MONASH

WP 1/98

ISSN 1032-3813
ISBN 0 7326 1038 9

MONASH UNIVERSITY

ANCORA · IMPARO

U.S. DEFICIT SUSTAINABILITY: A NEW APPROACH BASED ON MULTIPLE ENDOGENOUS BREAKS

Gael M. Martin

Working Paper 1/98
March 1998

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

U.S. DEFICIT SUSTAINABILITY: A NEW APPROACH BASED ON MULTIPLE ENDOGENOUS BREAKS

Gael M. Martin

Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3168, Australia. Phone: 61 3 99051189; Fax: 61 3 99055474; Email: Gael.Martin@BusEco.monash.edu.au

March, 1998

Abstract

Recent empirical work has questioned the consistency of U.S. fiscal policy with an intertemporal budget constraint. Empirical results have tended to indicate that the deficit process has undergone at least one structural shift during recent decades, with the deficit becoming either unsustainable or sustainable in only a weak sense in the post-shift period. In this paper, we re-examine sustainability using a new approach, based on a cointegration model with multiple endogenous breaks. A Bayesian methodology is applied, incorporating Markov chain Monte Carlo simulators. In contrast to previous analyses, we find evidence of a sustainable deficit process over the 1947 to 1992 period, despite the occurrence of breaks during the 1970's and 1980's.

1 Introduction

Over the past decade or so, empirical work has focussed on whether or not U.S. fiscal policy is consistent with a intertemporal budget constraint, i.e. is sustainable in the long run. A fiscal policy which implies that debt will explode over time at a faster rate than the growth rate of the economy is obviously not sustainable. Hence, the emphasis has been on assessing the statistical properties of debt. Analyses have concentrated on both the univariate properties of debt (see Hamilton and Flavin, 1986 and Wilcox, 1989) and the cointegration properties of revenues and expenditure (see Trehan and Walsh, 1988 and 1991, Hakkio and Rush, 1991, Haug, 1991, Tanner and Lui, 1994 and Quintos, 1995) with an important aspect of certain analyses being the incorporation of structural breaks. Overall, results have suggested that the U.S. deficit process has undergone a shift in recent times, with the deficit being either unsustainable or sustainable in only a weak sense in the post-shift period.

In this paper, we re-examine the sustainability issue using a new approach, based on a cointegration model for U.S. government revenue and interest inclusive expenditure, with allowance made for multiple endogenous breaks in both the intercept and slope parameters. The methodology produces simultaneous inferences about the presence of cointegration, the value of the cointegrating parameters and the size and timing of shifts in the relationship. Thus, inference about cointegration is not conditional on pre-imposed breakpoints, nor is inference about the breakpoints conditional on the presence of cointegration. Rather, a (potentially) cointegrating relationship, with possible deterministic shifts is estimated over the whole sample period and one set of results derived. The inferential approach is Bayesian, with results being based on Markov chain Monte Carlo (MCMC) posterior simulators.[1]

Previous methodological papers most closely related to the present paper include the following. Zivot and Phillips (1994) provide a Bayesian approach to unit root/structural break inference, while De la Croix and Lubrano (1996) conduct Bayesian cointegration/structural break inference on a restrictive form of model.[2] An MCMC sampling approach to single and multiple changepoint inference respectively in a variety of models, including stationary regression models, is used in Carlin, Gelfand and Smith (1992) and Stephens (1994). Chib (1997) applies an MCMC approach to univariate models in which multiple changes are modelled via Markov switching processes. In terms of classical work, Gregory and Hansen (1996a and b), extend the standard residual-based cointegration tests to accommodate a single shift in the cointegrating parameters at some unknown time point, whilst Bai and Perron (1995) consider estimation of multiple breaks in a standard linear regression model. Hendry and Clements (1998) consider the implications for forecasting of multiple breaks in cointegration models, as well as introducing the concept of 'co-breaking', in which deterministic structural breaks are removed by taking linear combinations of variables.

The paper is organized as follows. In Section 2 we provide a brief motivation of the use of a cointegration framework to analyze deficit sustainability. Section 3 outlines the model used to

[1] Computer programs written using the GAUSS software are available from the author on request.

[2] These authors allow for a single break in the slope coefficient only. As shown in Section 3, our model allows for *multiple* changes in *both* the intercept and slope parameters plus general autoregressive processes for both error terms in the model.

accommodate both cointegration and multiple parameter shifts. In Section 4, the joint posterior density function is defined, whilst Section 5 briefly describes the MCMC numerical procedures used to produce empirical results on sustainability, with details of the algorithms given in appendices. Section 6 reports the empirical results, based on quarterly U.S. data from 1947(2) to 1992(3). The results provide evidence of cointegration between revenue and interest inclusive expenditure, as well as evidence to suggest that the pre-shift coefficient on expenditure is unity. As such, strong sustainability of the deficit process, prior to any shift, is indicated. Shifts in both the intercept and slope of the deficit process are estimated as occurring in 1975(2), 1985(1) and 1987(1). The most substantial shifts occur in the intercept term, effecting a net increase of 50% in the level of the regression. The slope shifts, on the other hand, are small and almost offsetting. A stronger result in this direction occurs when the model is reparameterized to allow for slope shifts only. In that case, small slope shifts which effect almost no net change from an initial value close to unity, are estimated as occurring in 1975(1), 1984(4) and 1987(1), implying a maintenance of strong sustainability over the full period.

2 A Deficit Model

In this section, we motivate the use of a cointegration framework in testing for deficit sustainability. We follow closely the model development in Quintos (1995), and refer the reader to that paper for a more detailed discussion.

We begin with the government's one-period budget constraint, given by

$$\Delta B_t = GI_t - R_t \tag{1}$$

where B_t is the real market value of federal debt, $GI_t = G_t + r_t B_{t-1}$ is real government expenditure inclusive of interest payments in period t, R_t is real tax revenues in period t and r_t is the real interest rate, assumed to be stationary around a mean r. The quantity in (1) defines the real *interest inclusive deficit.* Further defining

$$E_t = G_t + (r_t - r)B_{t-1} \tag{2}$$

we can express debt as

$$B_t - (1+r)B_{t-1} = E_t - R_t \tag{3}$$

or

$$B_t = (\frac{1}{1+r})(R_{t+1} - E_{t+1}) + (\frac{1}{1+r})B_{t+1} \tag{4}$$

Solving for B_t in (4) via forward substitution yields

$$B_t = \sum_{j=0}^{\infty}(\frac{1}{1+r})^{j+1}(R_{t+j+1} - E_{t+j+1}) + \lim_{j\to\infty}(\frac{1}{1+r})^{j+1}B_{t+j+1}. \tag{5}$$

Defining $E_t(.)$ as an expectation conditional on information at time t, intertemporal budget balance, or deficit sustainability, holds if and only if

$$\lim_{j\to\infty} E_t(\frac{1}{1+r})^{j+1}B_{t+j+1} = 0, \tag{6}$$

since this implies that the current value of outstanding government debt is equal to the present value of future budget surpluses. In words, the deficit is sustainable if and only if the stock of debt held by the public is expected to grow no faster on average than the mean real rate of interest, which can be viewed as a proxy for the growth rate of the economy.[3]

The cointegration framework can be motivated in a number of different ways. (See Trehan and Walsh, 1988 for a particularly thorough motivation). Quintos (1995) proceeds by taking first differences in (5), yielding

$$\begin{aligned}\Delta B_t &= GI_t - R_t \\ &= \sum_{j=0}^{\infty}(\frac{1}{1+r})^{j+1}(\Delta R_{t+j+1} - \Delta E_{t+j+1}) + \lim_{j\to\infty}(\frac{1}{1+r})^{j+1}\Delta B_{t+j+1}.\end{aligned}$$

She then associates sustainability with the condition

$$\lim_{j\to\infty} E_t(\frac{1}{1+r})^{j+1}\Delta B_{t+j+1} = 0, \tag{7}$$

which, in turn, imposes conditions on the statistical properties of the interest inclusive deficit.[4] Quintos shows that (7) holds when ΔB_t is either stationary *or* nonstationary, with the rate at which the zero limit is approached being slower in the latter case.[5] She refers to these two cases as defining 'strong' and 'weak' sustainability respectively. With GI_t and R_t viewed as $I(1)$ processes, the cointegration properties of these two processes can be used to shed light on (7). Defining a regression model of the form

$$R_t = \alpha + \beta GI_t + u_t, \tag{8}$$

Quintos demonstrates that:

1. The deficit is *'strongly' sustainable* if and only if the $I(1)$ processes R_t and GI_t are cointegrated and $\beta = 1$.

2. The deficit is *'weakly' sustainable* if $0 < \beta < 1$, irrespective of the cointegration status of R_t and GI_t.

3. The deficit is *unsustainable* if $\beta \leq 0$, again irrespective of the cointegration status of R_t and GI_t.[6]

Strong sustainability, in turn, means that the budget constraint holds and, at the same time, the undiscounted debt process, B_t, is $I(1)$. Weak sustainability means that the constraint holds,

[3]Expectations are assumed to be formed rationally by the government's creditors and, therefore, to be consistent with the assumed generating process for B_t.

[4]A more detailed derivation of the link between the present value constraint and the behaviour of ΔB_t can be found in Trehan and Walsh (1988).

[5]ΔB_t is allowed to be mildly explosive, as long as the explosive roots behave as $\exp(C/T)$, with $C > 0$.

[6]Note that $\beta > 1$ is not consistent with a deficit, since revenues are growing at a faster rate than (interest inclusive) expenditures.

but with B_t exploding at a rate which is less than the growth rate in the economy, as approximated by the mean real interest rate. As Quintos points out, although the latter situation is consistent with sustainability, it may well have implications for the ability of the government to market its debt and is, therefore, the less desirable scenario. An unsustainable deficit is one which implies that B_t is exploding at a rate equal to or in excess of the growth rate in the economy, such that the intertemporal budget constraint in (6) is violated.

All previous cointegration analyses of the U.S. deficit process have been based on a model of the form of (8), with both full-sample and sub-sample results being produced. Using U.S. annual data for the period 1890 to 1986, Trehan and Walsh (1988) find cointegration over the full period and a value for β close to unity, suggesting strong sustainability for the full period. Haug (1991) effectively draws the same conclusion, based on quarterly U.S. data for 1960(1) to 1987(4).[7] Using quarterly U.S. data extending from 1950(2) to 1988(4), Hakkio and Rush (1991) determine sub-periods over which estimation is performed via an arbitrary imposition of breakpoints in 1964 and 1976 respectively. They conclude that the deficit process has effectively undergone a shift, since the post-1964 and post-1976 results suggest a lack of cointegration, whereas the full-sample results support cointegration.[8]. Based on quarterly U.S. data from 1947(2) to 1992(3), Quintos (1995) finds evidence of cointegration over the full period, but a value for β between zero and one, suggesting weak sustainability. Applying the parameter stability test of Hansen (1992) to the (assumed) cointegrating relation, she then imposes the estimated breakpoint in order to define relevant sub-samples. Since the Hansen test suggests two possible shiftpoints, at 1975(2) and 1980(4) respectively, Quintos performs sub-period analysis based on each. The evidence supports strong sustainability in both pre-break periods, but only weak sustainability in both post-break periods, including a lack of cointegration. Applying rank constancy tests, Quintos finds some evidence that the cointegration status of the relationship also changes in the early 1980's. Based on quarterly U.S. data for 1950(1) to 1989(4), Tanner and Lui (1994) impose a dummy variable for a level shift in their cointegration model at both a pre-defined point, 1982(1), and at a point determined via the univariate pre-test procedure of Christiano (1992), namely 1981(4). With this level dummy incorporated, they find cointegration and acceptance of the null hypothesis that β is equal to unity.

As is evident from this outline of previous work, analysis of the sustainability issue has both lead to conflicting results and been based on a conglomeration of cointegration and structural break methods. The approach adopted in this paper, being an integrated cointegration/structural break methodology, is ideally suited to tackle this type of empirical problem. In particular, by allowing for multiple shifts in both level and slope parameters, we are able to pinpoint whether or not shifts in the deficit process have occurred, the nature of the shifts and whether or not the shifts impinge upon conclusions regarding sustainability.

[7]Haug (1991) actually tests for cointegration between B_{t-1} and $(R_t - G_t)$. A finding of cointegration, with cointegrating vector $[r\ , -1]$ is equivalent to a finding of stationarity in ΔB_t. Athough he finds evidence of cointegration, he does not formally test the value of the coefficient on B_{t-1}.

[8]They conclude that the deficit in recent times is *unsustainable,* since they do not invoke Quintos' definition of weak sustainability.

3 The Cointegration/Structural Break Model

To accommodate m shifts in the parameters of the cointegration model in (8), we define the following model:

$$\begin{aligned} R_t &= \alpha_1 + (\alpha_2 - \alpha_1)i(r_1)_t + \cdots + (\alpha_m - \alpha_{m-1})i(r_m)_t + \\ &\quad \beta_1 GI_t + (\beta_2 - \beta_1)GI_t(r_1)_t \cdots + (\beta_m - \beta_{m-1})GI_t(r_m)_t + u_t, \\ GI_t &= GI_{t-1} + v_t, \end{aligned} \tag{9}$$

where

$$i(r_k)_t = \begin{cases} 0 & for \quad t \leq r_k \\ 1 & for \quad t > r_k \end{cases}, \quad k = 1, 2, \ldots, m$$

$$GI(r_k)_t = \begin{cases} 0 & for \quad t \leq r_k \\ GI_t & for \quad t > r_k \end{cases}, \quad k = 1, 2, \ldots, m.$$

The only formal restriction on the number of breaks, m, is that the supports of the associated r_k do not overlap and that each support contains at least two sample points. In practice, however, sensible inferences would also entail that the value of m is not too large relative to the size of the sample. In particular, the specification of a large number of parameter shifts would strain the definition of cointegration to an unreasonable extent.

The error processes are specified as having the following autoregressive representations:

$$\begin{aligned} \Phi(L)u_t &= \varepsilon_t \\ \Psi(L)v_t &= \eta_t, \end{aligned} \tag{10}$$

where $\Phi(L)$ and $\Psi(L)$ are defined as finite order polynomials in the lag operator L,

$$\begin{aligned} \Phi(L) &= \sum_{i=1}^{p} \phi_i L^i \\ \Psi(L) &= \sum_{i=1}^{q} \psi_i L^i, \end{aligned} \tag{11}$$

with the restriction that $\phi_0 = \psi_0 = 1$, and $e_t = (\varepsilon_t, \eta_t)'$ is a disturbance vector, assumed to be bivariate Normal of the form

$$[e]_{t=1}^{n} \sim N(0, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}), \tag{12}$$

with Σ a (2×2) symmetric positive definite matrix.

The roots of $\Psi(L)$ in (11) are assumed to all lie outside the unit circle, such that η_t is a stationary process and x_t an $I(1)$ process as a consequence. In examining the cointegration properties of R_t and GI_t we follow Phillips (1991a, b and 1993) and Zivot and Phillips (1994) in reparameterizing $\Phi(L)u_t$ as:

$$\begin{aligned} \Phi(L)u_t &= u_t - \phi_1 u_{t-1} - \cdots - \phi_p u_{t-p} \\ &= u_t - \rho_1 u_{t-1} - \sum_{i=2}^{p} \rho_i (u_{t-(i-1)} - u_{t-i}) \end{aligned}$$

where:

$$\begin{aligned}\rho_1 &= \phi_1 + \phi_2 + \cdots + \phi_p \\ &\text{and} \\ \rho_k &= -\sum_{i=k}^{p} \phi_i.\end{aligned}$$

The parameter ρ_1 is the appropriate parameter for capturing the long-run behavior of u_t, as it is the key parameter in determining the behavior of the spectrum of u_t at the origin. Inference concerning the presence of cointegration between the two $I(1)$ processes, R_t and GI_t, can be based on the marginal posterior density function for ρ_1, with a test of cointegration being based upon a comparison of $\Pr(\rho_1 < 1)$ and $\Pr(\rho_1 \geq 1)$.

The marginal posterior mass functions for the breakpoints, r_k, provide the basis for estimating the timing of the parameter shifts. The marginal densities for the $(\alpha_{k+1} - \alpha_k)$ and the $(\beta_{k+1} - \beta_k)$ provide the basis for both point and interval estimates of the magnitude of the intercept and slope shifts respectively. Such inferences complement those about the r_k. Conditional on (9) being viewed as a cointegrating relation, α_1 and β_1 represent the parameters of the pre-shift cointegrating relationship, with point and interval estimates of them being produced via their respective marginal posteriors. Estimates of these parameters and of the magnitudes of the shifts can be used, in turn, to produce estimates of the intercepts and slope values associated with each regime implied by the breaks. The alternative approach of parameterizing the model so that the intercept and slope in each regime were directly estimated was not taken, since the presence of shifts per se is as of as much the focus of the paper as anything else. Hence, the marginal posteriors for the shift magnitudes, in conjunction with the marginal breakpoint posteriors, is the more natural output of the analysis.[9]

The sharp null of $H_0 : \beta_1 = 1$ can be tested against the alternative $H_1 : \beta_1 \neq 1$ by determining whether or not the value of unity is contained in an appropriate Highest Posterior Density (HPD) interval for β_1.[10] Conditional on acceptance of $\beta_1 = 1$, evidence of small and offsetting values for the $(\beta_{k+1} - \beta_k)$ can be viewed as evidence in favor of $\beta_1 = 1$ for the full period. Note that this part of the analysis is necessarily informal. The sharp nulls of $(\beta_{k+1} - \beta_k) = 0$ for all $k = 1, 2, \ldots, m$ cannot be formally tested, since, as explained in the following section, these subspaces (of measure zero) are effectively viewed as being deleted from the support of the joint posterior. To test H_0 against the one-sided alternative $H_1 : 0 < \beta_1 < 1$, we apply a posterior odds test, with the odds ratio estimated from the simulation output in the manner of Chib (1995).

4 Posterior Distribution Specification

The model defined by (9) to (12) has two potential identification problems associated with it. First, for all $k = 1, 2, \ldots, m$, when $\alpha_{k+1} - \alpha_k = \beta_{k+1} - \beta_k = 0$, r_k is not identified, since $GI(r_k)_t$ drops

[9] We note that the method is applicable to a multiple regression relationship, with more than one $I(1)$ regressor and with all parameters undergoing change simultaneously. The partial stuctural change scenario of Bai and Perron (1995) can also be accommodated.

[10] An HPD interval is an interval with the specified probability coverage, whose inner density ordinates are not exceeded by any density ordinates outside the interval.

out of the model. In a classical context, in which $\alpha_{k+1} - \alpha_k = \beta_{k+1} - \beta_k = 0$ is a null hypothesis to be tested, inference involves nonstandard distribution theory, as a consequence of the parameter r_k being unidentified under the null.[11] In a Bayesian framework, in which a joint posterior distribution for the full parameter set is to be defined over some appropriate support, the manifestation of the lack of identification of r_k is a posterior which is "flat" in the direction of r_k in this subspace of the support. Although not improper, due to the fact that the change point r_k is defined over a *finite* interval of the sample period, such a distribution is not, in principle, a sensible basis for inference. However, the fact that the subspace has measure zero means that the exact lack of identification can formally be eliminated by deleting the subspace from the support of the density, with no impact being felt on the final results.[12]

The second identification problem concerns the pre-shift cointegrating parameters, with α_1 and β_1 being unidentified when there is an exact unit root in the error term u_t. The nature of this problem is detailed in Martin (1996) and is very similar to that discussed in Martin and Martin (1997). As in that work, and in the spirit of Kleibergen and van Dijk (1994) and Chao and Phillips (1996), we offset the identification problem via an information matrix-based prior.[13] This prior serves to offset not only the exact lack of identification of α_1 and β_1, but also the distortion to inferences resulting from the near lack of identification in the region around $\rho_1 = 1$. The near identification problem manifests itself in terms of asymptoting behavior in the marginal density of ρ_1 in the region around the singularity at $\rho_1 = 1$. This feature of the ρ_1 marginal, in turn, distorts inferences by strongly favoring a lack of identification irrespective of the true cointegration status of the model. Full derivation of the prior is provided in Martin (1996). As is made clear therein, due to the nature of the identification problem, the information matrix-based prior is defined for $\beta = (\alpha_1, \alpha_2 - \alpha_1, \ldots \alpha_{m+1} - \alpha_m, \beta_1, \beta_2 - \beta_1, \ldots \beta_{m+1} - \beta_m)'$ conditional on $\{r, \rho, \psi\}$, where $r = (r_1, r_2, \ldots r_m)'$, $\rho = (\rho_1, \rho_2, \ldots, \rho_p)'$ and $\psi = (\psi_1, \psi_2, \ldots, \psi_q)'$.

In total then, assuming a priori independence between Σ and the remaining parameters, a priori independence between r, ρ and ψ respectively, specifying a standard Jeffreys prior for Σ and making the simplifying assumption of uniform priors for each of r, ρ and ψ, we have a joint prior for the full parameter set $\Omega = \{\Sigma, \beta, r, \rho, \psi\}$ of the form

$$\begin{aligned} p(\Omega) &= p(\Sigma) \times p(\beta|r, \rho, \psi) \times p(r) \times p(\rho) \times p(\psi) \\ &\propto |\Sigma|^{-3/2} \left|X'X\right|^{1/2}. \end{aligned} \tag{13}$$

The matrix X is defined as

$$X = [i^*, i(r_1)^*, \ldots i(r_m)^*, GI^*, GI(r_1)^*, \ldots GI(r_m)^*, GI^{**}],$$

where the columns of X are, in turn, defined as follows. GI^* denotes the n-dimensional vector for $\Phi(L)GI_t = GI_t^* = GI_t - \rho' GI_{t-1}^{(p)}$, where

$$GI_{t-1}^{(p)} = (GI_{t-1}, \Delta GI_{t-1}, \ldots, \Delta GI_{t-p+1})',$$

[11] For a recent discussion of this of issue see Andrews (1993) and Hansen (1996).

[12] In the context of the sampling method, this simply involves discarding any simulated values for $(\alpha_{k+1} - \alpha_k)$ and $(\beta_{k+1} - \beta_k)$ which are simultaneously equal to zero.

[13] De la Croix and Lubrano (1996) also encounter this identification problem, tackling it in a way which amounts to a Bayesian version of the Engle/Granger two-step procedure.

whilst $GI(r_k)^*, k = 1, 2, \dots m$, denotes the n-dimensional vector for $\Phi(L)GI(r_k)_t = GI(r_k)^*_t = GI(r_k)_t - \rho' GI(r)^{(p)}_{t-1}$, where

$$GI(r_k)^{(p)}_{t-1} = (GI(r_k)_{t-1}, \Delta GI(r_k)_{t-1}, \dots, \Delta GI(r_k)_{t-p+1})'.$$

Defining $i_t = 1$ for all t, i^* denotes the n-dimensional vector for $\Phi(L)i_t = i^*_t = i_t - \rho' i^{(p)}_{t-1}$, where

$$\begin{aligned} i^{(p)}_{t-1} &= (i_{t-1}, i_{t-1} - i_{t-2}, \dots, i_{t-p+1} - i_{t-p})' \\ &= (i_{t-1}, 0, \dots, 0)'. \end{aligned}$$

It follows that $i^*_t = 1 - \rho_1$ for all t. Finally, $i(r_k)^*, k = 1, 2, \dots m$, denotes the n-dimensional vector for $\Phi(L)i(r_k)_t = i(r_k)^*_t = i(r_k)_t - \rho' i(r)^{(p)}_{t-1}$, where

$$i(r_k)^{(p)}_{t-1} = (i(r_k)_{t-1}, \Delta i(r_k)_{t-1}, \dots, \Delta i(r_k)_{t-p+1})',$$

whilst GI^{**} denotes the n-dimensional vector for $\Psi(L)\Delta GI_t = GI^{**}_t = \Delta GI_t - \psi' \Delta GI^{(q)}_{t-1}$ where

$$\Delta GI^{(q)}_{t-1} = (\Delta GI_{t-1}, \Delta GI_{t-2}, \dots, \Delta GI_{t-q})'.$$

Given the assumption of Normality, the likelihood function for Ω is

$$L(\Omega|R, GI) \propto |\Sigma|^{-n/2} \exp\{(\frac{-1}{2})\text{tr}(\Sigma^{-1}S), \tag{14}$$

where R and GI respectively denote the $n-$ dimensional vectors of observations on R_t and GI_t and $S = \sum_{t=1}^{n} e_t e_t'$. Applying the joint prior as defined in (13), the joint posterior density[14] is thus given by

$$p(\Omega|R, GI) \propto |\Sigma|^{-(n+3)/2} \exp\{(\frac{-1}{2})\text{tr}(\Sigma^{-1}S)\} \times |X'X|^{1/2}, \tag{15}$$

with Ω defined on the complement of F^*, where $F^* = F \cap \{\Omega;\ \beta_{k+1} - \beta_k = 0 \cap \alpha_{k+1} - \alpha_k = 0; k = 1, 2, \dots m\}$, $F = \mathbb{S}^{\text{PDS}} \times \mathbb{R}^{2(m+1)} \times \Pi_{k=1}^{m} \mathbb{Z}_j^{[n_{k-1}, n_k)} \times \mathbb{R}^p \times \mathbb{R}^q$ and $\mathbb{Z}_k^{[n_{k-1}, n_k)}$ denotes the set of integers in the semi-closed interval $[n_{k-1}, n_k)$, $k = 1, 2, \dots m$. The $[n_{k-1}, n_k)$ define non-overlapping intervals which span the interval $[1, n-1]$ and which contain at least two data points.[15]

5 Numerical Procedures

5.1 Discussion

The aim of this section is to outline the numerical procedures used to produce inferences on deficit sustainability, as based on the joint posterior density function in (15). Empirical results are based on estimates of the pertinent marginal posterior densities associated with (15) as well as on an estimate of the posterior odds ratio for testing $H_0 : \beta_1 = 1$ against $H_1 : 0 < \beta_1 < 1$. Estimation of the odds ratio involves estimating, under both H_0 and H_1, ordinates of (15), ordinates of the likelihood function in (14) and ordinates of the joint prior in (13). Section 5.2 briefly outlines the numerical algorithm used to estimate the marginals, whilst Section 5.3 outlines the way in which the odds ratio is estimated.

[14] The use of the term *density* is loose, since the r_k are discrete.

[15] In practice, the $[n_{k-1}, n_k)$ intervals do not have to span the full interval $[1, n-1]$, but can be defined over the parts of the sample in which the search for breaks is to be concentrated.

5.2 Estimation of the Marginal Posteriors

The MCMC algorithm used to estimate the marginal posteriors is a hybrid of the Gibbs and Metropolis Hastings (MH) sampling schemes. (See Tierney, 1994, Chib and Greenburg, 1996 and Geweke, 1997, for details of these algorithms). Briefly, Gibbs sampling involves an iterative generation of random drawings from all of the conditional (posterior) distributions associated with the joint posterior. So long as certain conditions are satisfied by both the joint posterior and the induced conditionals, these drawings represent a realization of a Markov chain with equilibrium distribution equal to the joint posterior. Once convergence to the equilibrium distribution has occurred, continued iterative simulation from the conditionals ultimately produces a sample of T $(\Sigma^{(i)}, \beta^{(i)}, r^{(i)}, \rho^{(i)}, \psi^{(i)})$ values from the joint posterior. A Gibbs-based scheme thus potentially involves simulation from the conditional posteriors for all parameter blocks included in Ω. There are two problems associated with the use of these full conditionals, both of which concern the conditional of β. First, the β conditional is flat in the direction of α_1 in the subspace $\rho_1 = 1$ and nearly flat in the surrounding subspace. Even if $\rho_1 = 1$ were to be deleted from the support, the flatness of the conditional in the surrounding subspace is enough to cause the Gibbs-based scheme to become 'stuck' in the region. Formally, the scheme is nearly reducible (see Tierney, 1994). Secondly, the univariate conditionals for the intercept *shifts,* as implied by the joint conditional for β, although well-defined, have been found to be very dispersed. Simulation from highly dispersed conditionals has been found to impact badly on the accuracy with which the marginals are estimated, and thus deemed to be undesirable.

The solution adopted here is to integrate out Σ and β analytically from (15) using the known integrating constants for the inverted Wishart and multivariate Student t densities respectively, and to produce simulations from the posterior distribution for (r, ρ, ψ). Marginal densities for the elements β (and Σ, if required) can then be backed out in the manner to be described.

The form of the joint posterior distribution of (r, ρ, ψ) is detailed in Appendix A. The natural 'blocking' of the Gibbs sampler in this case is in terms of r_1, $r_2, \dots r_m$, ρ and ψ, which defines, in turn, the conditional distributions on which the scheme is to be based. These conditionals are all nonstandard and those of ρ and ψ (potentially) multivariate. We use an MH scheme to simulate from the conditional distributions of ρ and ψ respectively at the ith iteration of the Gibbs sampler. Since the conditional mass functions of each of the breakpoints r_j, $j = 1, 2, \dots, m$ are one-dimensional they are amenable to simulation via the inverse cumulative distribution function technique (termed Griddy Gibbs by Tanner, 1993). Details of the MH scheme and of the convergence properties of the hybrid Gibbs/MH scheme are given in Appendix A.

Once the simulated values $(r^{(i)}, \rho^{(i)}, \psi^{(i)})$, $i = 1, 2, \dots, T$ have been produced via the hybrid algorithm, the marginal density of ρ_1 and marginal mass functions of the r_j can be estimated as finite mixtures. For ρ_1, for example, this estimate is given by

$$p(\widehat{\rho_1 | R, GI}) = (1/T) \sum_{i=1}^{T} p(\rho_1 | r^{(i)}, \rho_{\overline{1}}^{(i)}, \psi^{(i)}, R, GI), \tag{16}$$

where $\rho_{\overline{1}}$ denotes the elements in ρ other than ρ_1. Since $p(\rho_1 | r, \rho_{\overline{1}}, \psi, R, GI)$ is known only up to a constant, (16) requires one-dimensional numerical normalization of the T component densities.

The MCMC sample can also be used to produce, say, the estimate

$$p(\widehat{\beta_1|R, GI}) = (1/T)\sum_{i=1}^{T} p(\beta_1|r^{(i)}, \rho^{(i)}, \psi^{(i)}, R, GI), \tag{17}$$

with $p(\beta_1|r^{(i)}, \rho^{(i)}, \psi^{(i)}, R, GI)$ denoting the univariate Student t density for β_1 implied by the multivariate conditional for β, conditional on values for r, ρ and ψ from the *ith* iteration of the MCMC algorithm. Since $p(\beta_1|, r, \rho, \psi, R, GI)$ is of a standard form, (17) requires no numerical normalization of the component densities.

5.3 Estimation of the Posterior Odds Ratio

5.3.1 Discussion

In the order to test the null hypothesis of $H_0 : \beta_1 = 1$ against $H_1 : 0 < \beta_1 < 1$, we need to construct

$$\begin{aligned} \frac{P(H_0|R, GI)}{P(H_1|R, GI)} &= \frac{P(H_0)\int_\Omega [L(\Omega|R, GI, H_0)p(\Omega|H_0)]\, d\Omega}{P(H_1)\int_\Omega [L(\Omega|R, GI, H_1)p(\Omega|H_1)]\, d\Omega} \\ &= \frac{P(H_0)}{P(H_1)} \times \frac{p(R, GI|H_0)}{p(R, GI|H_1)} \\ &= \text{Prior Odds} \times \text{Bayes Factor} \end{aligned}$$

An analytical treatment of the Bayes factor is not feasible in the present situation. Various numerical alternatives could be used (see Geweke, 1997, for an outline). The method of Chib (1995), translated to the present problem, is based on recognizing that for H_i, $i = 0, 1$, the marginal likelihood under H_i is given by

$$p(R, GI|H_i) = \frac{L(\Omega|R, GI, H_i)p(\Omega|H_i)}{p(\Omega|R, GI, H_i)},$$

for any point in Ω. However, whilst the likelihood function and prior can be evaluated at any point in Ω, the posterior density cannot be, since its nonstandard nature renders computation of the necessary integrating constant infeasible. Chib's suggestion is to *estimate* the joint posterior ordinate $p(\Omega|R, GI, H_i)$, for some given $\Omega = \Omega^*$ by exploiting both knowledge of the conditional and marginal posteriors into which the joint posterior can be decomposed, and the output of the MCMC scheme which, by construction, produces simulated values from the joint posterior. With the likelihood and prior evaluated exactly at this same Ω^*, an estimate of the marginal likelihood under H_i, $i = 1, 2$, is given by

$$p(\widehat{R, GI}|H_i) = \frac{L(\Omega^*|R, GI, H_i)p(\Omega^*|H_i)}{p(\Omega^*|\widehat{R, GI}, H_i)}$$

and an estimate of the posterior odds ratio thus attainable.

The simulation output used to estimate the joint posterior ordinate $p(\Omega\ |R, GI, H_i)$ must be produced with H_0 and H_1 imposed respectively. With these hypotheses both involving restrictions on β_1, the analytical treatment of β as multivariate Student t, used in integrating out β and basing a simulation scheme on the posterior for (r, ρ, ψ) is not legitimate. In the following section, we describe a simple reparameterization of the model which enables the appropriate simulation output to be produced. We describe the use of this output in estimating the odds ratio, relegating details of the simulation scheme itself to Appendix B.

5.3.2 Estimation of a Posterior Odds Ratio for $H_0 : \beta_1 = 1$ versus $H_1 : 0 < \beta_1 < 1$

What is required here is a simulation scheme which involves the estimation of the endogenous shift points, but which avoids simulation from the problematic intercept/intercept shift conditionals. The approach taken is to demean the data to accommodate a non-zero intercept and allow for slope shifts only. This invokes a reduced parameter set $\underline{\Omega} = (\Sigma, \underline{\beta}, r, \rho, \psi)$, where $\underline{\beta} = (\beta_1, \beta_2 - \beta_1, \ldots, \beta_{m+1} - \beta_m)'$. The limitation of this approach is that a constant intercept is imposed for the full sample period.

Based on this parameterization, we require the evaluation of

$$\begin{aligned} L(\underline{\Omega}|R, GI) &= (2\pi)^{-n} |\Sigma|^{-n/2} \exp\{\frac{-1}{2}\mathrm{tr}(\Sigma^{-1}S)\} \\ p(\underline{\Omega}) &= k_1 \times |\Sigma|^{-3/2} \left|X'X\right|^{1/2} \text{ and} \\ p(\underline{\Omega}|R, GI) &= k_2 \times |\Sigma|^{-(n+3)/2} \times \\ &\quad \exp\{(\frac{-1}{2})\mathrm{tr}(\Sigma^{-1}S)\} \left|X'X\right|^{1/2}. \end{aligned}$$

at some $\underline{\Omega}^*$, where the matrix X is now defined as

$$X = [GI^*, GI(r_1)^*, \ldots GI(r_m)^*, GI^{**}].$$

Calculation of k_1 would involve both an arbitrary truncation of the parameter space and multidimensional integration. Our solution to this problem is to define the same prior (and, hence, the same implicit truncation) under both H_0 and H_1, so that the arbitrary integrating constants cancel.

Simulation of $\underline{\Omega}$ occurs under both H_0 and H_1, with $\beta_1 = 1$ and $0 < \beta_1 < 1$ imposed respectively. Details of the hybrid Gibbs/MH scheme used are given in Appendix B. Once the simulated values have been obtained, the posterior odds ratio is estimated as follows. First, both the likelihood and (unnormalized) prior function are evaluated at some $\Omega^* = (\Sigma^*, \underline{\beta}^*, r^*, \rho^*, \psi^*)$. Typically, Ω^* represents some high density point, like a mean or a mode. Secondly, the joint posterior is decomposed appropriately, with each component density evaluated at Ω^*. For example, in the case of the deficit example, in which $r = (r_1, r_2, r_3)'$ is specified and a preliminary model selection procedure chooses $\rho = \rho_1$ and $\psi = \psi_1$, we have

$$\begin{aligned} p(\underline{\Omega}^*|R, GI) &= p(\Sigma^*, \underline{\beta}^*, \psi_1^*, \rho_1^*, r^*|R, GI) \\ &= p(\Sigma^*|\underline{\beta}^*, \psi_1^*, \rho_1^*, r^*, R, GI) \\ &\quad \times p((\beta_2^* - \beta_1^*)|(\beta_3^* - \beta_2^*), (\beta_4^* - \beta_3^*), \beta_1^*, \psi_1^*, \rho_1^*, r^*, R, GI) \\ &\quad \times p((\beta_3^* - \beta_2^*)|(\beta_4^* - \beta_3^*), \beta_1^*, \psi_1^*, \rho_1^*, r^*, R, GI) \\ &\quad \times p((\beta_4^* - \beta_3^*)|\beta_1^*, \psi_1^*, \rho_1^*, r^*, R, GI) \end{aligned}$$

$$\times p(\beta_1^*|\psi_1^*, \rho_1^*, r^*, R, GI)$$
$$\times p(\psi_1^*|\rho_1^*, r^*, R, GI) \times p(\rho_1^*|r^*, R, GI)$$
$$\times p(r_1^*|r_2^*, r_3^*, R, GI) \times p(r_2^*|r_3^*, R, GI) \times p(r_3^*|R, GI).$$

The conditional for Σ is known in its entirety (as Inverted Wishart) and thus able to be evaluated automatically at Σ^*. The marginal $p(r_3|R, GI)$ can be estimated as a finite mixture from the output of the simulation scheme and evaluated at r_3^*. The intervening conditionals must be handled differently. The one-dimensional conditional $p((\beta_2 - \beta_1)|.)$ can be normalized numerically and evaluated at $(\beta_2^* - \beta_1^*)$. The remaining conditionals can be estimated as finite mixtures from additional applications of the same simulation scheme, but with the appropriate sets of parameters fixed. For instance, the conditional $p((\beta_3 - \beta_2)|.)$ is estimated as

$$\begin{aligned} & p(\widehat{(\beta_3 - \beta_2)}|.) \\ = \quad & (1/T)\sum_{i=1}^{T} p((\beta_3 - \beta_2)|\Sigma^{(i)}, (\beta_2 - \beta_1)^{(i)}, (\beta_4 - \beta_3)^*, \beta_1^*, \psi_1^*, \rho_1^*, r^*, R, GI). \end{aligned}$$

In this application of the scheme, only Σ, $\beta_2 - \beta_1$ and $\beta_3 - \beta_2$ are allowed to vary, with the remaining parameters fixed at the pre-specified values. Each additional application of the simulation scheme is a reduced version of the original scheme and hence faster, as well as involving no additional computer code. The conditional for β_1 is degenerate with value one under H_0 and restricted to the support $0 < \beta_1 < 1$ under H_1.

6 Empirical Results on Sustainability

6.1 Data and Model Specification

We produce results based on the same data set as used in Quintos (1995).[16]. The data set comprises quarterly U.S. data on real revenues and real government expenditure inclusive of interest paid on debt, over the period 1947(2) to 1992(3). Details of the construction of the data set can be found in the Quintos paper, as can results pertaining to the univariate time series properties of the series in question. Figure 1 plots the data over the 1947(2) to 1992(3) period. It is clear that the two series are essentially moving together, but with some sort of shifts occurring in the relationship from the mid 1970's onwards. We initially allowed for two shifts, with the sub-periods over which the shifts were "searched for" ranging from 1970(4) to 1983(1) and 1983(3) to 1989(3) respectively. However, the marginal mass function for the second break point ascribed large probability mass to two time periods, namely 1985(1) and 1987(1). As a consequence we re-estimated the model with three breaks, over the respective sub-periods of 1973(3) to 1980(4), 1981(2) to 1985(3)) and 1986(1) to 1990(4).

A lag length of one for both $\Phi(L)$ and $\Psi(L)$ is selected from lag lengths ranging up to four, via a preliminary model selection procedure based on the posterior information criterion (PIC) criterion of Phillips and Ploberger (1994). Given the imposition of a unit root in GI_t, the lag length for $\Psi(L)$ is chosen from alternative AR specifications for the series ΔGI_t. In choosing the lag length for the

[16]The data is publicly availaible at the Journal of Business and Economic Statistics web site.

autoregression in u_t, the PIC procedure is applied to the residuals resulting from OLS estimation of R_t on GI_t. Since the first 6 observations are used for lag specification in the preliminary model selection, final results are based on data from 1948(4) to 1992(3).[17]

6.2 Empirical Results

To highlight the impact of allowing for endogenous breaks, we begin by presenting in Table 1 results for the full sample period with *no* breaks estimated, in which case the cointegration model defined in (9) reduces to

$$\begin{aligned} R_t &= \alpha_1 + \beta_1 GI_t + u_t, \\ GI_t &= GI_{t-1} + v_t, \end{aligned} \tag{18}$$

with the errors still as defined in (11) and (12). Details of the MCMC scheme applied to this model are given in Appendix C. Once again, since direct estimation of an intercept is problematic, the data is demeaned and an intercept indirectly accommodated. For comparison, we report the β_1 estimates and t ratios based on the Fully Modified Least Squares (FMOLS) procedure of Phillips and Hansen (1990). As with the Bayesian method, the classical estimate is based on demeaned data. We include a PIC posterior odds test of the sharp null of $\rho_1 = 1$, in order to complement the inference based on the marginal posterior of ρ_1. As noted in Phillips (1993), with a flat prior specified on ρ_1, there is a tendency for the probability of cointegration to be slightly larger than if a true noninformative prior on ρ_1 were specified. The PIC test, on the other hand, as explained in Phillips and Ploberger (1994), represents an objective Bayesian assessment of the unit root hypothesis, with no inherent tendency to conclude either in favor of or against the presence of a unit root. If prior odds of unity are specified, the criterion leads to rejection of the unit root hypothesis if PIC < 1.[18] Both the PIC and ADF results are based on OLS residuals.

The results in Table 1 provide overwhelming evidence of cointegration over the full sample period, with the marginal density of ρ_1 ascribing 100% probability to $\rho_1 < 1$ and both the PIC and ADF tests leading to rejection of a unit root in the residuals. However, both the Bayesian and classical point estimates of β_1 are clearly less than one, with the classical t test leading to rejection of $H_0 : \beta_1 = 1$. The 95% HPD interval for β_1 excludes unity. The posterior odds ratio for testing H_0 against the alternative $H_0 : 0 < \beta_1 < 1$ is less than one, implying rejection of the null hypothesis.[19]

[17]Although not reported here, we have conducted elsewhere extensive Monte Carlo experiments to guage the repeated sampling performance of certain aspects of the methodology applied in this paper (see Martin, 1996). In order to base conclusions on a reasonable number of replications, each experiment is based on a model with only a single slope shift accommodated and the lag polynomials specified as: $\Phi(L) = 1 - \rho_1 L$ and $\Psi(L) = 1$. This parameterization allows the marginal densities to be estimated via a combination of analytical and low-dimensional numerical integration, which, in turn, is much faster (for the lower dimensional parameterization) than the MCMC method. This parameterization also enables comparison with the classical cointegration test of Gregory and Hanson (1996a), which allows for a single shift only. Classical estimates of the slope parameter and the slope shift are obtained by applying Fully Modified Least Squares (FMOLS) to the model with a breakpoint, as implied by the preliminary Gregory and Hansen test, imposed. In summary, the Bayesian inferences display accurate finite sample behaviour, superior, in the main, to the classical alternatives. Although not directly applicable to the multiple break scenario, these results give some cause for confidence in the sampling behaviour of the extended methodology.

[18]The PIC criterion, as described in Phillips and Ploberger (1994), leads to rejection of the unit root null if PIC > 1. We report the reciprocal of the statistic.

[19]The posterior odds are based on prior odds of unity. Comparison of the odds ratio with one implies a symmetric loss function, with equal prior expected loss ascribed to both Type 1 and Type 2 errors. The lower bound of $\beta_1 = 0$

These results suggest that the deficit is only weakly sustainable over the full period, a conclusion which tallies with that of Quintos (1995), who also finds evidence of full sample cointegration and a slope coefficient less than unity.[20]

Table 1: No Breaks Estimated

Cointegration Tests		Results for β_1	
$Pr(\rho_1 < 1)$	1.00	β_1 Mode	0.79
PIC	0.02	95% HPD Interval	(0.65,0.92)
ADF	-3.39	PO $(H_0 : \beta_1 = 1)$	0.11
		FMOLS	0.78
		t Test $(H_0 : \beta_1 = 1)$	-11.71

The results in Table 1 are, however, clearly deficient, given the evidence of structural shifts in Figure 1. Table 2 reports the results of estimation based on Model (9). Graphs of the corresponding marginal density/mass functions are presented in Figures 2 to 4. As mentioned earlier, initial exploratory analysis indicated the presence of three shift points. Thus (9) is estimated with $m = 3$ imposed, over the ranges described above. Once again, the evidence in favor of cointegration over the full period, with parameter shifts now accommodated, is conclusive, although less overwhelming than in the no-breaks case. The marginal mass functions for the r_j pinpoint breaks in 1975(2), 1985(1) and 1987(1). The two latter breaks are ascribed more than 70% probability, whilst the mass function for r_1 ascribes a total probability mass of almost 100% to 1975(2) and 1975(1) together. The modal point estimates indicate that the most substantial shifts occur in the intercept term. With the implied intercept after the third break being 0.91, the results suggest a net increase of 50% in the level of the regression over the sample period. However, interpretation of the point estimates of the intercept terms must be tempered by the fact that the variability in the mass functions is very high, with associated HPD intervals being very wide.

Estimation of the slope and slope shift parameters is much more accurate. The point estimate of the pre-shift β_1 is 0.95, with the 95% HPD interval easily covering one. This latter fact, in combination with HPD intervals for the slope shifts which cover zero, can be viewed as evidence in favor of a slope of unity over the full sample period. Consideration of the modal point estimates alone suggest that the net effect of the breaks is to effect a slight downward shift in the slope, with

has no qualitative impact on the results, with few simulated values of β_1 being less than or equal to zero. This implies that there is no point in constructing an odds ratio for $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 > 0$.

[20]Note that Quintos also includes a polynomial time trend in her regression, the order of which is not reported. This, in addition to her use of the full 1947(2) to 1992(3) sample may explain why her estimate of β_1 differs from ours.

the implied slope coefficient after the third break being 0.90. The main point to glean here is that the most substantial shifts would appear to be confined to the *level* of the deficit series and not the slope. With these level shifts accommodated, the evidence for a slope shift and, hence, a change in the nature of the sustainability, is not strong. This result tallies with that of Tanner and Lui (1994).

Table 2: Intercept and Slope Shifts

$Pr(\rho_1 < 1) = 0.96$

Structural Break Inference

Parameter	Mode	95% HPD Interval
α_1	0.61	(-0.46,5.42)
$(\alpha_2 - \alpha_1)$	-0.13	(-3.23,3.65)
$(\alpha_3 - \alpha_2)$	-1.06	(-8.66,5.54)
$(\alpha_4 - \alpha_3)$	1.50	(-4.87,8.21)
β_1	0.95	(-0.07,1.27)
$(\beta_2 - \beta_1)$	-0.03	(-0.54,0.43)
$(\beta_3 - \beta_2)$	0.08	(-0.58,0.86)
$(\beta_4 - \beta_3)$	-0.09	(-0.75,0.50)
r_1	1975(2)	
r_2	1985(1)	
r_3	1987(1)	

The marginal density/mass functions resulting from estimation of the model which accommodates three slope shifts only, are presented in Figures 5 and 6. These functions, plus the associated results in Table 3, other than the posterior odds ratio, involve unrestricted estimation of β_1. As is clear, with a constant level imposed, the probability ascribed to a slope break occurring in 1975(1) is almost 100%. The 70% probability mass ascribed to a break in 1985(1) in Figure 2c is, in Figure 5c, shared between 1984(4) and 1985(1), with the former time period having the higher probability. A slightly lower probability mass of 57% is now assigned to a breakpoint in 1987(1). As the results in Table 3 indicate, the HPD intervals are consistent with zero change in a pre-shift slope parameter of unity. The point estimates of the slope shifts, like those reported in Table 2, are small in magnitude. In this case they are almost totally offsetting, implying that the pre-shift slope of unity is essentially maintained throughout the full period.

The posterior odds ratio, in which β_1 is restricted to the interval $0 < \beta_1 < 1$ under the alternative, gives substantial support to the null hypothesis that $\beta_1 = 1$. Given the evidence in favor of cointegration and the evidence of small and offsetting slope shifts, this odds test result effectively represents acceptance of the null of strong sustainability against the alternative of weak sustainability, for the full sample period.[21]

Table 3: Slope Shifts Only Estimated

$Pr(\rho_1 < 1) = 0.92$

Structural Break Inference

Parameter	Mode	95% HPD Interval
β_1	0.93	(-0.06,1.30)
$(\beta_2 - \beta_1)$	-0.10	(-0.15,-0.05)
$(\beta_3 - \beta_2)$	0.04	(-0.07,0.07)
$(\beta_4 - \beta_3)$	0.05	(-0.04,0.08)
r_1	1975(1)	
r_2	1984(4)	
r_3	1987(1)	
PO $(H_0 : \beta_1 = 1)$	6.94	

7 Conclusions

This paper has presented new results relating to the sustainability of the U.S. deficit. The results indicate that the relationship over the 1947 to 1992 sample period is a cointegrating one, with three shifts having occurred, in 1975, 1984/5 and 1987 respectively. The precise timing of the breakpoints depends on the parameterization of the model (i.e. whether or not intercept shifts are accommodated) However, apart from one-quarter variations, the shift point estimation is invariant to the parameterization. In all cases, the probability mass assigned to either one particular period, or two adjacent periods, is substantial, usually well in excess of 50%. The most substantial shifts appear to have occurred in the level of the regression, although the level parameters are not accurately estimated by the model. On the other hand, the slope estimation is precise, as measured by marginal posterior variability, and indicates that the initial, pre-break situation of strong sustainability is maintained, despite small deviations, throughout the full sample period. The results are all jointly produced and, as such, are not subject to the usual pre-test biases. Further, they are based on the full sample and therefore, not affected by the degrees of freedom problems encountered in previous sub-sample analyses.

[21]For interest, the posterior odds ratio for the test of $H_0 : \beta_1 = 1$ against $H_1 : \beta_1 \neq 1$ was also calculated. The computed value of 8.85 gives strong support to the null.

Appendix A: MCMC scheme for Section 5.2

The joint posterior density for (r, ρ, ψ) is given by

$$p(r, \rho, \psi | R, GI) \propto (GI^{**'}GI^{**})^{-1/2} D^{-(n-2(m+1))/2}, \tag{19}$$

where D is defined as

$$D = E - F'C^{-1}F.$$

The matrix C is symmetric of dimension $(2(m+1) \times 2(m+1))$, with upper diagonal elements given by:

$$\begin{aligned}
C_{11} &= (i^{*'}i^{*})(GI^{**'}GI^{**}) - (i^{*'}GI^{**})^2, \\
C_{12} &= (i^{*'}i(r_1)^{*})(GI^{**'}GI^{**}) - (i^{*'}GI^{**})(i(r_1)^{*'}GI^{**}) \\
&\vdots \\
C_{1m+1} &= (i^{*'}i(r_m)^{*})(GI^{**'}GI^{**}) - (i^{*'}GI^{**})(i(r_m)^{*'}GI^{**}) \\
C_{22} &= (i(r_1)^{*'}i(r_1)^{*})(GI^{**'}GI^{**}) - (i(r_1)^{*'}GI^{**})^2 \\
C_{23} &= (i(r_1)^{*'}i(r_2)^{*})(GI^{**'}GI^{**}) - (i(r_1)^{*'}GI^{**})(i(r_2)^{*'}GI^{**}) \\
&\vdots \\
C_{2m+1} &= (i(r_1)^{*'}i(r_m)^{*})(GI^{**'}GI^{**}) - (i(r_1)^{*'}GI^{**})(i(r_m)^{*'}GI^{**}) \\
&\vdots \\
C_{(m+2)(m+2)} &= (GI^{*'}GI^{*})(GI^{**'}GI^{**}) - (GI^{*'}GI^{**})^2 \\
&\vdots \\
C_{2(m+1)2(m+1)} &= (GI(r_m)^{*'}GI(r_m)^{*})(GI^{**'}GI^{**}) - (GI(r_m)^{*'}GI^{**})^2.
\end{aligned}$$

The scalar E is defined by

$$E = (R^{*'}R^{*})(GI^{**'}GI^{**}) - (R^{*'}GI^{**})^2$$

and the $2(m+1)$ vector F defined by the elements

$$\begin{aligned}
F_1 &= (R^{*}i^{*})(GI^{**'}GI^{**}) - (R^{*'}GI^{**})(i^{*'}GI^{**}), \\
F_2 &= (R^{*'}i(r_1)^{*})(GI^{**'}GI^{**}) - (R^{*'}GI^{**})(i(r_1)^{*'}GI^{**}) \\
&\vdots \\
F_{m+1} &= (R^{*'}i(r_m)^{*})(GI^{**'}GI^{**}) - (R^{*'}GI^{**})(i(r_m)^{*'}GI^{**}) \\
F_{m+2} &= (R^{*'}GI^{*})(GI^{**'}GI^{**}) - (R^{*'}GI^{**})(GI^{*'}GI^{**}) \\
&\vdots \\
F_{2(m+1)} &= (R^{*'}GI(r_m)^{*})(GI^{**'}GI^{**}) - (R^{*'}GI^{**})(GI(r_m)^{*'}GI^{**}),
\end{aligned}$$

where R^* denotes the n-dimensional vector for $\Phi(L)R_t = R_t^* = R_t - \rho' R_{t-1}^{(p)}$, with

$$R_{t-1}^{(p)} = (R_{t-1}, \Delta R_{t-1}, \ldots, \Delta R_{t-p+1})'.$$

The conditional densities on which the Gibbs-based scheme are to be based are $p(|r, \psi, R, GI)$, $p(r|\rho, \psi, R, GI)$, $p(r_1|r_2, \ldots, r_m, \rho, \psi, R, GI), \ldots, p(r_m|r_1, \ldots, r_{m-1}, \rho, \psi, R, GI)$, and $p(\psi|r, \rho, R, GI)$, all of which are defined as the density in (19) viewed as a function of the argument of interest. All densities are non-standard in form. The MH schemes for ρ and ψ respectively are both based upon Normal candidate distributions. In the case of ρ for example, the mean and covariance matrix for the candidate distribution are, respectively, the modal value of $\ln(p(\rho|.))$, $\hat{\rho}$ say, and $(-\partial^2 \ln p(\rho|.)/\partial\rho\partial\rho')|_{\hat{\rho}}$, where $p(\rho|.)$ denotes the conditional posterior function for ρ given $\psi^{(i-1)}$ and $r^{(i-1)}$. The corresponding parameterization occurs in the case of ψ.

Both the joint distribution for (r, ρ, ψ) and its induced conditionals satisfy the sufficient conditions for the outer Gibbs chain to be simply ergodic. (See Tierney, 1994, Chib and Greenburg, 1996 and Geweke, 1997 for details of these conditions). Given the use of Normal candidate distributions in the MH subchains, the latter are based on bounded weight functions and, therefore, uniformly ergodic for the relevant conditionals. The average acceptance rate of the subchains is approximately 90%. With T denoting the final number of sample values on which inferences are produced we simulate a total of $M + (10 \times T)$ iterations for the outer Gibbs chain. We use $T = 3000$ and $M = 100$, as well as performing 20 iterations of the Metropolis subchain before taking a value as a realization from the relevant conditional density. This simulation scheme is conservative, in that the estimated marginals differ only negligibly from those based on a smaller number of iterations.

Appendix B: MCMC scheme for Section 5.3

The vectors GI and R now denote the vectors of demeaned observations on GI_t and R_t respectively and X is as defined in Section 5.3.2. Under H_1 the full set of conditional densities induced by the joint density for $\underline{\Omega} = (\Sigma, \underline{\beta}, r, \rho, \psi)$ are as follows:

$$p(\Sigma|\underline{\beta}, r, \rho, \psi, R, GI) \propto |\Sigma|^{-(n+3)/2} \exp\{\frac{-1}{2} tr(\Sigma^{-1}S)\}, \quad (20)$$

$$p(\underline{\beta}|\Sigma, r, \rho, \psi, R, GI) \propto \exp\{\frac{-1}{2\sigma_{11.2}}(\underline{\beta} - \tilde{\underline{\beta}})A_2(\underline{\beta} - \tilde{\underline{\beta}})\};\ 0 < \beta_1 < 1, \quad (21)$$

$$p(r_1|r_2, \ldots, r_m, \underline{\beta}, \Sigma, \rho, \psi, R, GI) \propto \exp\{\frac{-1}{2} tr(\Sigma^{-1}S)\}, \quad (22)$$

$$\vdots$$

$$p(r_m|r_1, \ldots, r_{m-1}, \underline{\beta}, \Sigma, \rho, \psi, R, GI) \propto \exp\{\frac{-1}{2} tr(\Sigma^{-1}S)\}, \quad (23)$$

$$p(\rho|\underline{\beta}, \Sigma, r, \psi, R, GI) \propto \exp\{\frac{-1}{2\sigma_{11.2}}(\rho - \tilde{\rho})G_2(\rho - \tilde{\rho})\} \times |X'X|^{1/2}, \quad (24)$$

$$\begin{aligned} p(\psi|\underline{\beta}, \Sigma, r, \rho, R, GI) \quad &\propto \quad \exp\{\frac{-1}{2\sigma^2_{22.1}}(\psi - \tilde{\psi})B_2(\psi - \tilde{\psi})\} \\ &\quad \times (GI^{**'}GI^{**})^{-1/2} |X'X|^{1/2} \end{aligned} \quad (25)$$

with $\underline{\tilde{\beta}} = A_2^{-1}A_1$, $\tilde{\psi} = B_2^{-1}B_1$ and $\tilde{\rho} = G_2^{-1}G_1$ and the additional notation used defined as follows:

$$\begin{aligned} A_1 &= X^{*\prime}[R - (\sigma_{12}/\sigma_{22})GI^{**}], \\ A_2 &= X^{*\prime}X^*, \\ X^* &= [GI^*, GI(r_1)^*, \ldots, GI(r_m)^*] \\ B_1 &= V_q'[(\Delta GI - (\sigma_{12}/\sigma_{11})(R - X^*\beta)], \\ B_2 &= V_q'V_q, \\ G_1 &= U_p'[(R - \widetilde{X}\beta) - (\sigma_{12}/\sigma_{22})GI^{**}], \\ \widetilde{X} &= [GI, GI(r_1), \ldots GI(r_m)] \\ G_2 &= U_p'U_p, \\ \sigma_{11.2} &= \sigma_{11} - \sigma_{12}^2/\sigma_{22} \text{ and} \\ \sigma_{22.1} &= \sigma_{22} - \sigma_{12}^2/\sigma_{11}. \end{aligned}$$

Further notation used is: $U_p = (u^{(1)}, u^{(1)} - u^{(2)}, \ldots, u^{(p-1)} - u^{(p)})$ and $V_q = (\Delta GI^{(1)}, \Delta GI^{(2)}, \ldots, \Delta GI^{(q)})$, with $u^{(j)} = (R_{1-j} - \widetilde{X}_{1-j}\beta, R_{2-j} - \widetilde{X}_{2-j}\beta, \ldots, R_{n-j} - \widetilde{X}_{n-j}\beta)'$, $j = 1, 2, \ldots, p$, $\Delta GI^{(j)}$ denoting the n dimensional vector for ΔGI_{t-j}, $j = 1, 2, \ldots, p$, and ΔGI being the vector for ΔGI_t.

The conditional for Σ is Inverted Wishart and thus able to be simulated from directly. The densities of ψ and ρ are proportional to the product of a Normal kernel and the prior factor, viewed as a function of ψ and ρ respectively. We simulate from these conditionals indirectly via an MH algorithm, with the Normal kernel used as the basis for the candidate density. In this case, the ratio of the target to the candidate density collapses to the (unnormalized) prior factor. Since this factor is bounded over any finite range for ρ (or ψ) the condition for uniform ergodicity for the candidate subchains is satisfied.. The average acceptance rate for the MH subchains is approximately 75% for ρ and approximately 98% for ψ. The conditionals for r_1, $r_2, \ldots, r_m$ are, once again, amenable to the Griddy Gibbs approach. The conditional density for $\underline{\beta}$ is $(m + 1)-$dimensional Normal truncated from above at $\beta_1 = 1$ and below at $\beta_1 = 0$. This means that $\underline{\beta}$ is simulated from the $(m+1)-$dimensional Normal defined as in (21), but with values for β_1 beyond the range $0 < \beta_1 < 1$ discarded.

Under H_0, the conditionals and simulation scheme remain the same except as regards β. With $\beta_1 = 1$ imposed, $\beta^d = (\beta_2 - \beta_1, \ldots, \beta_{m+1} - \beta_m)'$ is simulated from the $m-$dimensional Normal defined by

$$p(\beta^d|\beta_1 = 1, \Sigma, r, \rho, \psi, R, GI) \propto \exp\{\frac{-1}{2\sigma_{11.2}}(\beta^d - \tilde{\beta}^d)A_2(\beta^d - \tilde{\beta}^d)\},$$

where $\tilde{\beta}^d = A_2^{-1}A_1$ and notation is redefined as

$$\begin{aligned} A_1 &= X^{*\prime}[R^* - (\sigma_{12}/\sigma_{22})GI^{**} - GI^*], \\ A_2 &= X^{*\prime}X^* \text{ and} \\ X^* &= [GI(r_1)^*, \ldots GI(r_m)^*]. \end{aligned}$$

Appendix C: MCMC scheme for model (18)

The same type of MCMC scheme as outlined in Appendix A could be applied to the joint posterior for (ρ, ψ), obtained by analytically integrating out Σ and $\beta = (\alpha_1, \beta_1)'$ from the joint for $(\Sigma, \beta, \rho, \psi)$, defined, in turn, as (15) with $X = (GI^*, GI^{**})$ and an appropriate redefinition of the remaining terminology. However, it is computationally faster to base a scheme on the full set of conditionals defined by the parameter set $(\Sigma, \beta, \rho, \psi)$. With the data demeaned in order to cater for an intercept term, the four conditionals for β_1, Σ, ρ and ψ respectively are simulated from, with an MH algorithm, once again, being used to cater for the non-standard nature of the conditionals for ρ and ψ. It is the nature of this MH scheme, in comparison with the MH scheme described in Appendix A which produces the computational gains. If required, a marginal for α_1 can be backed out of the simulation output in the same fashion as described in (17) in the text for β_1.

With GI and R again denoting demeaned observation vectors, the conditionals induced by the joint posterior for $(\Sigma, \beta_1, \rho, \psi)$ are

$$p(\Sigma|\beta_1, \rho, \psi, R, GI) \propto |\Sigma|^{-(n+3)/2} \exp\{\frac{-1}{2} tr(\Sigma^{-1}S)\}, \tag{26}$$

$$p(\beta_1|\Sigma, \rho, \psi, R, GI) \propto \exp\{\frac{-1}{2\sigma_{11.2}(A_2'A_2)^{-1}}(\beta_1 - \tilde{\beta}_1)^2\}, \tag{27}$$

$$p(\rho|\beta_1, \Sigma, \psi, R, GI) \propto \exp\{\frac{-1}{2\sigma_{11.2}}(\rho - \tilde{\rho})G_2(\rho - \tilde{\rho})\} \times |X'X|^{1/2}, \tag{28}$$

$$\begin{aligned} p(\psi|\beta_1, \Sigma, \rho, R, GI) \quad \propto \quad & \exp\{\frac{-1}{2\sigma_{22.1}^2}(\psi - \tilde{\psi})B_2(\psi - \tilde{\psi})\} \\ & \times |X'X|^{1/2}, \end{aligned} \tag{29}$$

with $\tilde{\beta}_1 = A_1/A_2$, $\tilde{\psi} = B_2^{-1}B_1$ and $\tilde{\rho} = G_2^{-1}G_1$, X redefined as $X = (GI^*, GI^{**})$ and the additional notation used defined as follows:

$$\begin{aligned} A_1 &= GI^{*\prime}[R - (\sigma_{12}/\sigma_{22})GI^{**}], \\ A_2 &= GI^{*\prime}GI^*, \\ B_1 &= V_q'[(\Delta GI - (\sigma_{12}/\sigma_{11})(R - GI^*\beta_1)], \\ B_2 &= V_q'V_q, \\ G_1 &= U_p'[(R - GI\beta_1) - (\sigma_{12}/\sigma_{22})GI^{**}], \\ G_2 &= U_p'U_p, \\ \sigma_{11.2} &= \sigma_{11} - \sigma_{12}^2/\sigma_{22} \text{ and} \\ \sigma_{22.1} &= \sigma_{22} - \sigma_{12}^2/\sigma_{11}. \end{aligned}$$

Further, U_p and V_q are as defined in Appendix B, but with $u^{(j)} = (R_{1-j} - GI_{1-j}\beta_1, R_{2-j} - GI_{2-j}\beta_1, \ldots, R_{n-j} - GI_{n-j}\beta_1)'$, $j = 1, 2, \ldots, p$. The same type of simulation scheme as described in Appendix B is appropriate here, apart from the deletion of all references to the r_j and associated $\beta_{k+1} - \beta_k$, $k = 1, 2, \ldots, m$.

References

[1] Andrews, D.W.K. (1993), 'Tests for parameter instability and structural change with unknown change point', *Econometrica,* 61, No. 4, 821-856.

[2] Bai, J. and P. Perron (1995), 'Estimating and testing linear models with multiple structural changes', *Manuscript.*

[3] Carlin, B.P., Gelfand, A.E. and A.F.M. Smith (1992), 'Hierarchical Bayesian analysis of change-point problems', *Applied Statistics,* 41, 389-405.

[4] Chao, J.C. and P.C.B. Phillips (1996), 'Bayesian posterior distributions in limited information analysis of the simultaneous equations model', *Mimeo, Cowles Foundation, Yale University.*

[5] Chib S. (1995), 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association,* 90, No. 432, 1313.

[6] Chib, S. (1997), 'Estimation and comparison of multiple change point models', forthcoming *Journal of Econometrics.*

[7] Chib, S. and E. Greenburg (1996), 'Markov chain Monte Carlo simulation methods in Econometrics', *Econometric Theory,* 12, 409-431.

[8] Christiano, L.J. (1992), 'Searching for a break in U.S. GNP', *Journal of Business and Economic Statistics,* 10, No. 3, 237-250.

[9] Geweke, J. (1997), 'Using simulation methods for Bayesian econometric models: inference, development and communication', Paper prepared for *the Australasian Meeting of the Econometric Society, Melbourne.*

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

[11] Gregory, A.W. and B.E. Hansen (1996b), 'Tests for cointegration in models with regime and trend shifts', *Oxford Bulletin of Economics and Statistics,* 58, No. 3, 555-561.

[12] Hakkio, C.S. and M. Rush (1991), 'Is the Budget Deficit "too large" ?', *Economic Enquiry,* 29, 429-445.

[13] Hamilton, J.D. and M.A. Flavin (1986), 'On the limitation of government borrowing: a framework for empirical testing', *The American Economic Review,* 76, 808-819.

[14] Hansen, B.E. (1992), 'Tests for parameter instability in regressions with I(1) processes', *Journal of Business and Economic Statistics,* 10, No. 3, 321-335.

[15] Hansen, B.E. (1996), 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica,* 64, No. 2, 414-430.

[16] Haug, A.A. (1991), 'Cointegration and government borrowing constraints: evidence for the United States', *Journal of Business and Economic Statistics,* 9, No. 1.

[17] Hendry, D.F. and M.P. Clements (1998), 'Forecasting non-stationary economic time series', *Manuscript.*

[18] Kleibergen, F., and H.K. van Dijk (1994), 'On the shape of the likelihood/posterior in cointegration models', *Econometric Theory,* 10, 514-551.

[19] De la Croix, D. and M. Lubrano (1996), 'Are interest rates responsible for unemployment in the eighties? A Bayesian analysis of cointegrated relationship with a regime shift', *Advances in Econometrics,* 11, Part B, 147-184

[20] Martin, G.M. (1996), 'Bayesian inference in models of cointegration: methods and applications', *Unpublished Ph.D. Thesis, Monash University.*

[21] Martin, G.M. and V.L. Martin (1997), 'Private and Public Consumption Expenditure Substitutability: Bayesian Estimates for the G7 Countries', *Monash University Working Paper 4/97.*

[22] Phillips, P.C.B. (1991a), 'To criticize the critics: an objective Bayesian analysis of stochastic trends', *Journal of Applied Econometrics,* 6, 333-364.

[23] Phillips, P.C.B. (1991b), 'Bayesian routes and unit roots: de rebus prioribus semper est disputandum', *Journal of Applied Econometrics,* 6, 435-474.

[24] Phillips, P.C.B. (1993), 'The long-run Australian consumption function re-examined: an empirical exercise in Bayesian inference', *Cowles Foundation Paper No. 825, Yale University.*

[25] Phillips, P.C.B. and B.E. Hansen (1990), 'Statistical inference in instrumental variables regression with I(1) processes', *Review of Economic Studies,* 57, 99-125.

[26] Phillips, P.C.B. and W. Ploberger (1994), 'Posterior odds testing for a unit root with databased model selection', *Econometric Theory,* 10, 774-808.

[27] Quintos, C.E. (1995), 'Sustainability of the deficit process with structural shifts', *Journal of Business and Economic Statistics,* 13, No. 4, 409-417.

[28] Stephens, D.A. (1994), 'Bayesian retrospective multiple changepoint identification', *Applied Statistics,* 43, No. 1, 159-178.

[29] Tanner, M.A. (1993), *Tools for statistical inference, Second edition,* Springer-Verlag, New York.

[30] Tanner, E. and P. Liu (1994), 'Is the budget deficit 'too large': some further evidence', *Economic Inquiry,* 32, 511-518.

[31] Tierney, L. (1994), 'Markov chains for exploring posterior distributions', *The Annals of Statistics,* 22, No.4, 1701-1762.

[32] Trehan, B. and C.E. Walsh (1988), 'Common trends, the government budget constraint, and revenue smoothing', *Journal of Economic Dynamics and Control*, 12, 425-444.

[33] Trehan, B. and C.E. Walsh (1991), 'Testing intertemporal budget constraints: theory and applications to U.S. federal budget and current account deficits', *Journal of Money, Credit and Banking*, 23, No. 2.

[34] Wilcox, D.W. (1989), 'The sustainability of government deficits: implications of the present-value borrowing constraint', *Journal of Money, Credit and Banking*, 21, No.3.

[35] Zivot, E. and P.C.B. Phillips (1994), 'A Bayesian analysis of trend determination in economic time series', *Econometric Reviews*, 13, No. 3, 291-336.

Figure 1: Real U.S. Govt. Exp. and Rev. : 1947(2)-1992(3)

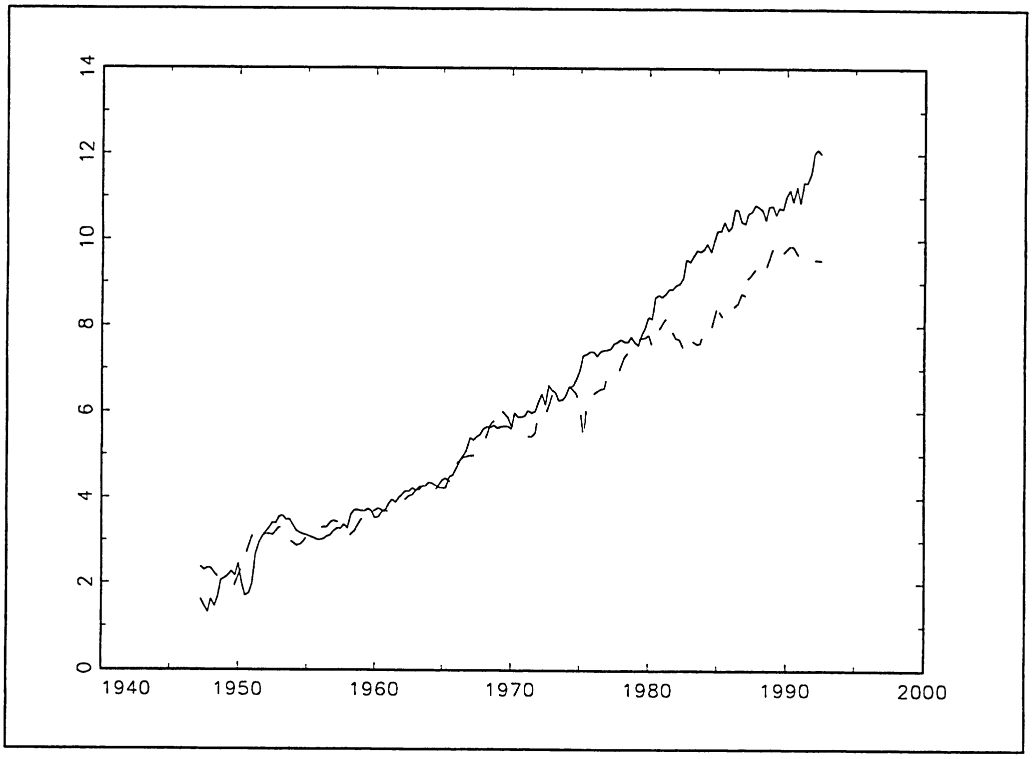

Figure 2: Marginal posterior functions for ρ_1, r_1, r_2 and r_3

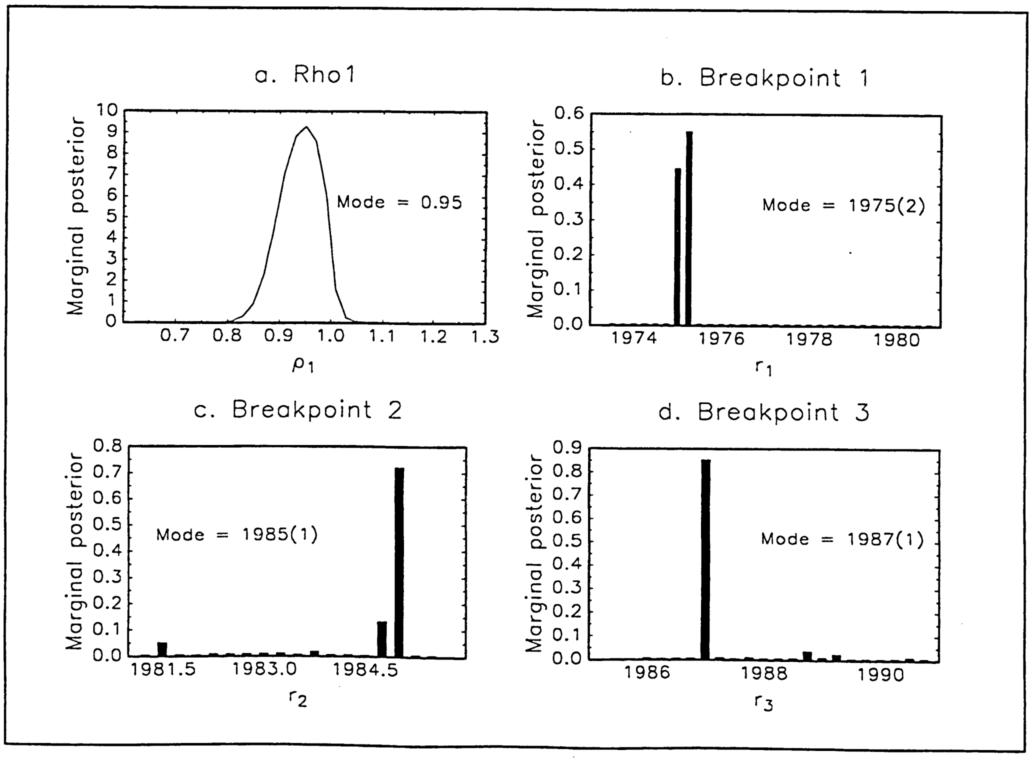

Figure 3: Marginal posterior densities for α_1 and the intercept shifts

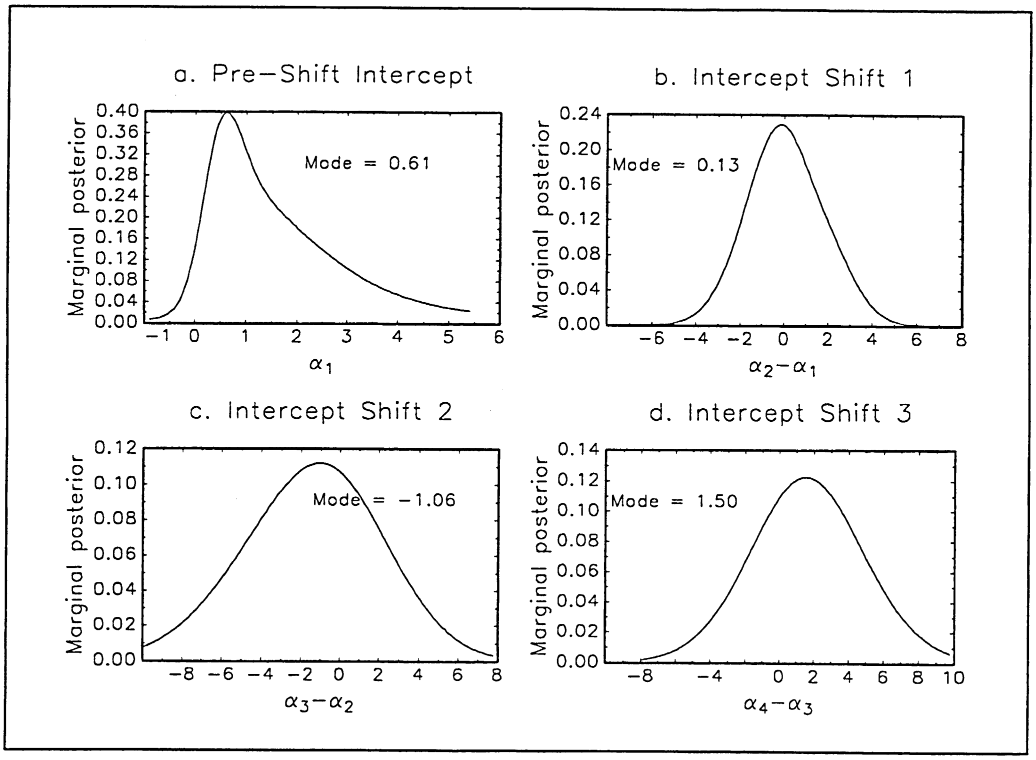

Figure 4: Marginal posterior densities for β_1 and the slope shifts

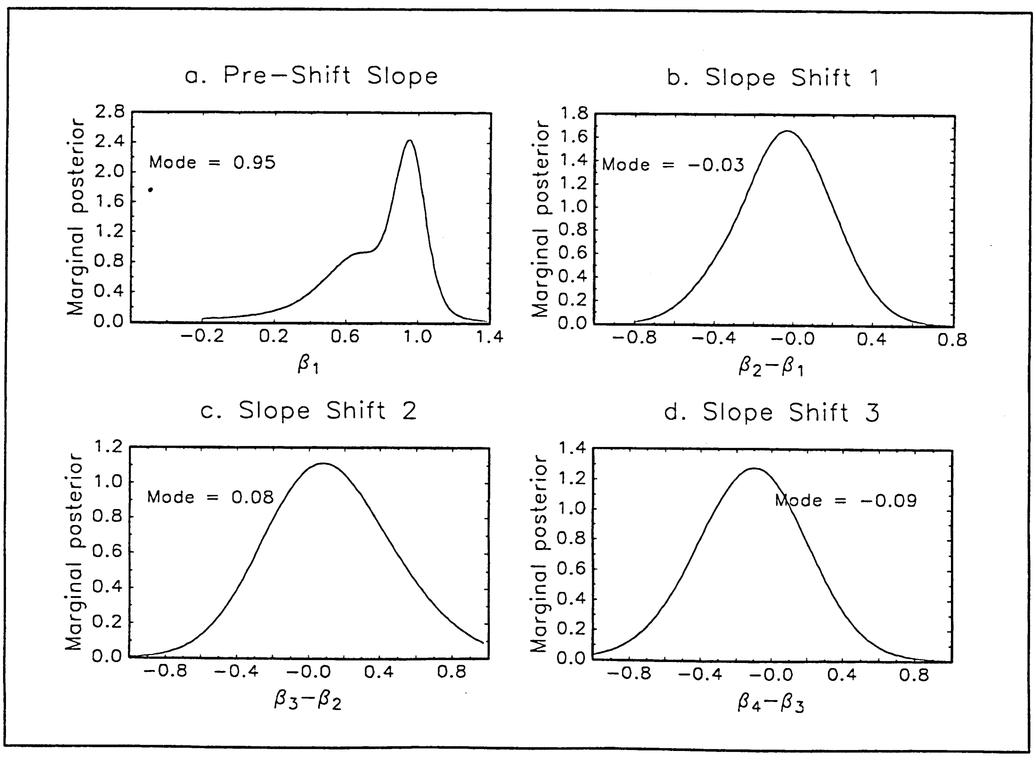

Figure 5: Marginal functions for ρ_1, r_1, r_2 and r_3 (no intercept shifts)

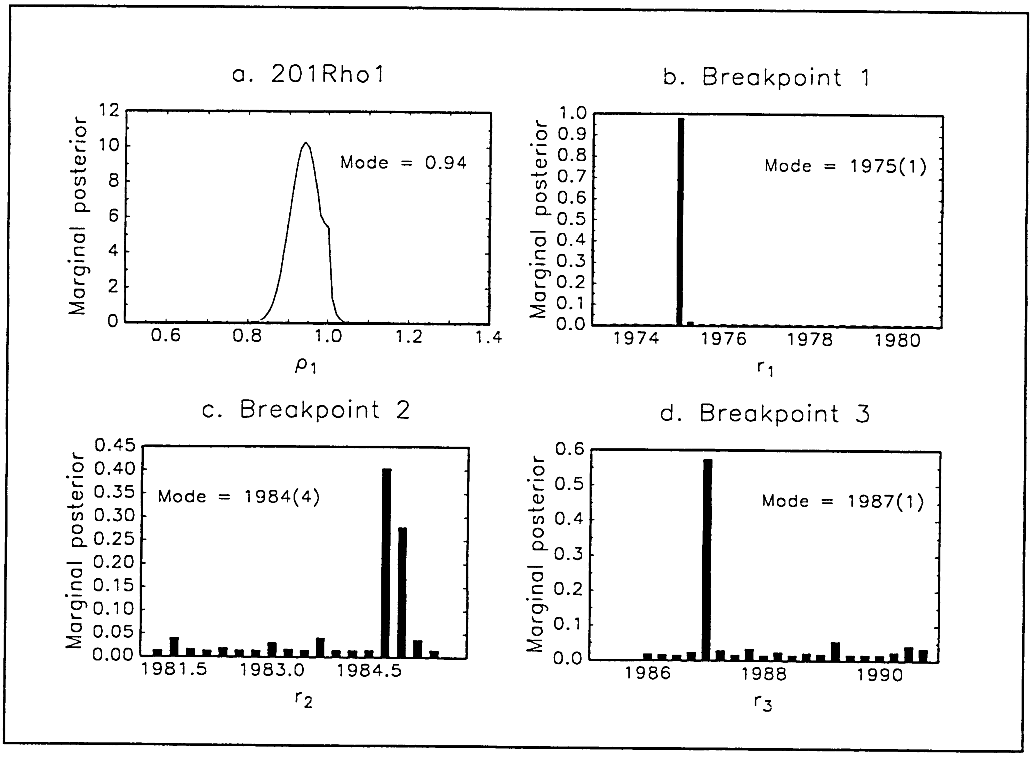

Figure 6: Marginal densities for β_1 and the slope shifts (no intercept shifts)