Quasi-Likelihood Ratio Tests for Cointegration, Cobreaking, and Cotrending*

Josep Lluís Carrion-i-Silvestre[†]
University of Barcelona

Dukpa Kim[‡]
Korea University

02/26/2018

Abstract

We consider a set of variables with two types of non-stationary features, stochastic trends and broken linear trends. We develop tests that can determine whether a linear combination of these variables exists under which the non-stationary features can be canceled out. We provide a comprehensive treatment of cointegration testing with structural breaks that is missing in the literature. In particular, the first test we develop can determine whether stochastic trends can be eliminated and thus whether cointegration holds, regardless of whether structural breaks in linear trends are eliminated. The second test can determine whether both stochastic trends and breaks in linear trends are simultaneously removed and thus whether cointegration and cobreaking simultaneously hold. The third test can determine whether not only breaks in linear trends but also linear trends themselves are eliminated along with stochastic trends and thus whether both cointegration and cotrending hold. We provide the asymptotic null distributions of proposed test statistics and some Monte Carlo simulation results to assess the adequateness of our asymptotic distributions in samples with common sizes.

JEL Classification Number: C22
Keywords: cointegration, cobreaking, cotrending, multiple structural breaks

[†]AQR-IREA research group. Department of Econometrics, Statistics and Applied Economics, Faculty of Economics and Business, University of Barcelona, Av. Diagonal, 690, 08034 Barcelona, (carrion@ub.edu).

[‡]Corresponding author, Department of Economics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841 Korea (dukpakim@korea ac.kr)

1 Introduction

Early developments in the time series literature focus on distinguishing between a difference stationary process and a trend stationary process. In the recent unit root literature, one line of research focuses on the case in which a stochastic trend and a structural break in the deterministic function coexist. For example, see Perron (1989); Carrion-i-Silvestre, Kim, and Perron (2009); Harris, Harvey, Leybourne, and Taylor (2009); Harvey, Leybourne, and Taylor (2012, 2013); and Kim and Perron (2009). One the motivation for this type of research is that a stochastic trend and a break in the deterministic function are separate objects and thus the existence of one implies nothing about the existence of the other. However, popular unit root tests are developed assuming either the presence or absence of a structural break. These unit root tests do not lead to correct inference when the assumption regarding structural breaks fails. Thus, the goal in the recent unit root literature is to devise a unit root test robust to the presence or absence of structural breaks.

The issues analyzed in this paper are closely related to this recent unit root literature. Suppose that a researcher finds, for example via the unit root tests proposed by the aforementioned authors, that the variables under investigation have both a stochastic trend and a broken linear trend. Her goal is to test whether there is a stable relationship among those variables. Can she use any of the tests proposed in the literature for this purpose (e.g., cointegration tests allowing for a structural break or structural break tests for cointegration models)? If not, what is the alternative? This paper intends to answer these questions.

Most methods proposed in the econometrics literature assume that variables have either a stochastic trend (cointegration models) or a structural break (cobreaking or cotrending models), but not both. Even cointegration models with structural breaks mostly assume that breaks are exogenously given only to the cointegration equation and do not affect individual variables. However, we are interested in the case in which a structural break in the cointegration equation exists because breaks in individual variables do not cancel one another out. In Section 2, we discuss in greater detail what statistical issues arise when existing tests are applied to our problem. Additionally, see Section 5 for the Monte Carlo simulation results supporting our argument.

The traditional definition of cointegration says that a set of trending variables is stochastically cointegrated if a certain linear combination of the variables yields a stationary process around a linear trend, and that it is deterministically cointegrated if that linear combination also eliminates the linear trend.[1] Given variables with both a stochastic trend and a broken linear trend, a linear combination of these variables is decomposed into a deterministic time function and a stochastic term. We consider this set of variables to be cointegrated if the stochastic term is stationary regardless of the shape of the deterministic function. Hence, our cointegration refers to stochastic cointegration in the traditional sense. We consider it to be cobreaking if the deterministic time function has no break and consider it to be cotrending if the deterministic time function is a constant (i.e., the linear combination eliminates not only all breaks but also the linear trend itself). The traditional deterministic cointegration holds when our cointegration and cotrending jointly hold, but not when our cointegration and cobreaking jointly hold.[2]

When cointegration does not hold, there is no stable relationship regardless of whether cobreaking or cotrending holds. Hence, a test that distinguishes between cointegration and no cointegration regardless of cobreaking and cotrending is of interest. We call this type of test a robust cointegration test because it is robust to the existence or absence of breaks. The robust cointegration test we develop uses the null hypothesis of cointegration and is designed to control size regardless of cobreaking and cotrending. Rejection

---

[1] See Campbell and Perron (1991) and Ogaki and Park (1997), for example.

[2] Our cobreaking and cotrending are special cases of Chapman and Ogaki's (1993) cotrending or Hendry and Mizon's (1998) cobreaking. Hatanaka and Yamada's (2003) cotrending refers to the case in which there is no slope change while intercept changes are allowed. Bierens (1997, 2000) considers cotrending of nonlinear time functions that are more general than broken linear trends. We restrict our analysis to broken linear trends to remain focused, but also to give coverage to wide empirical studies based on linear cointegration analysis with breaks.

of the robust cointegration test implies no stable relationship, while non-rejection implies cointegration with cobreaking and cotrending still being unfounded. To test for stronger relationships, we also develop a joint test for cointegration and cobreaking and another for cointegration and cotrending.

All three tests are obtained in a unified framework, which extends Jansson's (2005) point optimal invariant test. We consider two models, one with shifts in the mean and the other with intercept shifts in the linear trend. We analyze both the known and unknown break date cases.

The rest of the paper is organized as follows. Section 2 presents the models and then compares the new tests with other tests in the literature. Section 3 offers test statistics and their asymptotic distributions. Section 4 discusses endogenous regressors. Section 5 contains some Monte Carlo experiment results. Section 6 has an empirical illustration of U.S. budget sustainability. Section 7 concludes. The critical values of our tests are reported in Supplementary Appendix I, and all mathematical derivations are collected in Supplementary Appendix II. Additional Monte Carlo simulation results are provided in Supplementary Appendix III.

2 Model

We assume that the observation (y_t, x_t') is generated by

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \alpha^{y\prime} \\ \alpha^{x\prime} \end{pmatrix} d_t + \begin{pmatrix} y_t^0 \\ x_t^0 \end{pmatrix}, \tag{1}$$

where y_t is a scalar random variable, x_t is a $p_x \times 1$ random vector, d_t is a $p_d \times 1$ vector of deterministic functions in time, α^y and α^x are coefficient matrices, and y_t^0 and x_t^0 are stochastic components. We consider two specifications for d_t. That is, $d_t' = [DU_t(T_0), \ldots, DU_t(T_m)]$ in Model I and $[DU_t(T_0), \ldots, DU_t(T_m), t)]$, in Model II, where $DU_t(T_i) = 1$ for $t > T_i$ and 0 elsewhere, and $T_0 = 0$. The deterministic function has m breaks in each model. Model I has mean changes and Model II has a linear trend with shifts in the intercept.[3] The coefficient matrices α^y and α^x are defined as

$$\alpha^{y\prime} = \begin{cases} (\mu_0^y, \ldots, \mu_m^y), & \text{in Model I} \\ (\mu_0^y, \ldots, \mu_m^y, \psi_0^y), & \text{in Model II} \end{cases} \quad \text{and} \quad \alpha^{x\prime} = \begin{cases} (\mu_0^x, \ldots, \mu_m^x), & \text{in Model I} \\ (\mu_0^x, \ldots, \mu_m^x, \psi_0^x), & \text{in Model II.} \end{cases}$$

Hence, μ_i^y and μ_i^x denote the intercepts in the $(i+1)^{th}$ regime, and ψ_0^y and ψ_0^x are the slopes. Let $\pi = (\pi_1, \ldots, \pi_m)$ be a vector of break fractions where each π_i satisfies $T_i = [\pi_i T]$ with $[c]$ being the integer part of c. In what follows, it will be important to distinguish between the true break dates and break dates used in the regression equation. We will use $(T_1^0, \ldots, T_m^0)$ and $\pi^0 = (\pi_1^0, \ldots, \pi_m^0)$ where $T_i^0 = [\pi_i^0 T]$ to denote the true break dates and true break fractions. We use the following assumption regarding the break dates.

Assumption 1 $T_i^0 = [\pi_i^0 T]$, and $\pi_i^0 - \pi_j^0 \geq a > 0$ for all $j < i$ where $\pi_0^0 = 0$ and $\pi_{m+1}^0 = 1$.

The stochastic components y_t^0 and x_t^0 are assumed to be integrated of order 1, I(1). Hence y_t and x_t have both stochastic and broken deterministic trends. Our main focus is to test whether a certain linear combination of (y_t, x_t') becomes more stable in the sense that stochastic trends, broken deterministic trends, or both cancel one another out. Such a linear combination, if it exists, is assumed to be unique up to scalar multiplication and to always include y_t. We write this linear relationship in a regression form as

$$y_t = \beta' x_t + \alpha' d_t + v_t, \quad \text{for } t = 1, \ldots, T, \tag{2}$$

---

[3] A model with a slope change is of interest. Unlike a shift in the intercept, a slope change can be located consistently even in the presence of a stochastic trend, which leads to a different statistical procedure. We explore this model in a separate paper.

with $v_t = y_t^0 - \beta' x_t^0$ and $\alpha' = (\alpha^y - \alpha^x \beta)' = (\mu_0, \ldots, \mu_m)$ in Model I and $(\mu_0, \ldots, \mu_m, \psi_0)$ in Model II. We make the following assumptions regarding the random components.

Assumption 2 *(i) $v_t = v_{1t} + v_{2t}$; (ii) $\Delta v_{1t} = \varepsilon_t - \theta \varepsilon_{t-1}$ with $v_{10} = \varepsilon_0 = 0$, $\varepsilon_t = \sum_{i=0}^{\infty} c_i \eta_{t-i}$ with $\sum_{i=0}^{\infty} i \| c_i \| < \mathcal{C}$, $\omega_\varepsilon^2 = (\sum_{i=0}^{\infty} c_i)^2 > 0$ and $\eta_t \sim i.i.d.(0, 1)$; (iii) $u_t^x \equiv \Delta x_t^0 = \sum_{i=0}^{\infty} G_i \eta_{t-i}^x$ with $\sum_{i=0}^{\infty} i \| G_i \| < \mathcal{C}$, $G_\infty = \sum_{i=0}^{\infty} G_i$ is of full column rank, $\eta_t^x \sim i.i.d.(0, I_{p_x})$, and η_t and η_t^x are independent; (iv) $v_{2t} = \sum_{j=-\infty}^{\infty} u_{t-j}^{x\prime} \gamma_{-j}$ with $\sum_{j=-\infty}^{\infty} \| \gamma_j \| < \mathcal{C}$.*

Δ denotes the first difference operator. The disturbance ε_t is stationary with its long-run variance being $\omega_\varepsilon^2 > 0$. Hence, v_{1t} can be either I(0) or I(1), depending on the value of θ. The stochastic part of the regressors, x_t^0, is I(1) and the full column rank condition for G_∞ excludes the case where cointegration holds among the elements in x_t^0 without y_t^0. v_{2t} is always stationary and captures the correlation between the regressors and the regression error.

Assumption 3 $\| \psi_0^x \| > 0$.

Assumption 3 implies that the linear trend exists in at least one element of x_t. As explained below, the existence of a linear trend in x_t affects the limiting distribution of the test statistics. This assumption will be used whenever needed. We also use the following definitions.

1. (y_t, x_t') is cointegrated (CI) if and only if $\theta = 1$.

2. (y_t, x_t') is cobreaking (CB) if and only if $\mu_1 = \cdots = \mu_m = 0$.

3. (y_t, x_t') is cotrending (CT) if and only if it is CB and $\psi_0 = 0$ in Model II.

The next diagram summarizes the data generating processes (DGP) that we consider.[4]

	CB ($\mu_i = 0$, $\forall i \geq 1$)		No CB
	CT ($\psi_0 = 0$)	No CT ($\psi_0 \neq 0$)	($\mu_i \neq 0$, $\exists i \geq 1$)
CI ($\theta = 1$)	1	2	3
No CI ($\theta \neq 1$)	4	5	6

As explained, if there is no CI, the variables under consideration do not possess any stable relationship, irrespective of CB and CT. Hence, a test that can distinguish DGPs 1, 2, and 3 from DGPs 4, 5, and 6 will be useful. This motivates us to develop a robust CI test. To show a stronger relationship than CI, it is natural to test for the joint null hypothesis of CI and CB (DGPs 1 and 2) or the joint null of CI and CT (DGP 1). Hence, we also develop a joint test for CI and CB and a joint test for CI and CT. In the following, we discuss what issues one can encounter when tests that are available in the literature are used as a robust CI test or a joint test for CI and CB/CT.

First, the CI tests developed in line with unit root tests such as Phillips and Ouliaris (1990) take the null hypothesis of no CI (DGP 4 and/or DGP5) and are designed to have power against the alternative of CI (DGPs 1 and 2). To parallel the unit root tests, these tests should be called no CI tests. It is well known that these tests do not have much power against the alternative of CI with no CB (DGP 3), which is certainly a stronger relationship than no CI. Hence, modified tests such as the Gregory and Hansen (1996) test are proposed. The Gregory and Hansen test is specifically designed to have proper power against all of DGPs 1, 2, and 3, and thus rejections from this test are often regarded as supporting CI. Furthermore, because of the ability to detect all of DGPs 1, 2, and 3, the Gregory and Hansen test is even regarded as robust CI tests.

---

[4] Model I assumes that there is no linear trend and thus CB implies CT.

However, the null hypothesis employed is not just no CI (DGPs 4, 5, and 6). It should be described as no CI but CB (DGPs 4 and 5 only). It is not well known how this test would behave when neither CI nor CB is true (DGP 6). If it does not control size under DGP 6, it cannot be used as a robust CI test.

The Gregory and Hansen test is similar in statistical nature to the Zivot and Andrews (1992), Perron and Vogelsang (1992), and Perron (1997) unit root tests. Kim and Perron (2009) show that the probability the Zivot and Andrews test rejects the null of a unit root varies dramatically across the kinds and magnitudes of breaks if both a unit root and a break in the deterministic function exist. The asymptotic null distribution of the Zivot and Andrews test is derived with assuming that there is no break. Hence, if there is a break in the data, it affects the finite sample distribution. Similar properties are expected for the Gregory and Hansen test, which we demonstrate via Monte Carlo experiments. Rejecting the Gregory and Hansen test should not be regarded as supporting CI. The rejection may have been possible only due to a break.

Inoue (1999) considers a cointegrating rank test that allows for the breaking of deterministic trends. His treatment of breaks is similar to Gregory and Hansen's. The null assumes cointegration rank at most r with cobreaking deterministic components, while the alternative assumes cointegration rank greater than r with no cobreaking deterministic trends. This approach is not consistent in treating breaks, since breaks appear or disappear depending on whether we are under the null or the alternative hypothesis but also depending on the step of the sequential testing procedure. In the first step the null does not consider a structural break, but the alternative does; if we reject the null hypothesis, the alternative hypothesis of the first step becomes the null hypothesis of the second step with imposing that there is no break.

While our tests are based on the single equation cointegration model, some robust CI rank tests are developed in the context of the vector error correction model. Johansen, Mosconi, and Nielsen (2000); Saikkonen and Lütkepohl (2000); and Trenkler, Saikkonen, and Lütkepohl (2007) consider cointegrating rank tests that allow for breaks in deterministic trends. Their tests are robust to the existence of breaks in the deterministic components, but they assume known break dates. Lütkepohl, Saikkonen, and Trenkler (2004) extend Saikkonen and Lütkepohl (2000) to allow for a break at an unknown date. The cointegration rank is assumed to be at most r under the null and greater than r under the alternative. When testing for the null that $r = 0$, the test controls size irrespective of cobreaking, and the test can be regarded as a robust CI test. Harris, Leybourne, and Taylor (2016) also develop a robust cointegration rank test. Their model allows for a change in the slope of the trend at an unknown date.

The tests proposed by Carrion-i-Silvestre and Sansó (2006) and Arai and Kurozumi (2007) can be understood in a similar manner, although the testing strategy for these tests is the opposite of the Gregory and Hansen type tests. These tests are extensions of Shin's (1994) CI test to allow for breaks in the deterministic time function. They take the null of CI and are designed to have proper power against no CI regardless of CB/CT (DGPs 4, 5, and 6). However, they require breaks to exist under the null. Hence, these tests control size only when there is no CB (DGP 3). These tests are expected, though not shown, to have a well defined limiting distribution in the simultaneous presence of CI and CB (DGPs 1 and 2). Thus, they can be made robust to CB/CT simply by using bigger percentiles of the two limiting distributions obtained under 'DGPs 1 and 2' and 'DGP 3'.[5]

Many CI tests with the null of CI can be used as a joint test because they are general specification tests with non-trivial power not only under DGPs 4 and 5 but also under DGPs 3 and 6. This group includes the ones proposed by Shin (1994), Leybourne and McCabe (1993), Jansson (2005), and Kurozumi and Arai (2008). Since these tests are developed without considering the possibility of no CB, they leave some room for power improvement under DGPs 3 and 6 while they tend to have good power under DGPs 4 and 5. One good example is the non-monotonic power function of the Shin test. The Shin test is an extension of the KPSS test proposed by Kwiatkowski et al. (1992), which is the same as Perron's (1991) Q test. As is well

---

[5]In fact, this is how we develop our robust CI test. We derive two limiting distributions of our test statistic under the case of CB and no CB and suggest use of the bigger percentiles of the two distributions.

documented in Vogelsang (1999) and Perron (2005), these stationarity tests have a non-monotonic power function, which means that the power can drop even to zero as the magnitude of the break gets large. This occurs because the long-run variance estimate is contaminated by the breaks that are not accounted for in the construction of those tests. The bigger the unaccounted breaks are, the more severely the test loses power. In Section 5, we demonstrate, via Monte Carlo experiments, that the Shin test exhibits a non-monotonic power function when CB is violated severely (DGPs 3 and 6).

Structural break tests developed for CI models can be understood in a similar way. This group includes Hansen (1992), Kuo (1998), Hao (1996), Seo (1998), Hansen (2003), Qu (2007), and Kejriwal and Perron (2010). The null hypothesis of these tests is the joint of CI and CB (DGPs 1 and 2), which is the same as the CI tests in the previous group. The difference is that these tests are designed to have power against no CB (DGP 3) instead of no CI (DGPs 4 and 5). However, these tests are also general specification tests and have power under DGPs 4 and 5.

With the joint null of CI and CB/CT, it is not obvious why tests should be designed to have power against only one type of violation. We show that our joint tests, developed considering both types of violation, have better power properties than the aforementioned CI tests or structural break tests in CI models.

As mentioned in the introduction, many of the aforementioned tests assume that there is no break in each variable. Only the relationship among the variables is allowed to change. We deal with a different situation in which breaks in the relationship equation are caused by breaks in each variable, and aim at giving a comprehensive answer regarding what to do when practitioners need to test for a potential long-run relationship with variables that are I(1) non-stationary with structural breaks. In principle, we cannot simply ignore breaks detected in the univariate analysis. These breaks may or may not be present in the relationship equation due to possible CB, and this must be assessed in a rigorous way. Furthermore, the case in which only the relationship equation has breaks is a special case of our model that sets $\mu_i^x = 0$ for all i in (1), and thus our results hold regardless of the origin of breaks.

3 Test Statistics and Asymptotic Distributions

3.1 Known Break Dates

The first testing problem concerns the null hypothesis of CI ($\theta = 1$) against the alternative of no CI ($\theta = \bar{\theta} < 1$). We devise our test statistic as a quasi-likelihood ratio by assuming that $\varepsilon_t \sim IID\ N(0, \sigma_\varepsilon^2)$. In addition, we assume that the regressor vector x_t is exogenous (i.e., $v_{2t} = 0$ in Assumption 2). The treatment of endogenous regressors is postponed to the next section for the sake of expositional simplicity. The model in (2) in matrix notation becomes

$$y = D\alpha + X\beta + \Psi_\theta^{1/2}\varepsilon, \tag{3}$$

where $y = (y_1, \ldots, y_T)'$, $D = (d_1, \ldots, d_T)'$, $X = (x_1, \ldots, x_T)'$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$, and $\Psi_\theta^{1/2}$ is a lower triangular matrix with 1 on the diagonal and $1 - \theta$ off the diagonal. Since the goal is to develop a robust CI test, we focus on tests that are invariant to location shifts in the conditional mean of y. Following Jansson (2005), the maximal invariant we consider is

$$P'y \sim N(0, \sigma_\varepsilon^2 P'\Psi_\theta P), \tag{4}$$

where P is defined such that $PP' = M = I_T - Z(Z'Z)^{-1}Z'$ with $Z = [D, X]$ and $P'P = I_{T-p_d-p_x}$. The maximal invariant depends on θ and the break dates used in D. The log likelihood of the maximal invariant

is given as

$$\ell_T(\theta, \pi, \sigma_\varepsilon^2) = -\frac{1}{2}\log|P'\Psi_\theta P| - \frac{1}{2\sigma_\varepsilon^2}y'P(P'\Psi_\theta P)^{-1}P'y \tag{5}$$

$$= -\frac{1}{2}\log|Z'\Psi_\theta^{-1}Z| - \frac{1}{2\sigma_\varepsilon^2}y'\left(\Psi_\theta^{-1} - \Psi_\theta^{-1}Z\left(Z'\Psi_\theta^{-1}Z\right)^{-1}Z'\Psi_\theta^{-1}\right)y.$$

The point optimal invariant (POI) test rejects for large values of

$$-2\left(\ell_T(1, \pi^0, \sigma_\varepsilon^2) - \ell_T(\bar\theta, \pi^0, \sigma_\varepsilon^2)\right), \tag{6}$$

which defines our infeasible robust CI test. Let the log likelihood of y be $\ell_T(\alpha, \beta|\theta, \pi^0, \sigma_\varepsilon^2)$. The log likelihood of the maximal invariant given in (5) is nothing but $\ell_T(\theta, \pi^0, \sigma_\varepsilon^2) = \max_{\alpha,\beta}\ell_T(\alpha, \beta|\theta, \pi^0, \sigma_\varepsilon^2)$, and the POI test in (6) is the same as $-2\left(\max_{\alpha,\beta}\ell_T(\alpha, \beta|1, \pi^0, \sigma_\varepsilon^2) - \max_{\alpha,\beta}\ell_T(\alpha, \beta|\bar\theta, \pi^0, \sigma_\varepsilon^2)\right)$. Since CB and CT can be expressed as a set of linear restrictions on α, natural tests for the joint hypotheses are obtained by extending the principle of likelihood ratio.

The test statistic for the joint null of CI and CB is given by

$$-2\left(\max_{\alpha,\beta \ s.t. \ R_{cb}\alpha=0}\ell_T(\alpha, \beta|1, \pi^0, \sigma_\varepsilon^2) - \max_{\alpha,\beta}\ell_T(\alpha, \beta|\bar\theta, \pi^0, \sigma_\varepsilon^2)\right) + c_{cb},$$

where $R_{cb}\alpha = (\mu_1, \ldots, \mu_m)'$ and $c_{cb} = m\log(T)$.

The test statistic for the joint null of CI and CT in Model II is given by

$$-2\left(\max_{\alpha,\beta \ s.t. \ R_{ct}\alpha=0}\ell_T(\alpha, \beta|1, \pi^0, \sigma_\varepsilon^2) - \max_{\alpha,\beta}\ell_T(\alpha, \beta|\bar\theta, \pi^0, \sigma_\varepsilon^2)\right) + c_{ct},$$

where $R_{ct}\alpha = (\mu_1, \ldots, \mu_m, \psi_0)'$ and $c_{ct} = \log\left(\psi_0^{x'}\psi_0^x / \left(\psi_0^{x'}G_\infty^{-1}\psi_0^x\right)^2\right) + (m+2)\log(T)$. The constants c_{cb} and c_{ct} are included to warrant proper convergence of the test statistics.[6] The first term in c_{ct} can be estimated consistently from the first difference of x_t.

Let $\hat\omega_\varepsilon^2$ be a consistent estimator for ω_ε^2, the long-run variance of ε_t, and let $\hat c_{ct}$ be such that $\hat c_{ct} - c_{ct} = o_p(1)$. Then, using the elements defined above, the feasible robust CI test can be computed as

$$\mathbb{Q}_r = -2\left(\ell_T(1, \pi^0, \hat\omega_\varepsilon^2) - \ell_T(\bar\theta, \pi^0, \hat\omega_\varepsilon^2)\right),$$

the feasible joint test for CI and CB is given by

$$\mathbb{Q}_{cb} = -2\left(\max_{\alpha,\beta \ s.t. \ R_{cb}\alpha=0}\ell_T(\alpha, \beta|1, \pi^0, \hat\omega_\varepsilon^2) - \max_{\alpha,\beta}\ell_T(\alpha, \beta|\bar\theta, \pi^0, \hat\omega_\varepsilon^2)\right) + c_{cb},$$

and the feasible joint test for CI and CT in Model II is given by

$$\mathbb{Q}_{ct} = -2\left(\max_{\alpha,\beta \ s.t. \ R_{ct}\alpha=0}\ell_T(\alpha, \beta|1, \pi^0, \hat\omega_\varepsilon^2) - \max_{\alpha,\beta}\ell_T(\alpha, \beta|\bar\theta, \pi^0, \hat\omega_\varepsilon^2)\right) + \hat c_{ct}.$$

To describe the limiting distribution of the test statistics, we introduce the following notations. Let $du(r, \pi_i)$ and $b(r)$ be the limit counterpart of $DU_t(T_i)$ and B_t so that $du(r, \pi_i) = 1(r > \pi_i)$ and $b(r) = r$. We

---

[6]These constants originate from the log term of the Gaussian log-likelihood function, which involves trending regressors. See the proofs in Supplementary Appendix II to see how these constants are calculated.

also rewrite θ and $\bar{\theta}$ using local-to-unity parameters, so that $\theta = 1 - \lambda/T$ and $\bar{\theta} = 1 - \bar{\lambda}/T$. $V(r)$ and $W(r)$ are independent standard Wiener processes of dimensions one and p_x, respectively. We denote the first n elements of $W(r)$ by $W_n(r)$. We define $V^\lambda(s) = V(s) + \lambda \int_0^s V(r)dr$, $V_{\bar{\lambda}}^\lambda(s) = V^\lambda(s) - \bar{\lambda} \int_0^s e^{-\bar{\lambda}(s-r)} V^\lambda(r)dr$, and $Q_{\bar{\lambda}}^\pi(s) = Q^\pi(s|p,n) - \bar{\lambda} \int_0^s e^{-\bar{\lambda}(s-r)} Q^\pi(r|p,n)dr$, where $Q^\pi(s|p,n) = (du(s,\pi_0), \ldots, du(s,\pi_{p-1}), W_n(s)')'$ in Model I and $(du(s,\pi_0), \ldots, du(s,\pi_{p-1}), b(s), W_n(s)')'$ in Model II. Finally, $\Phi_1(\lambda, \bar{\lambda}) = 2\bar{\lambda} \int_0^1 V_{\bar{\lambda}}^\lambda(s)\, dV^\lambda(s) - \bar{\lambda}^2 \int_0^1 V_{\bar{\lambda}}^\lambda(s)^2\, ds$ and $\Phi_2(\lambda, \bar{\lambda}, \pi; p, n) = [\int_0^1 Q_{\bar{\lambda}}^\pi(s)\, dV_{\bar{\lambda}}^\lambda(s)]'[\int_0^1 Q_{\bar{\lambda}}^\pi(s) Q_{\bar{\lambda}}^{\pi\prime}(s)\, ds]^{-1}[\int_0^1 Q_{\bar{\lambda}}^\pi(s)\, dV_{\bar{\lambda}}^\lambda(s)]$ $- \log|\int_0^1 Q_{\bar{\lambda}}^\pi(s) Q_{\bar{\lambda}}^{\pi\prime}(s)\, ds|$.

The next theorem states the asymptotic distribution of the $\mathbb{Q}_r$ statistic.

Theorem 1 *Let $\theta = 1 - \lambda/T$, $\bar{\theta} = 1 - \bar{\lambda}/T$. Then, under Assumptions 1 $\sim$ 2 with $v_{2t} = 0$, we have*

$$\mathbb{Q}_r \Rightarrow \varphi_r(\lambda, \bar{\lambda}, \pi^0) = \Phi_1(\lambda, \bar{\lambda}) - \Phi_2(\lambda, 0, \pi^0; m+1, p_x) + \Phi_2(\lambda, \bar{\lambda}, \pi^0; m+1, p_x).$$

With π^0 known, the test statistic is exactly invariant to α, the CB and CT parameters. This implies that CB and CT may or may not hold under both the null and alternative hypotheses. Hence, the test is robust to the presence or absence of breaks. Note also that Assumption 3 is not needed here. Theorem 1 states the asymptotic distribution of $\mathbb{Q}_r$ not only under the null ($\theta = 1$) but also under the local alternative hypothesis ($\bar{\theta} < 1$ or $\bar{\lambda} > 0$). Further, the limiting distribution depends on the composition of the deterministic trend (d_t), the number of stochastic regressors (p_x), the break fractions (π^0), and the local-to-unity parameter ($\bar{\lambda}$). The usual recommendation in the econometrics literature is to choose $\bar{\lambda}$ so that the local asymptotic power curve is tangent to the theoretical power envelope with 50% power, which means that the choice of $\bar{\lambda}$ will depend on the break dates. This unnecessarily complicates the testing procedure with unknown break dates. Instead, we propose using $\bar{\lambda}$ that makes the average power over possible combinations of break fractions achieve 50% when $\lambda = \bar{\lambda}$ so that $\bar{\lambda}$ does not depend on break positions.

The suggested values for $\bar{\lambda}$ are reported in Supplementary Appendix I. The average power is obtained from all combinations of break fractions in one decimal place that satisfy Assumption 1 with $a = 0.2$. With this choice of $\bar{\lambda}$, $\mathbb{Q}_r$ is still a POI test with the local power curve tangent to the envelope around 50% rather than exactly at 50%. Using the suggested values of $\bar{\lambda}$, we report selected percentiles of $\varphi(0, \bar{\lambda}, \pi^0)$ in Supplementary Appendix I for one and two structural breaks cases.[7] The next two theorems state the limiting distribution of the $\mathbb{Q}_{cb}$ and $\mathbb{Q}_{ct}$ statistics under their respective null hypotheses.

Theorem 2 *Let $\theta = 1$ and $\bar{\theta} = 1 - \bar{\lambda}/T$.*
(i) If CB holds in Models I and II, we have under Assumptions 1 $\sim$ 2 with $v_{2t} = 0$,

$$\mathbb{Q}_{cb} \Rightarrow \varphi_{cb}(0, \bar{\lambda}, \pi^0) \equiv \Phi_1(0, \bar{\lambda}) - \Phi_2(0, 0, \pi^0; 1, p_x) + \Phi_2(0, \bar{\lambda}, \pi^0; m+1, p_x).$$

(ii) If CT holds in Model II, we have under Assumptions 1 $\sim$ 3 with $v_{2t} = 0$,

$$\mathbb{Q}_{ct} \Rightarrow \varphi_{ct}(0, \bar{\lambda}, \pi^0) \equiv \Phi_1(0, \bar{\lambda}) - \Phi_2(0, 0, \pi^0; 1, p_x - 1) + \Phi_2(0, \bar{\lambda}, \pi^0; m+1, p_x).$$

As can be seen, the limiting distribution of the test statistics depends on $\bar{\lambda}$, on the number and position of the structural breaks (π^0), and on the number of stochastic regressors (p_x). Selected percentiles of $\varphi_{cb}(0, \bar{\lambda}, \pi^0)$ and $\varphi_{ct}(0, \bar{\lambda}, \pi^0)$ are approximated by simulation, as reported in Supplementary Appendix I. Note that Assumption 3 is needed for Theorem 2 (ii). When Assumption 3 does not hold, there is no linear trend in x_t and the limiting distribution of $\mathbb{Q}_{ct}$ holds with $\Phi_2(0, 0, \pi^0; 1, p_x - 1)$ in the above theorem replaced by $\Phi_2(0, 0, \pi^0; 1, p_x)$ defined for Model I.

---

[7] In creating all asymptotic critical values, a Wiener process is approximated with 2,000 steps, and the number of replications is 20,000.

3.2 Unknown Break Dates

We now address the case of unknown break dates, in which analysts decide to use estimated break dates to compute the proposed test statistics. Unfortunately, when there are only intercept shifts, the break dates can be estimated from neither the levels nor the first differences of (y_t, x_t'). They can be estimated from the relationship equation in (2) only if CB does not hold while CI holds. This is a general feature shared by all proposals in the literature, including that of Lütkepohl, Saikkonen, and Trenkler (2004). The only way to overcome this issue is to assume that the break magnitude is large so that it is increasing with T, making the level shift non-negligible in the limit. While we do not pursue this approach in this paper, we suggest below a conservative approach that avoids assuming a particular magnitude of the structural breaks, covering the possible situation of uncertainty about whether the structural breaks have occurred.

We treat the break dates as additional parameters to be estimated. The break dates will be searched over the set $\Pi(m) = \{(\pi_1, \ldots, \pi_m) | \pi_i - \pi_j \geq a > 0 \text{ for all } 0 \leq j < i \leq m+1 \text{ where } \pi_0 = 0 \text{ and } \pi_{m+1} = 1\}$. Then, the feasible test statistics that can be computed for the unknown break dates case are

$$
\hat{\mathbb{Q}}_r = -2 \left(\max_{\alpha,\beta,\pi} \ell_T(\alpha, \beta, \pi | 1, \hat{\omega}_\varepsilon^2) - \max_{\alpha,\beta,\pi} \ell_T(\alpha, \beta, \pi | \bar{\theta}, \hat{\omega}_\varepsilon^2) \right)
$$

$$
\hat{\mathbb{Q}}_{cb} = -2 \left(\max_{\alpha,\beta \ s.t. \ R_{cb}\alpha=0} \ell_T(\alpha, \beta | 1, \pi, \hat{\omega}_\varepsilon^2) - \max_{\alpha,\beta,\pi} \ell_T(\alpha, \beta, \pi | \bar{\theta}, \hat{\omega}_\varepsilon^2) \right) + c_{cb},
$$

and

$$
\hat{\mathbb{Q}}_{ct} = -2 \left(\max_{\alpha,\beta \ s.t. \ R_{ct}\alpha=0} \ell_T(\alpha, \beta | 1, \pi, \hat{\omega}_\varepsilon^2) - \max_{\alpha,\beta,\pi} \ell_T(\alpha, \beta, \pi | \bar{\theta}, \hat{\omega}_\varepsilon^2) \right) + \hat{c}_{ct},
$$

where maximization with respect to π is taken over $\Pi(m)$.

It is worth mentioning that, in principle, it could be possible to define two different estimators for the break fractions, depending on whether the estimates are obtained under the null or alternative hypotheses, and using the notation in (5),

$$
\hat{\pi} = \arg\max_{\pi} \ell_T(1, \pi, \hat{\omega}_\varepsilon^2) \quad \text{and} \quad \check{\pi} = \arg\max_{\pi} \ell_T(\bar{\theta}, \pi, \hat{\omega}_\varepsilon^2). \tag{7}
$$

The consistency of break fraction estimator $\hat{\pi}$ at rate T is well established in various structural break models when CI holds; for instance, see Carrion-i-Silvestre and Sansó (2006). The next lemma shows the asymptotic equivalence of $\hat{\pi}$ and $\check{\pi}$.

Lemma 1 *Let $\theta = 1$ and $\bar{\theta} = 1 - \bar{\lambda}/T$. Assume that CB does not hold so that $\mu_i \neq 0$ for all i. Then, under Assumptions 1 ∼ 2 with $v_{2t} = 0$, we have $\|\hat{\pi} - \check{\pi}\| = o_p(T^{-1})$ in Models I and II.*

Taking into account these elements, let us first state in the following theorem the asymptotic null distributions of the robust CI test statistics for the unknown break date case.

Theorem 3 *Let $\theta = 1$ and $\bar{\theta} = 1 - \bar{\lambda}/T$. Suppose that Assumptions 1 ∼ 2 with $v_{2t} = 0$ hold.*
(i) If $\|\hat{\pi} - \pi^0\| = O_p(T^{-1})$ and $\|\hat{\pi} - \check{\pi}\| = o_p(T^{-1})$, we have in Models I and II

$$
\hat{\mathbb{Q}}_r \Rightarrow \varphi_r(0, \bar{\lambda}, \pi^0),
$$

where $\varphi_r(0, \bar{\lambda}, \pi^0)$ is as defined in Theorem 1.
(ii) If CB holds, we have

$$
\hat{\mathbb{Q}}_r \Rightarrow \varphi_r^{\max}(0, \bar{\lambda}, m) \equiv \Phi_1(0, \bar{\lambda}) - \max_{\pi \in \Pi(m)} \Phi_2(0, 0, \pi; m+1, p_x) + \max_{\pi \in \Pi(m)} \Phi_2(0, \bar{\lambda}, \pi; m+1, p_x).
$$

8

The result in (i) pertains to the case of no CB (DGP 3) and the result in (ii) pertains to the case of CB (DGPs 1 and 2). As can be seen, the asymptotic distribution of $\hat{\mathbb{Q}}_r$ is the same as that of $\mathbb{Q}_r$ with no CB, while it takes a different form with CB. This makes the empirical use of these statistics difficult, since a practitioner might not be sure about the presence of structural breaks. To control size over all of DGPs 1, 2, and 3, we propose using a conservative approach by taking the larger of the two critical values coming from these distributions.

When CB is known to hold, Jansson's (2005) test is the POI test. Naturally one might consider a test procedure that pretests for intercept shifts and chooses the optimal CI test depending on the pretest result. The only available candidate for the pretest is a structural break test for the intercept of a CI equation, because intercept shifts are not detectable without CI. As explained earlier, however, such a structural break test rejects with a probability approaching one the null of no break as long as there is no CI, irrespective of CB. The intended power gain then fails to materialize because the $\hat{\mathbb{Q}}_r$ test is selected over Jansson's test even when there is CB with no CI. Thus, we do not pursue such a sequential method.

Selected percentiles of $\varphi_r^{\max}(0, \bar{\lambda}, m)$ are reported in Supplementary Appendix I. When multiple breaks are involved, intermediate cases not covered by the above theorem exist. They occur when some breaks cancel out but others do not, and they correspond to no CB in our definition since there is at least one break appearing in the relationship. Hence, some break date estimators are consistent for existing breaks while the other break date estimators remain random. The limiting distribution of the test statistic will then be intermediate of the two distributions given by the above theorem. We do not explicitly derive these distributions, although they can be obtained from the results of this paper. Because our goal is to control size irrespective of CB, we only need to analyze the extreme cases that can yield the greatest critical value. See also the Monte Carlo simulation results in Section 5.

One practical issue is that the two break date estimates $\hat{\pi}$ and $\check{\pi}$ can be substantially different in practice. In such a case, one suggestion is to use only $\hat{\pi}$ in the construction of $\hat{\mathbb{Q}}_r$

$$\widetilde{\mathbb{Q}}_r = -2 \left(\ell_T(1, \hat{\pi}, \hat{\omega}_\varepsilon^2) - \ell_T(\bar{\theta}, \hat{\pi}, \hat{\omega}_\varepsilon^2) \right).$$

Then, $\hat{\mathbb{Q}}_r \geq \widetilde{\mathbb{Q}}_r$ and the test becomes only conservative.[8] Next, we provide the asymptotic null distributions for the joint tests.

Theorem 4 Let $\theta = 1$ and $\bar{\theta} = 1 - \bar{\lambda}/T$.
(i) If CB holds in Models I and II, we have under Assumptions $1 \sim 2$ with $v_{2t} = 0$,

$$\hat{\mathbb{Q}}_{cb} \Rightarrow \varphi_{cb}^{\max}(0, \bar{\lambda}, m) \equiv \Phi_1(0, \bar{\lambda}) - \Phi_2(0, 0, \pi^0; 1, p_x) + \max_{\pi \in \Pi(m)} \Phi_2(0, \bar{\lambda}, \pi; m+1, p_x).$$

(ii) If CT holds in Model II, we have under Assumptions $1 \sim 3$ with $v_{2t} = 0$,

$$\hat{\mathbb{Q}}_{ct} \Rightarrow \varphi_{ct}^{\max}(0, \bar{\lambda}, m) \equiv \Phi_1(0, \bar{\lambda}) - \Phi_2(0, 0, \pi^0; 1, p_x - 1) + \max_{\pi \in \Pi(m)} \Phi_2(0, \bar{\lambda}, \pi; m+1, p_x).$$

Selected percentiles of $\varphi_{cb}^{\max}(0, \bar{\lambda}, m)$ and $\varphi_{ct}^{\max}(0, \bar{\lambda}, m)$ are reported in Supplementary Appendix I.[9] While an optimal property is difficult to establish for the joint test, we note that $\hat{\mathbb{Q}}_{cb}$ consists of two terms,

---

[8] The tests proposed by Carrion-i-Silvestre and Sansó (2006) and by Arai and Kurozumi (2007) are comparable with our tests. The main difference is that they require that breaks exist when the break dates are unknown. Hence, no asymptotic distribution comparable with our $\varphi_r^{\max}(0, \bar{\lambda}, m)$ has been derived for these tests.

[9] Supplementary Appendix I reports the percentiles of $\varphi_{cb}^{\max}(0; \bar{\lambda}, m)$ and $\varphi_{ct}^{\max}(0; \bar{\lambda}, m)$ for $m = 0$ as well. These correspond to the percentiles of regular cointegration tests that do not consider broken deterministic trends.

$\hat{\mathbb{Q}}_{cb} = Q_1 + \hat{\mathbb{Q}}_r$ where

$$Q_1 = -2 \left(\max_{\alpha,\beta \ s.t. \ R_{cb}\alpha=0} \ell_T(\alpha, \beta | 1, \pi, \hat{\omega}_\varepsilon^2) - \max_{\alpha,\beta,\pi} \ell_T(\alpha, \beta, \pi | 1, \hat{\omega}_\varepsilon^2) \right) + c_{cb},$$

which is similar to the sup-type structural break test in the cointegrated equation explored by Hansen (1992) and Kejriwal and Perron (2010). Andrews and Ploberger (1994) show that the sup-type structural break test in general is not optimal although it is asymptotically admissible. Similarly, $\hat{\mathbb{Q}}_{ct} = Q_2 + \hat{\mathbb{Q}}_{cb}$ where

$$Q_2 = -2 \left(\max_{\alpha,\beta \ s.t. \ R_{ct}\alpha=0} \ell_T(\alpha, \beta | 1, \pi, \hat{\omega}_\varepsilon^2) - \max_{\alpha,\beta \ s.t. \ R_{cb}\alpha=0} \ell_T(\alpha, \beta | 1, \pi, \hat{\omega}_\varepsilon^2) \right) + (\hat{c}_{ct} - c_{cb}),$$

which is a test for the significance of the linear trend.

One issue in using the $\hat{\mathbb{Q}}_{cb}$ and $\hat{\mathbb{Q}}_{ct}$ tests is the choice of m. We now write $\hat{\mathbb{Q}}_{cb}(m)$ and $\hat{\mathbb{Q}}_{ct}(m)$ for $\hat{\mathbb{Q}}_{cb}$ and $\hat{\mathbb{Q}}_{ct}$ to signify the assumed number of breaks. When the break dates are unknown, it is unlikely that the number of breaks is known. A natural way to proceed is to extend the principle of likelihood ratio by treating m as another parameter. This would simply require computing $\hat{\mathbb{Q}}_{cb}(m)$ or $\hat{\mathbb{Q}}_{ct}(m)$ from a range of m values and taking the maximum. Since the percentiles of $\hat{\mathbb{Q}}_{cb}(m)$ or $\hat{\mathbb{Q}}_{ct}(m)$ are increasing in m, the maximum would be mostly determined by the $\hat{\mathbb{Q}}_{cb}(m)$ or $\hat{\mathbb{Q}}_{ct}(m)$ statistic with the largest m value, failing to deliver the intended power property. Hence, we consider the following modified statistic given by

$$\hat{\mathbb{Q}}_{cb}^{D\,\max} \equiv \max_{0 \le m \le M} \beta_m^{-1} \left[\hat{\mathbb{Q}}_{cb}(m) - \alpha_m \right] \quad \text{and} \quad \hat{\mathbb{Q}}_{ct}^{D\,\max} \equiv \max_{0 \le m \le M} \beta_m^{-1} \left[\hat{\mathbb{Q}}_{ct}(m) - \alpha_m \right],$$

with α_m and β_m to be defined below. These tests are related to the WDmax test proposed by Bai and Perron (1998). The following theorem states the asymptotic null distribution of the $\hat{\mathbb{Q}}_{cb}^{D\,\max}$ and $\hat{\mathbb{Q}}_{ct}^{D\,\max}$ statistics.

Theorem 5 *Let* $\theta = 1$ *and* $\bar{\theta} = 1 - \bar{\lambda}/T$.
(i) If CB holds in Models I and II, we have under Assumptions 1 $\sim$ 2 with $v_{2t} = 0$,

$$\hat{\mathbb{Q}}_{cb}^{D\,\max} \Rightarrow \varphi_{cb}^{D\,\max} \equiv \max_{0 \le m \le M} \beta_m^{-1} \left[\varphi_{cb}^{\max}(0, \bar{\lambda}_m, m) - \alpha_m \right],$$

where $\varphi_{cb}^{\max}(0, \bar{\lambda}_m, m)$ *is as defined in Theorem 4.*
(ii) If CT holds in Model II, then we have under Assumptions 1 $\sim$ 3 with $v_{2t} = 0$,

$$\hat{\mathbb{Q}}_{ct}^{D\,\max} \Rightarrow \varphi_{ct}^{D\,\max} \equiv \max_{0 \le m \le M} \beta_m^{-1} \left[\varphi_{ct}^{\max}(0, \bar{\lambda}_m, m) - \alpha_m \right],$$

where $\varphi_{ct}^{\max}(\lambda, \bar{\lambda}_m, m)$ *is as defined in Theorem 4.*

Depending on α_m and β_m, $\hat{\mathbb{Q}}_{cb}^{D\,\max}$ and $\hat{\mathbb{Q}}_{ct}^{D\,\max}$ will show different power characteristics. We set α_m to be the asymptotic 95th percentile of $\varphi_{cb}^{\max}(0, \bar{\lambda}_m, m)$ or $\varphi_{ct}^{\max}(0, \bar{\lambda}_m, m)$ and β_m to be the difference between the asymptotic 99th and 95th percentiles. Our choice of α_m and β_m is made to make the critical values at 5% and 1% size become zero and one, respectively, for each m.[10] Using this choice of α_m and β_m, selected percentiles of $\varphi_{cb}^{D\,\max}$ and $\varphi_{ct}^{D\,\max}$ are reported in Supplementary Appendix I.

---

[10]To see this, let $cv_{95}(m)$ and $cv_{99}(m)$ be the 95th and 99th percentiles. Then,

$$
\begin{aligned}
P\left(\beta_m^{-1}(\hat{\mathbb{Q}}_{cb}(m) - \alpha_m) > 0 \right) &= P\left((\hat{\mathbb{Q}}_{cb}(m) - cv_{95}(m))) > 0 \right) = 0.05, \\
P\left(\beta_m^{-1}(\hat{\mathbb{Q}}_{cb}(m) - \alpha_m) > 1 \right) &= P\left((\hat{\mathbb{Q}}_{cb}(m) - cv_{95}(m)) > (cv_{99}(m) - cv_{95}(m)) \right) = 0.01.
\end{aligned}
$$

4 Endogenous regressors

We now discuss endogenous regressors. While the approach taken by Jansson (2005) is one possibility, we use the dynamic ordinary least squares (DOLS) by Saikkonen (1991). From Assumption 2, let $v_{2t} = \sum_{j=-k}^{k} u_{t-j}^{x\prime} \gamma_{-j} + \omega_t$ with $\omega_t = \sum_{|j|>k} u_{t-j}^{x\prime} \gamma_{-j}$. The model in matrix notation is given by

$$y = D\alpha + X\beta + U_X\gamma + \omega + \Psi_\theta^{1/2}\varepsilon,$$

where $\gamma = (\gamma_{-k}', \ldots, \gamma_k')'$ $\omega = (\omega_1, \ldots, \omega_T)'$ and $U_X = [U_{-k+1}^x, \ldots, U_{k+1}^x]$ with $U_s^x = (u_s^x, \ldots, u_{T+s-1}^x)'$. Let $\Delta X = [\Delta X_{-k+1}, \ldots, \Delta X_{k+1}]$ and $\Delta D = [\Delta D_{-k+1}\alpha^x, \ldots, \Delta D_{k+1}\alpha^x]$ with $\Delta X_s = (\Delta x_s, \ldots, \Delta x_{T+s-1})'$ and $\Delta D_s = (\Delta d_s, \ldots, \Delta d_{T+s-1})'$. Since $U_X = \Delta X - \Delta D$, the model is equivalent to

$$y = D\alpha + X\beta + \Delta X\gamma + \phi + \omega + \Psi_\theta^{1/2}\varepsilon,$$

where $\phi = -\Delta D\gamma$. When we ignore ϕ and ω, the maximal invariant of interest is $P_W'y$ where $P_W P_W' = I_T - W(W'W)^{-1}W' = M_W$ with $W = [Z, \Delta X]$ and $Z = [D, X]$. This is the same as the maximal invariant in (4) except for ΔX. The log likelihood of $P_W'y$ is given as

$$\ell_T(\bar{\theta}, \pi, \sigma_\varepsilon^2) = -\frac{1}{2}\log\left|W'\Psi_{\bar{\theta}}^{-1}W\right| - \frac{1}{2\sigma_\varepsilon^2}y'\left(\Psi_{\bar{\theta}}^{-1} - \Psi_{\bar{\theta}}^{-1}W\left(W'\Psi_{\bar{\theta}}^{-1}W\right)^{-1}W'\Psi_{\bar{\theta}}^{-1}\right)y.$$

We make the following assumption, which is a standard in the literature.

Assumption 4 *As $T \to \infty$, $k \to \infty$ such that $k^2/T \to 0$ and $T^{1/2}\sum_{|j|\geq k}\left\|\gamma_j\right\| \to 0$.*

A finite number of leads and lags of Δx_t augmented in the regression will fix endogeneity only partially due to ω. However, we increase the number of leads and lags as the sample size grows making ω negligible. Assumption 4 provides the relevant conditions on the number of leads and lags, k. The condition that $\sum_{|j|\geq k}\left\|\gamma_j\right\| = o(T^{-1/2})$ regulates the lower bound for k.[11] Note that the lower bound depends on the decay rate of $\left\|\gamma_j\right\|$. For example, this condition permits k to increase at a logarithmic rate, if $\left\|\gamma_j\right\|$ decays exponentially. The ϕ term also has no effect asymptotically. For the robust CI test, we show in Supplement Appendix II that under Assumptions $1 \sim 4$,

$$
\begin{aligned}
&-2\left(\ell_T(1, \pi^0, \hat{\omega}_\varepsilon^2) - \ell_T(\bar{\theta}, \pi^0, \hat{\omega}_\varepsilon^2)\right) && (8)\\
&= \log|Z'Z| + \frac{1}{\hat{\omega}_\varepsilon^2}\varepsilon_\theta'\left(I - Z(Z'Z)^{-1}Z'\right)\varepsilon_\theta\\
&\quad - \log\left|Z'\Psi_{\bar{\theta}}^{-1}Z\right| - \frac{1}{\hat{\omega}_\varepsilon^2}\varepsilon_\theta'\left(\Psi_{\bar{\theta}}^{-1} - \Psi_{\bar{\theta}}^{-1}Z\left(Z'\Psi_{\bar{\theta}}^{-1}Z\right)^{-1}Z'\Psi_{\bar{\theta}}^{-1}\right)\varepsilon_\theta + o_p(1),
\end{aligned}
$$

which implies that the effects of additional regressors ΔX are asymptotically negligible and the asymptotic distributions obtained in the previous section continue to hold. Extensions for joint tests are straightforward. One only needs to add ΔX to each regression in the construction of the test statistics.

---

[11] In fact, a weaker condition $k^{1/2}\sum_{|j|\geq k}\left\|\gamma_j\right\| \to 0$ can be used, which allows the use of information criteria; see Kejriwal and Perron (2008).

5 Monte Carlo Simulation

The DGP is the same as (1). Recall that $\alpha = \alpha^y - \alpha^x \beta$. We set α^x to be a matrix of $1/2$, and β to be a $p_x \times 1$ vector of $1/p_x$s: $\beta = (1/p_x, \ldots, 1/p_x)'$. α^y is determined in such a way that α can have a desired value. Recall that $\alpha' = (\mu_0, \ldots, \mu_m, \psi_0)$. We always set $\mu_0 = 0$ as the simulation results are invariant. Hence, $\mu_i = 0$ $(i \neq 0)$ corresponds to CB and $\psi_0 = 0$ to CT. x_t^0 is generated by accumulating independent standard normal variates. y_t^0 is created as a sum of $\beta' x_t^0$ and $v_{1t} = \sum_{j=1}^t (\varepsilon_j - \theta \varepsilon_{j-1})$ with ε_t being an independent standard normal variate and $v_{10} = \varepsilon_0 = 0$. Hence $\theta = 1$ corresponds to CI. We also consider only exogenous regressors by setting $v_{2t} = 0$.[12] The number of stochastic regressors is set at $p_x = 1$ and the number of repetitions is 5,000 for the size and 1,000 for the power. We simulate the following six cases.

DGP 1 (CI + CT) $\theta = 1$, $\mu_i = 0$, $\psi_0 = 0$.

DGP 2 (CI + CB) $\theta = 1$, $\mu_i = 0$, $\psi_0 = 0.3$.

DGP 3 (CI + No CB) $\theta = 1$, $\mu_i = 1.0 \sim 3.0$ in steps of 1.0 and $6 \sim 12$ in steps of 3, $\psi_0 = 0.3$.

DGP 4 (No CI + CT) $\theta = 0, 0.25, 0.5$ and $1.05 \sim 1.25$ in steps of 0.1, $\mu_i = 0$, $\psi_0 = 0$.

DGP 5 (No CI + CB) $\theta = 0, 0.25, 0.5$ and $1.05 \sim 1.25$ in steps of 0.1, $\mu_i = 0$, $\psi_0 = 0.3$.

DGP 6 (No CI + No CB) $\theta = 1.05 \sim 1.25$ in steps of 0.1 with $\mu_i = 10(\theta - 1)$, and $\theta = 0$ with $\mu_i = 3 \sim 9$ in steps of 3, $\psi_0 = 0.3$.

To focus on the case of unknown break dates, we compute $\hat{\mathbb{Q}}_r(m)$, $\hat{\mathbb{Q}}_{cb}(m)$, and $\hat{\mathbb{Q}}_{ct}(m)$ for $m = 1$ and 2, $\varphi_{cb}^{D\,\max}$, and $\varphi_{ct}^{D\,\max}$ with $M = 2$. For all tests, the long-run variance is computed by a heteroskedasticity and autocorrelation consistent covariance estimator with the quadratic spectral kernel, for which the bandwidth parameter is selected using the Andrews (1991) data dependent method with an AR(1) approximation. In doing so, the residuals obtained under the null are used.

Table 1. Finite Sample Sizes

(i) DGP 1 (CI+CT)

Model	T	$\hat{\mathbb{Q}}_r(1)$	$\hat{\mathbb{Q}}_{cb}(1)$	$\hat{\mathbb{Q}}_{ct}(1)$	$\hat{\mathbb{Q}}_r(2)$	$\hat{\mathbb{Q}}_{cb}(2)$	$\hat{\mathbb{Q}}_{ct}(2)$	$\hat{\mathbb{Q}}_{cb}^{D\,\max}$	$\hat{\mathbb{Q}}_{ct}^{D\,\max}$
I	120	.044	.032	n.a.	.027	.021	n.a.	.039	n.a.
	360	.042	.034	n.a.	.027	.028	n.a.	.040	n.a.
II	120	.035	.027	.028	.029	.015	.013	.045	.040
	360	.035	.038	.036	.027	.032	.024	.041	.037

(ii) DGP 3 (CI+No CB, $\mu_i = 2.0$)

		Model I		Model II	
π^0	T	$\hat{\mathbb{Q}}_r(1)$	$\hat{\mathbb{Q}}_r(2)$	$\hat{\mathbb{Q}}_r(1)$	$\hat{\mathbb{Q}}_r(2)$
(0.5)	120	.062	.033	.060	.044
	360	.057	.036	.061	.055
(0.3, 0.7)	120	.968	.058	.847	.071
	360	.999	.058	.997	.064

---

[12] See Supplementary Appendix III for the case of endogenous regressors both with and without the DOLS corrections discussed in Section 4.

Table 1 has the finite sample sizes of our tests obtained using the asymptotic critical values for the norminal size of 5%. In (i), the results are obtained under DGP 1 (CI + CT) with one intercept change for each variable but no break in the relationship equation due to CB. For the robust CI tests $\hat{\mathbb{Q}}_r(m)$, the bigger percentiles of $\varphi_r^{\max}(0, \bar{\lambda}, m)$ and $\varphi_r(0, \bar{\lambda}, \hat{\pi})$ should be used. However, we use the bigger percentiles of $\varphi_r^{\max}(0, \bar{\lambda}, m)$ and $\varphi_r(0, \bar{\lambda}, \pi^0)$ for simplicity. As can be checked in Supplementary Appendix I, the percentiles of $\varphi_r(0, \bar{\lambda}, \pi)$ do not vary much with π and the simplification rarely affects the empirical sizes. Overall, our tests show rejection rates close to or slightly smaller than the nominal 5% due to the use of asymptotic critical values, so that better results would be expected if finite sample critical values were used instead.

In (ii), the results for the $\hat{\mathbb{Q}}_r(m)$ test are obtained under DGP 3 (CI + No CB) with $\mu_i = 2.0$. Since there is no CB, we consider two different combinations of break fractions, 0.5 and (0.3, 0.7). When the number of breaks specified by $\hat{\mathbb{Q}}_r(m)$ is greater than or equal to the true number of breaks, the robust CI test controls size near the nominal 5%, as desired. For example, $\hat{\mathbb{Q}}_r(1)$ controls size properly only for the case of one break and $\hat{\mathbb{Q}}_r(2)$ does so for up to two breaks.

When there are more breaks than specified by $\hat{\mathbb{Q}}_r(m)$, the robust CI test rejects with a large probability. This is a natural and desirable feature, because the existence of any additional break beyond those specified under the null hypothesis simply implies an unstable relationship.

Table 2. Power Comparison, Null Rejection Probabilities

(i) Model I

	θ	μ_1	$\hat{\mathbb{Q}}_r(1)$	$\hat{\mathbb{Q}}_{cb}(1)$	$\hat{\mathbb{Q}}_{cb}^{D\,\max}$	Shin	KP	GH	LST
DGP 1	1.0	0.0	.05	.05	.05	.05	.05	1.0	.93
DGP 3	1.0	1.0	.07	.81	.82	.72	.84	1.0	.93
		2.0	.06	1.0	1.0	.86	1.0	1.0	.89
		3.0	.09	1.0	1.0	.85	1.0	1.0	.87
		6.0	.09	.99	1.0	.66	1.0	1.0	.70
		9.0	.07	.83	1.0	.30	1.0	1.0	.59
		12.0	.07	.76	1.0	.14	1.0	1.0	.54
DGP 4	0.0	0.0	1.0	.95	1.0	.34	.77	.05	.05
	0.25		1.0	.92	1.0	.27	.75	.43	.13
	0.50		1.0	.95	1.0	.36	.80	.97	.41
	1.05		.09	.20	.24	.24	.22	1.0	.94
	1.15		.46	.76	.78	.58	.68	1.0	.91
	1.25		.77	.93	.94	.69	.82	1.0	.85
DGP 6	1.05	0.5	.11	.40	.47	.40	.42	1.0	.92
	1.15	1.5	.67	.93	.94	.72	.86	1.0	.89
	1.25	2.5	.88	.99	.99	.70	.92	1.0	.82
	0.0	3.0	1.0	.92	1.0	.35	.78	.05	.05
		6.0	1.0	.93	1.0	.34	.81	.11	.03
		9.0	1.0	.94	1.0	.33	.85	.28	.05

Table 2 reports size corrected null rejection probabilities under DGPs $1 \sim 6$ with $m = 1$ and $T = 120$. We also simulate the Gregory and Hansen test, the Shin test, the Lütkepohl, Saikkonen, and Trenkler test

13

and the Kejriwal and Perron test. To compute the long-run variance, residuals obtained under the null hypothesis are used for the Shin test, while those under the alternative hypothesis are used for the Gregory and Hansen and the Kejriwal and Perron tests. The VAR of order one is used for the Lütkepohl, Saikkonen, and Trenkler test. The Gregory and Hansen test (GH test) is selected to represent the group of no CI tests allowing for breaks; the Shin test to represent the group of CI tests using the joint null of CI and CB/CT; the Lütkepohl, Saikkonen, and Trenkler test (LST test) to represent a robust CI rank test; and the Kejriwal and Perron test (KP test) to represent the group of structural break tests for CI models. The KP test we simulate is the one with one break and is thus directly comparable with the $\hat{\mathbb{Q}}_{cb}(1)$ test.

Under DGP 1 (CI + CT), all tests except for the GH test and the LST test show 5% rejection rates because the finite sample critical values are obtained from DGP1. The finite sample critical values for the GH test and the LST tests are obtained under DGP 4 with $\theta = 0$. Since there is cointegration under DGP 1, the GH test and the LST test reject with a large probability.

DGP 2 (CI + CB) applies to Model II. Under DGP 2, CT does not hold while the CI and CB still hold. As such, the $\hat{\mathbb{Q}}_{ct}(1)$ and $\hat{\mathbb{Q}}_{ct}^{D\,\max}$ tests in Model II correctly detect the breakdown of CT under DGP 2. The other tests exhibit rejection probabilities similar to the ones under DGP 1.

Under DGP 3 (CI + No CB), the rejection rate of the $\hat{\mathbb{Q}}_r(1)$ test is slightly over 5% but stable across the values of μ_1. This is again due to the fact that we are not using the conservative approach. The $\hat{\mathbb{Q}}_{cb}(1)$, $\hat{\mathbb{Q}}_{cb}^{D\,\max}$, and KP tests show similar powers, while the Shin test falls short. This is very encouraging for the $\hat{\mathbb{Q}}_{cb}(1)$ test because the KP test is specifically developed to detect DGP 3. Note that the rejection rate of the Shin test drops as μ_1 increases in both models, which is the well known non-monotonic power phenomenon. The GH test always rejects under DGP 3 and shows the intended ability to detect CI regardless of CB/CT. The LST test also rejects with a large probability but is less powerful than the GH test.

Under DGP 4 (No CI + CT), the picture is quite different. As expected, the rejection rate of the $\hat{\mathbb{Q}}_r(1)$ test increases as θ deviates from one, which shows that the robust CI test has the intended power. All of the $\hat{\mathbb{Q}}_{cb}(1)$, $\hat{\mathbb{Q}}_{ct}(1)$, $\hat{\mathbb{Q}}_{cb}^{D\,\max}$, and $\hat{\mathbb{Q}}_{ct}^{D\,\max}$ tests also show the same pattern. Under DGP 4, CB/CT hold but these joint tests correctly reject the null hypothesis due to the violation of CI. Among these tests, the joint tests for CI and CT generally have the greatest power, the joint tests for CI and CB the second greatest, and the robust CI tests the least. The power difference is natural given the different null hypotheses of these tests. The $\hat{\mathbb{Q}}_{cb}^{D\,\max}$ test has greater power than the $\hat{\mathbb{Q}}_{cb}(1)$ test, and the $\hat{\mathbb{Q}}_{ct}^{D\,\max}$ test has greater power than the $\hat{\mathbb{Q}}_{ct}(1)$ test. This seems to be due to the fact that the $\hat{\mathbb{Q}}_{cb}^{D\,\max}$ and $\hat{\mathbb{Q}}_{ct}^{D\,\max}$ tests are also based on $\hat{\mathbb{Q}}_{cb}(0)$ and $\hat{\mathbb{Q}}_{ct}(0)$, which direct power only toward the violation of CI.[13] The Shin test does not show a non-monotonic power function. However, it is much less powerful than the other tests unless the DGP is very close to the null (θ close to one). It is interesting to see that the KP test is often more powerful than the Shin test even though it is designed as a structural break test. However, the KP test is still far less powerful than our joint tests. The rejection rate for the GH test and for the LST test is 5% when $\theta = 0$ since the critical value is obtained here. Their rejection rates are supposed to be about 5% under the entire DGP 4 but they increase to 100% when θ is close to one. We can write from the Beveridge-Nelson decomposition that $v_t = (1-\theta)\sum_{j=1}^{t}\varepsilon_j + \theta\varepsilon_t$, and thus $(1-\theta)^2$ stands for the variance of the innovation to the stochastic trend of v_t. When θ is close to one, the random walk portion of v_t can be masked by the stationary component $\theta\varepsilon_t$ and the data might look as if there were CI, which makes the GH test falsely reject the null of no CI. The size of the LST test is also close to 5% only when $\theta = 0$.[14]

DGP 5 (No CI + CB) applies to Model II. The rejection rates for the $\hat{\mathbb{Q}}_{ct}(1)$ and $\hat{\mathbb{Q}}_{ct}^{D\,\max}$ tests jump to one, because these tests are rejecting the null for the violated CT as well. The rejection rates for the other

---

[13] In general, the $\hat{\mathbb{Q}}_r(m)$, $\hat{\mathbb{Q}}_{cb}(m)$, and $\hat{\mathbb{Q}}_{cb}(m)$ tests exhibit lower power as m gets bigger, when there is no break in the data. Some relevant Monte Carlo simulation results are reported in Supplementary Appendix III.

[14] Another issue for the LST test is that it assumes the DGP for $(\Delta y_t^0, \Delta x_t^{0\prime})'$ to be a VAR of finite order. However, under DGPs 4~6, $(\Delta y_t^0, \Delta x_t^{0\prime})'$ is a moving average of order one with no equivalent finite order VAR representation unless $\theta = 0$.

tests are similar to those under DGP 4.

Table 2. Power Comparison, Null Rejection Probabilities, continued

(ii) Model II

	θ	μ_1	ψ_1	$\hat{\mathbb{Q}}_r(1)$	$\hat{\mathbb{Q}}_{cb}(1)$	$\hat{\mathbb{Q}}_{ct}(1)$	$\hat{\mathbb{Q}}_{cb}^{D\,\max}$	$\hat{\mathbb{Q}}_{ct}^{D\,\max}$	Shin	KP	GH	LST
DGP 1	1.0	0.0	0.0	.05	.05	.05	.05	.05	.05	.05	1.0	.99
DGP 2	1.0	0.0	0.3	.05	.06	1.0	.06	1.0	.03	.05	1.0	.99
DGP 3	1.0	1.0	0.3	.08	.38	1.0	.34	1.0	.22	.38	1.0	.98
		2.0		.09	.96	1.0	.94	1.0	.66	.96	1.0	.99
		3.0		.11	1.0	1.0	1.0	1.0	.85	1.0	1.0	.98
		6.0		.08	1.0	1.0	1.0	1.0	.27	1.0	1.0	.88
		9.0		.09	1.0	1.0	1.0	1.0	.02	1.0	1.0	.73
		12.0		.09	1.0	1.0	1.0	1.0	.01	1.0	1.0	.60
DGP 4	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	.30	.79	.05	.05
	0.25			1.0	1.0	1.0	1.0	1.0	.15	.71	.53	.14
	0.50			1.0	1.0	1.0	1.0	1.0	.36	.75	.98	.53
	1.05			.07	.13	.14	.12	.13	.11	.12	1.0	.98
	1.15			.31	.56	.64	.59	.65	.45	.46	1.0	.97
	1.25			.61	.84	.88	.85	.89	.61	.65	1.0	.95
DGP 5	0.0	0.0	0.3	1.0	1.0	1.0	1.0	1.0	.31	.77	.06	.04
	0.25			1.0	1.0	1.0	1.0	1.0	.15	.70	.52	.13
	0.50			1.0	1.0	1.0	1.0	1.0	.33	.75	.98	.48
	1.05			.08	.14	1.0	.15	1.0	.12	.11	1.0	.99
	1.15			.31	.57	1.0	.59	1.0	.42	.43	1.0	.97
	1.25			.67	.87	1.0	.88	1.0	.61	.67	1.0	.95
DGP 6	1.05	0.5	0.3	.09	.22	1.0	.21	1.0	.16	.20	1.0	.99
	1.15	1.5		.46	.83	1.0	.83	1.0	.61	.72	1.0	.97
	1.25	2.5		.83	.97	1.0	.97	1.0	.69	.89	1.0	.94
	0.0	3.0		1.0	1.0	1.0	1.0	1.0	.29	.80	.07	.04
		6.0		1.0	1.0	1.0	1.0	1.0	.29	.82	.13	.04
		9.0		1.0	1.0	1.0	1.0	1.0	.25	.89	.33	.03

Lastly, under DGP 6 (No CI + No CB), neither CI nor CB holds. The rejection rates for the $\hat{\mathbb{Q}}_{ct}(1)$ and $\hat{\mathbb{Q}}_{ct}^{D\,\max}$ tests are still one due to their ability to detect CT breakdown. The $\hat{\mathbb{Q}}_{cb}(1)$ and $\hat{\mathbb{Q}}_{cb}^{D\,\max}$ perform better than the Shin and KP tests with a good margin. The rejection rate for the GH test is increasing in μ_1 even when $\theta = 0$. As mentioned in Section 2, this is because the assumption of CB made under the null of the GH test breaks down. With this result, rejecting the GH test should not be interpreted as supporting CI. The rejection might have resulted from failed CB while CI does not hold. Then, not rejecting the null of the GH test supports no CI with CB while rejection does not support any particular DGP. This is a very serious drawback because neither rejection nor non-rejection implies any meaningful relation among the

variables under the test. The LST test is relatively free of this problem. When $\theta = 0$, the rejection rate is stable below 5% for all values of μ_1.

6   Empirical Illustration

The global economic crisis has created tensions in macroeconomic aggregates that must be monitored to drive economic recovery. Fiscal deficit sustainability has been an area in which economists have focused their attention, since other macroeconomic imbalances (e.g., current account imbalances or the debt of the economy) depend on government fiscal policy. Directly connected to the budget sustainability is the issue of whether the deficit series itself – defined as the difference between government revenues and expenditures – or any other linear combination of revenues and expenditures is I(0) stationary. Hence, one strategy that has been followed in the economic literature to assess fiscal deficit sustainability relies on discovering the strongest possible relationship between revenues and expenditures using cointegration analysis.
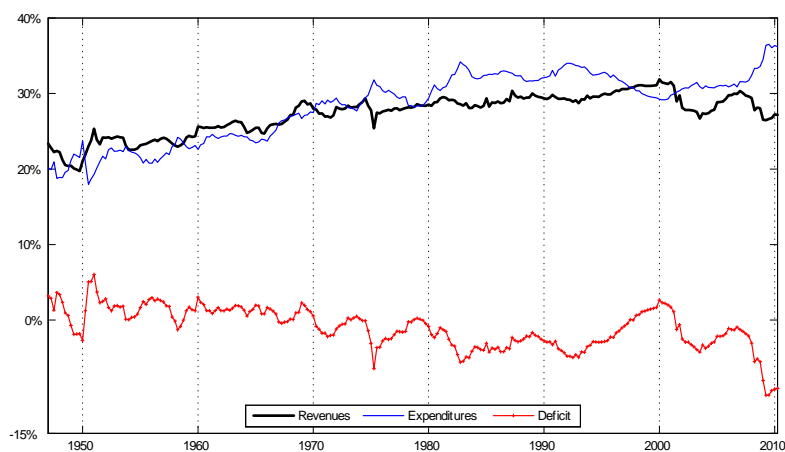

Figure 1: US Government Revenues/GDP, Expenditures/GDP, and Deficit/GDP

Earlier studies in the literature have already applied various statistical tests to analyse the existence of an equilibrium relationship between revenues and expenditures, and the results are quite mixed – see, for example, Haug (1991), Liu and Tanner (1994), Quintos (1995), Martin (2000), Bajo-Rubio et al. (2008), and Holmes (2010), among others. For instance, Holmes (2010) applies Gregory and Hansen's (1996) test statistics and, although the overall results are mixed, some evidence of cointegration is claimed upon rejecting the null hypothesis of no CI. However, this does not imply that there is CI, because the rejection might result from the presence of a structural break under the null hypothesis, which is ruled out in the framework used by Gregory and Hansen (1996). In this regard, the robust CI test that has been proposed in this paper should be of interest, since it is designed to control size regardless of the presence of structural breaks.

The study conducted in Bajo-Rubio et al. (2008) is based on Shin's (1994) cointegration test, which does not reject the null of deterministic cointegration, the joint of CI and CT in terms of our definition. With this non-rejection, they consider the system to be cointegrated and continue to apply Bai and Perron's (1998) multiple structural break test and find four regimes. Instead of Bai and Perron's (1998) test, a better alternative which has become available would be Kejriwal and Perron's (2010) multiple structural breaks test; note that this proposal has been specifically developed for cointegrating equations. As explained earlier, non-rejection of the Shin test combined with rejection of the structural break test does not imply piecewise cointegration. Both the Shin test and the structural break test should be viewed as a joint test of CI/CB or

CI/CT, and non-rejection of one test and rejection of the other test should be interpreted as only conflicting evidence due to the different power characteristics. As above, the joint tests that are proposed in this paper should be of interest, since they offer more reliable power over various unstable relationships.

In this section, we conduct an empirical illustration focusing on US fiscal deficit sustainability using cointegration analysis that considers the possible existence of structural breaks. Figure 1 depicts quarterly US government revenues, expenditures and deficit as percentages of GDP from 1947.I to 2010.II; the deficit series is obtained as the difference between the revenues and the expenditures.

Table 3. Structural Break Tests and Unit Root Tests

(i) One structural break

	Model II				Model III			
	Revenues		Expenditures		Revenues		Expenditures	
	Test	$\hat{\pi}_1$	Test	$\hat{\pi}_1$	Test	$\hat{\pi}_1$	Test	$\hat{\pi}_1$
PY	7.62**	.32	-0.22	.75	7.73**	.75	-0.07	.74
ZA	-4.09	.30	-4.19	.76	-5.16**	.84	-3.65	.52
CKP MZ_α	-19.28**	.86	-11.08	.37	-12.04	.45	-11.08	.76
CKP MSB	0.15**	.86	0.21	.37	0.19	.45	0.21	.76
CKP MZ_t	-2.98**	.86	-2.35	.37	-2.34	.45	-2.27	.76

(ii) Two structural breaks

	Model II				Model III			
	Revenues		Expenditures		Revenues		Expenditures	
	Test	$(\hat{\pi}_1, \hat{\pi}_2)$	Test	$(\hat{\pi}_1, \hat{\pi}_2)$	Test	$(\hat{\pi}_1, \hat{\pi}_2)$	Test	$(\hat{\pi}_1, \hat{\pi}_2)$
LP	-6.78**	(.30, .86)	-4.95	(.30, .76)	-8.09**	(.33, .86)	-5.09	(.55, .92)
CKP MZ_α	-13.01	(.45, .86)	-10.63	(.37, .49)	-20.63	(.45, .86)	-12.60	(.49, .77)
CKP MSB	0.18	(.45, .86)	0.22	(.37, .49)	0.16	(.45, .86)	0.19	(.49, .77)
CKP MZ_t	-2.39	(.45, .86)	-2.30	(.37, .49)	-3.21	(.45, .86)	-2.42	(.49, .77)

Note: PY stands for the Perron and Yabu (2009) test statistic, ZA for the Zivot and Andrews (1992) test statistic, LP for the Lumsdaine and Papell (1997) test statistic and CKP for the test statistics in Carrion-i-Silvestre et al. (2009). The superscript ** indicates rejection of the null hypothesis of either no structural break (PY test statistic) or unit root (ZA, LP and CKP test statistics) at the 5% level of significance, using the critical values provided in the corresponding paper.

Before we apply the proposal that has been designed in this paper, we proceed with the order of integration analysis of the revenues and expenditures series, testing for the presence of structural breaks. Throughout this section, one asterisk refers to rejection at the 10% level of significance, two asterisks at the 5% level of significance, and three asterisks at the 1% level of significance.

Table 3 presents the Perron and Yabu (2009) $Sup-ExpW_T$ statistic to test the presence of one structural break regardless of the order of integration of time series. Results are reported for two different model specifications according to whether the structural break affects the level of the time trend (Model II) or both, the level and the slope of the time trend – hereafter, Model III. Although Model II is the specification that matches the model setup that has been considered in this paper, results using Model III are also reported to assess the robustness of the conclusions at this stage of the analysis. Regardless of the model that is used, the null hypothesis of no structural break is rejected at the 5% level of significance for revenues, but not for expenditures – note that the estimated break fraction is almost the same for both time series for Model III. Consequently, there is evidence that the presence of structural breaks should be accounted for when testing the order of integration of the time series, especially for revenues.

The order of integration analysis that is carried out also bases on the specifications given by Models II and III allowing for up to two structural breaks. The unit root hypothesis cannot be rejected for expenditures either by the Zivot and Andrews (1992) test – ADF test with one structural break – or by the Lumsdaine and Papell (1997) test – ADF test with two structural breaks, so that expenditures can be characterised as an I(1) non-stationary variable.[15] Evidence is not clear-cut for revenues, since the unit root hypothesis is rejected in all cases except for Model II with one structural break. However, it should be born in mind that Zivot and Andrews (1992) and Lumsdaine and Papell (1997) proposals make an asymmetric treatment of structural breaks since structural breaks are allowed only under the alternative hypothesis. This feature might influence statistical inference. In order to overcome this limitation, we compute the modified unit root tests in Carrion-i-Silvestre et al. (2009) that control for the presence of structural breaks both under the null and alternative hypotheses. The use of these statistics leads to characterize expenditures as an I(1) non-stationary variable, regardless of the model specification and number of structural breaks, which reinforces the evidence obtained so far for this variable. Except for Model II with one structural break, revenues are also characterized as an I(1) non-stationary variable. The discrepancy concerning revenues is solved if we take into account the fact that the BIC information criterion selects the two structural breaks specification.

Table 4. U.S. Government Budget Sustainability

(i) Robust CI tests, Model II, Full Sample (1947.I~2010.II)

leads & lags	$\hat{\mathbb{Q}}_r(1)$	$\hat{\mathbb{Q}}_r(2)$
1	15.08***	16.24***
$\hat{\pi}$	(.78)	(.32, .79)
3	14.45***	14.71***
$\hat{\pi}$	(.78)	(.32, .79)
5	13.15***	13.64***
$\hat{\pi}$	(.78)	(.32, .79)
7	12.55***	14.08***
$\hat{\pi}$	(.78)	(.33, .78)

Note: The numbers in parentheses are the break dates as fractions of

the sample size estimated under the null of CI.

(ii) Joint tests, Model II, Full Sample (1947.I~2010.II)

leads & lags	Joint CI/CB			Joint CI/CT		
	$\hat{\mathbb{Q}}_{cb}(1)$	$\hat{\mathbb{Q}}_{cb}(2)$	$\hat{\mathbb{Q}}_{cb}^{D\,\max}$	$\hat{\mathbb{Q}}_{ct}(1)$	$\hat{\mathbb{Q}}_{ct}(2)$	$\hat{\mathbb{Q}}_{ct}^{D\,\max}$
1	18.17***	23.13**	4.55***	33.59***	38.87***	6.45***
3	18.57***	23.58**	4.48***	29.83***	34.98***	4.98***
5	18.45***	23.59**	4.30***	26.76***	32.16***	4.17***
7	18.33***	23.62**	4.18***	25.61***	31.11***	3.90***

Let us now focus on the robust CI test and joint test statistics that have been proposed in this paper. We use only Model II, since an upward trend in each variable is clearly visible from Figure 1. In constructing the

---

[15] The order of the augmented Dickey-Fuller regression equation used for the computation of the unit root test statistics discussed in Zivot and Andrews (1992) and Lumsdaine and Papell (1997) has been selected using the t-sig criterion in Ng and Perron (1995), whereas it has been selected using the MAIC information criterion in Ng and Perron (2001) for the modified test statistics in Carrion-i-Silvestre et al. (2009). In all cases, we have allowed for up to $[12\,(T/100)^{1/4}]$ lags.

test statistics, we correct for the endogeneity bias via the DOLS device suggested by Saikkonen (1991), which augments the regression with leads and lags of the differenced stochastic regressors. Instead of selecting a particular number of leads and lags, we report our results for a range of values, from one to seven in steps of two. Finally, it should be mentioned that the main conclusion does not change at all with the choice of the left-hand side variable. Consequently, we only present the results obtained with the expenditures being the dependent variable and the revenues the regressor.

Table 4 (i) reports the results of the robust CI test for Model II. For each number of allowed breaks, we reject the null of CI at the 1% level of significance in all cases, which strongly suggests that there is no CI. Table 4 (ii) reports the test results of the joint tests in Model II. Using the relevant asymptotic critical values, we again reject the joint null of CI/CB or CI/CT, at least at the 5% level of significance, in all cases; note that rejection of the null hypothesis is found at the 1% level of significance in most cases. Taking all these elements into consideration, we can conclude that the test statistics that have been computed do not support any stable relationship between the revenues and the expenditures (i.e., cointegration is not found in this case).

7   Conclusion

We consider a set of variables that have both stochastic and broken linear trends. Of main interest is whether there is a stable relationship among the variables. Whenever a cointegration model with structural breaks is considered, one should carefully consider where the breaks originate. Our model deals with the case in which the relationship equation is subject to breaks because of breaks in each variable. When it is certain that each variable has no break but exogenous breaks are given to the relationship equation only, the traditional methods proposed in the literature might suffice. We extensively discuss what problems might emerge when the existing methods are applied to a set of variables with both stochastic and broken linear trends, and develop three new tests, a robust CI test, a joint test for CI and CB, and a joint test for CI and CT.

References

[1] Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative hypothesis, Econometrica 62:1383-1414.

[2] Arai, Y. and E. Kurozumi (2007). Testing for the null hypothesis of cointegration with a structural break, Econometric Reviews 26:705-739.

[3] Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes, Econometrica 66:47-78.

[4] Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models, Journal of Applied Econometrics 18:1-22.

[5] Bajo-Rubio, O., C. Diaz-Roldan, and V. Esteve (2008). US deficit sustainability revisited: a multiple structural change approach, Applied Economics 40:1609-1613.

[6] Bierens, H. J. (1997). Nonparametric cointegration analysis, Journal of Econometrics 77:379-404.

[7] Bierens, H. J. (2000). Nonparametric nonlinear co-trending analysis, with an application to inflation and interest in the U.S., Journal of Business and Economic Statistics 18:323-337.

[8] Campbell, J.Y. and P. Perron (1991). Pitfalls and opportunities: what macroeconomists should know about unit roots. NBER Macroeconomics Annual 6:141-201.

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

[10] Carrion-i-Silvestre, J. L. and A. Sansó (2006). Testing the null of cointegration with structural breaks, Oxford Bulletin of Economics and Statistics 68:623-646.

[11] Chapman, D. A. and M. Ogaki (1993). Cotrending and the stationarity of the real interest rate, Economics Letters 42:133-138.

[12] Gregory, A. W. and B. E. Hansen (1996). Residual-based tests for cointegration in models with regime shifts, Journal of Econometrics 70:99-126.

[13] Harris, D., D. Harvey, S. J. Leybourne, and A. M. Robert Taylor, (2009). Testing for a unit root in the presence of a possible break in trend, Econometric Theory 25:1545-1588.

[14] Harris, D., S. J. Leybourne and A. M. Robert Taylor (2016). Tests of the co-integration rank in VAR models in the presence of a possible break in trend at an unknown point, Journal of Econometrics 192:451-467.

[15] Hansen, B. E. (1992). Tests for parameter instability in regressions with I(1) processes, Journal of Business and Economic Statistics 10:321-335.

[16] Hansen, P. R. (2003). Structural changes in the cointegrated vector autoregressive model, Journal of Econometrics 114:261–295.

[17] Hao, K. (1996). Testing for structural change in cointegrated regression models: some comparisons and generalizations, Econometric Reviews 15:401–429.

[18] Harvey, D., S. J. Leybourne and A. M. Robert Taylor (2012). Unit root testing under a local break in trend, Journal of Econometrics 167:140-167.

[19] Harvey, D., S. J. Leybourne and A. M. Robert Taylor (2013). Testing for unit root in the possible presence of multiple trend breaks using minimum Dickey-Fuller statistics, Journal of Econometrics 177:265-284.

[20] Hatanaka, M. and H. Yamada (2003). Co-trending: a statistical system analysis of economic trends, Springer.

[21] Haug, A. A. (1991). Cointegration and government borrowing constraints: evidence for the United States, Journal of Business and Economic Statistics 9:97-101.

[22] Hendry, D. F. and G. E. Mizon (1998). Exogeneity, causality and co-breaking in economic policy analysis of a small econometric model of money in the UK, Empirical Economics 23:267-294.

[23] Holmes, M. J. (2010). Nonlinearities, co-trending and budget balance sustainability, Economic Analysis and Policy 40:369-376.

[24] Inoue, A. (1999). Tests of cointegrating rank with a trend-break, Journal of Econometrics 90:215-237.

[25] Jansson, M. (2005). Point optimal tests of the null hypothesis of cointegration, Journal of Econometrics 124:187-201.

[26] Johansen, S., R. Mosconi, and B. Nielsen (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend, Econometrics Journal 3:216-249.

[27] Kejriwal, M. and P. Perron (2008). Data dependent rules for selection of the number of leads and lags in the dynamic OLS cointegrating regression, Econometric Theory 24:1425-1441.

[28] Kejriwal, M. and P. Perron (2010). Testing for multiple structural changes in cointegrated regression models, Journal of Business and Economic Statistics 28:503-522.

[29] Kim, D. and P. Perron (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.

[30] King, M. L. (1980). Robust Test for spherical symmetry and their application to least squares regression, Annals of Statistics 8:1265-1271.

[31] Kurozumi, E. and Y. Arai (2008). Test for the null hypothesis of cointegration with reduced size distortion, Journal of Time Series Analysis 29:476-500.

[32] Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root, Journal of Econometrics 54:159-178.

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

[34] Leybourne, S. J. and B. P. M. McCabe (1993). A simple test for cointegration, Oxford Bulletin of Economics and Statistics 55:97-103.

[35] Liu, P. and E. Tanner (1994). Is the budget deficit "Too Large"?: some further evidence, Economic Inquiry 32:511-518.

[36] Lütkepohl, H., P. Saikkonen, and C. Trenkler (2004). Testing for the cointegrating rank of a VAR process with level shift at unknown time, Econometrica 72:647-662.

[37] Martin, G. (2000). US deficit sustainability: A new approach based on multiple endogenous breaks, Journal of Applied Econometrics 15:83-105.

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

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

[40] Ogaki, M. and J.Y. Park (1997). A cointegration approach to estimating preference parameters, Journal of Econometrics 82:107-134.

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

[42] Perron, P. (1991). A test for changes in a polynomial trend function for a dynamic time series. Mimeo. Princeton University Department of Economics.

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

[44] Perron, P. (2005). Dealing with structural breaks, Palgrave Handbook of Econometrics 1: Econometric Theory, K. Patterson and T.C. Mills (eds.), 278-352.

[45] Perron, P. and T. Yabu (2009). Testing for Shifts in Trend with an Integrated or Stationary Noise Component, Journal of Business and Economic Statistics 27:369-396.

[46] Perron, P. and T.J. Vogelsang (1992). Nonstationarity and level shifts with an application to purchasing power parity. Journal of Business & Economic Statistics 10:301-320.

[47] Phillips, P. C. B. and S. Ouliaris (1990). Asymptotic properties of residual based tests for cointegration, Econometrica 58:165-193.

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

[49] Seo, B. (1998). Tests for structural change in cointegrated systems, Econometric Theory 14:222–259.

[50] Qu, Z. (2007). Searching for cointegration in a dynamic system, Econometrics Journal 10:580–604.

[51] Quintos, C. E. (1995). Sustainability of the deficit process with structural shifts, Journal of Business and Economic Statistics 13:409-417.

[52] Saikkonen, P. and H. Lütkepohl (2000). Testing for the cointegrating rank of a VAR process with structural shifts, Journal of Business and Economic Statistics 18:451-464.

[53] Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration, Econometric Theory 10:91-115.

[54] Trenkler, C., P. Saikkonen, and H. Lütkepohl (2007). Testing for the cointegrating rank of a VAR process with level shift and trend break, Journal of Time Series Analysis 29:331-358.

[55] Vogelsang, T. (1999). Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series, Journal of Econometrics 88:283-299.

[56] 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 and Economic Statistics 10:251-270.