Economics Bulletin

Volume 33, Issue 4

Is the US current-account deficit sustainable? The importance of structural breaks in testing sustainability

Dimitris Hatzinikolaou
University of Ioannina, Department of Economics

Theodore Simos
University of Ioannina, Department of Economics

Agathi Tsoka
University of Ioannina, Department of Economics

Abstract

This paper improves the test for deficit sustainability developed by Hatzinikolaou and Simos (2013, henceforth HS) by taking into account structural breaks when deriving critical values for the test. Using quarterly data on the US current-account deficit, 1947.1-2012.2, we find that taking into account a structural break when testing for cointegration and when estimating a Box-Jenkins model for the derivation of the critical values renders the HS test more powerful than when ignoring the break. Thus, although HS failed to reject sustainability using their full sample period, 1947.1-2010.1, with the updated sample we can reject sustainability at the 10% level.

Citation: Dimitris Hatzinikolaou and Theodore Simos and Agathi Tsoka, (2013) "Is the US current-account deficit sustainable? The importance of structural breaks in testing sustainability", *Economics Bulletin*, Vol. 33 No. 4 pp. 2817-2827.
Contact: Dimitris Hatzinikolaou - dhatzini@cc.uoi.gr, Theodore Simos - tsimos@cc.uoi.gr, Agathi Tsoka - agathatsoka@yahoo.gr.
Submitted: March 15, 2013. **Published:** November 05, 2013.

Economics Bulletin, 2013, Vol. 33 No. 4 pp. 2817-2827

1. Introduction

In 1997-1998 the United States (US) current-account deficit (CAD) started to deteriorate and has since reached unprecedented levels, hence the current debate on its sustainability. This structural break is shown in Figure 1. Possible explanations include, first, the decline of the personal saving rate from over 5% at the beginning of the 1990s to nearly zero in 1998 and the ensuing "consumption boom," which increased imports. Second, the financial and exchange rate crises in Asia, Russia, and Brazil from the middle of 1997 to early 1999 contributed to an inflow of foreign capital into the US, "the safe haven," and caused the US dollar to appreciate, thus worsening the US CAD. Third, in the late 1990s there occurred a technological shift in the US, which increased productivity and investment spending. The return on US assets increased relative to that on foreign assets, thus stimulating capital inflow, appreciating the dollar, and worsening the US CAD; see Hervey and Merkel (2000, pp. 2-3), Holman (2001, pp. 7-14), and Pakko (1999, p. 15).

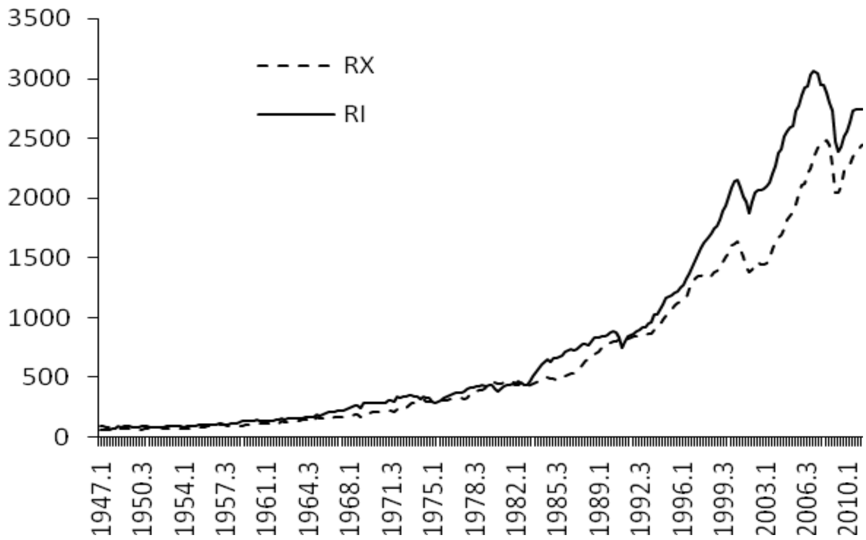

Figure 1. Real imports (*RI*) and exports (*RX*), *inclusive* of income payments and receipts

A country's CAD is said to be sustainable if there is no incentive for the country to default on its international debt. Technically, according to a standard definition, this implies that the country´s present discounted value of the international debt tends to zero, which means that the country´s expected future current-account surpluses equal, in present-value terms, to the market value of its current international debt. According to a more stringent definition, the CAD is sustainable if both the discounted debt converges to zero and the undiscounted debt is bounded.

Using the latter definition, Hatzinikolaou and Simos (2013, henceforth HS) recently developed a test for deficit sustainability and applied it to the US quarterly budget- and current-account deficit. Somewhat surprisingly, although they strongly rejected the sustainability of the US CAD for the sub-sample periods 1967.1-1989.4 and 1973.2-1998.4, they failed to reject it for their full sample period, 1947.1-2010.1.

This paper extends the work of HS by taking into account structural breaks when deriving critical values for the test. We find that taking into account the 1997-1998 break in the US CAD, by including a dummy variable in the cointegrating regression and in the Box-Jenkins models when deriving the critical values, renders the HS test more powerful than when ignoring the break. Other things equal, including the dummy variable reduces the standard error of estimate of the Box-Jenkins model, the critical values generated by it, and the *p*-values of the test. Thus, using an updated sample, 1947.1-2012.2, we can reject sustainability at the 10% level. Section 2 describes the data, section 3 estimates a "levels relationship" between exports and imports, section 4 implements the HS test, and section 5 concludes.

2. The data

We use the same variables as those in HS, namely: (1) X_t = real exports of goods and services plus income receipts from the rest of the world (compensation, interest, and dividends paid to US residents by foreigners, plus reinvested earnings on US direct investment abroad); (2) M_t = real imports of goods and services plus income payments plus real taxes and transfers paid to the rest of the world (net); (3) $XGNP_t = X_t/RGNP_t$, where $RGNP_t$ = real gross national product; (4) $MGNP_t = M_t/RGNP_t$; (5) $XPOP_t = X_t/POP_t$, where POP_t = population; (6) $MPOP_t = M_t/POP_t$; (7) $DEF_t = M_t - X_t$; (8) $DEFPOP_t = DEF_t/POP_t$; and (9) $DEFGNP_t = DEF_t/RGNP_t$. The variables X_t, M_t, and $RGNP_t$ are expressed in billions of 2005 dollars and are seasonally adjusted, whereas POP_t is in thousands of persons (mid-period). For this set of definitions, our sample consists of quarterly data for the period 1947.1-2012.1. Alternatively, when income payments and receipts are excluded from the above definitions, as in Husted (1992), the sample period is 1947.1-2012.2, and the $RGNP_t$ is replaced by $RGDP_t$ = real gross domestic product (Source: US Department of Commerce, Bureau of Economic Analysis).

We begin by testing the variables for unit-roots. Table 1 reports the results. Generally, the tests that ignore the presence of breaks show clearly that all of the above variables are I(1). Evidence for stationarity of some series is produced only by the Lee-Strazicich (2003, 2004) test, which takes into account breaks.

3. Empirical model and estimation

Following HS (section 3), we test sustainability by testing the hypotheses of cointegration (or, more generally, a "levels relationship") and H_0: $a = 0$, $b = 1$ based on the following equation:

$$y_{1t} = a + by_{2t} + \varepsilon_t, \qquad (1)$$

where y_{1t} = real exports and y_{2t} = real imports in levels or in ratios to *RGNP* (or *RGDP*) or to population, inclusive or exclusive of income payments and receipts. In each case, we use several methods to test for cointegration with structural breaks.

We begin by applying the Gregory and Hansen (1996a, b) tests to Equation (1). Note that these tests are sensitive to the choice of maximum lag length, k. Assuming $k = 24$, we find that three (out of six) pairs of variables form cointegrating regressions with a "level shift" at the beginning of 1998 (see Table 2). This evidence is consistent with Figure 1 and with the three explanations discussed in the Introduction.

Taking into account the information given in the previous paragraph, our strategy is to consider cointegration with a structural break in 1997:4 for all the six pairs of variables (y_{1t}, y_{2t}) listed in Table 2, and to use the dummy variable $D97_t$, which takes on the value of 1 for $t \geq$ 1997:4, and the value of zero otherwise. This strategy is justified by Johansen's trace test as well as by the "bounds test" (*BT*) of Pesaran, *et al*. (2001); see Table 2. In particular, when $D97_t$ is included among the regressors, these two tests provide strong evidence for cointegration, but in most cases this evidence vanishes if $D97_t$ is dropped, thus leading to *un*sustainability. This result demonstrates the importance of taking into account structural breaks when testing for sustainability.

Economics Bulletin, 2013, Vol. 33 No. 4 pp. 2817-2827

Table 1. Unit-root tests on US exports, imports, and current-account deficit series

Test / Series	PP_μ	PP_τ	$KPSS_\mu$	$KPSS_\tau$	*LS* one crash	*LS* two crashes	*LS* one break	*LS* two breaks	I(0) or I(1)?
Panel A. Current-account deficit series *inclusive* of income payments and receipts									
M_t	1.67	-1.06	4.40***	1.18***	-1.28	-1.33	-3.86	-5.12* (1996:3)	I(1)
X_t	2.42	-0.65	4.49***	1.19***	-1.32	-1.38	-4.57** (1988:2)	-5.74** (1984:4, 2005:1)	I(0)
$MGNP_t$	0.62	-1.75	4.71***	1.11***	-1.78	-1.88	-3.61	-5.13* (1996:2)	I(1)
$XGNP_t$	1.34	-2.55	4.85***	1.05***	-1.77	-1.86	-3.70	-4.63	I(1)
$MPOP_t$	1.07	-1.45	4.55***	1.15***	-1.64	-1.72	-3.71	-4.81	I(1)
$XPOP_t$	1.76	-1.35	4.67***	1.18***	-1.75	-1.83	-4.59** (1988:2)	-5.37* (1980:3, 2005:2)	I(0)
DEF_t	-1.18	-1.75	3.11***	0.62***	-2.96	-3.08	-4.72** (1996:4)	-6.40*** (insignif. dummies)	I(0)
$DEFPOP_t$	-1.41	-1.81	3.00***	0.55***	-3.08	-3.23	-4.54** (1996:4)	-6.00*** (insignif. dummies)	I(0)
$DEFGNP_t$	-2.25	-2.36	2.38***	0.42***	-3.30* (2001:4)	-3.45	-3.62	-4.59	I(1)
Panel B. Current-account deficit series *exclusive* of income payments and receipts									
M_t	2.41	-0.69	4.46***	1.21***	-1.06	-1.11	-3.59	-4.83	I(1)
X_t	3.68	-0.05	4.63***	1.24***	-1.03	-1.07	-4.69** (1984:3)	-5.58* (1987:3, 2001:2)	I(0)
$MGDP_t$	1.07	-1.46	4.71***	1.14***	-1.77	-1.88	-3.73	-4.65	I(1)
$XGDP_t$	1.72	-2.84	4.88***	0.98***	-1.08	-1.21	-2.47	-3.39	I(1)
$MPOP_t$	1.63	-1.18	4.61***	1.18***	-1.34	-1.43	-3.58	-4.77	I(1)
$XPOP_t$	2.69	-1.06	4.80***	1.22***	-1.19	-1.27	-4.26* (1981:4)	-5.11	I(1)
DEF_t	-0.51	-1.55	3.47***	0.77***	-2.53	-2.74	-4.68** (1997:2)	-6.33*** (2002:3)	I(0)
$DEFPOP_t$	-0.89	-1.66	3.40***	0.69***	-2.73	-2.81	-4.24* (1997:2)	-8.08*** (1990:3, 2001:4)	I(0)
$DEFGDP_t$	-1.97	-2.31	2.80***	0.52***	-2.90	-3.04	-3.25	-4.17	I(1)

Notes: (1) ***, **, * indicate significance at the 1%, 5%, and 10% level; (2) a series is taken to be I(0) only if there is evidence against the unit-root hypothesis at the 5% level; (3) the subscripts μ and τ indicate "intercept-but-no-trend" and "intercept-plus-trend," respectively; (4) *LS* denotes a Lee and Strazicich (2003, 2004) test; the break dates are given in parentheses, except when the values of the test statistic or the dummy variables are insignificant; (5) the tests provide strong evidence that the first differences of all the variables are I(0), so, for space considerations, test values are not reported for the first differences; (6) all the tests have been implemented by the econometric program *RATS* 7.0; (7) all data are expressed in real terms and are seasonally adjusted.

Table 2. Three cointegration tests

Test Regression	*GH* (*C*)	*GH* (*C* | *T*)	*GH* (Full break)	*BT*	*Trace* for H_0: $r = 0$ (lag length = ?)
Panel A: Income payments and receipts are included					
X_t on M_t	-4.39^{*} (1998:3)	-4.30	-4.50	11.59^{***}	21.68^{***} (lag length = 5)
$XGNP_t$ on $MGNP_t$	-3.72	-3.72	-3.50	11.96^{***}	21.55^{***} (lag length = 3)
$XPOP_t$ on $MPOP_t$	-4.16	-3.92	-4.14	9.59^{***}	19.87^{***} (lag length = 6)
Panel B: Income payments and receipts are excluded					
X_t on M_t	-4.71^{**} (1998:2)	-4.64	-4.57	15.25^{***}	31.12^{***} (lag length = 3)
$XGDP_t$ on $MGDP_t$	-3.70	-4.68	-3.70	18.90^{***}	24.98^{***} (lag length = 3)
$XPOP_t$ on $MPOP_t$	-4.52^{*} (1998:1)	-4.21	-4.24	15.71^{***}	22.75^{***} (lag length = 6)

Notes: (1) In all three tests, the null hypothesis (H_0) is "no cointegration"; (2) ***, **, * indicate rejection of H_0 at the 1%, 5%, and 10% level, respectively; (3) *GH* (*C*), *GH* (*C* | *T*), and *GH* (Full break) stand for Gregory-Hansen's "level shift," "level shift with trend," and "full break" models; maximum lag length was set equal to 24; the break point is given in parentheses; (4) *BT* stands for the Pesaran, *et al*. (2001) "bounds test" for a "levels relationship," where the maximum lag length was set equal to 8, and insignificant lags were dropped; standard errors are robust to heteroscedasticity and serial correlation; critical values are obtained from Table CI(iii) Case III of Pesaran, *et al*. (2001, p. 300); (5) these *BT* regressions do not include trend, but include the dummy variable $D97_t$, defined as $D97_t = 1$ for $t \geq 1997.4$, and 0 otherwise, which allows for a level shift; we assume that the presence of $D97_t$ in these regressions does not affect the critical values of the "bounds test," since the fraction of the observations where $D97_t = 1$ is only 0.22 (see Pesaran, *et al*., 2001, p. 307, Footnote 17); (6) *Trace* is Johansen's trace statistic; in every pair (y_{1t}, y_{2t}), the right-hand side variable (y_{2t}) is treated as weakly exogenous, based on the test calculated by the program, so the value of *Trace* is reported only for the cointegration rank (r) = 0; when it is significant, we conclude that there is one cointegrating vector (r = 1); critical values are simulated by the program *CATS in RATS* (see Dennis, 2006, p. 142); (7) the lag length used in each application of the Johansen procedure is given in parentheses underneath the value of *Trace*; (8) the dummy $D97_t$ is included in every regression, to allow a change in the intercept (a) of Equation (1).

Thus, we estimate the equation

$$y_{1t} = a + by_{2t} + cD97_t + \varepsilon_t \qquad (2)$$

and test the hypothesis H_0: $a = 0$, $b = 1$. We use two methods: (1) Johansen's method, implemented by the program *CATS in RATS* (Dennis, 2006); and (2) the method of Pesaran, *et al*. (2001). Table 3 reports the results.

As an illustration of the Johansen procedure, consider the first pair of variables in Panel A of Table 3, ($y_{1t} = X_t$, $y_{2t} = M_t$). We first determine the lag length (k) using a testing-down procedure: beginning with k = 12 lags, we test for 11 vs. 12 lags, then 10 vs. 12, then 10 vs. 11, etc. (see Dennis, 2006, pp. 140-141). In this way, we choose k = 5. Next, after imposing k = 5, we test for weak exogeneity of y_{1t} and y_{2t} using the test provided by the program (Dennis, 2006, pp. 12, 74-75). The p-value of the test for y_{1t} is 0.000, whereas that for y_{2t} is 0.282, so we treat y_{2t} as weakly exogenous. In the context of this restricted model, the value of the trace test statistic for the hypothesis that the cointegration rank is zero (r = 0) is 21.68 with p-value = 0.000, so we reject r = 0 and set r = 1.

Table 3. Estimation of Eqs. (2) & (3) by the methods of Johansen and of Pesaran *et al*. (2001)

	Estimation of Equation (2) by the Johansen procedure			Estimation of Equation (3) by the method of Pesaran, *et al*. (2001)		
Panel A: US current account *inclusive* of income payments and receipts						
Regression	X_t on M_t	$XGNP_t$ on $MGNP_t$	$XPOP_t$ on $MPOP_t$	X_t on M_t	$XGNP_t$ on $MGNP_t$	$XPOP_t$ on $MPOP_t$
$\hat{a}$	-81.86** (0.047)	-0.018*** (0.007)	-0.0004** (0.031)	-52.72** (0.013)	-0.176*** (0.001)	-0.0003*** (0.004)
$\hat{b}$	1.155*** (0.000)	1.136*** (0.000)	1.173*** (0.000)	1.074*** (0.000)	1.154*** (0.000)	1.082*** (0.000)
$\hat{c}$	-665.40*** (0.000)	-0.044*** (0.001)	-0.002*** (0.000)	-515.14*** (0.000)	-0.046*** (0.000)	-0.002*** (0.000)
p-values for H_0: *a*=0, *b*=1 and H_0: $b \geq 1$	0.107 0.966	0.009*** 0.936	0.090* 0.966	0.032** 0.842	0.004*** 0.982	0.008*** 0.857
p-values for *LM*(1) & *LM*(2)	0.843 0.879	0.101 0.785	0.054* 0.901	0.368 0.425	0.495 0.812	0.083* 0.228
p-value for *RESET*	–	–	–	0.0004***	0.0001***	0.0001***
Panel B: US current account *exclusive* of income payments and receipts						
Regression	X_t on M_t	$XGDP_t$ on $MGDP_t$	$XPOP_t$ on $MPOP_t$	X_t on M_t	$XGDP_t$ on $MGDP_t$	$XPOP_t$ on $MPOP_t$
$\hat{a}$	-110.21*** (0.011)	-0.025*** (0.001)	-0.0004** (0.019)	-57.92*** (0.003)	-0.023*** (0.000)	-0.0003*** (0.002)
$\hat{b}$	1.274*** (0.000)	1.196*** (0.000)	1.172*** (0.000)	1.034*** (0.000)	1.207*** (0.000)	1.056*** (0.000)
$\hat{c}$	-864.70*** (0.000)	-0.051*** (0.000)	-0.002*** (0.000)	-538.95*** (0.000)	-0.052*** (0.000)	-0.002*** (0.000)
p-values for H_0: *a*=0, *b*=1 and H_0: $b \geq 1$	0.024** 0.989	0.001*** 0.955	0.063* 0.916	0.011** 0.626	0.001*** 0.979	0.006*** 0.716
p-values for *LM*(1) & *LM*(2)	0.327 0.573	0.914 0.189	0.027** 0.642	0.911 0.733	0.724 0.554	0.656 0.898
p-value for *RESET*	–	–	–	0.0000***	0.1174	0.0000***

Notes: (1) ***, **, * indicate significance at the 1%, 5%, and 10% level, respectively; (2) the numbers in parentheses underneath coefficient estimates are the *p*-values for the χ^2 statistic for the hypothesis that the true coefficient is zero; (3) *LM*(1) and *LM*(2) are the standard Breusch-Godfrey *LM* tests for serial correlation of orders 1 and 2; (4) *RESET* refers to the hypothesis that the squared fitted value of the dependent variable is insignificant when added to the set of regressors; (5) in all the six applications of the Johansen procedure, the hypothesis of absence of ARCH is strongly rejected at the 1% level (the specific test values are not reported here for space considerations); (6) normality is also strongly rejected in both methods for every pair of variables, except for ($y_{1t} = XGDP_t$, $y_{2t} = MGDP_t$), in which case the *p*-value of the Bera-Jarque test is 0.43 when Johansen's method is used and 0.51 when the method of Pesaran, *et al*. (2001) is used; (7) when the latter method is used, an *LM* test also rejects homoscedasticity at the 1% level in every case, but this is no source of concern, since the method is robust to heteroscedasticity (and to autocorrelation).

The diagnostic tests indicate that: (1) there is no autocorrelation in the residuals, since the *p*-values of the *LM* statistics for autocorrelation of order 1 and 2 produced by the program are 0.843 and 0.879; and the same can be said about the other cases, except for the pair ($XPOP_t$, $MPOP_t$) in both panels of Table 3, where there exists some evidence of autocorrelation; (2) there exist strong ARCH effects (the *p*-values of the two *LM* statistics produced by the program are both 0.000), a result that persists for every pair of variables, so we do not report it in Table 3; and (3) the normality assumption is also strongly rejected (*p*-value = 0.000) for all pairs of variables, except for ($XGDP_t$, $MGDP_t$), in which case the *p*-value of the test produced by the program is 0.429. According to Gonzalo (1994), however, normality is not crucial for the Johansen procedure. Thus, since the ARCH effects do not introduce biases, the coefficient estimates and the tests of the various hypotheses produced by the Johansen procedure and reported in Table 3 are considered to be reliable.

Next, as an illustration of the method of Pesaran, *et al*. (2001), consider again the first pair of variables in Panel A of Table 3. After dropping the insignificant variables, we estimate the equation

$$\Delta y_{1t} = \beta_0 + \beta_1 y_{1t-1} + \beta_2 y_{2t-1} + \beta_3 D97_t + \varphi_1 \Delta y_{1t-1} + \psi_0 \Delta y_{2t} + \Sigma_{i=3}^{4} \psi_i \Delta y_{2t-i} + \psi_8 \Delta y_{2t-8} + \varepsilon_t. \quad (3)$$

The parameters of interest, i.e., *a*, *b*, and *c* of Equation (2), can be recovered from Equation (3) by setting $\Delta y_{1t-i} = \Delta y_{2t-i} = 0$, $i = 0, 1, 3, 4, 8$, and then leading the equation by one period and solving for y_{1t}. The result is $a = -\beta_0/\beta_1$, $b = -\beta_2/\beta_1$, and $c = -\beta_3/\beta_1$. Thus, in the context of Equation (3), testing the hypothesis H_0: $a = 0$, $b = 1$ amounts to testing H_0: $\beta_0 = 0$, $\beta_1 + \beta_2 = 0$; testing H_0: $c = 0$ amounts to testing H_0: $\beta_3 = 0$; etc. Note also that the "bounds test" of cointegration (see Table 2) is a standard *F*-test of the hypothesis $\beta_1 = \beta_2 = 0$ in Equation (3), but with critical values obtained from Table CI(iii) Case III of Pesaran *et al*. (2001, p. 300).

The diagnostic tests indicate that: (1) there is no autocorrelation in the residuals, since the *p*-values of the standard Breusch-Godfrey *LM* statistics for autocorrelation of order 1 and 2 are 0.368 and 0.425; and the same can be said about the other pairs of variables, except for the pair ($XPOP_t$, $MPOP_t$) in panel A of Table 3, where there exists some evidence of autocorrelation at the 10% level; (2) an *LM* test (not reported in Table 3) rejects homoscedasticity at the 1% level in every case, but this is no source of concern, since the method is robust to heteroscedasticity (and to autocorrelation); (3) except for the pair of variables ($XGDP_t$, $MGDP_t$), in which case the *p*-value of the Bera-Jarque test of normality (not reported in Table 3) is 0.51, this test strongly rejects normality (*p*-value = 0.000) in every other case; since our sample is fairly large, however, we invoke the central limit theorem to overcome this problem; (4) except for the pair ($XGDP_t$, $MGDP_t$), in which case the *p*-value of the standard *RESET* is 0.12, this test strongly rejects the chosen specification in every other case (see the last row of each panel of Table 3). Notice, however, that the equation generally passes the autocorrelation tests, which are often thought of as specification tests. In addition, the results that emerge when the *RESET* rejects are similar to those when it does not reject, and are also similar to those produced by the Johansen procedure, which we have already deemed reliable. Thus, we consider all the coefficient estimates and all the tests of the hypotheses reported in Table 3 reliable.

We now follow the steps listed in section 3 of HS to test sustainability. First, since for every pair of variables (y_{1t}, y_{2t}) there is cointegration (see Table 2) and the hypothesis H_0: $a = 0$, $b = 1$ can be rejected (see Table 3), we test the following two left-sided hypotheses separately: (i) H_0: $b \geq 1$ against H_1: $b < 1$; and (ii) H_0: $a \geq 0$ against H_1: $a < 0$. Table 3 shows that in no case can the hypothesis H_0: $b \geq 1$ be rejected,[1] whereas the hypothesis H_0: $a \geq 0$ can be rejected in every case at the 5% level. In addition, the two-sided hypothesis H_0: $c = 0$ can be rejected at the 1% level. Since all the estimates of c are negative, the last two results lead to the inference that the parameter a is negative, and became even more negative after 1997:4. This is case 3b in the testing procedure of HS (see their section 3), so we need to test their condition (9). In the next section, we calculate new critical values for the HS test, since we have an updated sample and evidence of a structural break in 1997:4.

4. The HS test with structural breaks

First, note that we will not apply the HS test to the definitions *DEFGNP* and *DEFGDP*, which seem to be I(1) (see Table 1), because the test requires that the deficit series be I(0). Next, note that the critical values reported in the first column of Table 5 of HS, where the definitions of CAD *include* income payments and receipts, have been produced using an ARMA(3, 9) model without the AR(2) and MA(5) terms for *DEF*; and an ARMA(4, 7) model without the AR(2), AR(3), and MA(1)-MA(5) terms for *DEFPOP*. With the HS sample, 1947.1-2010.1, whenever the dummy $D97_t$ is included as an additional regressor at the stage of identification of an ARMA model, it is not found to be statistically significant at the 10% level, so it is dropped.

With the updated sample, 1947.1-2012.1, we, too, are unable to beat the above ARMA(3, 9) model for *DEF* and are also unable to reject sustainability. As for *DEFPOP*, if $D97_t$ is ignored, we, too, are unable to beat the above ARMA(4, 7) model, but with the updated sample we can now reject sustainability at the 10% level (see the first column of our Table 4). If, on the other hand, $D97_t$ is included in the candidate ARMA models, then we could choose an alternative model, namely ARMA(4, 6) without the AR(2), MA(1), MA(2), and MA(4) terms, and with $D97_t$ as an additional regressor, which is now marginally significant at the 10% level (p-value = 0.0998). In this case, we can also reject sustainability at the 10% level (see the second column of our Table 4).

For the definitions of CAD that *exclude* income payments and receipts, and for their full-sample period (1947.1-2010.1), HS used an ARMA(4, 4) model without the MA(2) term for *DEF*, and an ARMA(5, 4) model without the AR(2) and the MA(1)-MA(2) terms for *DEFPOP*. If $D97_t$ is included, it is significant at the 10% level in the case of *DEF* (p-value = 0.071), but insignificant in the case of *DEFPOP* (p-value = 0.111). HS did not reject sustainability, regardless of whether $D97_t$ is included or not.[2] Note, however, that including $D97_t$ in the ARMA model reduces its standard error of estimate and the critical values generated by it, thus leading to rejections more easily; and that this is true even in the case of *DEFPOP* where $D97_t$ is not statistically significant. Thus, from a different point of view, we reach the same conclusion as that of the previous section: taking into account structural breaks when testing for sustainability makes the test more powerful.

With the updated sample, 1947.1-2012.2, if we use the above ARMA models for the definitions of CAD that *exclude* income payments and receipts, we, too, are unable to reject sustainability. Using a slightly better model for *DEFPOP* than the above ARMA(5, 4) model, however, we are able to reject sustainability at the 10% level. In particular, based on the standard criteria of model selection (see HS, section 4), we choose an ARMA(4, 3) model without the AR(2) and MA(2) terms.

[1] According to the standard Hakkio and Rush (1991) test for sustainability, since for every pair of variables considered here there is evidence for cointegration (see Table 2) and $b \geq 1$, the hypothesis of even *strong* sustainability cannot be rejected. As we will see below, however, the HS test rejects sustainability at the 10% level.

[2] HS did not report these results in their paper, but referred to an appendix available upon request.

Table 4 reports critical values (*CV*s) as well as the values of the HS test statistic (*TS*) calculated from the actual data, along with its *p*-value, only for the ARMA models that produce rejections. These values have been calculated in the same way as in HS, so, to save space, we will not describe the methodology here any further.

According to the results of Table 4, sustainability is rejected at the 10% level in almost every case. Note, in particular, that when $D97_t$ is included in the above ARMA(4, 3) model for *DEFPOP*, where it is significant at the 10% level (*p*-value = 0.084), the *p*-value of *TS* is smaller than when $D97_t$ is ignored (see the last two columns of Table 4). Thus, for example, if we use Equation (12) of HS and an 8.5% level of significance, then including $D97_t$ leads to rejection, whereas ignoring it leads to *non*-rejection. Once again, taking into account the structural break increases the power of the test.

Table 4. Values of the HS test statistic (*TS*) calculated from the actual data for two definitions of CAD, and 1%, 5%, and 10% critical values (*CVs*)

DEFPOP, *inclusive* of income payments and receipts; ARMA(4, 7) without AR(2), AR(3), MA(1)-MA(5) terms; $D97_t$ is *not* included; data 1947.1-2012.1	*DEFPOP*, *inclusive* of income payments and receipts; ARMA(4, 6) without AR(2), MA(1), MA(2), and MA(4) terms; $D97_t$ *is* included; data 1947.1-2012.1	*DEFPOP*, *exclusive* of income payments and receipts; ARMA(4, 3) without AR(2) and MA(2) terms; $D97_t$ *is* included; data 1947.1-2012.2	*DEFPOP*, *exclusive* of income payments and receipts; ARMA(4, 3) without AR(2) and MA(2) terms; $D97_t$ is *not* included; data 1947.1-2012.2
Equation (12) of HS:	Equation (12) of HS:	Equation (12) of HS:	Equation (12) of HS:
q = 10: *TS* = 4.17* [0.084], *CVs* = 7.80, 5.05, 3.83	q = 10: *TS* = 4.17* [0.055], *CVs* = 6.66, 4.29, 3.25	q = 10: *TS* = 4.63* [0.081], *CVs* = 8.67, 5.60, 4.19	q = 10: *TS* = 4.63* [0.092], *CVs* = 9.21, 5.94, 4.43
q = 7: *TS* = 4.84* [0.083], *CVs* = 8.95, 5.82, 4.41	q = 7: *TS* = 4.84* [0.054], *CVs* = 7.66, 4.95, 3.76	q = 7: *TS* = 5.40* [0.081], *CVs* = 10.01, 6.49, 4.86	q = 7: *TS* = 5.40* [0.091], *CVs* = 10.64, 6.89, 5.14
q = 4: *TS* = 6.08* [0.081], *CVs* = 11.14, 7.26, 5.51	q = 4: *TS* = 6.08* [0.053], *CVs* = 9.51, 6.17, 4.69	q = 4: *TS* = 6.79* [0.080], *CVs* = 12.47, 8.13, 6.09	q = 4: *TS* = 6.79* [0.091], *CVs* = 13.27, 8.63, 6.45
Equation (14) of HS:	Equation (14) of HS:	Equation (14) of HS:	Equation (14) of HS:
TS = 1.67* [0.093], *CVs* = 3.74, 2.24, 1.61	*TS* = 2.09* [0.097], *CVs* = 6.46, 3.05, 2.04	*TS* = 1.76 [0.130], *CVs* = 6.05, 3.14, 2.12	*TS* = 1.57 [0.150], *CVs* = 6.09, 3.15, 2.14

Notes: (1) ***, **, and * denote significance at the 1%, 5%, and 10% level; (2) the critical values (*CVs*) are derived from simulations with 50,000 replications; (3) *DEFPOP* is the real current account deficit in per capita terms; (4) Equations (12) and (14) of HS give two alternative estimators of the spectrum; (5) q is the number of autocovariances of the deficit series taken into account in the first estimator of the spectrum, Equation (12) of HS, which employs the Bartlett kernel; (6) the number in square brackets immediately after the value of *TS* is its *p*-value; (7) results for the ratios of the current-account deficit to *RGNP* or *RGDP* are not presented, because our unit-root tests of Table 1 suggest that these ratios may be I(1), so the HS test may not be applicable in these cases; (8) results for the definition *DEF* are not presented either, because sustainability is not rejected when this definition of CAD is used.

5. Conclusion

This paper extends the recent work of Hatzinikolaou and Simos (2013), who failed to reject the hypothesis of sustainability of the US current-account deficit using their test and their full-sample period, 1947.1-2010.1. The present paper, which uses the same test, but an updated sample, 1947.1-2012.2, demonstrates the importance of taking into account structural breaks when testing sustainability. The evidence suggests that a structural break occurred around the end of 1997, when the US current-account deficit started to deteriorate. Some possible reasons are cited in section 1.

We find that taking into account the structural break when testing for cointegration and when estimating a Box-Jenkins model that is used to derive critical values for the HS test renders the test more powerful than when ignoring the break. Note in particular that, other things equal, including the dummy variable for the break as an additional regressor in the "best" Box-Jenkins model reduces its standard error of estimate, the critical values generated by it, and the *p*-values of the test. As a result, with the updated sample, the HS test rejects sustainability at the 10% level. Thus, at the 10% level, the conclusion we reach in the present paper is that the US may have difficulty in marketing its foreign debt in the long run.

REFERENCES

Dennis, J.G. (2006) *CATS in RATS*: *Cointegration Analysis of Time Series*, v. 2, Estima: Evanston, IL.

Gonzalo, J. (1994) "Five alternative methods of Estimating long-run equilibrium relationships" *Journal of Econometrics* **60**, 203-233.

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

--- (1996) "Tests for cointegration in models with regime and trend shifts" *Oxford Bulletin of Economics and Statistics* **58**, 555-560 (b).

Hakkio, C.S. and M. Rush (1991) "Is the budget deficit "too large?" *Economic Inquiry* **29**, 429-445.

Hatzinikolaou, D. and T. Simos (2013) "A new test for deficit sustainability and its application to US data" *Empirical Economics* **45**, 61-79.

Hervey, J.L. and L.S. Merkel (2000) "A record of current account deficit: causes and implications" Federal Reserve Bank of Chicago *Economic Perspectives* **14**, 3-12.

Holman, J.A. (2001) "Is the large US current account deficit sustainable?" *Economic Review*, Federal Reserve Bank of Kansas, 1st Quarter, 5-23.

Husted, S. (1992) "The emerging U.S. current account deficit in the 1980s: a cointegration analysis" *The Review of Economics and Statistics* **74**, 159-166.

Lee, J. and M.C. Strazicich (2003) "Minimum Lagrange Multiplier Unit Root Test with Two Structural Breaks" *Review of Economics and Statistics* **85**, 1082-1089.

--- (2004) “Minimum LM Unit Root Test with One Structural Break” http://econ.appstate.edu/RePEc/pdf/wp0417.pdf

Pakko, M.R. (1999) “The US trade deficit and the ‘new economy’” Federal Reserve Bank of St. Louis, *Review*, September/October, 11-19.

Pesaran, M.H, Y. Shin, and R.J. Smith (2001) “Bounds testing approaches to the analysis of level relationships” *Journal of Applied Econometrics* **16**, 289-326.