IS THE BUDGET DEFICIT "TOO LARGE?"

CRAIG S. HAKKIO and MARK RUSH*

Yes. Specifically, we find that recent spending and taxing policies of the government—if continued—violate the government's intertemporal budget constraint. As a result, government spending must be reduced and/or tax revenues must be increased. These conclusions are based on tests of whether government spending and revenue are cointegrated. In addition to examining real spending and revenue, we also normalize these variables by real GNP and population. For a growing economy, these normalized measures are perhaps more pertinent. We also test and find support for the hypothesis that deficits have become a problem only in recent years.

I. INTRODUCTION

The question—"Is the government's budget deficit too large?"—has received an immense amount of public attention. Recently, economists have begun to study this question more intensively in a series of articles that explicitly investigate the government's intertemporal budget constraint.[1] In one of the first papers to adopt this approach, Hamilton and Flavin [1986] examined twenty-three years of annual data—1962 to 1984—to determine whether the deficit follows a stationary stochastic process. They conclude that it does and, therefore, that the data are consistent with the government's intertemporal budget constraint.

Wilcox [1987] and Kremers [1989] have extended Hamilton and Flavin's work. Wilcox allows for stochastic variation in the real interest rate and also tests for parameter instability; Kremers provides evidence suggesting that Hamilton and Flavin's unit root test was misspecified. Both authors claim to reverse Hamilton and Flavin's finding that the deficit is stationary in recent years. They conclude that the government's behavior is inconsistent with its intertemporal budget constraint.

Taking a slightly different approach are Trehan and Walsh [1988; 1990], Smith and Zin [1988], and Haug [undated]. These authors focus directly on government

* Craig S. Hakkio is an assistant vice president and economist at the Federal Reserve Bank of Kansas City. Mark Rush is an associate professor at the University of Florida. Part of this paper was written while Mark Rush was a Visiting Scholar at the Federal Reserve Bank of Kansas City and while Craig S. Hakkio was at the Council of Economic Advisers. We thank the Financial Institutions Center at the University of Florida for helping support this research. Anindya Banerjee, Sean Becketti, Andy Filardo, Larry Kenny, V. Vance Roley, David Wilcox, and participants of workshops at the Federal Reserve Bank of Kansas City, the Board of Governors of the Federal Reserve System, and the University of Illinois made useful comments. We thank Michael Cox for providing us with the data on the market value of Federal debt. Mike Grace supplied valuable research assistance. Several anonymous referees supplied many useful comments; their value-added is evident on almost every page. The views expressed herein are solely those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Kansas City, nor of the Federal Reserve System, nor of the Council of Economic Advisers.

1. Earlier, Buchanan and Wagner [1977] used a public choice approach to argue that without major structural change—in particular, a constitutional amendment requiring a balanced budget—the budget deficit was too large and would remain so. However, Barro [1979; 1986], using a "traditional" economic framework concluded that until at least 1985 there is no evidence of structural change in the factors that determine the size of the deficit. Although mentioned by Barro, none of these studies took account of the government's intertemporal budget constraint.

Economic Inquiry
Vol. XXIX, July 1991, 429–445

spending and revenue, and use newly developed tests for cointegration. Trehan and Walsh [1988] use yearly U.S. data from 1890 to 1983; Smith and Zin use monthly Canadian data from 1946 to 1984; and Haug uses quarterly U.S. data from 1960 to 1986. All three papers reach similar conclusions: the government's behavior is consistent with its intertemporal budget constraint.

We also focus on cointegration of government spending and revenue, but our approach differs from the above authors in several ways. First, unlike Haug, and Smith and Zin, who implicitly assume that the (expected) real interest rate is constant, we allow for fluctuations in the interest rate. Second, unlike Trehan and Walsh, we use several different sample periods to test the view that deficits have become a problem only in recent years. Our data run from 1950:II to 1988:IV; we test for cointegration over the whole sample and over subsamples that run from 1964:I to 1988:IV (one hundred observations) and from 1976:III to 1988:IV (fifty observations). Finally, in addition to examining revenue and spending directly, we also normalize these variables using real GNP and population. This is an important extension beyond previous work since McCallum [1984], among others, deems these ratios—per capita spending and revenue, and spending and revenue as a fraction of GNP—as more pertinent for a growing economy.

Foreshadowing our results, we find evidence indicating that *recent* spending and taxing policies of the government—if continued—violate the government's intertemporal budget constraint. This contrasts with the more sanguine view provided by authors who do not concentrate on recent policies. We conclude that to be consistent with its intertemporal budget constraint, eventually government spending must be reduced and/or tax revenue must be increased.

The next section of the paper examines the implications of the government's intertemporal budget constraint for government spending and revenue. The second section discusses the theory of cointegration, while the third section presents the empirical results. The fourth section presents the results from using an alternative measure of revenue. Finally, the last section summarizes the paper's main conclusions.

II. THE GOVERNMENT'S INTERTEMPORAL BUDGET CONSTRAINT

Each period the Federal government faces a budget constraint. For ease of exposition, assume that all government bonds have a one period maturity. Then, the government's one-period budget constraint can be written as

$$(1) \qquad G_t + (1 + i_t)B_{t-1} = R_t + B_t .$$

In equation (1), B_t is the funds raised by issuing new debt; R_t is the government's revenue; i_t is the (one-period) interest rate; and G_t is the value of government purchases of goods and services plus transfer payments. Notice that G_t does not include interest payments on the debt; these interest payments, as well as the principal that must be repaid, are the second term on the expenditure side. At this point in the discussion, the variables in equation (1) can be nominal, real, or "deflated" by population or real GNP.[2] As we will discuss below, the choice depends on auxiliary assumptions.

2. The interpretation of the interest rate in equation (1) depends on how government spending and revenue are measured. When variables are nominal, i_t is the nominal interest rate; when variables are real, i_t is the real interest rate; when variables are real per real GNP, i_t is the real interest rate minus the rate of growth of real GNP; and when variables are real per capita, i_t is the real interest rate minus the rate of population growth. In addition, the real value of B_{t-1} is the nominal value (at time t–1) divided by the price level at time t–1.

The budget constraint in equation (1) pertains to period t; there is a similar constraint for periods $t+1$, $t+2$, and so on. By solving equation (1) "forward," these single period budget constraints can be combined to yield the government's intertemporal budget constraint. Letting r_t be the discount factor, the solution is given in equation (2):

$$(2) \qquad B_0 = \sum_{t=1}^{\infty} r_t(R_t - G_t) + \lim_{n\to\infty} r_n B_n$$

where

$$r_t = \prod_{s=1}^{t} \beta_s \text{ and } \beta_s = 1/(1 + i_s).$$

The crucial element in the intertemporal budget constraint is the last term, $\lim(r_n B_n)$, where the limit is taken as $n \to \infty$. When this limit equals zero, the intertemporal budget constraint merely asserts that the outstanding stock of bonds, B_0, equals the present value of the government's budget surpluses. Equation (2) shows that the "correct" budget surplus excludes interest payments as an expenditure. In particular, equation (2) shows that the conventional definition of the budget deficit, total spending minus tax revenue, $G_t + i_t B_{t-1} - T_t$, is not the economically relevant definition (T_t is tax revenue).

As discussed more extensively by Hamilton and Flavin [1986], Barro [1987], and McCallum [1984], the limiting value of $(r_n B_n)$ must equal zero in order to rule out the possibility of the government financing its deficit by issuing new debt. If the limit term does not equal zero, the government is bubble-financing its expenditures, in which old debt that matures is financed by issuing new debt. We interpret the assertions that the government's deficit is "too large" as claiming that the government is using this sort of Ponzi scheme to finance its deficit.

Thus, we try to determine if the data are consistent with the condition that $\lim(r_n B_n) = 0$. To do this, we ask the following question: if R and G continue to follow the stochastic process that we estimate, is $\lim(r_n B_n) = 0$? In other words, this paper investigates the following condition:

$$(3) \qquad E[\lim_{n\to\infty}(r_n B_n) \mid R \text{ and } G \text{ follow historical stochastic process}] = 0.$$

This is obviously a difficult question since we often know things about the future that are not included in the historical record. For example, between now and 2030, the social security trust fund is predicted to run surpluses each year and to ultimately have over \$5 trillion in its accumulated fund. If we examined the social security trust fund using data between 1989 and 2030, we might conclude that the accumulated fund was going off to infinity. However, because of demographic changes (that are known today), the yearly surpluses are predicted to turn to deficits after 2030, and by 2050 the accumulated fund will be in deficit. The point is that predicted demographic changes may provide information not captured in the historical process. So too may be the case with government spending and revenue.[3] These observations present an unavoidable weakness in our paper. Of course, this drawback exists in all work that focuses on the time series behavior of data. Moreover, in the present case, we know of no extraneous factor that is predicted to affect R and G in the future, so

3. Suppose we use historical values of R, G, and i. In this case, i incorporates information about future values of R and G (since it is an asset price). Therefore, the analysis using historical values of R and G may provide little information about future R and G; it is the current value of i that provides information about the "correct" values of future R and G.

it may be that this is not a major problem for our results.[4]

While the intertemporal budget constraint is generally written as equation (2), we need an alternative equation to derive testable implications[5]. Assume that the interest rate is stationary, with unconditional mean equal to i. (This means we cannot analyze the government's budget constraint in nominal terms, since the stationarity of the nominal interest rate is questionable.) Add and subtract iB_{t-1} to both sides of equation (1) to obtain

$$(4) \qquad E_t + (1+i)B_{t-1} = R_t + B_t$$

where $E_t = G_t + (i_t - i)B_{t-1}$. Equation (4) holds every period. Solving equation (4) "forward" yields

$$(5) \quad B_{t-1} = \sum_{j=0}^{\infty} \beta^{j+1}(R_{t+j} - E_{t+j}) + \lim_{j\to\infty} \beta^{j+1} B_{t+j}.$$

where

$$\beta = 1/(1+i).$$

After much laborious mathematical manipulation, equation (5) can be rewritten as

$$(6) \quad G_t + i_t B_{t-1} = R_t + \sum_{j=0}^{\infty} \beta^{j-1}(\Delta R_{t+j} - \Delta E_{t+j}) + \lim_{j\to\infty} \beta^{j+1} B_{t+j} \; .$$

For ease of exposition, let GG denote total government spending on goods and services, transfer payments, and interest on the debt; GG equals the left-hand side of equation (6).

Assume that R and E are non-stationary, so that ΔR_t and ΔE_t are stationary. In particular, assume that R and E follow random walks with drift:

$$R_t = \alpha_1 + R_{t-1} + \varepsilon_{1t}$$

$$E_t = \alpha_2 + E_{t-1} + \varepsilon_{2t}.$$

In this case, equation (6) can be rewritten as

$$(7) \qquad GG_t = \alpha + R_t + \lim_{j\to\infty} \beta^{j+1} B_{t+j} + \varepsilon_t$$

where

$$\alpha \equiv \Sigma\beta^{j-1}(\alpha_1 - \alpha_2) = [(1+i)/i](\alpha_1 - \alpha_2)\,,$$

and $\varepsilon_t \equiv \Sigma\beta^{j-1}(\varepsilon_{1t} - \varepsilon_{2t})$.

Equation (7) forms the basis of the hypothesis tests in this paper. First, assume the limit term in equation (7) is zero, and rewrite equation (7) as a regression equation:

$$R_t = a + b\, GG_t + \varepsilon_t.$$

The null hypothesis is $b = 1$ and ε_t stationary. In other words, if GG and R are non-stationary, the null hypothesis is that $b = 1$ and that GG and R are cointegrated.[6]

4. The Gramm-Rudman-Hollings law does imply something about future values of the budget deficit and, therefore, something about the difference between R and G. However, the law may prove to be ineffective.

5. Other researchers frequently have proceeded by taking an expected value in equation (1) and iterating it forward to arrive at a stochastic version of equation (2). This procedure is not strictly correct, since equation (1) is an accounting identity. Hence, (1) *must* hold for all realizations of G_t and R_t, not just the "average" or "expected" realizations. Although this point has no apparent impact on the validity of the other work, equation (3) and the surrounding discussion present what we consider to be the appropriate approach.

6. Notice that if, say, GG is non-stationary while R is stationary, there is no long-run relationship between GG and R. Intuitively, this implies the government is violating its intertemporal budget constraint because GG tends to grow while R does not. More formally, in this case the $\hat{b}$ converges to zero. Therefore, the limit term, for which we derive an expression (equation (8)), does not equal zero. This violates the intertemporal budget constraint.

When GG and R are non-stationary, cointegration is a necessary condition for the government to obey its present value budget constraint. However, the condition $b = 1$ is not, strictly speaking, a necessary condition for the government's budget constraint to hold. To see this, substitute $\hat{a} + \hat{b}GG_t$ for R_t in equation (1), assume for simplicity that $i_t = i$ for all t, and iterate forward to obtain

$$B_{t+j} = \sum_{k=0}^{j} [1 + (1 - \hat{b})i]^{j-k} S_{t+k} + [1 + (1 - \hat{b})i]^j B_{t-1}$$

where S_t equals spending $([1 - \hat{b}]G_t - \hat{a})$. Using this equation, the limit term in equation (7) can be written as

$$(8)\quad \lim_{j\to\infty} \left(\sum_{k=0}^{j} \{[1 + (1 - \hat{b})i]^{j-k} / (1 + i)\}^{j+1} S_{t+k} + \{[1 + (1 - \hat{b})i]^j / (1 + i)^{j+1}\} B_{t-1} \right).$$

This term will equal zero as long as $0 < \hat{b} < 1$.

Although the limit term equals zero, the limit of the undiscounted value of B equals infinity when $\hat{b} < 1$. This introduces a new element in the interpretation of the intertemporal budget constraint. As the undiscounted value of B gets large, the incentive for the government to default becomes large, especially when revenue and spending are expressed relative to real GNP or population. If $\hat{b} < 1$ (and revenue and spending are measured relative to, say, GNP), the real value of debt relative to GNP diverges to infinity. Thus, as Barro [1979], McCallum [1984], and Kremers [1988; 1989] have noted, the incentive to default grows and the government will face increasing difficulty in marketing its debt. Therefore, although $\hat{b} < 1$ is consistent with a strict interpretation of the government's intertemporal budget constraint, it is inconsistent with the requirement that debt to GNP (or debt per capita) must be finite and hence may be inconsistent with the government's ability to market its debt.

Thus, we focus on two issues: (1) Are GG and R cointegrated? That is, is ε_t stationary in equation (7)? and (2) is $\hat{b} = 1$? The first condition is necessary; the second condition is probably necessary.

Cointegration of GG and R is consistent with McCallum's [1984, 129–130] discussion of the government's intertemporal budget constraint. McCallum argued that a constant, positive deficit (excluding interest payments) cannot be financed entirely by bond sales; however, a constant, positive deficit (including interest payments) can be financed entirely by bond sales. The difference between GG and R is the deficit including interest payments. Therefore, the focus on GG and R—rather than G and R—seems appropriate.

III. COINTEGRATION

Cointegration is a relatively new statistical concept, pioneered by Granger [1983], Granger and Weiss [1983], and Engle and Granger [1987]. Cointegration is a property possessed by some non-stationary time series data. Basically, two variables are said to be cointegrated when each variable taken separately is non-stationary, yet a linear combination is stationary.

More precisely, consider two time series—R and GG. Assume that both R and GG are non-stationary and need to be differenced once to induce stationarity. In general, most linear combinations of R and GG, such as $R_t - aGG_t = v_t$, are also non-stationary. However, there may be a number, b, such that $R_t - bGG_t = u_t$ is stationary. In this case, R and GG are said to be cointegrated of order (1,1), with cointegration vector (1, $-b$).[7] Thus, if R and GG are cointegrated with vector (1, -1), they cannot drift too far apart because their difference, $R_t - GG_t = u_t$, is stationary. If they are not cointegrated, however, they will, with probability one, drift arbitrarily far apart since their difference can—and will—take on arbitrarily large values. Engle and Granger have expressed the intuition behind this idea as: "An individual economic variable, viewed as time series, can wander extensively yet some pairs of series [if they are cointegrated] may be expected to move so that they do not drift too far apart" [1987, 251].

Engle and Granger have proposed seven tests for cointegration. Six of the tests come in pairs; in our empirical section we report only the "augmented" versions of the tests. Based on simulations, Engle and Granger report critical values for one hundred observations and Engle and Yoo report critical values for fifty observations. The null hypothesis is no cointegration for all the tests; that is, a large test statistic rejects the null and "accepts" the alternative of cointegration.

All of the tests advocated by Engle and Granger involve estimating the so-called "equilibrium" or "cointegration" regression

$$R_t = a + bGG_t + \mu_t.$$

The first three tests focus on whether the estimated residuals, u_t, are stationary; the last four tests examine whether GG and R obey an "error-correction" process.

7. The first term of the order of cointegration pertains to the number of times it is necessary to difference the individual data series to attain stationarity. The second term is the reduction in the number of times it is necessary to difference the linear combination to achieve stationarity. Thus, taken separately, R_t and GG must be differenced once to achieve stationarity, but the linear combination $R_t - bGG_t$ need not be differenced at all to achieve stationarity, a reduction of one from what was required for R_t and GG_t.

The first test involves the Durbin-Watson statistic from the equilibrium regression. If the Durbin Watson is "big," the two series are cointegrated because the residuals are stationary.

The first pair of tests use Dickey-Fuller regressions, as described in Dickey and Fuller [1979], to test whether the estimated time series of the residuals from the equilibrium regression has a unit root: if there is a unit root, then GG and R are not cointegrated. The first test of the pair estimates the regression

$$\Delta u_t = -\rho u_{t-1} + e_t$$

and examines the significance of the estimated ρ. If the estimated ρ equals zero, u is non-stationary and so GG and R are not cointegrated; however, if ρ is significantly different from zero, u is stationary, and the hypothesis of cointegration is "accepted." The second test of the pair, the augmented Dickey-Fuller test (ADF), includes additional lags of Δu in the regression.

The next pair of tests use the fact that cointegrated variables can always be written in an error correction form. The first test of the pair requires estimation of two equations:

$$\Delta R_t = c_1 + \beta_1 u_{t-1} + \upsilon_t$$

$$\Delta GG_t = c_2 + \beta_2 u_{t-1} + \gamma \Delta R_t + \nu_t\,.$$

If the estimated β_1 and β_2 are jointly significant, then GG and R have an error-correction representation and are therefore cointegrated. The second test of the pair, the augmented restricted VAR test—ARVAR—includes additional lags of ΔGG and ΔR in the regression.

The last pair of tests is based on a vector autoregression in which the levels of GG and R are allowed to enter; cointegration restrictions are not imposed on the model.

The first test of the pair requires the estimation of two equations:

$$\Delta R_t = c_1 + \beta_{11} R_{t-1} + \beta_{12} GG_{t-1} + \upsilon_t$$

$$\Delta GG_t = c_2 + \beta_{21} R_{t-1} + \beta_{22} GG_{t-1} + \gamma \Delta R_t + \nu_t.$$

The unrestricted VAR test, UVAR, tests for the joint significance of the estimated values of β_{11}, β_{12}, β_{21}, and β_{22}; under the null hypothesis, these should be zero—levels should not enter the set of equations. The second test of the pair, the augmented unrestricted VAR test—AUVAR—includes additional lags of ΔGG and ΔR in the regression.

Engle and Granger's tests for cointegration may not be appropriate if the non-stationary variables have a drift. Therefore, we also used the tests proposed by Stock and Watson [1988], which are appropriate in this case.[8] The Stock and Watson test calculates the number of common stochastic trends in GG and R. Intuitively, if GG and R are cointegrated, there is one common stochastic trend; if they are not cointegrated, then there must be two stochastic trends. The null hypothesis is two stochastic trends (not cointegrated) and the alternative is one common stochastic trend (cointegrated).

IV. EMPIRICAL RESULTS

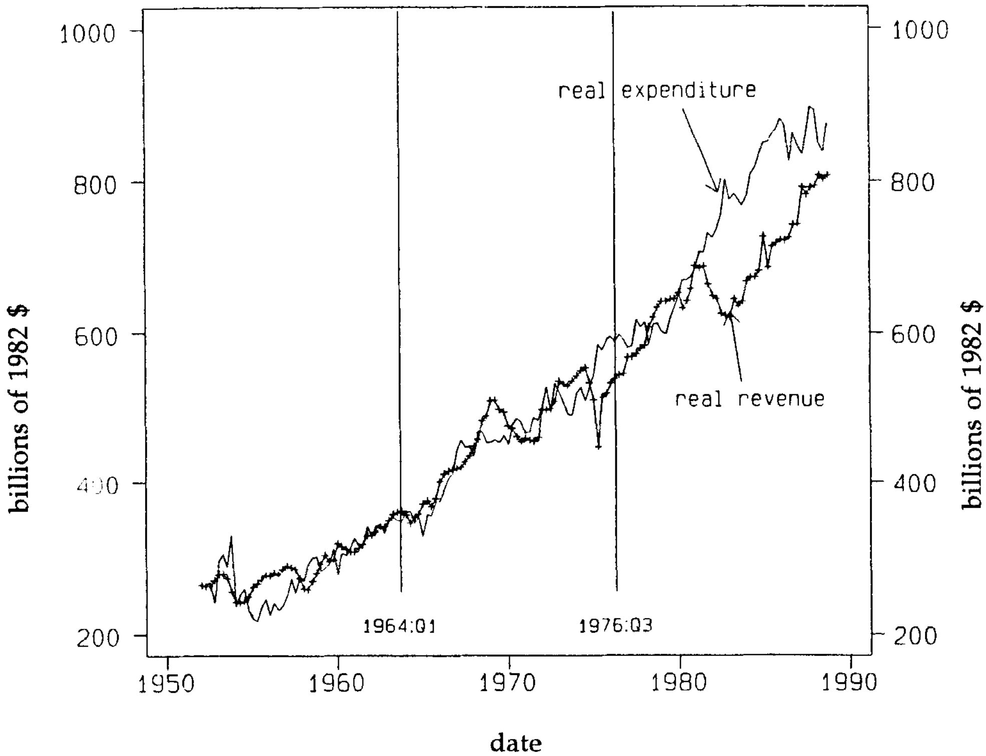
Two hypotheses form the basis of our empirical testing: (1) Are GG and R cointegrated, and (2) is $\hat{b} = 1$? However, there are several ways that GG and R can be measured. They can be measured in nominal terms, real terms, real per capita terms, and real per real GNP terms. The choice depends, in part, on the assumptions that were made in deriving the empirical implications. The assumption that the interest rate is stationary rules out nominal magnitudes since nominal interest rates are not stationary. In arguing that $\hat{b} < 1$ is inconsistent with a broad interpretation of the government's intertemporal budget constraint, we focused on real variables relative to GNP or relative to population. Thus, we present results for three cases: R and GG in real terms, deflated by real GNP, and deflated by population.

To get a feel for the data, Figures 1-3 plot (1) real revenue and spending, (2) real revenue and spending relative to real GNP, and (3) real revenue and spending relative to population. The vertical lines at 1964:I and 1976:III indicate the two places we split the sample. As can be seen in Figures 1 and 3, real spending and revenue, and real spending and revenue per capita, definitely are trending upwards over the sample period. Consequently, the Stock and Watson tests, which allow for a drift in the variables, are perhaps more appropriate for these two sets of variables.

The necessary first step in testing whether the budget deficit is "too large" is to determine if revenue and spending, in their various forms, are non-stationary. To this end, we used Dickey-Fuller [1979] unit root tests for the variables when measured in level form and when measured as first differences. The results are contained in Table I. For all our sample periods we "accept" the hypothesis of non-stationarity for real expenditures and revenues in levels and when deflated by population. The case of expenditures and revenues deflated by GNP gives a more complicated set of results. Using the entire sample period, we reject the hypothesis of non-stationarity for both expenditures and revenues (that is, expenditures and revenues are stationary); for the one hundred-observation period we reject non-stationarity

8. That is, Engle and Granger calculate their critical values using Monte Carlo simulations assuming the data generating process (DGP) is a random walk without drift. Stock and Watson give critical values for several DGP's, one of which is a random walk with drift. Several of our series are better modeled with a drift than without. Thus, we thought it important to use both sets of tests. As it turns out, using both tests makes little difference since the results from the Stock/Watson test generally agree with those from the Engle/Granger tests.

FIGURE 1
Real Revenue and Real Expenditure



for revenues but not expenditures (that is, expenditures are non-stationary and revenues are stationary); finally, for the fifty-observation period we fail to reject non-stationarity for both expenditures and revenues (that is, expenditures and revenues are non-stationary).[9] The general trend we observe is that these data are becoming "more non-stationary" in recent years.[10]

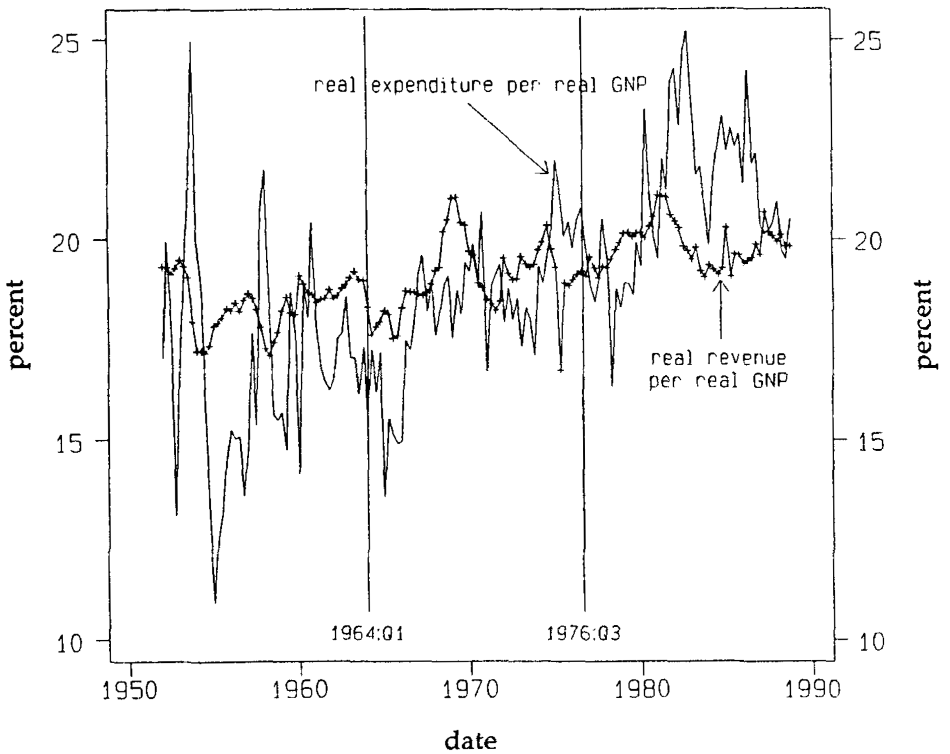
When measured as first differences all the variables are stationary. Therefore we conclude that cointegration tests for real spending and revenue and for real spending and revenue deflated by population are appropriate. However, these tests are appropriate only for the fifty-observation sample period when deflating with real GNP. For the entire period, the stationarity of both variables makes cointegration tests unnecessary and for the one hundred-observation period the difference in the order of integration makes the tests incorrect.

The results of testing whether government spending and revenue are cointegrated are contained in Table II. Three sample periods are included and four test statistics for cointegration are presented, along with the 5 percent critical values. In addition, Stock and Watson's

9. Notice that for the one hundred-observation case, the different orders of integration for spending and revenue allow us to immediately conclude that the government is violating its intertemporal budget constraint.

10. However, this observation is mitigated by the fact that Dickey-Fuller tests generally have low power.

FIGURE 2
Real Revenue and Real Expenditure Relative to Real GNP

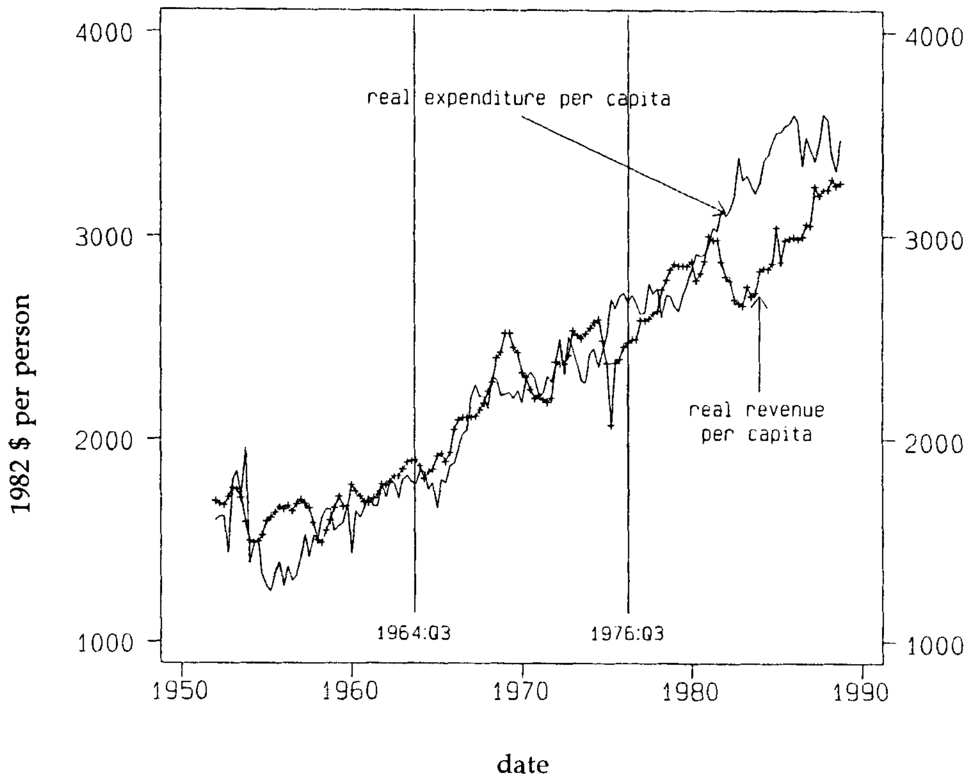


q_c test statistic for two common stochastic trends is reported. The null hypothesis for all tests is non-cointegration—which is equivalent to two stochastic trends in the Stock/Watson formulation; the alternative hypothesis is cointegration—which is equivalent to one common stochastic trend. A large test statistic, therefore, rejects non-cointegration and "accepts" cointegration. Critical values for the fifty-observation period are available only for the Durbin-Watson statistic and the augmented Dickey-Fuller statistic.

Although the results are not in total agreement, the bulk of the evidence suggests that if the entire sample—1950:II to 1988:IV—is used, government spending and revenue measured in real terms and per capita appear to be cointegrated. For the one hundred-observation period, 1964:I to 1988:IV, all of the tests save the Dickey-Fuller fail to reject lack of cointegration (more casually, most of the tests suggest the data are not cointegrated). Even more strikingly, however, *all* of the tests fail to reject lack of cointegration in the last subperiod—1976:III to 1988:IV. This indicates that the behavior of spending and revenue may well have changed recently, within, say, the last one hundred quarters. This result rationalizes much of the current worry about the budget deficit, since it indicates that the current processes generating spending and revenue—if continued—violate the government's intertemporal budget constraint.

Further evidence suggesting that the intertemporal budget constraint is vio-

FIGURE 3
Real Revenue and Real Expenditure Relative to Population



lated comes from the equilibrium regression results, reported at the bottom of Table II. As West [1988] points out, if two non-stationary variables are cointegrated, ordinary least squares estimates are normally distributed and consistent when the variables have a drift.[11] However, the estimated standard errors reported by computer programs are incorrect. West derives a correction that can be used to calculate a consistent estimate of the asymptotic variance-covariance matrix even if the disturbances are autocorrelated and conditionally heteroskedastic. Table II reports the West-corrected standard errors. It is clear that $\hat{b}$ is *always* significantly less than one, even when revenue and spending are measured relative to real GNP and population. This condition may be inconsistent with the government's ability to market its debt in the long run.

A related test is reported in Table I. The section of the table labeled "Budget deficit" tests whether the budget deficit is stationary. This test is related to a test of $b = 1$, because the budget deficit constrains the parameters of the cointegrating regression to be $a = 0$ and $b = 1$. The results are generally consistent. The budget deficit is

11. That is, if a non-stationary series X_t is cointegrated with another non-stationary series Y_t with $\Delta X_t \neq 0$ and $\Delta Y_t \neq 0$, then the estimated cointegration vector, say $\hat{b}$, is asymptotically normally distributed. In fact, the estimated cointegration factor, b, is "super consistent" in the sense that it converges to its true value at a faster rate—T—than normal least squares estimates, which converge at rate $T^{1/2}$.

TABLE I
Tests for a Random Walk

	Levels		First Difference	
	ADF stat	Model lags	ADF stat	Model lags
	Entire Sample			
Expenditures				
Real	.61	1	-16.14*	0
Real per GNP	-4.33*	0	-15.62*	0
Real per capita	-.12	1	-17.26*	0
Revenue				
Real	.53	0	-12.91*	0
Real per GNP	-3.13*	0	-13.82*	0
Real per capita	-.10	0	-12.31*	0
Budget deficit				
Real	-2.29	1	-14.81*	0
Real per GNP	-4.10*	5		
Real per capita	-2.47	1	-15.71*	0
	100 Observations			
Expenditures				
Real	-.31	0	-11.68*	0
Real per GNP	-2.36	1	-15.61*	0
Real per capita	-.84	0	-11.89*	0
Revenue				
Real	-.34	0	-11.09*	0
Real per GNP	-3.22*	0	-12.30*	0
Real per capita	-1.00	0	-10.80*	0
Budget deficit				
Real	-2.18	0	-11.07*	0
Real per GNP	-2.96*	1		
Real per capita	-2.34	0	-11.19*	0
	50 Observations			
Expenditures				
Real	-.74	0	-7.53*	0
Real per GNP	-2.80	0	-9.52*	0
Real per capita	-1.01	0	-7.63*	0
Revenue				
Real	-0.66	0	-8.68*	0
Real per GNP	-1.94	1	-10.22*	0
Real per capita	-1.16	0	-8.51*	0
Budget deficit				
Real	-1.48	0	-7.06*	0
Real per GNP	-2.96*	0		
Real per capita	-1.51	0	-7.10*	0

Note: "ADF" is the augmented Dickey-Fuller statistic for testing whether a variable has a unit root. "Model lags" indicates the form of the ADF statistic, in particular the number of lagged dependent variables. The null hypothesis is nonstationarity. Therefore, a "large" negative test statistic means reject a random walk and accept the alternative of stationarity; a "small" negative statistic (or positive) means "accept" the null of a random walk. Critical values are -2.89 for one hundred-observations and -2.93 for fifty-observations.

*means the variable is stationary (more formally, reject the null hypothesis of non-stationary).

TABLE II
Tests for Cointegration

Stat/Period	Real	Real per GNP	Real per capita
DW			
ENTIRE SAMPLE	0.40*	—	0.46*
100 OBSERVATIONS	0.34	—	0.35
50 OBSERVATIONS	0.33	0.69	0.32
DF/ADF			
ENTIRE SAMPLE	-3.77* [0]	—	-3.95* [0]
100 OBSERVATIONS	-3.02* [0]	—	-3.06* [0]
50 OBSERVATIONS	-2.05 [4]	-1.85 [1]	-1.61 [0]
RVAR/ARVAR			
ENTIRE SAMPLE	8.76 [1]	—	9.30 [1]
100 OBSERVATIONS	7.55 [0]	—	7.29 [0]
UVAR/AUVAR			
ENTIRE SAMPLE	18.97* [1]	—	20.26* [1]
100 OBSERVATIONS	15.11 [0]	—	15.43 [0]
q_c (Stock-Watson)			
ENTIRE SAMPLE	-32.72*	—	-37.98*
100 OBSERVATIONS	-18.16	—	-19.28
Equilibrium Regression: *Revenue = a + b(Spending) + u*			
Estimated *a*'s			
ENTIRE SAMPLE	88.34	—	$0.70*10^{-3}$
	(18.73)	—	$(0.10*10^{-3})$
100 OBSERVATIONS	129.95	—	$0.83*10^{-3}$
	(31.04)	—	$(0.17*10^{-3})$
50 OBSERVATIONS	198.77	0.19	$0.13*10^{-2}$
	(87.67)	(0.02)	$(0.05*10^{-2})$
Estimated *b*'s			
ENTIRE SAMPLE	0.78	—	0.68
	(0.04)	—	(0.04)
100 OBSERVATIONS	0.72	—	0.65
	(0.05)	—	(0.06)
50 OBSERVATIONS	0.63	0.01	0.51
	(0.12)	(0.11)	(0.16)

Note: DW denotes the cointegrating regression Durbin-Watson statistic; DF/ADF denotes the Dickey-Fuller statistic (zero lags) or the augmented Dickey-Fuller statistic (positive lags); RVAR/ARVAR denotes the restricted VAR statistic (zero lags) or the augmented restricted VAR statistic (positive lags); UVAR denotes the unrestricted VAR statistic (zero lags) or the augmented unrestricted VAR statistic (positive lags). The number in brackets next to the test statistics equals the number of lags used in the regression. The number in parentheses below the equilibrium regression coefficients is the West-standard error.

The null hypothesis is "non-cointegration," so a "large" test statistic rejects non-cointegration and "accepts" cointegration.

*means revenue and spending are cointegrated (more formally, * means we can reject that revenue and spending are not cointegrated).

TABLE II continued
Tests for Cointegration

Five percent critical values for the statistics are:

	system is 1st order 50 obs	100 obs	system is 4th order 50 obs	100 obs
DW	0.78	0.386	1.03	0.282
DF	3.67	3.37	3.29	3.05
ADF	3.67	3.17	3.29	3.17
RVAR		13.6		22.4
ARVAR		11.8		12.3
UVAR		18.6		40.3
AUVAR		17.9		22.0
Stock-Watson		-23.1		

stationary only when it is measured relative to real GNP.[12]

IV. AN ALTERNATIVE MEASURE OF REVENUE

As we detail in the data appendix, to this point we have measured revenues and expenditures using data from the National Income and Product Accounts combined with a series on the market value of the debt and a five-year interest rate. A referee has correctly pointed out that using data from different sources implies that the period-by-period budget constraint, equation (1), does not hold as an identity.[13] One possible way around this problem is to define government revenue as a residual. That is, equation (1) can be rearranged as

$$R_t = G_t + (1+i_t)\, B_{t-1} - B_t, \tag{9}$$

and a new measure of revenue, call it R^r, defined as equal to the right side of equation (9). Thus, by definition, the period-by-period budget constraint holds as an identity and the cointegration of R_t^r and GG_t can be tested.

We have a couple of problems with this procedure. First, using equation (9) to define revenue essentially discards all the information in the National Income and Product Account measured revenue series. We have no a priori reason to believe that the National Income and Product Accounts series for expenditure, which is an important component of (9), is measured any more accurately than the series for revenue. Second, and perhaps more importantly, using the definition of GG_t, (9) can be rewritten as

$$R_t^r = GG_t + (B_{t-1} - B_t) \tag{10}$$

so that a regression of R_t^r on GG_t is actually a regression of GG_t on $GG_t+(B_{t-1}-B_t)$.

12. There is one anomaly: the budget deficit relative to real GNP—for fifty-observations—is stationary, even though expenditures and revenue are non-stationary. This means that expenditure and revenue are cointegrated because there is a linear combination of expenditure and revenue ($a = 0$ and $b = 1$) that is stationary. However, in Table II we found that expenditure and revenue relative to GNP, for fifty-observations, were not cointegrated. We have no explanation for this result other than the power of the test is low.

13. Of course, keep in mind that many accounting "identities" do not actually hold in reality. For instance, it is well known that there is a large "statistical discrepancy" in the current account, and the worldwide sum of all trade balances is rather far from zero. To check the importance of our "statistical discrepancy," over the whole period the budget constraint was out of balance by an average of only $32 billion. This is a small amount, approximately 10 percent of the average year's measure of revenue.

TABLE III
Tests for a Random Walk
Alternative Definition of Revenue—R^r

	Levels		First Difference	
	ADF stat	Model lags	ADF stat	Model lags
		Entire Sample		
Revenue				
Real	.87	5	-6.78*	4
Real per GNP	-3.78*	1	-19.31*	0
Real per capita	-.07	5	-7.02*	4
		100 Observations		
Revenue				
Real	.52	3	-12.19*	2
Real per GNP	-1.93	4	-10.97*	2
Real per capita	-.34	3	-12.76*	2
		50 Observations		
Revenue				
Real	-0.21	4	-6.76*	3
Real per GNP	-1.61	2	-8.63*	1
Real per capita	-.54	4	-6.74*	3

Note: "ADF" is the augmented Dickey-Fuller statistic for testing whether a variable has a unit root. "Model lags" indicates the form of the ADF statistic, in particular the number of lagged dependent variables. The null hypothesis is nonstationarity. Therefore, a "large" negative test statistic means reject a random walk and accept the alternative of stationarity; a "small" negative statistic (or positive) means "accept" the null of a random walk. Critical values are -2.89 for one hundred-observations and -2.93 for fifty-observations.

* means the variable is stationary (more formally, reject the null hypothesis of non-stationary).

Clearly these two series will be cointegrated if and only if $B_{t-1} - B_t$ is stationary. Of course the stationarity of $B_{t-1} - B_t$ is another implication of the government's intertemporal budget constraint, which we previously tested using National Income and Product Account data in Table I. Testing the stationarity of $B_{t-1} - B_t$ directly is preferable to testing it indirectly through the cointegration of GG and R^r.[14]

14. For instance, suppose that GG_t is a drifting random walk with a large variance while $B_{t-1} - B_t$ is a nondrifting random walk with a small variance. In finite samples, GG_t (and hence R_t^r) will be dominated by the large variance and drift so that GG_t and R_t^r will appear (falsely) to be cointegrated. Testing the stationarity of $B_{t-1} - B_t$ directly can avoid such a problem. We would like to thank Anindya Banerjee for helping us with this point.

However, even with these drawbacks, we deemed it interesting to determine how this new definition of revenue would affect our conclusions. As before, we first checked the stationarity of our new revenue series. The results from Dickey-Fuller tests are reported in Table III. The only deviation from our previous results was found for the one hundred-observation, real per GNP series. Previously we found this series to be stationary. Based on the residual definition of revenue, it appears to be non-stationary.

Next we estimated cointegrating regressions between GG_t and R_t^r. Table IV presents the results. Basically, the results using the residual definition are more supportive of cointegration. Indeed, it seems that the only real doubt about the cointegration of spending and revenue occurs when looking at the one hundred-observation data series. It is important to note, however, that the estimated b coefficients are, as before, always significantly less than 1.0. Thus, there still exists the potential for the government to experience difficulty in the long-run when it comes to marketing its debt.

V. CONCLUSIONS

Is the government's budget deficit "too large?" Our results suggest the answer is "yes." Two pieces of evidence support this conclusion. First, using our preferred definition of revenue, spending and revenue appear not to be cointegrated for a shorter sample period starting in 1964. Second, even though spending and revenue may be cointegrated for our entire sample period from 1950:II to 1988:IV, the estimated cointegration factor, $\hat{b}$, is significantly less than 1.0. This is true for both the measure of revenue based on the National Income and Product Accounts and the residual measure of revenue. As a result, government spending is growing more rapidly than government revenue. For example, the point estimate for real spending and measured revenue relative to population, estimated using the entire sample, indicates that for each dollar increase in per capita spending, revenues rise by only 0.68 cents. This clearly is not sustainable.

DATA APPENDIX

The data for government spending and revenue are the National Income and Product Accounts figures, in billions of dollars. Government spending equals Federal government expenditures. To calculate the interest payments, we suitably deflated (by the price level, or the price level and real GNP, or population growth rate) the previous period's market value of the privately held gross Federal debt. The market value is from Cox and Lown [1989]. We multiplied this by the appropriately "deflated" nominal interest rate, discussed in footnote 2. For the nominal interest rate, between 1953:II to the end of our sample we used the Treasury constant maturity five-year note rate; between 1950:I to 1953:I this series was unavailable, so we used the series on the interest rate on three- to five-year Treasury notes. We spliced the data using a one-year overlap from April 1953 to March 1954. These series were used because their five-year maturity approximates the average maturity of the public debt over our sample period.

Government revenue equals Federal government receipts. Because the Federal Reserve is an independent agency, we chose not to include it as part of the Federal government. Thus, we included Federal Reserve earnings as part of the government's receipts, but did not directly include the change in the monetary base. (Federal Reserve earnings are transferred to the Treasury. This makes them simply another source of revenue for the Federal government, and they are included in the data on Federal government receipts.) Treating the Federal Reserve as part of the Federal government would require us to exclude Federal Reserve earnings (for in this case the funds would be a transfer payment from one branch of the government to another) but to include the change in the monetary base (as an additional source of revenue.) However, inclusion or exclusion of the Federal Reserve has a negligible effect on our results.

Real GNP is from the National Income and Products Accounts, in billions of 1982 dollars. Real government spending and revenue equals nominal spending and revenue divided by the GNP implicit price deflator (1982 = 1.0). Population is mid-period estimates, from the Bureau of Economic Analysis National Income and Product Accounts Table 2.1.

TABLE IV
Tests for Cointegration
Alternative Definition of Revenue—R^r

Stat/Period	Real	Real per GNP	Real per capita
DW			
ENTIRE SAMPLE	1.57*	—	1.68*
100 OBSERVATIONS	1.59*	2.06*	1.66*
50 OBSERVATIONS	1.43*	1.81*	1.45*
DF/ADF			
ENTIRE SAMPLE	-4.28* [5]	—	-4.37* [5]
100 OBSERVATIONS	-2.88 [3]	-4.81* [6]	-2.85 [3]
50 OBSERVATIONS	-5.27* [0]	-6.51* [0]	-5.40* [0]
RVAR/ARVAR			
ENTIRE SAMPLE	9.19 [3]	—	10.94 [3]
100 OBSERVATIONS	8.17 [3]	11.23 [3]	8.87 [3]
UVAR/AUVAR			
ENTIRE SAMPLE	22.04* [3]	—	23.23* [3]
100 OBSERVATIONS	17.11 [3]	34.98* [3]	18.39 [3]
q_c (Stock-Watson)			
ENTIRE SAMPLE	-138.6*	-146.3*	-141.2*
100 OBSERVATIONS	-84.7*	-95.8*	-88.3*
Equilibrium Regression: *Revenue = a + b(Spending) + u*			
Estimated *a*'s			
ENTIRE SAMPLE	22.74	—	$0.17*10^{-3}$
	(6.41)	—	$(0.03*10^{-3})$
100 OBSERVATIONS	37.24	0.05	$0.24*10^{-3}$
	(10.78)	(0.01)	$(0.06*10^{-3})$
50 OBSERVATIONS	51.00	0.07	$0.35*10^{-2}$
	(34.88)	(0.01)	$(0.19*10^{-2})$
Estimated b's			
ENTIRE SAMPLE	0.94	—	0.92
	(0.01)	—	(0.01)
100 OBSERVATIONS	0.92	0.73	0.90
	(0.02)	(0.03)	(0.02)
50 OBSERVATIONS	0.90	0.64	0.86
	(0.05)	(0.06)	(0.06)

Note: DW denotes the cointegrating regression Durbin-Watson statistic; DF/ADF denotes the Dickey-Fuller statistic (zero lags) or the augmented Dickey-Fuller statistic (positive lags); RVAR/ARVAR denotes the restricted VAR statistic (zero lags) or the augmented restricted VAR statistic (positive lags); UVAR denotes the unrestricted VAR statistic (zero lags) or the augmented unrestricted VAR statistic (positive lags). The number in brackets next to the test statistics equals the number of lags used in the regression. The number in parentheses below the equilibrium regression coefficients is the West-standard error.

The null hypothesis is "non-cointegration," so a "large" test statistic rejects non-cointegration and "accepts" cointegration.

* means revenue and spending are cointegrated (more formally, * means we can reject that revenue and spending are not cointegrated).

TABLE IV continued
Tests for Cointegration
Alternative Definition of Revenue—R^r

Five percent critical values for the statistics are:

	system is 1st order		system is 4th order	
	50 obs	100 obs	50 obs	100 obs
DW	0.78	0.386	1.03	0.282
DF	3.67	3.37	3.29	3.05
ADF	3.67	3.17	3.29	3.17
RVAR		13.6		22.4
ARVAR		11.8		12.3
UVAR		18.6		40.3
AUVAR		17.9		22.0
Stock-Watson		-23.1		

REFERENCES

Barro, Robert J. "On the Determination of the Public Debt." *Journal of Political Economy*, October 1979, 940–72.

_____. "U.S. Deficits Since World War I." *Scandinavian Journal of Economics*, 88, March 1986, 195–222.

_____. *Macroeconomics*. New York: John Wiley and Sons, 1987.

Buchanan, James and R. Wagner. *Democracy in Deficit*. New York: Academic Press, 1977.

Cox, Michael and Cara Lown. "The Capital Gains and Losses on U.S. Government Debt: 1942–1987." *Review of Economics and Statistics*, February 1989, 1–14.

Dickey, David A. and Wayne A. Fuller. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association*, June 1979, 427–31.

Engle, Robert F. and Clive W. J. Granger. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica*, March 1987, 251–76.

Engle, Robert F. and Byung Sam Yoo. "Forecasting and Testing in Co-Integrated Systems." *Journal of Econometrics*, 35(1), 1987, 143–59.

Fuller, Wayne A. *Introduction to Statistical Time Series*. New York: John Wiley and Sons, 1976.

Granger, Clive W. J. "Cointegrated Variables and Error-Correction Models." Unpublished manuscript, University of California, San Diego, 1983.

Granger, Clive W. J. and A. A. Weiss. "Time Series Analysis of Error-Correction Models," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, edited by Samuel Karlin, Takeshi Amemiya, and Leo A. Goodman. New York: Academic Press, 1983, 255–78.

Hamilton, James D. and Marjorie A. Flavin. "On the Limitations of Government Borrowing: A Framework for Empirical Testing." *American Economic Review*, September 1986, 808-16.

Haug, Alfred A. "Cointegration and Government Borrowing Constraints: Evidence for the U.S." Unpublished manuscript, Ohio State University, undated.

Kremers, J. M. Jereon. "U.S. Federal Indebtedness and the Conduct of Fiscal Policy. *Journal of Monetary Economics*, March 1989, 219–38.

_____. "Long-Run Limits on the U.S. Federal Debt." *Economic Letters*, 28(3), 1988, 259–62.

McCallum, Bennett T. "Are Bond-Financed Deficits Inflationary? A Ricardian Analysis." *Journal of Political Economy*, February 1984, 123–35.

Smith, Gregor W. and Stanley E. Zin. "Testing a Government's Present-Value Borrowing Constraint." Unpublished manuscript, Queen's University, May 1988.

Stock, James H. and Mark W. Watson. "Testing for Common Trends." *Journal of the American Statistical Association*, December 1988, 1097–107.

Trehan, Bharat and Carl E. Walsh. "Testing Intertemporal Budget Constraints: Theory and Applications to U.S. Federal Budget and Current Account Deficits." Unpublished manuscript, Federal Reserve Bank of San Francisco, June 1990.

_____. "Common Trends, the Government Budget Constraint, and Revenue Smoothing." *Journal of Economic Dynamics and Control*, June/September 1988, 425–46.

U.S. Department of Commerce, Bureau of Economic Analysis. National Income and Product Accounts. (Data from this series are published in various issues of the *Survey of Current Business* or available from computer databases.)

West, Kenneth D. "Asymptotic Normality, When Regressors have a Unit Root." *Econometrica*, November 1988, 1397–417.

Wilcox, David W. "The Sustainability of Government Deficits: Implications of the Present-Value Borrowing Constraint." Board of Governors of the Federal Reserve System Working Paper No. 77, September 1987.