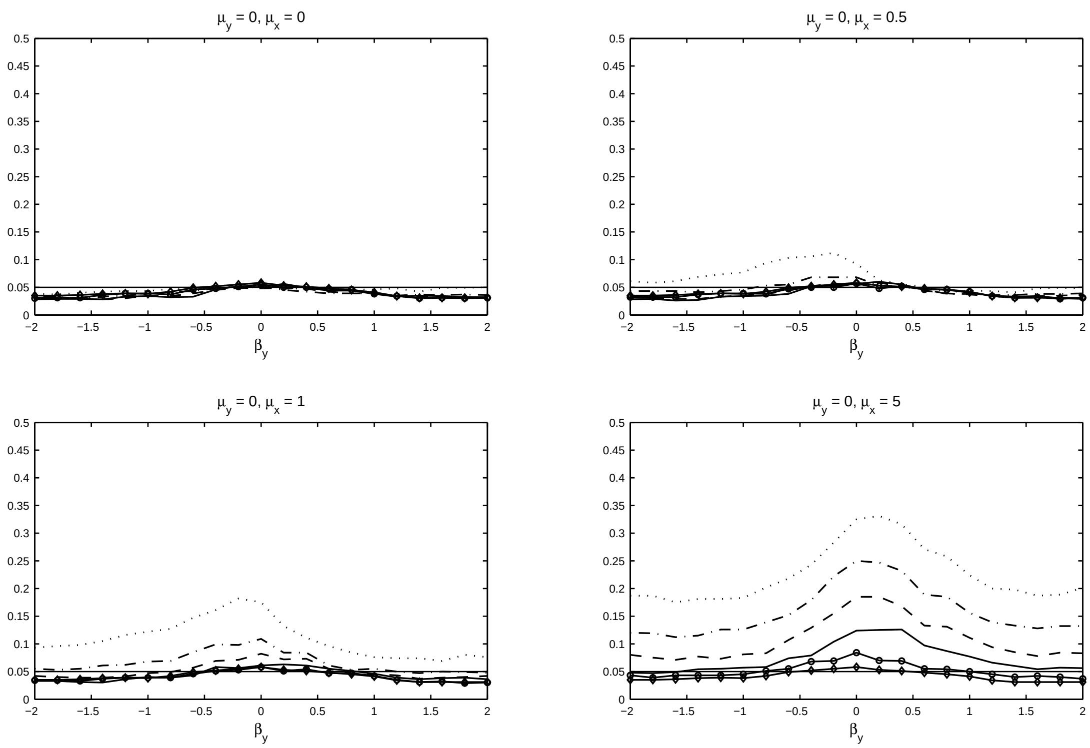

Figure 5: Empirical size for case 5-C, $T = 100$, $T_0^x = T_0 = 50$, and 5% nominal size. ADF-GLS($\hat{\delta}$): $-\diamond-$; and CADF-GLS($\hat{\delta}$) with $R^2 = 0$: $-\circ-$, $R^2 = 0.09$: —— , $R^2 = 0.25$: $--$, $R^2 = 0.49$: $-\cdot-$, $R^2 = 0.81$: $\cdots$.

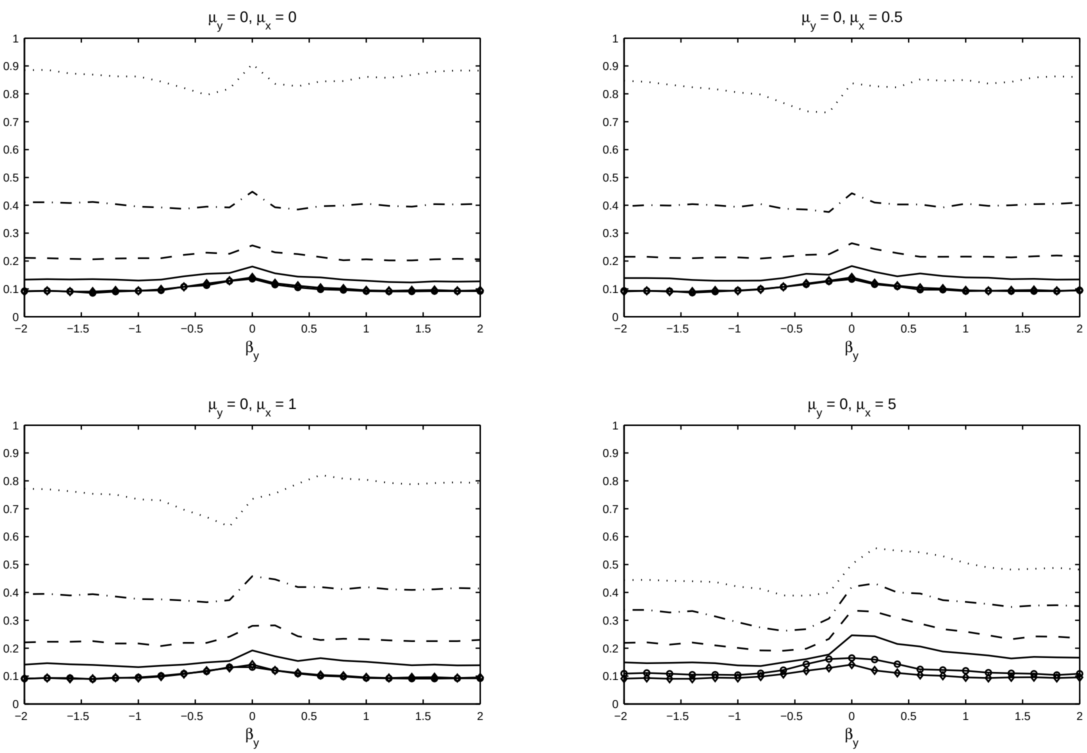

Figure 6: Empirical power for case 5-C, $T = 100$, $T_0^x = T_0 = 50$, and $\rho = 0.90$. ADF-GLS($\hat{\delta}$): $-\diamond-$; and CADF-GLS($\hat{\delta}$) with $R^2 = 0$: $-\circ-$, $R^2 = 0.09$: —— , $R^2 = 0.25$: $--$, $R^2 = 0.49$: $-\cdot-$, $R^2 = 0.81$: $\cdots$.