n° 9810

Feedback Covariates Unit Root Tests

An Application to the Sustainability of Fiscal Policy

Patrick Fève[(1)], Pierre-Yves Hénin[(2)] and Philippe Jolivaldt[(3)]

(1) Université de Nantes, LEN–C3E and CEPREMAP
(2) CEPREMAP and MAD–Université Paris I
(3) CEME and Université Paris I

We thank J. Adda, F. Collard, C. Gourieroux, S. Grégoir, C. Hurlin, F. Langot, P. N'Diaye and X. Guyon for fruitful comments. We thank H. Bohn for kindly providing the US data. A previous version of this paper has also benefited from discussions during presentation at the *Econometric Society European Meeting* in Toulouse 1997. All remaining mistakes are our own.

Tests de racine unitaire utilisant des covariables dans un processus de rétroaction

Une applicatiopn à la soutenabilité de la politique fiscale

Patrick Fève, Pierre-Yves Hénin and Philippe Jolivaldt

Résumé

Les tests de soutenabilité ignorent souvent la dynamique jointe des variables de flux et de stocks. Cet article montre qu'une telle pratique peut induire une mauvaise inference statistique, mais que la prise en compte de cette dynamique jointe conduit à d'importants gains de puissance. L'article introduit une nouvelle statistique, notée FADF, qui combine au mieux des effets de retour et les co–variations. Les distributions limites sont obtenues ainsi que les fonctions de puissance asymptotique. Un exercie de simulation et une application basée sur la soutenabilité de la dette américaine illustrent les potentialités de cette approche.

Feedback Covariates Unit Root Tests

An Application to the Sustainability of Fiscal Policy

Patrick Fève, Pierre-Yves Hénin and Philippe Jolivaldt

Abstract

Sustainability tests often ignore the joint dynamics of stock and flow variables. This paper shows that such a practice leads to inappropriate statistical inference, and that taking into account this joint dynamics when testing for unit root induces large power gains. This paper introduces a new statistics, denoted FADF (*Feedback Augmented Dickey Fuller*), which fully combines feedback effects and covariates. The paper derives the limit distribution of the test statistic and comments on the asymptotic power functions. A simulation study and an empirical application based on U.S. public debt sustainability illustrate the potential of this approach.

Keywords: sustainability, unit root, covariates, feedback mechanism.

JEL classification: C12, C22, H62

1 Introduction

Testing for unit root becomes omnipresent in time series econometrics. However, doubt was casted and a debate was engaged on both the usefulness and the power of unit root tests, following the twin interrogations:[1] "*unit root: do we know ? do we care* ?" . As a partial answer to these questions, the issue of debt sustainability may be quoted as the one for which the presence of a unit root actually cares. Persistent deficits and increases in public debt in most of OECD countries during the nineties explain the large body of empirical studies devoted to sustainability of fiscal policies. Following the pioneering work of Hamilton and Flavin (1986) for the US economy, many studies have raised this issue, extending the sample and looking at comparative evidences on various OECD countries.

However, these studies are far from agreeing on the findings. To some extent, contradictory results may reveal the diversity of national experiences, or may be the consequence of structural shifts, as consequences of changes in fiscal policy (see Haug (1995) and Quintos (1995)). Alternatively, they can be explained by the large variety of specifications and associated regression frameworks adopted in order to test for sustainability. A main approach of sustainability test is based on unit root test in the debt or deficit processes (see Hamilton and Flavin (1986) or Wilcox (1989)). As such, they suffer from the lack of power (see Campbell and Perron (1991), Schwert (1989)), especially in small sample.[2] Similar difficulties appear when the sustain-

[1] This is in the lines of the influential paper of Christiano and Eichenbaum (1990), and many related contributions.

[2] As mentioned by Hansen (1995), a refrain often heard in applied macroeconometric circles is that *unit root tests have low power*. This view may be particulary true if we consider unit root tests on stock variables without accounting for their joint dynamics with flows variables. The basis of this paper is in part to demonstrate that low power results from this ignorance.

ability tests are performed in a cointegration framework (see Hakkio and Rush (1991), Smith and Zin (1991) and Trehan and Walsh (1988), (1991)). Moreover, as mentioned by Wickens and Uctum (1993) and Bohn (1991b), cointegration is not a sufficient condition for sustainability. They suggest an alternative approach resulting from the following idea: if primary surplus does not respond positively to higher initial debt, government policy is not sustainable. This result simply implies that public debt explodes at a rate greater than the discount factor and so the expected discounted debt fails to converge, violating the transversality condition.[3]

Rather than discarding unit root and cointegration tests, one may reformulate them in order to increase power. An important advance in this way is provided by the *Covariate Augmented Dickey Fuller* (CADF) test proposed by Hansen (1995). He shows how the introduction of a stationary covariate may increase the power of the unit root test. Our basic idea in this paper is that more power may be gained by introducing a covariate informative on the adjustment mechanism toward some equilibrium value and not only on short term components unrelated to long run dynamics. We propose here a *feedback unit root test* (or *Feedback Augmented Dickey Fuller* test), denoted FADF. As a sustainability test, the FADF results from embedding the Bohn (1991b) and Wickens and Uctum (1993) auxiliary regression approach within an ADF framework. The null hypothesis corresponds to a joint restriction on the two equations defining the system. This leads to test for a unit root in the process of the variable of interest and for the nullity of the feedback mechanism in an auxiliary equation.

[3]It is worth noting that neither Bohn (1991b), nor Wickens and Uctum (1993), cope with these technical difficulties, which result from the fact that the debt is integrated of order one under the null of non–sustainability. In this paper, we will cope with this issue in a way similar, but not identical, to the one adopted by Cavanagh, Elliott and Stock (1995).

Out test improves previous unit root and cointegration tests in two ways. First, the basis of this statistical inference is to consider that a feedback effect occurs through the estimated parameter in the auxiliary equation. Second, using stationary covariates allows to increase the power of the statistics, as shown by Hansen (1995) in the univariate case. The paper first studies the properties of the maximum likelihood estimate in a system based equations. The limit distribution of the FADF *t-statistics* is shown to be a mixture of the standard normal and Dickey-Fuller distributions, whose weights are determined by a variance ratio and a correlation parameter between the structural disturbances. Another focus of the paper is the comparison with other approaches: univariate unit root test (Hamilton and Flavin (1986) , Wilcox (1989)) and single auxiliary regression approach (Wickens and Uctum (1993), Bohn (1991b)). We show that if the data are generated by the initial model under the null hypothesis of a unit root, none of these previous approaches leads to a correct size for the statistics, except in very particular cases. Moreover, the associated statistics display very low power.

The paper is organized as follows. Section 2 introduces the FADF statistic associated to our sustainability test. In section 3, we derive and comment the asymptotic distribution of this new unit root test statistic. Methodological issues are also addressed. Section 4 presents some simulations experiments. Section 5 illustrates the potential of this approach with US data. The last section concludes. Proofs are given in appendix.

2 Introducing FADF as a Sustainability Test

In this section, we will first review the justification of the unit root test for sustainability. We start from the following law of motion of the public debt

$$B_t = (1 + r_{t-1})B_{t-1} - S_t + \mathcal{E}_t \ ,$$

where B_t is the market value of the government debt at constant prices, r_{t-1} is the ex post real interest rate and S_t is the primary surplus at constant prices. Following Hamilton and Flavin (1986), we add an error term $\mathcal{E}_t$ which is supposed to capture the some *measurement errors* in the previous account identity.[4] Such measurement errors may appear because of the timing of purchases or sales of public bonds or because measurements of the market value of the government indebtedness are necessary imperfect.[5] Another source of errors comes from the assumption of constant real interest rate. This term allows for the fact that the series S_t and B_t do not verify an exact linear constraint each period (see Bohn (1991)). If the expected real return is assumed to be constant (see Hamilton and Flavin (1986), Trehan and Walsh (1988), (1991)), this error term represents an expectation error and it must be serially uncorrelated. If not, it may represent an approximation error on public debt identity and thus may be serially correlated.

2.1 Testing for sustainability

We consider here sustainability as the satisfaction of the intertemporal budget constraint (see Wilcox(1989)). Let q_t be the real discount factor from period t back to period zero

$$q_t = \prod_{i=1}^{t} (1 + r_{t-i})^{-1} \ ,$$

[4]The term *measurement errors* must be employed carefully, because if it implies that the public debt is subject to measurement errors, this term would result from the accumulation of these measurement errors. Moreover, this suggests another estimation methods, for example instrumental variables estimators. This induces the study of the asymptotic distribution of such estimators, which is beyond the scope of this paper.

[5]Beyond statistical adjustments, another rationale of such a residual is provided by Holzmann, Herve and Demmel (1996). This residual includes the following components: the valuation difference between nominal and market values, the valuation effect on government assets and liabilities of changes in the exchange rates, the net increase in other government assets (industrial participation, trade credit, ...). According to Holzmann, Herve and Demmel (1996), this residual accounts roughly for one third of the change of the debt, *i.e.* as large as half the importance of the reported deficits.

with $q_0 = 1$. Under the assumption that q_t is known at period t, we have

$$b_t = b_{t-1} - s_t + \varepsilon_t \ , \tag{1}$$

where $b_t = q_t B_t$, $s_t = q_t S_t$ and $\varepsilon_t = q_t \mathcal{E}_t$. In expected terms, the current value of the discounted debt equals the sum of expected future noninterest surpluses if the expectation of the discounted debt goes to zero in the limit

$$\lim_{T \to \infty} E_t b_{t+T} = 0 \ , \forall t \tag{2}$$

The sustainability condition (2) means that the value of the debt must grow at a slower rate than the discount factor, precluding *Ponzi* financing. As a consequence, it insures that the value of the debt equals the expected sum of future discounted surpluses

$$b_{t-1} = E_t \sum_{j=0}^{\infty} \{s_{t+j} - \varepsilon_{t+j}\} \ , \forall t \tag{3}$$

Sustainability has to be satisfied in expectations but testable implications can be obtained through further assumptions, such that the rules and conditions observed for a given sample period will continue to prevail. An approach frequently used in the literature is based on equation (3) (see Trehan and Walsh (1988)). Given the stochastic processes generating s_t and ε_t which are assumed to be exogenous with respect to b_{t-1}, equation (3) is solved. Such a practise is not relevant for statistical inference and leads to inappropriate statistical evaluation.[6] To see this suppose that the primary surplus and the error term in the debt identity are both *iid*. From (3), one has $b_{t-1} = s_t - \varepsilon_t$, which provides an alternative formulation, irrespective to the debt identity (1). The difference disappears if we impose the restriction $b_t = 0$ each period. A sufficient condition for non sustainability is thus $E\left[s_t / b_{t-1}\right] = 0$ and $b_{t-1} \neq 0$. In what follows, we use another approach based on the assumption that the surplus is an endogenous variable which depends on public debt.

[6] See Wickens and Uctum (1993) for a general discussion on this point.

2.2 The regression framework and empirical issues

Endogenous surplus can be simply obtained by the specification of a policy function of the government. The model proposed here assumes that the government adopts a debt smoothing behavior. Suppose that a government chooses a sequence of $\{b_{t+i}\}_{i=0}^{\infty}$ in order to minimize the following intertemporal loss function, conditional on its current information set

$$L = E_t \left\{ \sum_{i=0}^{\infty} \zeta^i \{\theta(b_{t+i} - \overline{b}_t)^2 + (b_{t+i} - b_{t+i-1})^2\} \right\} , \tag{4}$$

where ζ is the discount factor and $\overline{b}_t$ a *desired* level of the discounted debt. The amplitude of fluctuations around the target depend on the magnitude of the relative cost[7] parameter θ. The first order condition is a second-order difference equation of the form

$$\Delta b_t = \zeta E_t \Delta b_{t+1} - \theta(b_t - \overline{b_t}) \tag{5}$$

A stable and particular solution to (5) is given by the following forward looking equation

$$b_t = \rho b_{t-1} + (1-\rho)(1-\rho\zeta) E_t \sum_{i=0}^{\infty} (\zeta\rho)^i \overline{b}_{t+i} \tag{6}$$

where ρ is the stable root of the characteristic equation of (5), i.e. imposing the transervality condition on the discounted debt $\lim_{T\to\infty} E_t(\zeta\rho)^{-T} b_{t+T} = 0$. We assume for simplicity that the target $\overline{b}_t$ is *iid*. This assumption insures also that the effect of b_{t-1} on s_t correspond to $1-\rho$. From (6) and (1), one deduces

$$s_t = (1-\rho) b_{t-1} + u_t \tag{7}$$

where $u_t = \varepsilon_t - (1-\rho)(1-\rho\zeta)\overline{b_t}$. The more it is costly to deviate from $\overline{b}_t$, the more s_t must respond positively to inherited public debt. Plugging (7)

[7] For example, if $\theta \to \infty$, $\overline{b}_t$ represents the optimal level of public debt for each period.

into the debt identity yields

$$b_t = \rho b_{t-1} + \varepsilon_t - u_t \tag{8}$$

Equations (7) and (8) show the equivalence between the two following conditions:

i) the discounted debt follows a unit root process.

ii) the discounted primary surplus is insensitive to the inherited value of the discounted debt.

We propose to test for non-sustainability as a test for unit root in equations (7) and (8).Whereas Bohn (1991b) and Wickens and Uctum (1993) exploit the equation (7) and test for $1-\rho=0$, Wilcox (1989) only considers equation (8) and tests for a unit root in this equation . Our *feedback unit root* inference is based on the joint test of the null hypothesis $\rho = 1$ in the system based equation (7)–(8) (or equivalently (1)-(7)) against the alternative $\rho < 1$. Sustainability holds if the government responds to higher initial debt by increasing the primary surplus. This condition is obviously not sufficient for satisfying the transversality condition. Sufficient conditions are obtained if both ε_t and u_t follow zero mean stochastic processes.

3 Estimation and Inference

Inference on sustainability may be conducted in the general following framework, where y_t denotes the stock variable and x_t the related flow variable. Let $\{y_t, x_t\}$ be time series generated by

$$y_t = y_{t-1} - x_t + \varepsilon_t \tag{9}$$

$$x_t = (1-\rho)y_{t-1} + u_t \tag{10}$$

where the null hypothesis to be tested is $\rho = 1$.

Assumptions

A1: the sequence $\{\varepsilon_t, u_t\}$ is *iid* and normally distributed

A2: $E(\varepsilon_t/\underline{Y_{t-1}}) = 0$ and $E(u_t/\underline{Y_{t-1}}) = 0$, with $\underline{Y_{t-1}} = \{y_{t-1}, y_{t-2}, ..., y_0\}$

A3: $\sup_t E(\varepsilon_t^4) < \infty$ and $\sup_t E(u_t^4) < \infty$

A4: $x_0 = u_0 = 0$ and $y_0 = \varepsilon_0 - u_0 = 0$

Assumption A1 is introduced in order to simplify the asymptotic distribution of the statistics. Assumption A2 states that the errors terms are orthogonal to the past values of the regressors. Assumption A3 insures integrability conditions and A4 fixes the initial conditions to zero. We do not exclude contemporaneous correlation between the two error terms. The covariance matrix of $\{\varepsilon_t, u_t\}$ is given by

$$\Sigma_1 = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\ u} \\ \sigma_{\varepsilon\ u} & \sigma_u^2 \end{pmatrix} \equiv \sigma_u^2 \begin{pmatrix} \eta^2 & \tau\eta \\ \tau\eta & 1 \end{pmatrix}$$

where $\tau = Cov(\varepsilon_t, u_t)(\sigma_\varepsilon \sigma_u)^{-1}$ is the correlation parameter between ε_t and u_t, which verifies $|\tau| < 1$ and $\eta^2 = \sigma_\varepsilon^2 \sigma_u^{-2}$ is the variance ratio. In what follows, we will express the limit distribution of the test statistics in terms of these two parameters. From (9)-(10), we deduce the following reduced form

$$y_t = \rho y_{t-1} + \varepsilon_t - u_t \ , \tag{11}$$

$$x_t = (1-\rho) y_{t-1} + u_t \ , \tag{12}$$

where the covariance matrix of $\{\varepsilon_t - u_t, u_t\}$ is now given by

$$\Sigma_2 = \sigma_u^2 \begin{pmatrix} 1 + \eta^2 - 2\tau\eta & \tau\eta - 1 \\ \tau\eta - 1 & 1 \end{pmatrix}$$

A natural test statistic for the hypothesis of unit root in model (11)-(12) is the *t-statistic* defined as

$$t_{(\widehat{\rho}-1)} = \frac{\widehat{\rho} - 1}{s(\widehat{\rho} - 1)}$$

where $s(\widehat{\rho}-1)$ is the standard-error for $\widehat{\rho}-1$. We refer to this regression framework as *case 1*. In what follows, the *t-statistic* is denoted FADF.

3.1 Preliminary Asymptotic Results

The asymptotics provided here are related to a prototypical model, under a set of restrictive assumptions. This is partly because it makes the limit distributions easier to obtain[8]. The asymptotic theory is based on "local-to-unity" asymptotic, following Phillips (1987). The model (11)-(12) has a unit root when $\rho = 1$. This constitutes the null hypothesis of interest. We consider here local departure from the null hypothesis with setting

$$\rho = 1 - \frac{c}{T}$$

The null hypothesis holds when $c = 0$ and holds "locally" as $T \to \infty$ for $c \neq 0$. The asymptotic theory for near-integrated processes uses the diffusion representation. We define the functional $B_1^c(r)$ generated by the stochastic differential equation[9] $dB_1^c(r) = -cB_1^c(r) + dB_1(r)$. The Gaussian log likelihood of (11)-(12) is, up to a constant term

$$\ell(\rho, \Sigma_2) = -(T/2)\ln|\Sigma_2| - (1/2)\sum_{t=1}^{T}(\varepsilon_t - u_t; u_t)\Sigma_2^{-1}(\varepsilon_t - u_t; u_t)' \qquad (13)$$

To extract the relevant asymptotics, we first use the following properties.

Lemma 1 *We have under the local alternative* $c \neq 0$

$$\frac{1}{T}\sum y_{t-1}(\varepsilon_t - u_t) \implies \sigma_u^2(1 + \eta^2 - 2\eta\tau)\int_0^1 B_1^c dB_1(r)$$

[8] Clearly, to provide a more operational framework, some of the previous assumptions must be relaxed, in order to take into account serial correlation. The simulation experiments indicate also low size distortion in presence of autocorrelated errors after an appropriate correction.

[9] This functional is also known as an Ornstein-Uhlenbeck process, whose the solution is given by $B_1^c(r) = B_1(r) - c\int_0^1 \exp(-c(r-s))B_1(s)ds$.

$$\begin{aligned}
\frac{1}{T^2}\sum y_{t-1}^2 &\Longrightarrow \sigma_u^2(1+\eta^2-2\eta\tau)\int_0^1 \left\{B_1^c\right\}^2 dr \\
\frac{1}{T}\sum y_{t-1}u_t &\Longrightarrow \sigma_u^2(\eta\tau-1)\int_0^1 B_1^c dB_1(r) \\
&\quad +\sigma_u^2\eta(1-\tau^2)^{1/2}\int_0^1 B_1^c dB_2(r)
\end{aligned}$$

where B_2 is independent of B_1 and B_1^c.

Given the score function of (13) and using Lemma 1, we obtain the following asymptotic distribution for the FADF *t-statistic*

Theorem 1 *Under the local alternative $c \neq 0$, as $T \to \infty$*

$$t_{(\widehat{\rho}-1)} \Longrightarrow -\frac{c}{\xi}\left\{\int_0^1 \left\{B_1^c(r)\right\}^2 dr\right\}^{1/2} + \xi\frac{\int_0^1 B_1^c(r)dB_1(r)}{\left\{\int_0^1 \left\{B_1^c(r)\right\}^2 dr\right\}^{1/2}} + (1-\xi^2)^{1/2}\mathcal{N}(0,1)$$

Under the null hypothesis $c = 0$

$$t_{(\widehat{\rho}-1)} \Longrightarrow \xi\frac{\int_0^1 B_1(r)dB_1(r)}{\left\{\int_0^1 \left\{B_1(r)\right\}^2 dr\right\}^{1/2}} + (1-\xi^2)^{1/2}\mathcal{N}(0,1)$$

where $\xi = (1-\tau^2)^{1/2}(1+\eta^2-2\eta\tau)^{-1/2}$

The effect of parameter uncertainty has asymptotically no effect on the distribution of the FADF *t-statistic*, under both the local alternative and the null hypothesis. The distribution is also similar to the one obtained with a known covariance matrix of the residuals. The limiting distribution of $t_{(\widehat{\rho}-1)}$ depends under the local alternative and the null hypothesis on the nuisance parameter ξ . This parameter is consistently estimated by the sample correlation between $\widehat{\varepsilon}_t$ and $\widehat{u}_t$ and their variance ratio. This is because we can treat ξ as known for the purposes of asymptotic theory.

The mixture representation of the limit distribution of the *t-statistic* is a simple consequence of the independence of the Brownian motions B_1 and B_2. The distribution of the FADF *t-statistic* is a mixture of Dickey-Fuller type

and standard normal distribution.[10] As $\xi^2 \to 1$, we obtain a Dickey-Fuller type distribution, whereas $\xi^2 \to 0$ the normal distribution. The parameter ξ appears as a non–linear function of η and of τ. The complexity of this relation is illustrated in the figure 1. We have drawn the weights ξ and $(1-\xi^2)^{1/2}$ with τ in abscissa, for different values of η. At a first sight, it is difficult to obtain general interpretation. However, we observe that, when η grows, the value of ξ decreases. Except for η small, there is asymmetry in the weights.[11] When $\tau = \eta$ for values less than one, ξ is near one and the statistic is alike a Dickey-Fuller distribution. At the opposite, when τ is near -1 or 1, the normality is predominating, except for $\eta = 1$. This last situation arises also when η is very large. For instance, when $\eta = 10$, ξ never exceeds 0.1 and $(1-\xi^2)^{1/2}$ is approximately one.

These previous particular cases may be explained using the ML estimate of ρ. Because the uncertainty on η and τ does not matter asymptotically (see the appendix), the estimate $\widehat{\rho}$ of ρ has the following simple expression

$$\widehat{\rho} = 1 + \frac{\tau}{\eta}\frac{\sum y_{t-1}\Delta y_t}{\sum y_{t-1}^2} + \frac{\tau-\eta}{\eta}\frac{\sum y_{t-1}x_t}{\sum y_{t-1}^2}$$

If $\eta = \tau$, the second terms vanishes. This is because there exists now only one source of noise in the system based equations. One thus obtains limit distribution of univariate unit root test (see Phillips (1987))

$$t_{(\widehat{\rho}-1)} \Longrightarrow -c\left\{\int_0^1 \left\{B_1^c(r)\right\}^2 dr\right\}^{1/2} + \frac{\int_0^1 B_1^c(r)dB_1(r)}{\left\{\int_0^1 \left\{B_1^c(r)\right\}^2 dr\right\}^{1/2}}$$

When $\tau = 0$, the estimated value of ρ correspond to the single regression approach without covariates. The fact to ignore covariates induce a mixture

[10] The combination is not linear and the sum of the weights is not equal to one.

[11] An exception appears with $\eta = 1$. In this case, one has $\xi = (1+\tau)^{1/2}/\sqrt{2}$. As τ grows from -1 to 1, the limit distribution moves from a Dickey-Fuller distribution to a normal one.

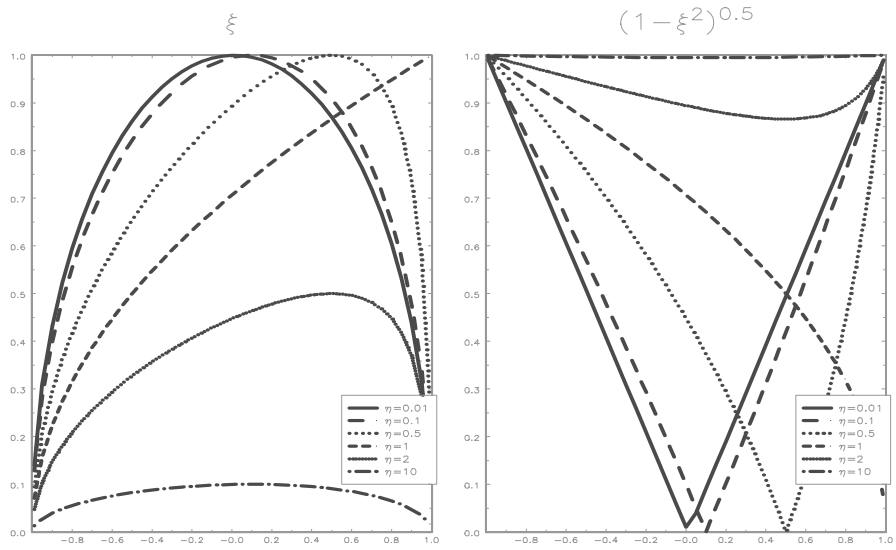

Figure 1: Values of the weights in case 1 (FADF *t-statistics*)

of Dickey-Fuller and normal distribution (see Cavanagh, Elliott and Stock (1995))

$$t_{(\widehat{\rho}-1)} \implies -c(1+\eta^2)^{1/2}\left\{\int_0^1 \{B_1^c(r)\}^2\,dr\right\}^{1/2} + (1+\eta^2)^{-1/2}\frac{\int_0^1 B_1^c(r)dB_1(r)}{\left\{\int_0^1 \{B_1^c(r)\}^2\,dr\right\}^{1/2}} + \eta(1+\eta^2)^{-1/2}\mathcal{N}(0,1)$$

The limit distribution of the FADF *t-statistic* shows that the large power gain are associated to a normal limit distribution. First, a small value of ξ implies a departure from Dickey-Fuller type distribution. Second, for $c > 0$ the limit distribution takes large negative values. For large sample size, one may expect significant power gains, even for a small value of c if $\xi \to 0$.[12]

[12]This can be easily illustrated from the previous limit distribution.. For η large, one has $-c(1+\eta^2)^{1/2} \to -\infty$, $(1+\eta^2)^{-1/2} \to 0$ and $\eta(1+\eta^2)^{-1/2} \to 1$. Under the null, the limit distribution is thus a normal one and under the alternative the limit distribution may take large negative values.

3.2 Comparison with other approaches

This section is devoted to methodological issues, looking at limit distributions of the statistics when statistical inference is based on misspecified regression models. Two types of misspecifications are considered in this paper.

Case 2 assumes that the variable y_t is integrated of order one and ignores cross equation restrictions

$$y_t = y_{t-1} + \varepsilon_t - u_t \tag{14}$$

$$x_t = (1-\rho)y_{t-1} + u_t \tag{15}$$

This approach ignores uncertainty on unit root in y_t , but takes into account for covariates. It corresponds to ML in cointegrated system (Phillips (1991)) or FM-OLS in a single regression framework (Phillips and Hansen (1990)). We look at the limit distribution of the *t-statistic* of $\widehat{\rho}-1$ in a single regression after correction for covariates

$$x_t = (1-\rho)y_{t-1} + \delta\Delta y_t + e_t$$

In what follows, the associated *t-statistic* is denoted COLS (*Covariates OLS*).

Case 3 assumes that the variable x_t is *iid*

$$y_t = \rho y_{t-1} + \varepsilon_t - u_t \tag{16}$$

$$x_t = u_t \tag{17}$$

In this approach, uncertainty on ρ in the auxiliary equation is ignored, but covariates are taken into account. It corresponds to the single regression approach to unit root tests with covariates proposed by Hansen (1995)

$$y_t = \rho y_{t-1} + \gamma x_t + v_t$$

In what follows, the associated *t-statistic* is denoted CADF (*Covariates ADF*).

In *case 2,* we obtain the following asymptotics for the COLS *t-statistic*:

Theorem 2 *Under the local alternative* $c \neq 0$*, as* $T \to \infty$ *, if* $\{y_t, x_t\}$ *follows the process (11)-(12) with* $\rho = 1 - \frac{c}{T}$*, then as* $T \to \infty$

$$t_{(\widehat{\rho}-1)} \Longrightarrow -c\frac{(\eta-\tau)}{(1-\tau^2)^{1/2}}\left\{\int_0^1 \{B_1^c(r)\}^2\, dr\right\}^{1/2} + \mathcal{N}(0,1)$$

Under the null hypothesis $c = 0$

$$t_{(\widehat{\rho}-1)} \Longrightarrow \mathcal{N}(0,1)$$

Two noticeable results occurs. First, under the null hypothesis, the limit distribution of the COLS *t-statistic* is a normal one. This distribution is free of the nuisance parameter ξ. This result corresponds to the limit distribution of maximum likelihood estimate (see Phillips (1991)) or FM-OLS estimate (see Phillips and Hansen (1990)). Second, it appears that the test statistic has no power if $\eta = \tau$, for values of η less than one in absolute value. Indeed, the single regression approach after correction for covariates is now $\Delta y_t + x_t = (1-\rho)y_{t-1} + \eta\sigma_u e_t$ where e_t is $iid(0,1)$ and $E(e_t, \Delta y_t) = 0$. This last equation correspond to the identity $y_t = y_{t-1} - x_t + \varepsilon_t$, which induces no power for the COLS *t-statistic*. In this case, the limit distribution is always a normal one, for any values of c . Moreover, if $\eta < \tau < 1$, the test statistic is biased. Power gain may be expected only if the variance ratio is larger than the correlation parameter.

We now look at the limit distribution in *case 3* (equations (16)-(17)). Thus, we obtain the following asymptotics for the CADF *t-statistic*:

Theorem 3 *Under the local alternative* $c \neq 0$*, as* $T \to \infty$ *, if* $\{y_t, x_t\}$ *follows the process (11)-(12) with* $\rho = 1 - \frac{c}{T}$*, then as* $T \to \infty$

$$t_{(\widehat{\rho}-1)} \Longrightarrow -\frac{c\tau}{\xi}\left\{\int_0^1 \{B_1^c(r)\}^2\, dr\right\}^{1/2} + \eta\xi\frac{\int_0^1 B_1^c(r)dB_1(r)}{\left\{\int_0^1 \{B_1^c(r)\}^2\, dr\right\}^{1/2}} + (1-\eta^2\xi^2)^{1/2}\mathcal{N}(0,1)$$

Under the null hypothesis $c = 0$

$$t_{(\widehat{\rho}-1)} \Longrightarrow \eta\xi \frac{\int_0^1 B_1(r)dB_1(r)}{\left\{\int_0^1 \{B_1(r)\}^2 dr\right\}^{1/2}} + (1-\eta^2\xi^2)^{1/2}\mathcal{N}(0,1)$$

The limit distribution is again a mixture of Dickey-Fuller and normal distribution. Under the null hypothesis of unit root, the weights are identical to *case 1* only in the particular case $\eta = 1$. Nevertheless, they are now an increasing function of η , *i.e.* we obtain a Dickey-Fuller type distribution when the variance of ε_t is larger than the one of u_t. If $\tau = 0$, the test statistic has now power. This is because the two residuals in (9)-(10) are independent and the unit root is maintained whatever the values of c. Indeed, the regression framework may be written as $y_t = \rho y_{t-1} - x_t + \varepsilon_t$ which correspond to equation (9). Moreover, if $\tau < 0$, the test statistic is biased.[13]

4 Simulation experiments

We present three types of experimental results. The first concerns the distributions of the FADF *t-statistic*. The second provides some results on local power functions. The third uses a size-power curve in order to study the effects of serial correlation.[14]

4.1 Critical values of FADF *t-statistic*

We provide critical values of the *t-statistic* in *case 1*. We select five sample sizes: $T = 25,\ 50, 100, 250, 500$. The asymptotic case is deduced from direct calculation of the Brownian functionals[15]. We have to choice different pairs

[13] In these two cases, a single regression approach without covariates seems to be more adequate.

[14] All these results are obtained, in each case, from 40,000 samples of various sizes, with *iid* Gaussian innovations. Computation are made with GAUSS, ver. 3.2.13.

[15] Our experience indicates that a Riemann approximation with a partition of $[0, 1]$ by 200 points appears sufficient to stabilize the results.

Table 1: Critical values of the t-statistic

	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\boldsymbol{\tau = -0.5}$	$\boldsymbol{\eta = 0.1}$			$\boldsymbol{\eta = 1}$			$\boldsymbol{\eta = 10}$		
$(\ \xi\ ;(1-\xi^2)^{1/2})$	$(0.82; 0.57)$			$(0.50; 0.87)$			$(0.08; 1.00)$		
T=25	-2.79	-2.01	-1.62	-2.75	-1.92	-1.51	-2.59	-1.77	-1.36
T=50	-2.66	-1.94	-1.59	-2.65	-1.89	-1.50	-2.48	-1.73	-1.35
T=100	-2.59	-1.93	-1.59	-2.55	-1.87	-1.50	-2.42	-1.69	-1.33
T=250	-2.58	-1.93	-1.59	-2.54	-1.87	-1.49	-2.39	-1.69	-1.33
T=500	-2.56	-1.92	-1.59	-2.51	-1.84	-1.48	-2.38	-1.68	-1.32
T: ∞	-2.54	-1.92	-1.58	-2.49	-1.83	-1.47	-2.34	-1.68	-1.31
$\boldsymbol{\tau = 0.1}$	$\boldsymbol{\eta = 0.1}$			$\boldsymbol{\eta = 1}$			$\boldsymbol{\eta = 10}$		
$(\ \xi\ ;(1-\xi^2)^{1/2})$	$(1.00; 0.00)$			$(0.74; 0.67)$			$(0.10; 0.99)$		
T=25	-2.77	-2.02	-1.65	-2.75	-1.96	-1.58	-2.60	-1.79	-1.39
T=50	-2.69	-2.00	-1.64	-2.63	-1.94	-1.57	-2.45	-1.73	-1.37
T=100	-2.58	-1.95	-1.62	-2.61	-1.92	-1.56	-2.40	-1.70	-1.33
T=250	-2.58	-1.94	-1.62	-2.57	-1.92	-1.56	-2.38	-1.68	-1.32
T=500	-2.57	-1.94	-1.62	-2.54	-1.91	-1.56	-2.35	-1.68	-1.32
T: ∞	-2.56	-1.94	-1.62	-2.54	-1.89	-1.55	-2.35	-1.68	-1.31
$\boldsymbol{\tau = 0.5}$	$\boldsymbol{\eta = 0.1}$			$\boldsymbol{\eta = 1}$			$\boldsymbol{\eta = 10}$		
$(\ \xi\ ;(1-\xi^2)^{1/2})$	$(0.91; 0.42)$			$(0.87; 0.50)$			$(0.09; 1.00)$		
T=25	-2.74	-2.02	-1.63	-2.78	-2.01	-1.62	-2.64	-1.81	-1.39
T=50	-2.64	-1.96	-1.61	-2.66	-1.98	-1.62	-2.46	-1.74	-1.36
T=100	-2.60	-1.95	-1.60	-2.61	-1.94	-1.61	-2.42	-1.71	-1.34
T=250	-2.59	-1.95	-1.60	-2.57	-1.94	-1.60	-2.44	-1.70	-1.33
T=500	-2.59	-1.94	-1.60	-2.57	-1.93	-1.60	-2.42	-1.68	-1.31
T: ∞	-2.54	-1.92	-1.59	-2.54	-1.92	-1.59	-2.35	-1.68	-1.31

for η, τ. For illustration, we take $\eta = 0.1, 1, 10$ and $\tau = -0.5, 0.1, 0.5$. These nine pairs provide a relatively large sample for the weights associated to the Dickey–Fuller and Normal distributions.

We give in table 1 the values of the weights corresponding to particular values of η and τ. The presentation is limited to 1%, 5%, 10% critical values. From table 1, we see that the conventional Dickey–Fuller critical values are inappropriate when the regression equations involves feedback covariates. For example, if we compare the case $(\eta = 0.1; \tau = 0.1)$ to the one $(\eta = 10; \tau = -0.5)$, it appears that the conventional critical values are too conservative[16].

[16]For interested readers, we have similar results for the cases 2 and 3. In fact, all the tables are available with τ varying from -0.95 to 0.95 by step of 0.05 and for critical values from 0.5% to 50% by step of 0.5%. Furthermore, we have similar experimental results for case 1 with a constant. When they correspond, results are in accordance with those of Hansen, called "demeaned critical values". See also Feve and Henin (1996) for the asymptotic distribution in case 1.

4.2 Local power of FADF, COLS and CADF statistics

We look at local power for each statistic. The power function[17] are computed for $\rho = 1 - c/T$ with $c = 1, 2, \ldots, 16$ and $T = 100$. From figure 2, one can see the large gains in power when $\tau = 0.1$ and $\eta = 10$. In this case, the weight ξ is closed to zero and the distribution is essentially Gaussian.

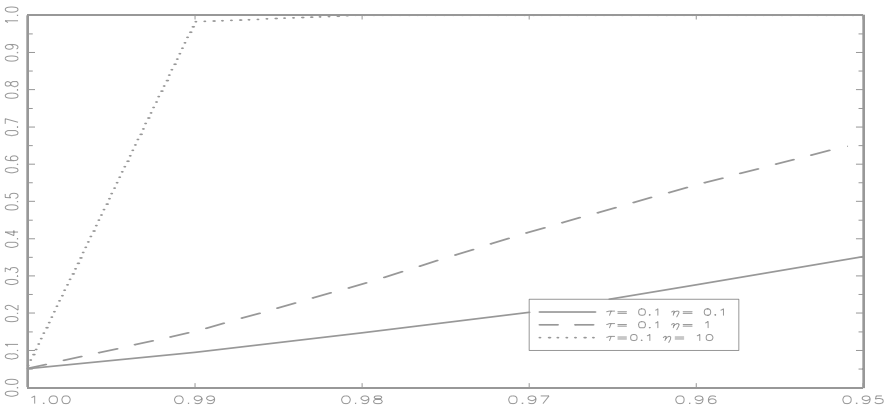

Figure 2: Power Curves for Local Alternatives (*case 1*)

Figure 3 displays local power for *case 2*. From theorem 2, we may expect some gains in power only if $\tau < \eta$. The pair $\tau = 0.7$ and $\eta = 0.95$ illustrates this result but with a relatively small gain in power. The situation $\tau = 0.7$ and $\eta = 0.7$ is an example of lack of power. Last, with $\tau = 0.95$ and $\eta = 0.70$, the test statistic remains dramatically biased.

Finally, figure 4 reports local power for *case 3*. The power depends only on the sign of τ. From theorem 3, increase in power is expected if $\tau > 0$. But even in this case, the gain in power is slow. At the opposite, for $\tau < 0$, here $\tau = -0.5$, the test is biased.

[17] As previously, the power functions were calculated for various pairs of η, τ from 40,000 samples with *iid* Gaussian innovations, using the critical values from table 1.

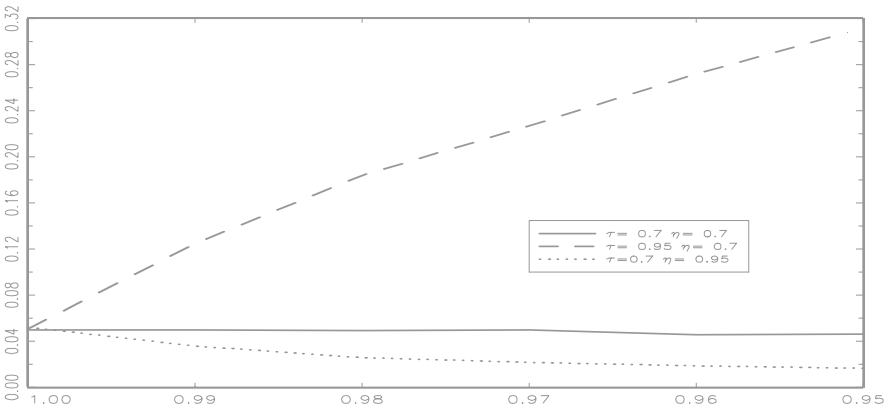

Figure 3: Power Curves for Local Alternatives (*case 2*)

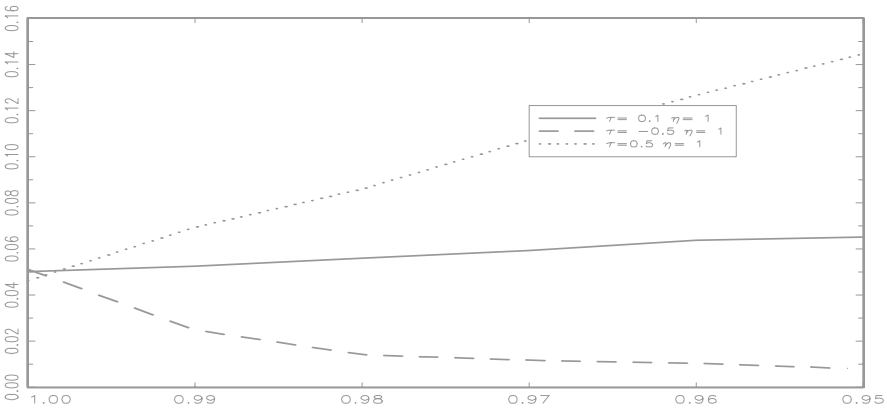

Figure 4: Power Curves for Local Alternatives (*case 3*)

4.3 P-Plot curves

Theorem 1 state that uncertainty on the covariance matrix of residuals does not matter for the asymptotic distribution of the FADF *t-statistic*. A simple way to evaluate this result is small sample is to compare the P-values of the FADF statistic with the ones obtained through univariate approaches. If $\eta = \tau$, the limit distribution of the *t-statistic* corresponds to a Dickey Fuller one, while if $\eta = 0$ it corresponds to the one of the single auxiliary regression approach. In order to locate size effects, we use a P-Plot (see Davidson and Mac Kinnon (1998)) which gives, in respect to the original critical values, the corresponding P-values for a *t-statistic* computed from an estimate of the covariance matrix. This figure takes in abscissa P-values, calculates the corresponding critical values and the associated P-values of the second statistic with this critical values[18]. A straight line extending from $(0, 0)$ to $(1, 1)$ will indicate a good agreement between the two statistics. We report in figure 5 two P-plots. Panel (a) report the FADF *t-statistic* versus the ADF statistic in the case $\eta = \tau$ and panel (b) the FADF *t-statistic* versus the single regression approach statistic in the case $\tau = 0$. These two figures indicate little variations in the P-values for a sample size equal to 100.

Our table is constructed under the restrictive assumption that $\{\varepsilon_t, u_t\}$ are *iid*. Such an assumption cannot be maintained in practise. To investigate potential size effects, we compute the FADF *t-statistic* in this way. We thus introduce serial correlation by including higher-order autoregressive terms in the regression. This is in fact an experimental verification (theoretical results needs to be done) that serial correlation can be treated, as in the augmented Dickey-Fuller method. We simulate a 2-order autoregressive process for y_t

[18] 215 P-values are used: from 0.001 to 0.01 (step 0.0001), from 0.01 to 0.985 (step 0.005) and from 0.99 to 0.999 (step 0.001).

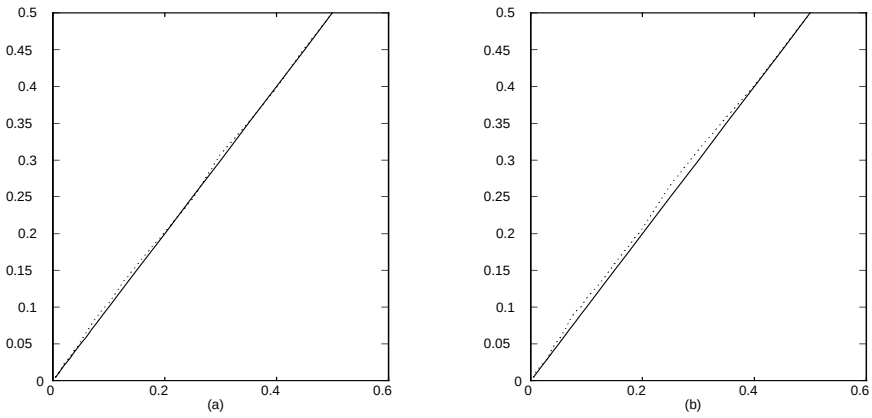

Figure 5: Size of FADF statistic. Panel (a) ADF versus FADF. Panel (b) Auxiliary Regression–OLS versus FADF

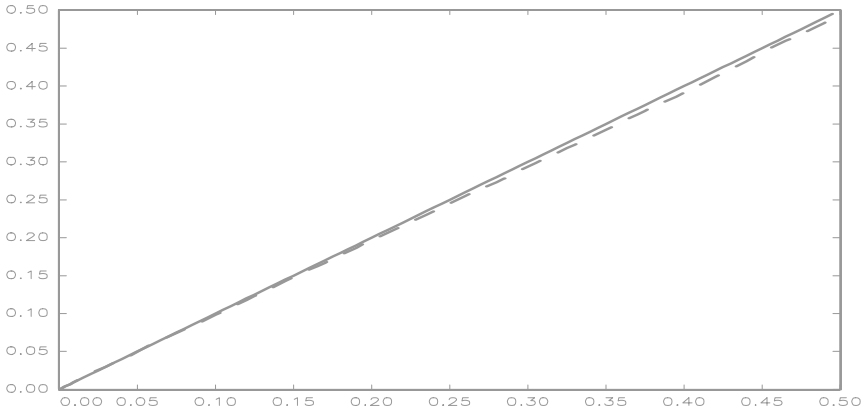

Figure 6: Size of the statistic with autocorrelated errors

with 1 and 0.5 as roots. We consider *case 1*, a size $T = 100$, and the pair $\eta = 0.1$ and $\tau = 0.5$ and we compare the associated critical values to the ones of the table 1. We use again the P-Plot curves to locate some size effects. Variation for the numerical values of the critical values appear mostly, only in the third decimal. Figure 6 illustrates this fact. This result seems to allow us to use table 1, even with serial correlation.

5 An application with US data

To illustrate the potential of our approach, we examine the sustainability of US government deficit. Data are taken from Bohn (1991a) and cover the period 1792-1988. The use of a long data set allows to prevent from high and persistent budget deficits. The empirical analysis is based on the following three series each of them being deflated by GDP: *i)* Federal outlays net of interest payments, *ii)* Federal receipts and *iii)* debt at the beginning of the fiscal year. The variables *i), ii)* and *iii)* are the empirical representation of the spending variable G_t , of the tax variable T_t (in a wide sense) and of the variable B_t. Figure 7 reports the public debt and the primary surplus $T_t - G_t$ in term of nominal GDP from 1792 to 1988. One can immediately see the impact of the three major wars (Civil War, World War I and World War II). These wars lead to sudden and large increases in debt and in outlays, which can be viewed here as sudden and significant negative values of primary surplus.

Following Wilcox (1989), the test for sustainability is based on the discounted debt and primary surplus as percentage of GDP. Thus, we must choose an arbitrary value for the discount factor. Empirical studies clearly indicate that the sample mean of the interest rate on government bonds does not exceed economic growth. Nevertheless, as mentioned by Bohn (1995),

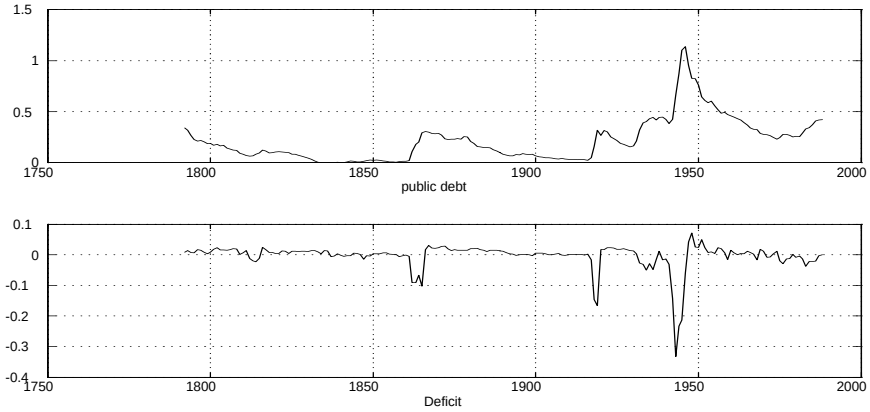

Figure 7: Public debt and Primary Surplus

this result does not provide evidence against a positive discount factor. The reason is that risky tax revenues must be discounted at a higher than the safe rate. In respect to the empirical findings of Bohn (1991a), we select an admissible range of values for the discount factor and we perform the unit root test for these different values. The discount factor is set at $0\%, 0.5\%, 1\%, 1.5\%, 2\%$ and 3% per year, following the suggestion of Bohn (1991a).[19] The system based equations is thus given by (7)-(8), where $b_t = (B_t/Y_t)/(1+r)^t$ and $s_t = (T_t - G_t/Y_t)/(1+r)^t$ for $t = 1, ..., T$. As aforementioned, the introduction of an error term into the discounted debt identity comes from the assumption of a constant real interest rate. This error may be also correlated because we do not assume constant expected real returns.

Before conducting statistical inference on our reference model, we must look at the time series properties of the errors terms ϵ_t and u_t under the

[19]The stationarity between public debt, taxes and government spending with a cointegration vector of the form (1,-1,-r) can be viewed as the stationarity of the error term in the auxiliary equation. Statistical inference on the parameter r provided in Bohn (1991a) also indicates that an admissible value for r lies between -2% and 15%. The point estimate subjected to restriction on the cointegrated vector is r = 2.9% and we keep this value as the maximum admissible value for the discount factor.

null and then select the bivariate time series model. Univariate unit root tests (ADF type) indicate that these two terms are both stationary. Nevertheless, it appears significant serial correlation. Thus, we select under the null of unit root, the number of lags to be introduced in the system based equation. We estimate a VAR model on Δb_t and s_t and we select the number of lags using a BIC criterion. The number of lags is equal to 6 and specification tests on VAR residuals indicate that assumption of no serial correlation cannot be rejected. We thus estimate the following representation for $r = \{0\%, 0.5\%, 1\%, 1.5\%, 2\%, 3\%\}$ per year:

$$\begin{aligned} \Delta b_t &= (\rho - 1) b_{t-1} + \sum_{i=1}^{4} \phi_{1i} \Delta b_{t-i} + \sum_{i=1}^{4} \phi_{2i} s_{t-i} + \varepsilon_t - u_t \\ s_t &= (1 - \rho) b_{t-1} + \sum_{i=1}^{4} \phi_{3i} \Delta b_{t-i} + \sum_{i=1}^{4} \phi_{4i} s_{t-i} + u_t \end{aligned}$$

Our previous simulation experiments indicate that the critical values can be used in presence of serial correlation if an adequate specification is introduced,*i.e.* if the lagged values of the dependent variables are introduced in the system based equations. Given the particular values of η and τ, we tabulated the critical values for each model (*cases 1,2,3*) using 2500 replications. Our results are in accordance with the previous experiments, *i.e* the critical values are comparable to the ones of table 1. Now, using these simulations under the null hypothesis, one can obtain the P-value associated to each *t-statistics* in each case. The results of unit root tests are presented in table 2. This table also reports the weight associated to the Dickey-Fuller type distribution. Obviously, this weight is zero in *case 2* (COLS), because the limit distribution is always a normal one. The unit root tests conducted in the three cases provides mixed results. Surprisingly, the CADF type test (*case 3*) leads to the rejection of sustainability for values of the discount factor which exceed 1% per quarter, whereas the two other cases lead to its accep-

Table 2: Unit Root Tests (*t- statistics*)

discount factor	0%	0.5%	1%	1.5%	2%	3%
FADF	-0.91	-1.56	-2.38	-3.18	-3.76	-4.40
weight	0.89	0.88	0.87	0.85	0.81	0.72
P-value	36.5%	11.2%	1.9%	0.1%	0%	0%
COLS	1.29	0.76	-0.21	-1.47	-2.71	-4.35
weight	0	0	0	0	0	0
P-value	90.9%	50.2%	39.8%	6.5%	0.4%	0%
CADF	-2.08	-2.02	-1.84	-1.65	-1.49	-1.35
weight	0.66	0.68	0.75	0.84	0.92	0.99
P-value	3.5%	4.1%	6.1%	9.8%	13.4%	16.8%

tance. Moreover, with low values of the discount factor, the COLS statistic does not reject the unit root hypothesis. The CADF appears less sensitive to the choice of the discount factor. For the values of the discount factor which exceed 0.5% per year, the FADF *t-statistic* leads to the rejection of the unit root hypothesis. Notice that the weight associated to the Dickey-Fuller type distribution decreases as the discount factor grows. Conversely, the weight in *case 3* increases with the discount factor and the null hypothesis cannot be rejected.

The potential of the FADF t-statistic may be illustrated using the power curves for local alternatives. To see this, we use the empirical results with $r = 1\%$, because the FADF, COLS and CADF *t- statistics* does not provide the same conclusions on dynamic properties of the public debt.

Figure 8 shows that FADF *t-statistic* performs better in terms of power than both the COLS and CADF statistics. These two statistics lead to accept too often the null hypothesis and the power gains appear very low even for large alternatives. For example, for $\rho = 0.95$, the power with the FADF *t-statistic* is equal to 88%, whereas the power is equal to 37% and 14%, using the COLS and CADF *t-statistics*. Therefore, empirical approaches

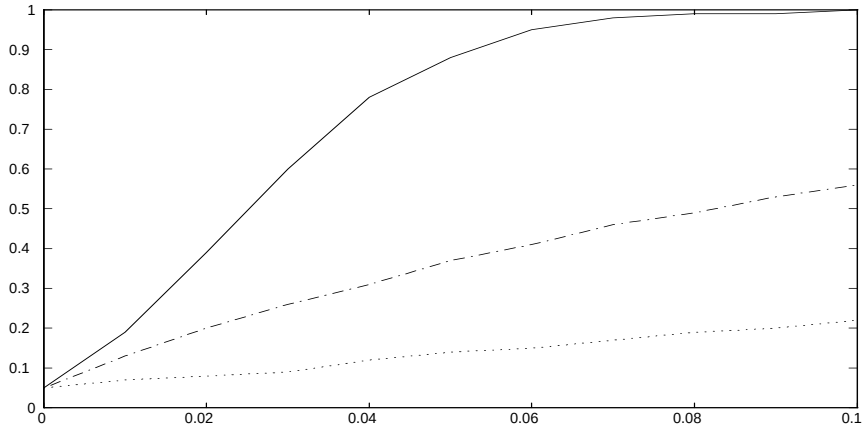

Figure 8: Power in % for local alternatives $\rho - 1$ (FADF solid line, COLS dashdotted line , CADF dotted line)

which ignore cross–equations restrictions may be potentially uninformative regarding the order of integration in discounted public debt.

6 Concluding remarks

The paper introduces a new test for unit root in a bivariate framework, accounting for cross-equations restrictions between an usual ADF type equation and an auxiliary equation informative on a feedback mechanism. This strategy and its associated FADF test statistic naturally nest two previously unrelated approaches, *i.e.* the usual ADF one and the auxiliary equation approach of Bohn (1991b) and Wickens and Uctum (1993).

Our analytical results, confirmed by simulation experiments on finite sample, indicate the potential gains in power provided by our system based strategy. We also point out the limits of alternative previous approaches, which ignore cross-equation restrictions and/or covariates. While providing a powerful sustainability test, the FADF test statistic will constitute a useful tool

for inference in other related situations. As examples, we may think about other stock-flow problems, such as inventories dynamics. Moreover, a famous relation as the *Phillips curve* may be seen as a feedback relation essential for unemployment stabilization. Testing for hysteresis in unemployment jointly as a unit root test in unemployment rate and the nullity of the *Phillips effect* of unemployment on wage dynamics may be performed using our FADF test statistic, as a straightforward application of our strategy. Beyond increase in statistical efficiency, economic interpretation will benefit from this strategy.

To be more convincing, our results must be extended in several ways. The paper provides basic asymptotic results under strong assumptions and the potentials of the FADF statistic are only addressed on the basis of simulation experiments. One main line of future progress is the introduction of regime shifts and its consequence on the distribution of the FADF statistic must be investigated. As suggested by Muellbauer (1996), changes in the persistence of macroeconomic variables are often related to shifts in policy feedback rules. Combining our approach with such elements will thus appear especially relevant.

Appendix

Proof of Lemma 1

The covariance matrix can be written as $\Sigma_2 = PP'$, using a Choleski decomposition, where the matrix P is given by

$$P = \sigma_u \begin{pmatrix} (1+\eta^2-2\eta\tau)^{1/2} & 0 \\ \frac{\eta\tau-1}{(1+\eta^2-2\eta\tau)^{1/2}} & \frac{\eta(1-\tau^2)^{1/2}}{(1+\eta^2-2\eta\tau)^{1/2}} \end{pmatrix}$$

Then the residuals from the regression equations can be written as

$$\begin{aligned} (\varepsilon_t - u_t) &= \sigma_u (1+\eta^2-2\eta\tau)^{1/2} e_{1t} \\ u_t &= \sigma_u \frac{\eta\tau-1}{(1+\eta^2-2\eta\tau)^{1/2}} e_{1t} + \sigma_u \frac{\eta(1-\tau^2)^{1/2}}{(1+\eta^2-2\eta\tau)^{1/2}} e_{2t} \end{aligned}$$

where e_{1t} and e_{2t} are two independent $iid\mathcal{N}(0,1)$. Using this decomposition, we have

$$\begin{aligned} \frac{1}{T}\sum y_{t-1}(\varepsilon_t - u_t) &\implies \sigma_u^2(1+\eta^2-2\eta\tau)\int_0^1 B_1^c dB_1(r) \\ \frac{1}{T^2}\sum y_{t-1}^2 &\implies \sigma_u^2(1+\eta^2-2\eta\tau)\int_0^1 \{B_1^c\}^2 dr \end{aligned}$$

To prove the last element of the lemma, one must remark

$$\frac{1}{T}\sum y_{t-1}u_t = \frac{\eta\tau-1}{(1+\eta^2-2\eta\tau)}\frac{1}{T}\sum y_{t-1}(\varepsilon_t - u_t) + \sigma_u \frac{\eta(1-\tau^2)^{1/2}}{(1+\eta^2-2\eta\tau)^{1/2}}\frac{1}{T}\sum y_{t-1}e_{2t}$$

Now, using the property

$$\frac{1}{T}\sum y_{t-1}e_{2t} \Longrightarrow \sigma_u(1+\eta^2-2\eta\tau)^{1/2}\int_0^1 B_1^c dB_2(r)$$

one obtains the limit distribution of $\frac{1}{T}\sum y_{t-1}u_t$. This completes the proof. ■

Proof of Theorem 1

The ML estimate $\widehat{\rho}$ of ρ is given by the score function

$$\sum \left[(-y_{t-1}; y_{t-1}) \begin{pmatrix} 1 & 1-\widehat{\eta}\widehat{\tau} \\ 1-\widehat{\eta}\widehat{\tau} & 1+\widehat{\eta}^2-2\widehat{\eta}\widehat{\tau} \end{pmatrix} \begin{pmatrix} \Delta y_t - (\widehat{\rho}-1)y_{t-1} \\ x_t - (1-\widehat{\rho})y_{t-1} \end{pmatrix} \right] = 0$$

which may be reduced to

$$\widehat{\rho} - 1 = \frac{\widehat{\tau}}{\widehat{\eta}}(\widehat{\rho}_1 - 1) + \left(1 - \frac{\widehat{\tau}}{\widehat{\eta}}\right)(\widehat{\rho}_2 - 1)$$

where $\widehat{\rho}_1 - 1$ and $\widehat{\rho}_2 - 1$ correspond to the single regression estimates, *i.e.* the OLS estimates of ρ in equations (11) and (12), respectively $\widehat{\rho}_1 - 1 = \left(\frac{1}{T}\sum \Delta y_t y_{t-1}\right)\left(\frac{1}{T}\sum y_{t-1}^2\right)^{-1}$ and

$\widehat{\rho}_2 - 1 = -\left(\frac{1}{T}\sum x_t y_{t-1}\right)\left(\frac{1}{T}\sum y_{t-1}^2\right)^{-1}$. The ML estimate of ρ corresponds to a weighted sum of two single OLS estimates, with weights given by the estimated values of η and τ

$$\begin{aligned}
\frac{\widehat{\tau}}{\widehat{\eta}} &= \frac{Cov(\widehat{\varepsilon}_t, \widehat{u}_t)}{\widehat{\sigma}_\varepsilon \widehat{\sigma}_u \widehat{\eta}} \\
&= \frac{\frac{1}{T}\sum \widehat{\varepsilon}_t \widehat{u}_t}{\frac{1}{T}\sum \widehat{\varepsilon}_t^2} \\
&= \frac{\frac{1}{T}\sum (\Delta y_t + x_t)(x_t - (1-\widehat{\rho}) y_{t-1})}{\frac{1}{T}\sum \widehat{\varepsilon}_t^2} \\
&= \frac{\frac{1}{T}\sum \Delta y_t x_t + \frac{1}{T}\sum x_t^2 - (1-\widehat{\rho})\frac{1}{T}\sum (\Delta y_t y_{t-1} + x_t y_{t-1})}{\frac{1}{T}\sum \widehat{\varepsilon}_t^2}
\end{aligned}$$

Now, given that under the local alternative $c \neq 0$

$$\begin{aligned}
\Delta y_t &= -\frac{c}{T} y_{t-1} + \varepsilon_t - u_t \\
x_t &= \frac{c}{T} y_{t-1} + u_t
\end{aligned}$$

we deduce

$$\begin{aligned}
\frac{1}{T}\sum \Delta y_t x_t &= -\frac{c^2}{T^3}\sum y_{t-1}^2 - \frac{c}{T^2}\sum y_{t-1} u_t + \frac{c}{T^2}\sum y_{t-1}(\varepsilon_t - u_t) + \frac{1}{T}\sum u_t(\varepsilon_t - u_t) \\
\frac{1}{T}\sum x_t^2 &= \frac{c^2}{T^3}\sum y_{t-1}^2 + \frac{2c}{T^2}\sum y_{t-1} u_t + \frac{1}{T}\sum u_t^2
\end{aligned}$$

From Lemma 1, it follows that $\frac{1}{T^2}\sum y_{t-1}^2$, $\frac{1}{T}\sum y_{t-1} u_t$ and $\frac{1}{T}\sum y_{t-1}(\varepsilon_t - u_t)$ are both $Op(1)$. As $T \to \infty$, $\frac{1}{T}\sum \Delta y_t x_t$, $\frac{1}{T}\sum x_t^2$ and $\frac{1}{T}\sum \widehat{\varepsilon}_t^2$ are consistent estimates of $\sigma_u^2(\eta\tau - 1)$, σ_u^2 and σ_ε^2 , respectively. Thus, we have

$$\frac{\widehat{\tau}}{\widehat{\eta}} = \frac{\tau}{\eta} + (\widehat{\rho} - 1)\frac{1}{\sigma_\varepsilon^2}\frac{1}{T}\sum (\Delta y_t y_{t-1} + x_t y_{t-1})$$

Plugging this expression into the ML estimate $\widehat{\rho}$ of ρ , one gets

$$\widehat{\rho} - 1 = \left\{\frac{\tau}{\eta}(\widehat{\rho}_1 - 1) + \left(1 - \frac{\tau}{\eta}\right)(\widehat{\rho}_2 - 1)\right\}\widehat{A}^{-1}$$

where

$$\begin{aligned}
\widehat{A} &= 1 - \frac{1}{\sigma_\varepsilon^2}\left(\frac{1}{T}\sum (\Delta y_t y_{t-1} + x_t y_{t-1})\right)((\widehat{\rho}_1 - 1) - (\widehat{\rho}_2 - 1)) \\
&= 1 - \frac{1}{\sigma_\varepsilon^2}\frac{\left(\frac{1}{T}\sum (\Delta y_t y_{t-1} + x_t y_{t-1})\right)^2}{\frac{1}{T}\sum y_{t-1}^2} \\
&\equiv 1 - \widetilde{A}
\end{aligned}$$

Under the local alternative, we have from Lemma 1

$$\begin{aligned}
\frac{1}{T}\sum \Delta y_t y_{t-1} &= -\frac{c}{T^2}\sum y_{t-1}^2 + \frac{1}{T}\sum y_{t-1}(\varepsilon_t - u_t) \\
&\Rightarrow -c\sigma_u^2(1 + \eta^2 - 2\eta\tau)\int_0^1 \{B_1^c\}^2 dr + \sigma_u^2(1 + \eta^2 - 2\eta\tau)\int_0^1 B_1^c dB_1(r)
\end{aligned}$$

and

$$\begin{aligned}\frac{1}{T}\sum x_t y_{t-1} &= \frac{c}{T^2}\sum y_{t-1}^2 + \frac{1}{T}\sum y_{t-1}u_t \\ &\Rightarrow c\sigma_u^2(1+\eta^2-2\eta\tau)\int_0^1 \{B_1^c\}^2\,dr + \sigma_u^2(\eta\tau-1)\int_0^1 B_1^c dB_1(r) \\ &\quad +\sigma_u^2\eta(1-\tau^2)^{1/2}\int_0^1 B_1^c dB_2(r)\end{aligned}$$

We directly deduce that

$$\frac{1}{T}\sum(\Delta y_t y_{t-1} + x_t y_{t-1}) \Rightarrow \sigma_u^2\eta\left\{(\eta-\tau)\int_0^1 B_1^c dB_1(r) + (1-\tau^2)^{1/2}\int_0^1 B_1^c dB_2(r)\right\}$$

It follows that $T\widetilde{A}$ admits the following asymptotics

$$T\widetilde{A} \Rightarrow \frac{\left\{(\eta-\tau)\int_0^1 B_1^c dB_1(r) + (1-\tau^2)^{1/2}\int_0^1 B_1^c dB_2(r)\right\}^2}{(1+\eta^2-2\eta\tau)\int_0^1 \{B_1^c\}^2\,dr}$$

$\widetilde{A}$ is thus superconsistent and $\widehat{A} \to 1$. The statistic $T(\widehat{\rho}-1)$ is now given by

$$T(\widehat{\rho}-1) = \frac{\tau}{\eta}T(\widehat{\rho}_1-1) + \left(1-\frac{\tau}{\eta}\right)T(\widehat{\rho}_2-1)$$

where the limit distributions of the statistics associated to the two OLS estimates of ρ are

$$\begin{aligned}T(\widehat{\rho}_1-1) &\Rightarrow -c + \frac{\int_0^1 B_1^c(r)dB_1}{\int_0^1 \{B_1^c(r)\}^2\,dr} \\ T(\widehat{\rho}_2-1) &\Rightarrow -c + \frac{1-\eta\tau}{1+\eta^2-2\eta\tau}\frac{\int_0^1 B_1^c(r)dB_1}{\int_0^1 \{B_1^c(r)\}^2\,dr} + \frac{\eta(1-\tau^2)^{1/2}}{1+\eta^2-2\eta\tau}\frac{\int_0^1 B_1^c(r)dB_2}{\int_0^1 \{B_1^c(r)\}^2\,dr}\end{aligned}$$

One deduces the limit distribution of the statistic $T(\widehat{\rho}-1)$

$$\begin{aligned}T(\widehat{\rho}-1) &\Rightarrow -c + \frac{\tau}{\eta}\frac{\int_0^1 B_1^c(r)dB_1}{\int_0^1 \{B_1^c(r)\}^2\,dr} \\ &\quad + \left(1-\frac{\tau}{\eta}\right)\left\{\frac{\kappa^2}{1-\eta\tau}\frac{\int_0^1 B_1^c(r)dB_1}{\int_0^1 \{B_1^c(r)\}^2\,dr} + \frac{\kappa(1-\kappa^2)^{1/2}}{1-\eta\tau}\frac{\int_0^1 B_1^c(r)dB_2}{\int_0^1 \{B_1^c(r)\}^2\,dr}\right\} \\ &\Rightarrow -c + \xi^2\frac{\int_0^1 B_1^c(r)dB_1}{\int_0^1 \{B_1^c(r)\}^2\,dr} + \xi(1-\xi^2)^{1/2}\frac{\int_0^1 B_1^c(r)dB_2}{\int_0^1 \{B_1^c(r)\}^2\,dr}\end{aligned}$$

where $\xi = (1-\tau^2)^{1/2}(1+\eta^2-2\eta\tau)^{-1/2}$. The ML standard error for the estimated value of ρ is

$$se(\widehat{\rho}) = \left\{\frac{1}{(1-\widehat{\tau}^2)\widehat{\sigma}_u^2}\sum y_{t-1}^2\right\}^{-1/2}$$

where $\widehat{\sigma}_u^2$ denotes the estimate of the residual variance

$$\widehat{\sigma}_u^2 = \frac{1}{T}\sum(x_t - (1-\widehat{\rho})y_{t-1})^2$$

The consistency of $\widehat{\rho}$ implies $\widehat{\sigma}_u^2 \to \sigma_u^2$ and $\widehat{\tau} \to \tau$. We thus deduce the following asymptotic

$$Tse(\widehat{\rho}) \Rightarrow \xi \left\{ \int_0^1 \left\{ B_1^c(r) \right\}^2 dr \right\}^{-1/2}$$

Then, from the continuous mapping theorem, one obtains

$$\begin{aligned} t(\widehat{\rho}-1) \quad \Rightarrow \quad & -\frac{c}{\xi} \left\{ \int_0^1 \left\{ B_1^c(r) \right\}^2 dr \right\}^{1/2} + \xi \frac{\int_0^1 B_1^c(r) dB_1}{\left\{ \int_0^1 \left\{ B_1^c(r) \right\}^2 dr \right\}^{1/2}} \\ & + (1-\xi^2)^{1/2} \frac{\int_0^1 B_1^c(r) dB_2}{\left\{ \int_0^1 \left\{ B_1^c(r) \right\}^2 dr \right\}^{1/2}} \end{aligned}$$

Since B_2 is independent of B_1 and B_1^c ,$\left\{ \int_0^1 B_1^c(r) dB_2 \right\} \left\{ \int_0^1 \left\{ B_1^c(r) \right\}^2 dr \right\}^{1/2}$ is distributed $\mathcal{N}(0,1)$ and is independent of B_1^c . This completes the proof ■

Proof of Theorem 2

The OLS estimates of ρ and δ are given by

$$\begin{pmatrix} 1-\widehat{\rho} \\ \widehat{\delta} \end{pmatrix} = \begin{pmatrix} \sum y_{t-1}^2 & \sum y_{t-1}\Delta y_t \\ \sum y_{t-1}\Delta y_t & \sum \Delta y_t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{t-1}x_t \\ \sum \Delta y_t x_t \end{pmatrix}$$

which provides

$$\begin{aligned} \widehat{\rho}-1 &= \frac{\sum y_{t-1}\Delta y_t \sum \Delta y_t x_t - \sum y_{t-1}x_t \sum \Delta y_t^2}{\sum y_{t-1}^2 \sum \Delta y_t^2 - \left(\sum y_{t-1}\Delta y_t \right)^2} \\ \widehat{\delta} &= \frac{\sum y_{t-1}^2 \sum \Delta y_t x_t - \sum y_{t-1}\Delta y_t \sum y_{t-1}x_t}{\sum y_{t-1}^2 \sum \Delta y_t^2 - \left(\sum y_{t-1}\Delta y_t \right)^2} \end{aligned}$$

$\sum y_{t-1}^2 \sum \Delta y_t x_t$ and $\sum y_{t-1}^2 \sum \Delta y_t^2$ are $Op(T^3)$ while $\sum y_{t-1}\Delta y_t \sum y_{t-1}x_t$ and $\left(\sum y_{t-1}\Delta y_t \right)^2$ are $Op(T^2)$, it follows that $\widehat{\delta}$ is a consistent estimate of $\delta = (\eta\tau - 1)(1 + \eta^2 - 2\eta\tau)$. From Lemma 1, we deduce the asymptotics for $\sum y_{t-1}\Delta y_t$, $\sum \Delta y_t x_t$, $\sum x_t y_{t-1}$ and $\sum \Delta y_t^2$. From the continuous mapping theorem, $T(\widehat{\rho}-1)$ admits the following asymptotic

$$T(\widehat{\rho}-1) \Rightarrow -c\frac{\eta(\eta-\tau)}{1+\eta^2-2\eta\tau} + \frac{\eta(1-\tau^2)^{1/2}}{1+\eta^2-2\eta\tau} \frac{\int_0^1 B_1^c(r) dB_2}{\int_0^1 \left\{ B_1^c(r) \right\}^2 dr}$$

The standard error of estimated value of ρ is given by

$$se(\widehat{\rho}-1) = \left\{ \widehat{\sigma}_e^2 \frac{\sum \Delta y_t^2}{\sum y_{t-1}^2 \sum x_t^2 - \left(\sum y_{t-1}x_t \right)^2} \right\}^{1/2}$$

where $\widehat{\sigma}_e^2$ denotes the estimate of the residual variance

$$\widehat{\sigma}_e^2 = \frac{1}{T} \sum \left(x_t - (1-\widehat{\rho}) y_{t-1} - \widehat{\delta}\Delta y_t \right)^2$$

The consistency of $\widehat{\rho}$ and $\widehat{\delta}$ implies $\widehat{\sigma}_e^2 \to \sigma_e^2 = \sigma_u^2\eta^2(1-\tau^2)(1+\eta^2-2\eta\tau)^{-1}$. We thus deduce the following asymptotic

$$Tse(\widehat{\rho}) \Rightarrow \frac{\eta(1-\tau^2)^{1/2}}{(1+\eta^2-2\eta\tau)^{1/2}} \left\{ \int_0^1 \{B_1^c(r)\}^2 \, dr \right\}^{-1/2}$$

Then, from the continuous mapping theorem, one obtains

$$t(\widehat{\rho}-1) \Rightarrow -c\frac{\eta-\tau}{(1-\tau^2)^{1/2}} \left\{ \int_0^1 \{B_1^c(r)\}^2 \, dr \right\}^{1/2} + \frac{\int_0^1 B_1^c(r)dB_2}{\left\{ \int_0^1 \{B_1^c(r)\}^2 \, dr \right\}^{1/2}}$$

Since B_2 is independent of B_1 and B_1^c ,$\left\{ \int_0^1 B_1^c(r)dB_2 \right\} \left\{ \int_0^1 \{B_1^c(r)\}^2 \, dr \right\}^{1/2}$ is distributed $\mathcal{N}(0,1)$ and is independent of B_1^c . This completes the proof ■

Proof of Theorem 3

The OLS estimates of ρ and γ are given by

$$\begin{pmatrix} \widehat{\rho}-1 \\ \widehat{\gamma} \end{pmatrix} = \begin{pmatrix} \sum y_{t-1}^2 & \sum y_{t-1}x_t \\ \sum y_{t-1}x_t & \sum x_t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{t-1}y_t \\ \sum x_t y_t \end{pmatrix}$$

which provides

$$\begin{aligned} \widehat{\rho}-1 &= \frac{\sum y_{t-1}\Delta y_t \sum x_t^2 - \sum y_{t-1}x_t \sum \Delta y_t x_t}{\sum y_{t-1}^2 \sum x_t^2 - \left(\sum y_{t-1}x_t\right)^2} \\ \widehat{\gamma} &= \frac{\sum y_{t-1}^2 \sum \Delta y_t x_t - \sum y_{t-1}\Delta y_t \sum y_{t-1}x_t}{\sum y_{t-1}^2 \sum x_t^2 - \left(\sum y_{t-1}xy_t\right)^2} \end{aligned}$$

From Lemma 1 and the proof of theorem 2, one deduces that $\widehat{\gamma}$ is a consistent estimate of $\gamma = \eta\tau - 1$. The limit distribution of $T(\widehat{\rho}-1)$ is the following

$$T(\widehat{\rho}-1) \Rightarrow -c\eta\tau + \vartheta^2 \frac{\int_0^1 B_1^c(r)dB_1}{\int_0^1 \{B_1^c(r)\}^2 dr} + \vartheta(1-\vartheta^2)^{1/2} \frac{\int_0^1 B_1^c(r)dB_2}{\int_0^1 \{B_1^c(r)\}^2 dr}$$

where $\vartheta = \xi\eta$. The standard error of estimated value of ρ is given by

$$se(\widehat{\rho}-1) = \left\{ \widehat{\sigma}_v^2 \frac{\sum x_t^2}{\sum y_{t-1}^2 \sum x_t^2 - \left(\sum y_{t-1}x_t\right)^2} \right\}^{1/2}$$

where $\widehat{\sigma}_v^2$ denotes the estimate of the residual variance

$$\widehat{\sigma}_v^2 = \frac{1}{T} \sum \left(\Delta y_t - (\widehat{\rho}-1)y_{t-1} - \widehat{\gamma}x_t\right)^2$$

The consistency of $\widehat{\rho}$ and $\widehat{\gamma}$ implies $\widehat{\sigma}_v^2 \to \sigma_v^2 = \sigma_u^2\eta^2(1-\tau^2)$. We thus deduce the following asymptotic

$$Tse(\widehat{\rho}) \Rightarrow \frac{\eta(1-\tau^2)^{1/2}}{(1+\eta^2-2\eta\tau)^{1/2}} \left\{ \int_0^1 \{B_1^c(r)\}^2 \, dr \right\}^{-1/2}$$

Then, from the continuous mapping theorem, one obtains

$$t_{(\widehat{\rho}-1)} \Longrightarrow -\frac{c\eta\tau}{\vartheta}\left\{\int_0^1 \{B_1^c(r)\}^2\,dr\right\}^{1/2} + \vartheta\frac{\int_0^1 B_1^c(r)dB_1(r)}{\left\{\int_0^1 \{B_1^c(r)\}^2\,dr\right\}^{1/2}} + (1-\vartheta^2)^{1/2}\frac{\int_0^1 B_1^c(r)dB_2}{\left\{\int_0^1 \{B_1^c(r)\}^2\,dr\right\}^{1/2}}$$

Since B_2 is independent of B_1 and B_1^c ,$\left\{\int_0^1 B_1^c(r)dB_2\right\}\left\{\int_0^1 \{B_1^c(r)\}^2\,dr\right\}^{1/2}$ is distributed $\mathcal{N}(0,1)$ and is independent of B_1^c . This completes the proof ■

References

[1] Bohn, H.(1991a) "Budget Balance Through Revenue or Spending Adjustments? Some Historical Evidence for the US", *Journal of Monetary Economics*, 28: 333–359.

[2] Bohn, H.(1991b) "On Testing the Sustainability of Budget Deficits in a Stochastic Environment", *working paper* no3, June, Center for Economic Studies. University of Munich.

[3] Bohn, H.(1995) "The Sustainability of Budget Deficits in a Stochastic Economy", *Journal of Money, Credit and Banking,* vol 27, No. 1, 257-271, February.

[4] Campbell, J. and Perron, P. (1991) "Pitfalls and Opportunities: What Macroeconomists should Known About Unit Root", *NBER Macroeconomics Annual*, MIT Press.

[5] Cavanagh, C. , G. Elliott and Stock, J. (1995) "Inference in Models with Nearly Integrated Regressors", *Econometric Theory*, 11, 1131–1147.

[6] Christiano, L. and Eichenbaum, M. (1990) "Unir Roots in Real GNP: Do we know and Do we care?" *Canergie Rochester Conference series on Public Policy*, 32: 7–62.

[7] Davidson, R. , and Mac Kinnon, J. (1998) "Graphical Methods for Investigating the size and Power of Hypothesis Tests" , *The Manchester School*, 66(1), 1–26.

[8] Feve, P. and Henin, P.Y. (1996) "Assessing Sustainability of Public Deficits in a Stochastic Environment" ,*mimeo*, CEPREMAP.

[9] Hakkio, G. and Rush, M. (1991) "Is the Budget Deficit too large?", *Economic Inquiry*, 29: 429–445.

[10] Hamilton, J. and Flavin, A. (1986) "On the Limitations of Government Borrowing: A Framework for Empirical Testing", *The American Economic Review*, 76 (4): 808–819, September.

[11] Hansen, B. (1995) "Rethinking the Univariate Approach to Unit Root Testing: using Covariates to Increase Power", *Econometric Theory*, 11: 1148–1171.

[12] Haug, A. (1995) "Has federal budget deficit policy changed in recent years?" *Economic Inquiry*, 104–118.

[13] Holzmann, R., Herve, Y. and Demmel, R. (1996) "The Maastrich Fiscal Criteria: Required but Ineffective" *Empirica*, 25–51.

[14] Muellbauer, J.(1996) "Income Persistence and MacroPolicy Feedbacks in the US" *Oxford Bulletin of Economic and Statistics*, 58 (4): 703–733.

[15] Phillips P. (1987) "Toward an Unified Asymptotic Theory for Autoregression" *Biometrika*, 74 :535–547.

[16] Phillips, P.C.B. (1991) "Optimal Inference in Cointegrated Systems" *Econometrica*, 59(2): 283–306, March.

[17] Phillips P. and Hansen B.(1990) " Statistical inference in instrumental variables regression with I(1) processes" *Review of Economic Studies*, 57:99–125.

[18] Quintos, C. (1995) "Sustainability of the Deficit Process with Structural Shifts", *Journal of Business Economic & Statistics*, 13(4): 409–417, October.

[19] Smith, G. and Zin, S. (1991) "Persistent Deficits and the Market Value of Government Debt", *Journal of Applied Econometrics*,6: 31–44.

[20] Schwert, G. (1989) "Test for Unit Root in Time Series Regression", *Journal of Business Economics & Statistics*,7: 147–159.

[21] Trehan, B. and Walsh, C. (1988) "Commond Trends, the Government Budget Constraint and Revenue Smoothing", *Journal of Economic Dynamics and Control*, 12:425–444,June/September

[22] Trehan, B. and Walsh, C. (1991) "Testing Intertemporal Bubget Constraint: Theory and Applications to US Federal Budget and Current Account Deficit", *Journal of Money Credit and Banking*, 23 (2): 206–223, May.

[23] Wickens M. and Uctum M. (1993) "The sustainability of current account deficits: a test of the US intertemporal budget constraint" *Journal of Economic Dynamics and Control*, 17(3):423–442, May.

[24] Wilcox, D. (1989) "The Sustainability of Government Deficits: Implications of the Present-Value Borrowing Constraint", *Journal of Money Credit and Banking*, 21 (3):291–306, August.