5

Time Series Implications of Present Value Budget Balance and of Martingale Models of Consumption and Taxes

by Lars Peter HANSEN, William ROBERDS and Thomas J. SARGENT

1. Introduction

Let $\{(r_t, p_t)\}$ be a covariance stationary process where $r_t - p_t$ is the *net surplus* at time t, k_t a level of assets (or debts) carried into period t, and $(1+\delta_t)$ the gross rate of return on those assets between t and $t+1$. Assume that there is a sequence of budget constraints

$$k_{t+1} = (1+\delta_t)\, k_t + r_t - p_t, \quad t = 0, 1, \ldots \tag{1.1}$$

with k_0 given. This paper studies the observable implications of some models which impose a terminal condition on assets $\{k_t\}$ which has the effect of converting (1.1) into an intertemporal budget constraint, one that asserts that for each $t \geq 0$, k_t equals present value of current and future surpluses, discounted at rates generated by the process $\{\delta_t\}$.

We study the implications of imposing two alternative assumptions on δ_t. For most of the paper, we focus on the first assumption, which is that $\delta_t = \delta$ for all $t \geq 0$. In Section 7, we briefly study a second assumption, that $E_{t-1}\, \delta_t = \delta$ for all $t \geq 0$. We characterize the restrictions on the bivariate process $\{(r_t, p_t)\}$ under both assumptions and characterize the restrictions on the trivariate process $\{(k_{t+1}, r_t, p_t)\}$ under the second assumption on δ_t. Presumably, an econometrician who possesses data on $\{(r_t, p_t)\}$, but not on $\{k_{t+1}\}$, would want to use the restrictions imposed on the bivariate process to arrive at a judgment of whether the data are consistent with present value budget balance.

We are motivated to obtain these characterizations because of our interest in two types of models, each of which incorporates a version of the intertemporal budget constraint induced by (1.1) and the terminal condition on $\{k_t\}$. The first type of model consists only of the intertemporal budget constraint, and seeks to test whether observations on the joint process $\{(r_t, p_t)\}$ or on the joint process $\{(k_{t+1}, r_t, p_t)\}$ satisfy the budget constraint. In this literature, represented by contributions by Hamilton and Flavin (1986), Hakkio and Rush (1986), Shim (1984), and Sargent (1987b), r_t has been interpreted as government expenditures, p_t as taxes and k_t as government debt. This literature can be interpreted as getting at the question either of whether the terminal constraint is operative (e.g., Hamilton and Flavin 1986), or whether the series on government expenditures r_t and tax collections p_t are measured in the way that is required to make the present value constraint implied by (1.1) hold (e.g., Shim 1984).[1] This literature also seeks to determine what time invariance and finite dimensional state restrictions on the joint stochastic process $\{(p_t, r_t)\}$ and the interest rate process δ_t must be imposed in order to give observable content to the present value budget constraint.

Under the assumption that $\delta_t = \delta$ for all t, it was shown by Sargent (1987b) that the present value restriction induces the restriction on a *particular* moving-average representation for $\{(r_t, p_t)\}$ that for each innovation, the present value of the response of $\{(r_t - p_t)\}$ is zero. But even if the budget restriction is true, it is possible to specify many other moving-average representations for $\{(r_t, p_t)\}$ that violate this restriction. Indeed, as we remark below, any vector autoregressive representation for $\{(r_t, p_t)\}$ *must* correspond to a moving-average representation that violates this restriction. However, moving-average representations are not unique. It turns out that for *any* jointly covariance stationary process $\{(r_t, p_t)\}$, there exist moving-average representations that *do* satisfy the restriction. These results characterize the sense in which even with ample time-invariance imposed, the present value restriction by itself is observationally empty and are described in Section 2.

The second type of model which includes a version of our present value budget constraint is a linear quadratic version of Hall's (1978) martingale model of consumption. In that model, δ_t is constant over time, p_t is consumption, r_t is labor income or an endowment shock, and k_t is the level of household nonhuman assets or the capital stock. A linear version of Barro's (1979) model of tax smoothing is isomorphic to Hall's, with δ again constant, p_t being tax collections, r_t being government expenditures, and k_t being government debt at time t. Present value budget balance is among the restrictions imposed by Hall and Barro's model. In Section 3, we characterize what, if anything, the present value budget balance restriction adds to the martingale restriction. We show that the conjunction of the martingale hypothesis with the present value hypothesis gives added content to the latter. In effect, the martingale restriction allows the econometrician to pin down one component of the information set of private agents, isolating an innovation to which the present value budget restriction on moving average responses does apply.

Section 4 extends the results of Section 3 by considering a richer class of models that, for the consumption interpretation of the models, assume a type of nonseparable preferences for consumption goods. This class of models permits aggregate consumption to be broken into several components, each of which is of different durability in the sense that it gives rise to a different time profile of service flows. We show that even in this richer framework, present value budget balance continues to impose an additional restriction over and above the martingale restrictions imposed by the Euler equations associated with the optimum problem. Section 5 focuses on a consumption interpretation of the model, though presumably there is also a *tax smoothing* interpretation, with different components of consumption being reinterpreted, *a la* Barro, as different components of government revenues.

Section 5 implements the Section 4 tests for U.S. data on consumption and income. The tests turn up no evidence against the present value budget balance restriction. It would be useful to perform such tests for data on U.S. government expenditures and tax receipts, but we do not execute those tests here.

Section 6 briefly uses a version of ideas introduced by Sims (1972a) to study how issues of approximation bear on the interpretation of tests of our Section 4 restrictions. We show that even though those restrictions have content, they are very tenuous because of the existence of a sequence of false models that satisfy the restriction, and that approximates arbitrarily well a process that is known to violate the restrictions. This means that in practice rejections of the restrictions hinge on adopting sufficiently parsimonious specifications for the observable stochastic processes.

Section 7 focuses on the restrictions imposed by present value budget balance on the joint processes $\{(k_{t+1}, r_t, p_t)\}$ and $\{(r_t, p_t)\}$ under the assumption that δ_t is stochastic but satisfies $E_{t-1}\,\delta_t = \delta$ for all t. The specification $E_{t-1}\,\delta_t = \delta$ does not restrict $\{(r_t, p_t)\}$ but does result in an *exact linear rational expectations model* for the joint process $\{(k_{t+1}, r_t, p_t)\}$. We briefly compare these latter restrictions with those proposed by Hamilton and Flavin (1986). In the next paper in this volume (Chapter 6), Roberds uses data on the U.S. Federal budget to test present value budget balance by using an exact linear rational expectations model for the process $\{k_{t+1}, r_t - p_t\}$.

2. Implications of Present-Value Budget Balance

Let $s_t = r_t - p_t$. Depending in the particular model at hand, $\{s_t\}$ is a stochastic process either of receipts minus expenditures or of expenditures minus receipts. There is a sequence of budget constraints:

$$k_{t+1} = (1+\delta)k_t + s_t \qquad \text{for } t = 0, 1, \ldots \tag{2.1}$$

where k_0 is an initial condition, k_t is a measure of an asset or debt stock at time t, and $(1+\delta)$ is the gross rate of return between time t and time $t+1$. This return is assumed to be constant. In a permanent income model for consumption, let k_t be a consumer's assets at the beginning of period t, r_t exogenous labor income, and p_t consumption. In a model of the government budget, we let k_t be the stock of goverment debt at the beginning of t, r_t the level of government expenditures, and p_t the level of government tax collections. When the initial condition k_0 is observed by an econometrician, $\{k_t\}$ can be generated using (2.1). However, we assume that this initial level is not observed, so that $\{k_t\}$ is not observed by the econometrician.

Without constraining the process $\{k_t\}$, (2.1) is evidently not restrictive. In fact, we can just use (2.1) to define k_{t+1} recursively as a function of k_t and s_t for any initial condition k_0 and any process $\{s_t\}$. We are interested, however, in situations in which there is a terminal constraint imposed on asset holdings that, in effect, converts (2.1) into a discounted present-value budget constraint. For example, suppose that (2.1) holds for $t = 0, 1, \ldots T$ and that k_{T+1} is constrained to be zero. Then we can write

$$k_0 = -\sum_{t=0}^{T} \lambda^{t+1} s_t \, , \tag{2.2}$$

where $\lambda \equiv 1/(1+\delta)$. Taking limits as T goes to infinity gives

$$k_0 = -\sum_{t=0}^{\infty} \lambda^{t+1} s_t \, , \tag{2.3}$$

where the infinite series on the right side is assumed to be mean-square convergent. We view (2.3) as the infinite horizon counterpart to the terminal condition that k_{T+1} be zero.

Let L^2 be the space of all scalar stochastic processes $\{x_t\}$ such that

$$E\sum_{t=0}^{\infty} \lambda^t (x_t)^2 < \infty. \tag{2.4}$$

Throughout all of our analysis, we maintain that $\{s_t\}$ is in L^2. This restriction is sufficient for the right side of (2.3) to be a well-defined mean-square limit. It can accommodate growth in $\{s_t\}$ as long as the growth is dominated appropriately by the discount factor λ. For example, let μ be a common growth factor for $\{s_t\}$. Multiply both sides by μ^{-t}, which gives

$$\mu^{-t} k_{t+1} = [(1+\delta)/\mu]\mu^{-t+1} k_t + \mu^{-t} s_t \, . \tag{2.5}$$

Let variables with * superscripts be scaled by μ^{-t} to remove the effect of the geometric growth, and let δ^* be constructed so as to satisfy $(1+\delta^*) = [(1+\delta)/\mu]$. Then (2.4) can be expressed as

$$k^*_{t+1} = (1+\delta^*)k^*_t + s^*_t \, , \tag{2.6}$$

which is a version of (2.1). Our subsequent analysis can be thought of as applying to the * variables, although for notational simplicity we will omit the *'s.

Alternatively, we can accommodate other forms of stochastic growth that can be eliminated by taking appropriate *quasi-differences*. Let $\alpha(L)$ be a quasi-differencing filter with a finite order ℓ. Apply $\alpha(L)$ to both sides of equation (2.1). In this case, let * variables denote variables to which $\alpha(L)$ has been applied. Then

$$k^*_{t+1} = (1+\delta)k^*_t + s^*_t \qquad \text{for } t = \ell, \ \ell+1, \ldots \, . \tag{2.7}$$

For convenience, we shift the starting point back from ℓ to zero, and again we omit the *'s.

Replicating the analysis leading up to (2.3) for any initial period t gives

$$k_t = -\sum_{\tau=0}^{\infty} \lambda^{\tau+1} s_{t+\tau} \, . \tag{2.8}$$

where the sequence $\{k_t\}$ is in L^2. In the remainder of this section, we focus on the following question: given a process $\{y_t\}$ that includes $\{s_t\}$ as one of its components, under what set of circumstances can we construct a process $\{k_t\}$ that satisfies (2.8) and that is *predetermined* in the sense that k_t depends only on random variables realized at time $t-1$ and earlier.

Consider an environment in which there is a covariance stationary, n-dimensional vector martingale difference sequence $\{w_t\}$ used to generate information in the economy. The time t information set J_t is generated by $w_t,\ w_{t-1},\ldots$ for each t.[2] We let H be a subspace of L^2 containing processes $\{x_t\}$ that are adapted to $\{J_t\}$ in the sense that x_t is in J_t for each t. The restriction that k_t be predetermined is formalized as the restriction that $\{k_{t+1}\}$ be in H.

For convenience, we suppose that $E(w_t w_t') = I$. The net surplus process $\{s_t\}$ (or some geometrically scaled or quasi-differenced version of this process) is the first component of an m-dimensional vector process $\{y_t\}$ that is assumed to be a time invariant, linear function of this martingale difference sequence:

<u>Assumption A1:</u> $y_t = C(L)w_t$ for $t = 0, 1, \ldots$ where $C(z) = \sum_{j=0}^{\infty} c_j z^j$ and $\sum_{j=0}^{\infty} |c_j|^2 < \infty$.

One possibility is that s_t is the only component of y_t, in which case m is one. More generally, y_t can contain other variables that are useful in forecasting future values of s_t and are observed by an econometrician. It is straightforward to show that when A1 is satisfied, $\{s_t\}$ is in H.

We shall use the idea that the stochastic process $\{y_t\}$ is *stochastically nonsingular* from the perspective of linear prediction theory. Informally, stochastic nonsingularity requires that no component of y_t can be expressed as a possibly infinite linear combination of current, past and future values of the other components. Formally, this requirement is stated in terms of the spectral density matrix of $\{y_t\}$. Assumption A1 implies that the following radial limit

$$C(\theta) \equiv \lim_{\eta \uparrow 1} C[\eta \exp(i\theta)] \tag{2.9}$$

exists for almost all θ in $(-\pi, \pi]$. Then the spectral density matrix for frequency θ is

$$S(\theta) \equiv C[\exp(-i\theta)]\, C[\exp(i\theta)]' \,. \tag{2.10}$$

In terms of the spectral density matrix, stochastic nonsingularity amounts to:

<u>Assumption A2:</u> $S(\theta)$ has rank m for almost all θ in $(-\pi, \pi]$.

We now deduce the restrictions on $\{y_t\}$ implied by the fact k_t as given by (2.8) is predetermined. Let the first row of $C(z)$ be denoted $\sigma(z)$. Then

$$s_t = \sigma(L)w_t \,. \tag{2.11}$$

It can be verified by substitution into (2.8) that a process $\{k_t\}$ that satisfies (2.8) is given by:[3]

$$\begin{aligned} &k_t = \kappa(L)w_t \\ &\text{where } \kappa(z) \equiv -\lambda\sigma(z)/(1-\lambda z^{-1}) = -\lambda z\sigma(z)/(z-\lambda) \,. \end{aligned} \tag{2.12}$$

However, in general, the function $\kappa(z)$ has a two-sided Laurent series expansion about $z = 0$ implying that k_t depends on current and future values of w_t. For this reason, k_t given by (2.12) may not be predetermined. The problem is that the function $\kappa(z)$ may have a pole (diverge to infinity) at $z = \lambda$. In the special case in which $\sigma(\lambda) = 0$, $\kappa(z)$ ceases to have a pole at λ and, in fact, $\kappa(z)$ has a one-sided power series expansion. In addition, $\kappa(0) = 0$ which guarantees that k_t depends only on past information. Thus, a necessary and sufficient condition for k_t to be predetermined relative to the sequence of information sets $\{J_t\}$ is that $\sigma(\lambda) = 0$, which we summarize as:

<u>Restriction R1:</u> $\sigma(\lambda) = 0$

We have thus established

<u>Proposition 1:</u> Suppose A1 is maintained and that economic agents have access to J_t at time t. Then $\{k_{t+1}\}$ given by (2.8) is in H if and only if R1 holds.

A version of this result was derived previously and interpreted by Sargent (1987b, pages 381–385).[4] Note that $\sigma(\lambda) = 0$ (or equivalently, $\sum_{j=0}^{\infty} \sigma_j \lambda^j = 0$) states the present value of the moving average coefficients of the surplus equals zero for each innovation in w_t.

In general, R1 rules out the possibility that the moving-average representation in A1 is a Wold representation of $\{y_t\}$. Recall that in a Wold representation the Hilbert spaces generated by $\{y_t,\ y_{t-1},\ \ldots\}$ and $\{w_t,\ w_{t-1},\ \ldots\}$ must be identical. When Assumption A2 is satisfied, a necessary and sufficient condition for these two spaces to be the same is given by the following condition:

<u>Restriction R2:</u> The rank of $C(z)$ is n for all $|z| < 1$.

Under A2, restrictions R1 and R2 cannot both be satisfied because R1 implies that $C(\lambda)$ can have at most rank $m-1$. Hence for R1 and R2 to be compatible, n must be less than m. If n is less than m, A2 is violated and the process $\{y_t\}$ is stochastically singular. We summarize this finding in the following:

Proposition 2: Suppose A1 and A2 are maintained. Then R1 cannot hold for any C satisfying R2.

This finding has the practical implication that one ought not to test R1 by estimating versions of Wold representations as is done, for example, when estimating vector autoregressions. This is true even though we have assumed that $\{y_t\}$ is covariance stationary and so possesses a Wold representation. Restriction R1 applies to a moving-average representation that is necessarily distinct from the Wold representation. It follows that one cannot test R1 by examining directly the impulse response functions from a vector autoregression.

Is there any way that R1 can be tested? We now show that without additional restrictions the answer is no. Suppose that $\{y_t\}$ satisfies A1 and A2. We know from the Wold Decomposition Theorem for covariance stationary processes that there exists a moving-average representation that satisfies rank condition R2. Hence it is an implication of A1 and A2 that

(2.13) $$y_t = C^*(L)w_t^*$$

where $C^*(z)$ satisfies assumption A2 and $\{w_t^*\}$ is an m-dimensional, covariance stationary white noise process with contemporaneous covariance matrix I.[5] In light of Proposition 2, C^* does not satisfy restriction R1.

To show that Restriction R1 is not testable, we demonstrate that it is always possible to build another moving-average representation, distinct from (2.13), that satisfies R1. Given $\{w_t^*\}$, first construct another m-dimensional serially uncorrelated process, say $\{w_t\}$, that depends on current and future values of $\{w_t^*\}$ and for which $Ew_tw_t' = I$. The process w_t satisfies:

(2.14) $$w_t = D(L^{-1})'w_t^* \text{ where } D(z) = \sum_{j=0}^{\infty} d_j z^j, \quad \sum_{j=0}^{\infty} |d_j|^2 < \infty$$

and

(2.15) $$D[\exp(i\theta)]\, D[\exp(-i\theta)]' = I \text{ for almost all } \theta.$$

Notice that $D[\exp(i\theta)]$ is a unitary matrix for almost all θ, and

(2.16) $$w_t^* = D(L)w_t$$

so that w_t^* depends only on current and past values of w_t. Consequently, the Hilbert space generated by w_t, w_{t-1}, ... is no smaller than the space generated by w_t^*, w_{t-1}^*, ... and in fact is often strictly larger. We can represent y_t in terms of w_t via:

(2.17) $$y_t = C(L)w_t \quad \text{where} \quad C(z) = C^*(z)D(z) .$$

In light of (2.17), we can show that Restriction R1 is not testable by establishing the existence of a function $D(z)$ satisfying (2.15) such that the first row of $C^*(\lambda)D(\lambda)$ is zero. We now propose two alternative ways in which this can be accomplished. One possibility is to construct $D(z)$ so that $D(\lambda)$ is a matrix of zeros. An example of such a $D(z)$ is

(2.18) $$D(z) = [(z-\lambda)/(1-\lambda z)]\, I .$$

This choice of $D(z)$ satisfies (2.15) because

(2.19) $$\frac{[\exp(i\theta)-\lambda]\;[\exp(-i\theta)-\lambda]}{[1-\lambda\exp(i\theta)]\;[1-\lambda\exp(-i\theta)]} = 1 \text{ for all } \theta.$$

A second possibility is to form an orthogonal matrix Q, the first column of which is a vector that is proportional to the first row of $C^*(\lambda)$ and has norm one. The remaining columns of Q are a set of $m-1$ orthonormal vectors that are orthogonal to the first row of $C^*(\lambda)$. Therefore all entries in the first row of $C^*(\lambda)Q$ are zeroes except for the first entry. The matrix Q is then used to build $D(z)$ as follows:

(2.20) $$D(z) = Q \begin{bmatrix} (z-\lambda)/(1-\lambda z) & 0 \\ 0 & I \end{bmatrix}$$

where the matrix I has $m-1$ rows and columns. Using (2.19), it is straightforward to show that $D(z)$ satisfies (2.15). Notice that the first row and column of $C^*(\lambda)D(\lambda)$ contain all zeroes.

An equivalent way to construct $C(z)$ is as follows. Form the process $\{k_t\}$ using equation (2.8) and a composite process $\{x_t\}$ where $x_t' \equiv [y_t', k_t]$. This composite process is stochastically singular because k_t is an infinite linear combination of current and future values of s_t. Nevertheless, it possess a Wold moving-average representation in terms of an m-dimensional vector white noise $\{w_t^+\}$:

(2.21) $$y_t = C^+(L)w_t^+ , \quad k_t = \kappa^+(L)w_t^+ ,$$

where linear combinations of x_t, $x_{t-1}, \ldots$ generate the same Hilbert space as linear combinations of w_t^+, w_{t-1}^+, ... Let $\sigma^+(z)$ be the first row of $C^+(z)$. Since k_t satisfies (2.8),

$$\kappa^+(z) = -z\lambda\sigma^+(z)/(z-\lambda)\ . \tag{2.22}$$

Note that k_t depends only on current and past values of w_t^+ because (2.21) is a Wold representation for $\{x_t\}$. Therefore, $\sigma^+(z)$ must satisfy Restriction R1. In Appendix A we show that $C^+(z) = C^*(z)D(z)$ where $D(z)$ is given by (2.20) can be used in (2.21) in forming a Wold representation for $\{x_t\}$.

Summarizing these results we have established:

Proposition 3: Suppose A1 and A2 are maintained. It is always possible to find a moving-average representation $y_t = C^+(L)w_t^+$ that satisfies R1.[6]

Thus, Proposition 3 shows that R1 is not testable without additional restrictions to aid in the identification of w_t. Proposition 2 demonstrates that one commonly used device for identifying w_t is inappropriate when $\{y_t\}$ has full rank. In the next two sections we consider alternative ways to identify components of moving-average representations that can be used in testing R1.

3. The Martingale Model

In this section we impose considerably more structure on the problem. First, we decompose the surplus process $\{s_t\}$ into two components, *payouts* and *receipts*. The receipt process is specified exogeneously, but the payout process is modeled as the optimal decision process from a quadratic optimization problem subject to constraint (2.3).[7] The objective function for this optimization problem is designed to imply a martingale model for the payout process. This leads to a version of Hall's (1978) model of consumption or Barro's (1979) model of taxation. We show that beyond the martingale characterization for payouts, present-value budget balance delivers an additional restriction that is testable so long as the discount factor is known *a priori*.

Let $\gamma(L)$ be a stationarity inducing transformation for a receipt process $\{r_t\}$. In our analysis we take $\gamma(z)$ to be either 1 or $1-z$ in cases in which the stochastic process for receipts has a unit root. Other specifications of $\gamma(z)$ can be explored by mimicking the analysis in this section.

Assumption A3: $\gamma(L)r_t = \rho(L)w_t$ where $\rho(z) = \sum_{j=0}^{\infty} \rho_j z^j$, $\sum_{j=0}^{\infty} |\rho_j|^2 < \infty$.

We let $\{p_t\}$ denote the payout process. The net surplus process $\{s_t\}$ of Section 2 is given by

$$s_t = r_t - p_t. \tag{3.1}$$

The payout process $\{p_t\}$ is determined as the solution to an optimization problem with the objective being to maximize:

$$-E\Big[\sum_{t=0}^{\infty} \lambda^t\,(p_t - b)^2\Big] \tag{3.2}$$

where $b > 0$ subject to constraint (2.1) with k_0 given. The processes $\{p_t\}$ and $\{k_{t+1}\}$ are restricted to be in H. The solution to this problem is described in Sargent (1987b) and Hansen (1987). The optimal decision processes for p_t and k_{t+1} satisfy:

$$\begin{aligned} p_t &= \delta k_t + (1-\lambda)E\Big(\sum_{j=0}^{\infty} \lambda^j r_{t+j} \mid J_t\Big) \\ k_{t+1} &= k_t + r_t - (1-\lambda)\,E\Big(\sum_{j=0}^{\infty} \lambda^j r_{t+j} \mid J_t\Big) \end{aligned} \tag{3.3}$$

It can be verified that constraint (2.1) is satisfied adding the two equations in (3.3) together and rearranging terms.

To deduce implications for $\{p_t\}$, take the first equation in (3.3) at time $t+1$, subtract the same equation at time t and then substitute for $k_{t+1} - k_t$ from the second equation. This yields the following result of Flavin (1981):

(3.4)

$$p_{t+1} - p_t = (1-\lambda)\,E\Big(\sum_{j=0}^{\infty} \lambda^j r_{t+j+1} \,|J_{t+1}\Big) - (1-\lambda)\,E\Big(\sum_{j=0}^{\infty} \lambda^j r_{t+j+1} \mid J_t\Big)\ .$$

Using a formula reported in Hansen and Sargent (1980a, 1981b), it follows that

$$p_t - p_{t-1} = [(1-\lambda)\rho(\lambda)/\gamma(\lambda)]w_t \quad t = 1,\ 2,\ \ldots \tag{3.5}$$

We now investigate implications of (3.5) for the process $\{s_t\}$. Stationarity in $\{s_t\}$ is induced by taking first-differences:

$$\begin{aligned} s_t - s_{t-1} &= r_t - r_{t-1} - p_t + p_{t-1} \\ &= \sigma(L)w_t \end{aligned} \tag{3.6}$$

where

$$\sigma(z) \equiv (1-z)\rho(z)/\gamma(z) - (1-\lambda)\rho(\lambda)/\gamma(\lambda). \tag{3.7}$$

To relate the present analysis to the analysis in Section 2, we assume that $\{y_t\}$ contains at least two components ($m \geq 2$), with the first component being $(1-L)s_t$ and the second component being $\gamma(L)r_t$. Notice that $(1-z)/\gamma(z)$ is either $(1-z)$ or 1. Evaluating $\sigma(z)$ at $z=\lambda$, it is evident that $\sigma(z)$ satisfies R1. Finally, consider the special case in which $\rho(\lambda)$ is zero. In this case $p_t - p_{t-1} = 0$ which implies that $\{y_t\}$ is stochastically singular. Therefore, this case is eliminated from consideration by A3.

More generally, we consider a payout process that satisfies:

$$p_t - p_{t-1} = \pi w_t \tag{3.8}$$

for some $\pi \neq 0$. Then $\{p_t\}$ is a martingale adapted to $\{J_t\}$. In this case, $\sigma(z)$ satisfies:

Restriction R3: $\sigma(z) = (1-z)\rho(z)/\gamma(z) - \pi$ for some row vector π.

Restriction R3 is the focus of Hall (1978) and Flavin (1981). Note that R1 and R3 imply (3.8) for $\pi = (1-\lambda)\rho(\lambda)/\gamma(\lambda)$. The question of interest in this section is whether R1 imposes any additional restrictions once R3 is satisfied. In other words, is the restriction

$$\pi = (1-\lambda)\rho(\lambda)/\gamma(\lambda) \tag{3.9}$$

testable?

To answer this question, we construct an orthogonal matrix Q as follows. Since A2 is satisfied, π cannot be zero. The first column of Q is $\pi'/|\pi|$ and the remaining columns contain n-1 orthonormal vectors that are orthogonal to π'. Then πQ is row vector with $|\pi|$ in the first position and zeros elsewhere. Define

(3.10)
$$\sigma^+(z) \equiv \sigma(z)Q\,,\ \rho^+(z) \equiv \rho(z)Q\,,\ \pi^+ \equiv \pi Q\,,\ \text{and}\ \ w_t^+ \equiv Q'w_t\,.$$

The Q transformation is introduced so that the first entry of w_t^+ is proportional to πw_t. Hence $p_t - p_{t-1}$ depends only on the first entry of w_t^+. Partition $\sigma^+(z) = [\sigma_1^+(z), \sigma_2^+(z)]$, $\rho^+(z) = [\rho_1^+(z), \rho_2^+(z)]$, $w_t^{+\prime} = [w_{1t}^+, w_{2t}^{+\prime}]$ where $\sigma_1^+(z)$, $\rho_1^+(z)$ and w_{1t}^+ each have one entry. Write

$$\begin{aligned} s_t - s_{t-1} &= \sigma_1^+(L)w_{1t}^+ + \sigma_2^+(L)w_{2t}^+ \\ \gamma(L)r_t &= \rho_1^+(L)w_{1t}^+ + \rho_2^+(L)w_{2t}^+ \\ p_t - p_{t-1} &= |\pi|w_{1t}^+ \quad . \end{aligned} \tag{3.11}$$

An equivalent representation of (3.9) is

$$(1-\lambda)\rho_1^+(\lambda)/\gamma(\lambda) = |\pi| \tag{3.12}$$

and

$$\rho_2^+(\lambda) = 0. \tag{3.13}$$

First, we investigate whether (3.12) is testable. We use (3.11) to compute the regression of $\gamma(L)r_t$ onto current and past values of $p_t - p_{t-1}$ or equivalently onto current and past values of w_{1t}^+. These regressions are:

$$\begin{aligned} \gamma(L)r_t &= \rho_1^+(L)w_{1t}^+ + e_t \\ &= \beta(L)\,(p_t - p_{t-1}) + e_t \end{aligned} \tag{3.14}$$

where $\beta(z) \equiv \rho_1^+(z)/|\pi|$ and the regression error e_t is given by

$$e_t = \rho_2^+(L)w_{2t}^+. \tag{3.15}$$

The function $\beta(z)$ is identifiable for $|z| < 1$.[8] Consequently, (3.12) is a testable restriction given knowledge of the discount factor λ.[9]

Restriction R4: $(1-\lambda)\beta(\lambda)/\gamma(\lambda) = 1$.

Proposition 4: Suppose A1–A3 are maintained. If R1 and R3 are satisfied for $\pi \neq 0$, then R4 is also satisfied.

In Section 2 we showed that, by itself, present-value budget balance imposes no testable restrictions on $\{y_t\}$, essentially because pertinent innovations to agent's information sets could not be identified. How then does it occur that, in conjunction with the martingale restriction (R3), the present-value budget balance restriction acquires the content summarized in R4? This content emerges because the martingale restriction allows us to deduce one component of the new information that arrives at time t to economic agents, namely, $p_t - p_{t-1}$. The present-value budget balance restriction then applies to this component.[10] In the analysis in Section 2, none of the components of the new information was identified. As a consequence, the derived restrictions were not testable.

The presence of an additional implication implied by present-value budget balance also has been noted by West (1988).[11] West derived a variance inequality that is robust to the misspecification of the information set of economic agents used to forecast future values of r_t.

Restriction R4 shares this robustness and in fact implies West's variance inequality. The converse is not true, however. West's variance inequality does not imply R4 so that, in principle, R4 can be used to construct a statistical test with additional power.[12]

Finally, we investigate whether R4 exhausts the testable implications of R1 when $\{p_t\}$ is a martingale (i.e., when R3 is satisfied). This amounts to assessing the empirical content of (3.13) $[\rho_2^+(\lambda) = 0]$. Relation (3.13), however, is not testable for the same reasons discussed in Section 2 that $\sigma(\lambda) = 0$ is not testable. To see this, form an m-1 dimensional vector $\hat{y}_t$ by taking the least squares regression error of y_t on the Hilbert space generated by current and past values of $p_t - p_{t-1}$. Since the first entry of y_t is a linear combination of $r_t - r_{t-1}$ and $p_t - p_{t-1}$, its forecast error is a linear combination of current and past values of $\gamma(L)r_t$. Hence the m-dimensional vector stochastic process of regression errors is stochastically singular. We simply eliminate the first entry when forming $\hat{y}_t$ which means that e_t is now the first entry of $\hat{y}_t$. The analysis in Section 2 applies to the $(m-1)$-dimensional process $\{\hat{y}_t\}$ with e_t playing the role of s_t. We thus have:

Proposition 5: Suppose A1–A3 are maintained. If R3 and R4 are satisfied, then it is always possible to find a moving-average representation $y_t = \hat{C}(L)\hat{w}_t$ that satisfies R1.

Proposition 5 offers a word of caution for statistical tests of the restrictions implied by R1 and R3. For a given parameterization of ρ^+, one might consider testing whether $\rho_2^+(\lambda) = 0$. Holding fixed the particular finite-dimensional parameterization of ρ_2^+, this restriction can be tested using, say, a likelihood ratio test. The message from Proposition 5 is that such a test is not particularly interesting because there will always exist an alternative, observationally equivalent parameterization of ρ_2^+ that satisfies the restriction by construction. Restriction R4, however, is not subject to this same criticism.

4. Nonseparable Preferences

A potentially important defect of the model used in Section 3 is that preferences for the payout are assumed to be time separable. While this may not be a bad model when the payout consists only of consumption goods that are nondurable, there is no natural measure of receipts (or say labor income) that is matched to nondurable consumption in the manner assumed in Section 3. In this section we extend the martingale model described in Section 3 to allow for the possibility that the payout can be divided into a vector of consumption goods, and that preferences for these consumption goods are not necessarily separable either across goods or over time. We then show that the present-value budget balance restriction R1 still has a testable implication.

In their analyses of permanent income models, Campbell (1987) and West (1988) assume, in effect, that nondurable consumption is a fixed fraction of total consumption. We adopt a different approach in which the total payout is divided and invested in one of ℓ possible ways. For instance, the alternative investments might entail expenditures on alternative consumption goods such as durables, nondurables, and services. Hence

$$p_t = \mu' c_t \tag{4.1}$$

where μ is an ℓ-dimensional vector of positive numbers. Bernanke (1985) used this strategy in modeling simultaneously two consumption goods classifications, durables and nondurables.

Following Telser and Graves (1972), Hansen (1987) and Eichenbaum and Hansen (1990), we model each of these investments as generating an intertemporal bundle of ℓ different services. The bundling is a device for modeling intertemporal nonseparabilities in preferences, such as consumption durability and habit persistence, and nonseparabilities across the different goods classifications. More precisely, a vector of consumption goods c_t generates a corresponding service vector $b_\tau c_t$ at time $t+\tau$ for $\tau \geq 0$ where:

Assumption A4: $B(z) \equiv \sum_{\tau=0}^{\infty} b_\tau z^\tau$ is continuous and nonsingular on the domain $\{|z| \leq 1\}$.[13]

At time t the total quantity of consumption services is given by

$$h_t \equiv \sum_{\tau=0}^{\infty} b_\tau c_{t-\tau} = B(L)c_t \tag{4.2}$$

where investments prior to time zero are taken as initial conditions. This specification accommodates the linear-quadratic durable goods consumption models of Mankiw (1982) and Bernanke (1985) as special cases, as well as linear-quadratic versions of the habit persistence model of Ryder and Heal (1973).

The preferences used to induce the optimal payout process are specified as follows in terms of the ℓ-dimensional service process:

$$-(1/2)E \sum_{t=0}^{\infty} \lambda^t (h_t - g)'(h_t - g) . \tag{4.3}$$

Here g is an ℓ-dimensional vector of satiation points. As in the Sections 2 and 3, we impose (2.1) and require $\{k_{t+1}\}$ to be in H. In addition, $\{p_t\}$ and the entries of the processes $\{c_t\}$ and $\{h_t\}$ are restricted to be in H. Finally, the receipt process $\{r_t\}$ is assumed to satisfy A3.

We solve this optimization problem using the method of Lagrange multipliers. Let mk_t be the Lagrange multiplier associated with (2.1), mp_t be the multiplier associated with (4.1) and mh_t the multiplier associated with (4.2) at time t. The multiplier processes $\{mk_t\}$, $\{mp_t\}$, and the components of the process $\{mh_t\}$ are restricted to be in H. The Lagrangean is

(4.4)
$$\mathcal{L} = E\sum_{t=0}^{\infty}\lambda^t\Big\{-(1/2)(h_t-g)\cdot(h_t-g)+mp_t(\mu' c_t - p_t)$$
$$+\,mh_t'[h_t - B(L)c_t] + mk_t[-k_{t+1}+(1+\delta)k_t + r_t - p_t]\Big\}\,.$$

The first-order conditions for p_t, k_t, c_t and h_t are:

$$-\,mp_t + mk_t = 0 \tag{4.5}$$
$$-\,mk_t + E(mk_{t+1}\mid J_t) = 0 \tag{4.6}$$
$$-\,E[B(\lambda L^{-1})' mh_t \mid J_t] + \mu mp_t = 0 \tag{4.7}$$
$$mh_t - (g - h_t) = 0\,. \tag{4.8}$$

The multiplier mk_t can be interpreted as the shadow valuation of the capital stock at the end of time t, while the multiplier mp_t is the (indirect) marginal utility of the payout at time t. First-order condition (4.5) indicates that these two multipliers should be equal. First-order condition (4.6) restricts the shadow valuation process for the capital stock to be a martingale. In light of (4.5) and (4.6), the marginal utility process for the total payout should be a martingale as in the model investigated in Section 3.

The multiplier mh_t is the time t marginal utility vector for services. First-order condition (4.7) relates the current (indirect) marginal utility for the payout to the current and expected future values of the marginal utility vector for services. This link reflects the technology for converting a payout today into services in current and future time periods.

We now use the connection between the marginal utilities to deduce the optimal process for $\{c_t\}$. First solve (4.7) for mh_t as a function of current and expected future values of mp_t. In light of A4, $B(\lambda L^{-1})$ has a one-sided forward inverse. Consequently,

$$mh_t = E\left\{[B(\lambda L^{-1})^{-1'}(\mu mp_t)\mid J_t]\right\}. \tag{4.9}$$

Since $\{mp_t\}$ is a martingale,

$$mh_t = B(\lambda)^{-1'}\mu mp_t \tag{4.10}$$

implying that that the vector of marginal utilities for services is also a martingale.[14] Substituting for mh_t from (4.8) and for h_t from (4.2) gives

$$B(L)c_t + B(\lambda)^{-1'}\mu mp_t - g = 0\,. \tag{4.11}$$

We compute c_t, k_{t+1}, and mp_t by solving the three equation system:

$$\begin{aligned} &B(L)c_t + B(\lambda)^{-1'}\mu mp_t - g = 0\\ &E(mp_{t+1}|J_t) - mp_t = 0\\ &k_{t+1} - (1+\delta)k_t - r_t + \mu' c_t = 0\,. \end{aligned} \tag{4.12}$$

Taking a linear combination of these three equations,

$$\begin{aligned} &(1-\lambda)\mu' B(\lambda)^{-1}[B(L)c_t + B(\lambda)^{-1'}\mu mp_t - g]\\ &+\lambda[\mu' B(\lambda)^{-1}B(\lambda)^{-1'}\mu]\,E[(-L^{-1}+1)mp_t|J_t]\\ &-(1-\lambda)\,[(1-\lambda^{-1}L)k_{t+1} - r_t + \mu' c_t] = 0\,. \end{aligned} \tag{4.13}$$

Rearranging terms gives

(4.14)
$$\begin{aligned} &(1-\lambda)\mu'[B(\lambda)^{-1}B(L) - I]c_t\\ &+[\mu' B(\lambda)^{-1}B(\lambda)^{-1'}\mu]\,E[(1-\lambda L^{-1})mp_t|J_t]\\ &-(1-\lambda)(1-\lambda^{-1}L)\,k_{t+1} + (1-\lambda)r_t - (1-\lambda)\mu' B(\lambda)^{-1}g = 0\,. \end{aligned}$$

Note that $(1-\lambda)\mu'[B(\lambda)^{-1}B(z) - I]$ can be factored

$$(1-\lambda)\mu'[B(\lambda)^{-1}B(z) - I] = (z-\lambda)\chi(z) \tag{4.15}$$

where $\chi(z)$ has a power series expansion for $|z|<1$, because $B(\lambda)^{-1}B(z) - I$ is a matrix of zeros when $z=\lambda$. Applying the forward operator $(1-\lambda L^{-1})^{-1}$ to (4.14) and taking expectations conditioned on J_t gives

$$\begin{aligned} &\chi(L)c_{t-1} + [\mu' B(\lambda)^{-1}B(\lambda)^{-1'}\mu]mp_t + \delta k_t\\ &+(1-\lambda)E(\sum_{j=0}^{\infty}\lambda^j r_{t+j}\mid J_t) - \mu' B(\lambda)^{-1}g = 0\,. \end{aligned} \tag{4.16}$$

Since A4 is satisfied, $[\mu' B(\lambda)^{-1} B(\lambda)^{-1'} \mu]$ is strictly positive. Solving for mp_t,

$$\begin{aligned} mp_t = & - \epsilon\chi(L)c_{t-1} \\ & - \epsilon\Big[\delta k_t + (1-\lambda)E\Big(\sum_{j=0}^{\infty} \lambda^j r_{t+j}|J_t\Big)\Big] \\ & + \epsilon\mu' B(\lambda)^{-1} g \end{aligned} \tag{4.17}$$

where

$$\epsilon \equiv 1/[\mu' B(\lambda)^{-1} B(\lambda)^{-1'} \mu] \ . \tag{4.18}$$

We can obtain recursive representations for c_t and k_{t+1} by substituting (4.17) into the first and third equations in (4.12). It can be verified by the reader that these recursive representations are the same as those reported in (3.3) for the special case in which $B(z)$ is 1.

What is of interest to us is the implied moving-average representation for $p_t - p_{t-1}$ in terms of current and past values of w_t. To deduce this representation, note from (4.17) that the one-step-ahead forecast error in mp_t is given by

$$mp_t - mp_{t-1} = -\epsilon(1-\lambda)\,[\rho(\lambda)/\gamma(\lambda)]w_t \ , \tag{4.19}$$

because c_{t-1} and k_t are in J_t. Differencing (4.11), we have that

$$B(L)\,[c_t - c_{t-1}] = [\epsilon(1-\lambda)B(\lambda)^{-1'}\,\mu\rho(\lambda)/\gamma(\lambda)]w_t \ . \tag{4.20}$$

We derive a moving-average representation for $c_t - c_{t-1}$ in a stochastic steady state by applying the inverse operator $B(L)^{-1}$:

$$c_t - c_{t-1} = B(L)^{-1}[\epsilon(1-\lambda)B(\lambda)^{-1'}\,\mu\rho(\lambda)/\gamma(\lambda)]w_t \ . \tag{4.21}$$

Therefore, in a stochastic steady state $\{c_t - c_{t-1}\}$ depends only on the scalar noise $\rho(\lambda)w_t$. While the consumption goods are not proportional, the stochastic process $\{c_t - c_{t-1}\}$ is stochastically singular. Premultiplying both sides of (4.21) by μ' gives:

$$p_t - p_{t-1} = \psi(L)\pi w_t \ , \tag{4.22}$$

where

$$\begin{aligned} \psi(z) &= \mu' B(z)^{-1} B(\lambda)^{-1'}\,\mu/[\mu' B(\lambda)^{-1} B(\lambda)^{-1'} \mu] \\ &\text{and} \ \ \pi = (1-\lambda)\rho(\lambda)/\gamma(\lambda) \ . \end{aligned} \tag{4.23}$$

Notice that $\psi(\lambda) = 1$ and that π is the same as in martingale model of Section 3. The scalar lag polynomial $\psi(L)$ occurs in (4.23) because of the nonseparabilities over time in the induced preferences for the payout. In contrast to Section 3, $\{p_t\}$ will not be a martingale unless $\psi(z) = 1$ for all z. Since A3 is maintained, $\rho(\lambda)$ and hence π is different from zero. This eliminates the possibility that the payout process is deterministic. As in Section 3, $\{p_t - p_{t-1}\}$ depends only on current and past values of a scalar noise πw_t.

To relate this model to the analysis in Section 2, we assume that an econometrician observes a process $\{y_t\}$ that satisfies A1 and A2. As in Section 3, the first component of y_t is $s_t - s_{t-1}$ and the second component is $\gamma(L)r_t$. While the econometrician is assumed to observe the payout $\{p_t\}$, observations on the vector process $\{c_t\}$ of components are not used in the analysis.[15] As in Section 3, we let

$$s_t - s_{t-1} = \sigma(L)w_t \ . \tag{4.24}$$

In light of (4.22) and (4.23), we impose the following restriction on σ:

Restriction R5: $\sigma(z) = (1-z)\rho(z)/\gamma(z) - \psi(z)\pi$ for some scalar $\psi(z)$ satisfying $\psi(z) = \sum_{j=0}^{\infty}\psi_j z^j$, $\sum_{j=0}^{\infty}|\psi|^2 < \infty$ and $\psi(\lambda) = 1$ and for some row vector π.

Notice that R5 is weaker than R3 used in Section 3. Restrictions R1 $[\sigma(\lambda) = 0]$ and R5 together imply that

$$\pi = (1-\lambda)\rho(\lambda)/\gamma(\lambda) \ . \tag{4.25}$$

We now investigate whether (4.25) is a testable restriction. To address this question, it is convenient to use the transformations given in (3.10) and to partition σ^+ and ρ^+ in the same manner. This results in

$$\begin{aligned} s_t - s_{t-1} &= \sigma_1^+(L)w_{1t}^+ + \sigma_2^+(L)w_{2t}^+ \\ \gamma(L)r_t &= \rho_1^+(L)w_{1t}^+ + \rho_2^+(L)w_{2t}^+ \\ p_t - p_{t-1} &= |\pi|\psi(L)w_{1t}^+ \ . \end{aligned} \tag{4.26}$$

Restriction (4.25) implies that

$$(1-\lambda)\rho_1^+(\lambda)/\gamma(\lambda) = |\pi| \tag{4.27}$$

and, by construction,

$$\rho_2^+(\lambda) = 0 \ . \tag{4.28}$$

First, we investigate whether (4.27) is testable. To address this question we decompose y_t as

$$y_t = y_{1t} + y_{2t} \tag{4.29}$$

where

$$y_{1t} = C_1^+(L)w_{1t}^+ \quad \text{and} \quad y_{2t} = C_2^+(L)w_{2t}^+ . \tag{4.30}$$

This decomposition can be constructed as follows. Note that linear combinations of current, past and future values of w_{1t}^+ generate the same Hilbert space as linear combinations of current, past and future values of $p_t - p_{t-1}$. Also note that since w_{2t}^+ is uncorrelated with w_{1t}^+, y_{1t} is the least squares regression of y_t onto current, past and future values of w_{1t}^+ and hence onto current, past and future values of $p_t - p_{t-1}$. This means that y_{2t} is the vector of regression errors. The function $\psi(z)$ may have zeros inside the unit circle of the complex plane.[16] As a consequence, linear combinations of current and past values of w_{1t}^+ may generate a strictly larger Hilbert space than linear combinations of current and past values of $p_t - p_{t-1}$. As such, the receipt process may Granger-cause the payout process in this model. For this reason, it is potentially important that future values of $p_t - p_{t-1}$ be included in the regression when forming y_{1t}.

In contrast to Section 3, the scalar noise w_{1t}^+ is not identifiable. Instead, we form a Wold representation of the stochastically singular (rank one) process $\{y_{1t}\}$,

$$y_{1t} = C_1^*(L)w_{1t}^* , \tag{4.31}$$

where $\{w_{1t}^*\}$ is scalar white noise with a unit variance, and where linear combinations of current and past values of w_{1t}^* generate the same Hilbert space as linear combinations of current and past values of y_{1t}.

As in the analysis in Section 2, we know that $\{w_{1t}^*\}$ and $\{w_{1t}^+\}$ must be related via

(4.32)
$$w_{1t}^+ = D_1(L^{-1})w_{1t}^* \quad \text{where } D_1(z) = \sum_{j=0}^{\infty} d_j(z)^j , \ \sum_{j=0}^{\infty} |d_j|^2 < \infty$$

and

$$D_1[\exp(i\theta)]D_1[\exp(-i\theta)] = 1 \quad \text{for almost all } \theta . \tag{4.33}$$

Therefore,

$$C_1^+(z) = C_1^*(z)D_1(z) . \tag{4.34}$$

The function $D_1(z)$ can have zeros inside the unit circle but it cannot have a zero at λ, because under A3, $|\pi|$ is different from zero and $\psi(\lambda)$ is one. Hence it is not possible for the first two entries of $C_1^+(\lambda)$ both to be zero under R5. The same must be true for $C_1^*(z)$. Let $\sigma_1^*(z)$ denote the first row of $C_1^*(z)$. An implication of (4.27) and (4.34) is

Restriction R6: $\sigma_1^*(\lambda) = 0$.

Since $\sigma_1^*(z)$ is identifiable up to a sign, R5 is testable given knowledge of the discount factor λ.

Proposition 6: Suppose that A1–A4 are maintained. Then R1 and R5 imply R6.

While R6 is designed to accommodate nonseparabilities in preferences, in general, neither $B(z)$ nor $\psi(z)$ can be identified. In contrast to Section 3, the noise w_{1t}^+ is not necessarily identified. Nevertheless the theoretical model imposes enough structure via R5 to guarantee that R1 is testable.

Finally, we investigate whether R6 exhausts all of the testable implications. This amounts to assessing the empirical content of (4.28) $[\rho_2^+(\lambda) = 0]$. Here we simply mimic the logic of Section 3 to conclude:

Proposition 7: Suppose that A1–A5 are maintained. If R5 and R6 are satisfied, then it is always possible to find a moving-average representation $y_t = \hat{C}(L)\hat{w}_t$ that satisfies R1.

The first entry of $\hat{w}_t$ can be chosen to be w_{1t}^* and the remaining components constructed as described in Sections 2 and 3.

Recall from (4.21) that the theoretical model studied in this section implies that the consumption goods all depend on a scalar noise process $\{\pi w_t\}$. In the analysis so far, we have assumed that the econometrician uses data only on total consumption $\{p_t - p_{t-1}\}$. An alternative approach is to assume that the components of consumption are measured with error and to model the observed income and consumption components as a dynamic factor model of the sort studied by Geweke (1977b), Sargent and Sims (1977), Geweke and Singleton (1981a, 1981b), Engle and Watson (1981) and Watson and Engle (1983) with a single factor. The present-value budget balance restriction would then apply to the single factor. The analysis in this section can be modified appropriately to apply to such a model.[17]

5. An Empirical Example

In this section we test R6 using aggregate post war US time series on consumption and labor income. We describe in turn the data, the parameterization for the underlying time series, the estimation method, and the empirical results.

5a. Data

We used aggregate data on total consumption and labor income from 1953:2 – 1984:4 supplied to us by Kenneth West. The original source for the labor income series is Blinder and Deaton (1985). West (1988) used consumption of nondurables and services excluding clothing and services. For this series to be comparable to the labor income series, West scaled the consumption series to reflect the fact that his measure of nondurable consumption is only a portion of total consumption. An advantage of proceeding in this fashion would be that preferences for this nondurable consumption good may be modeled plausibly as time separable. The analysis in Section 3 could then be exploited in studying present-value budget balance. We adopted a somewhat different approach. We used data on total consumption and tested restriction R6 derived in Section 4. Recall that the economic model proposed in Section 4 permits the *indirect* preferences for total consumption to be nonseparable over time. For this reason we were not compelled to remove the consumption of durable goods from total consumption.

5b. Parameterization

We modeled the bivariate consumption and income process using a two-factor specification. These factors are identified by assuming that the factors are mutually uncorrelated and that the second factor has no impact on consumption. In Section 4, we showed that a factor decomposition can always be obtained by representing the rank one process $\{y_{1t}\}$ in (4.29) and (4.30) in terms of a scalar time series, say $\{f_{1t}\}$.

For convenience, we model each factor as having an N^{th}-order univariate autoregressive representation:

$$f_{jt} = b_j(L) f_{jt-1} + v_{jt} \tag{5.1}$$

where $[1 - zb_j(z)]$ is an $(N+1)^{\text{th}}$-degree polynomial with zeros that are outside the unit circle of the complex plane. Total consumption is then modeled as a distributed lag of $\{f_{1t}\}$ and labor income as distributed

lags of both $\{f_{1t}\}$ and $\{f_{2t}\}$:

$$\begin{bmatrix} p_t - p_{t-1} \\ r_t - r_{t-1} \end{bmatrix} = \begin{bmatrix} a_{11}(L) & 0 \\ a_{21}(L) & a_{22}(L) \end{bmatrix} \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} \tag{5.2}$$

where $a_{ij}(z)$ is a N^{th}-degree polynomial. The shock process $\{v_t\}$ is serially uncorrelated with $E(v_t v_t') = I$.

To relate this setup to the analysis in Section 4, it is convenient to deduce the implied moving-average average representation for $\{y_t\}$ in terms of $\{v_t\}$. It follows from (5.1) and (5.2) that

(5.3)
$$\begin{aligned} p_t - p_{t-1} &= \big\{a_{11}(L)/[1 - Lb_1(L)]\big\} v_{1t} \\ r_t - r_{t-1} &= \big\{a_{21}(L)/[1 - Lb_1(L)]\big\} v_{1t} + \big\{a_{22}(L)/[1 - Lb_2(L)]\big\} v_{2t} \,. \end{aligned}$$

Subtracting the first equation in (5.3) from the second gives:

(5.4)
$$s_t - s_{t-1} = \big\{[a_{11}(L) - a_{21}(L)]/[1 - Lb_1(L)]\big\} v_{1t} + \big\{a_{22}(L)/[1 - Lb_2(L)]\big\} v_{2t} .$$

The moving-average representation given by (5.3) and (5.4) coincides with (4.26) when $v_{jt} = w_{jt}^+$. In this case

$$\begin{aligned} \sigma_1^+(z) &= [a_{11}(z) - a_{21}(z)]/[1 - zb_1(z)] \\ \rho_1^+(z) &= a_{21}(z)/[1 - zb_1(z)] \\ \sigma_2^+(z) &= \rho_2^+(z) = a_{22}(z)/[1 - zb_2(z)] \end{aligned} \tag{5.5}$$

Also $\psi(z)$ is proportional to $a_{11}(z)[1 - z\, b_1(z)]$. Recall that the time nonseparabilities in the induced preference ordering for $\{p_t\}$ are reflected in $\psi(z)$. Present-value budget balance implies that $\sigma_j^+(\lambda) = 0$ for $j = 1, 2$ which, in this case, is equivalent to

$$a_{11}(\lambda) = a_{21}(\lambda) \tag{5.6}$$

and

$$a_{22}(\lambda) = 0 \,. \tag{5.7}$$

For the reasons given in Section 4, it is not necessarily true that $v_{jt} = w_{jt}^+$. Furthermore, the shocks w_{jt}^+ are not identifiable without further restrictions. Recall from Section 4 that the inability to identify w_{2t}^{+} makes a test based on (5.7) uninteresting. Relation (5.6) is, however, potentially testable because the fundamental shock process $\{w_{1t}^{*}\}$ for $\{y_{1t}\}$ is identifiable. Furthermore, as long as $a_{11}(z)$ and $a_{21}(z)$ do not have common zeros, w_{1t}^{*} is equal to v_{1t} up to a sign convention. In this case (5.6) is equivalent to the Restriction R6. Thus we make (5.6) the focus point of our empirical analysis.

5c. Estimation Method

We tested (5.6) using the method of maximum likelihood assuming a Gaussian likelihood function. In evaluating the likelihood function, the time t components of both series were scaled by μ^{-t} where the growth rate $\log(\mu)$ was estimated. First-differences of the resulting series were taken and sample means removed. The scaled series were modeled as stationary processes using the two factor specification just described. We used a transformation suggested by Monahan (1984) to ensure that $b_j(z)$ has zeros outside the unit circle of the complex plane for $i = 1, 2$. The likelihood for the stationary model was evaluated using recursive state-space methods as described in Chapter 3. An extra (Jacobian) factor was included in the likelihood function to accommodate the scaling of the original time series.

5d. Empirical Results

The unrestricted model was fit for $N = 1, 2, 3$. The log-likelihood values for these cases are reported in Table 1.

Table 1: Log-Likelihood Values for the Unrestricted Parameterization

N	log-likelihood value
1	−897.21
2	−888.58
3	−878.92

Increasing N by one introduces five new parameters into the model. As seen in Table 1, the changes in the values of log-likelihood function are substantial from the vantage point of the likelihood ratio statistic. On the other hand, from the vantage point of the Schwarz (1978) criterion for model selection, these changes are not substantial. The Schwarz criterion was developed for a different estimation environment and is known to be conservative. Under a nonparameteric perspective in which a finite-dimensional parameterization is viewed as an approximation to an infinite parameter model, it is not apparent that

model selection based on either classical likelihood ratio inference or the Schwarz criterion can be justified. We did not explore larger values of N for reasons of numerical tractability and also because the analysis in the next section suggests that restriction (5.6) has very little content for generous parameterizations.

For each N, Table 2 gives the fraction of variation of each series that is explained by the first factor. Recall that, by construction, this fraction is one for consumption.

Table 2: Fraction of Variation Explained by the First Factor for the Unrestricted Parameterization

N	net surplus	labor income
1	.17	.47
2	.22	.53
3	.25	.51

Notice that for each of the parameterizations, about half of the variation in labor income can be attributed to the first factor. It is this variation that forms the basis for our test of present-value budget balance.

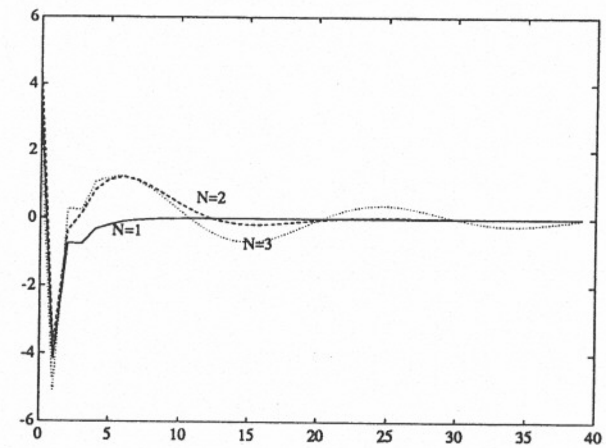

Figure 1. Impulse Responses of Net Surplus in the Unrestricted Models, N = 1, 2, 3.

Figure 1 gives the estimated impulse response function for the first difference of the net surplus process to a surprise movement in v_{1t}

for each of the three specifications of N. In other words, this figure reports the estimated coefficients of the power series expansion of $[a_{11}(z) - a_{21}(z)]/[1 - zb_1(z)]$. The pattern of the response is similar across the three different specifications. In all cases the original increase in the net surplus is more than offset after one time period. Figure 2 gives the estimated impulse response function for the first-difference in consumption and labor income separately for the $N = 3$ run. The peak responses for consumption and labor income both occur after one time period. The nontrival response of consumption to the first shock suggests that it is important to allow for nonseparablities in preferences. In fact, the coefficients of the power series expansion of $\psi(z)$ are proportional to the impulse response function for $\{p_t - p_{t-1}\}$. The similar response patterns of consumption and income might be taken to suggest a trivial model in which consumption and income are the same. Such an interpretation would be misleading, however, because the first shock accounts for all of the variability in consumption but only half of the variability in income.

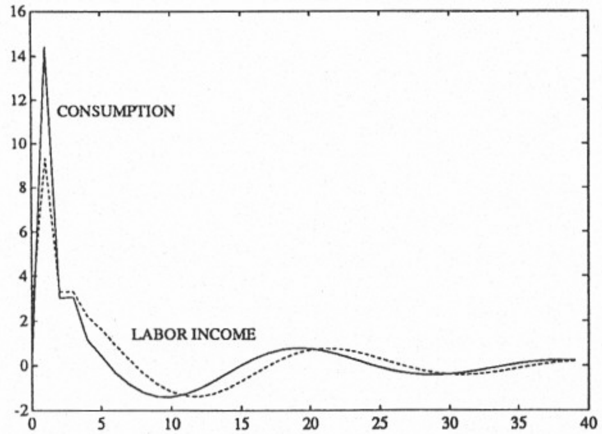

Figure 2. Impulse Response of Consumption and Labor Income in the Unrestricted N = 3 Model.

Restricted versions of the model were also estimated for $N = 1, 2, 3$. For these versions, the $a_{11}(z)$ polynomial was parameterized as follows. First we constructed the function

$$\psi(z) = \phi(z)/[1 - zb_1(z)] \tag{5.8}$$

where $\phi(z)$ is an N^{th}-degree polynomial. In order that $\psi(\lambda) = 1$, we required that

$$\phi(\lambda) = 1 - \lambda b_1(\lambda) . \tag{5.9}$$

The polynomial ϕ was parameterized as

$$\phi(z) = \sum_{j=0}^{N} \phi_j z^j , \tag{5.10}$$

where the ϕ_j's were treated as free parameters for $j = 1, 2, \ldots, N$ and ϕ_0 was chosen so as to satisfy (5.9). Finally, the polynomial $a_{11}(z)$ was constructed as

$$a_{11}(z) = a_{21}(\lambda)\phi(z)/[1 - \lambda b_1(\lambda)] . \tag{5.11}$$

Note that (5.6) is satisfied by construction, and that there are only N underlying parameters of $a_{11}(z)$, whereas in the unrestricted estimation there were $N + 1$ such parameters.

We estimated the restricted model for five different values of $\log(\lambda)$ ranging from $-.005$ to $-.025$.[18] The values of these likelihoods are reported in Table 3:

Table 3: Log-Likelihood Values for the Restricted Parameterization

			$-\log\lambda$		
N	.005	.010	.015	.020	.025
1	−897.76	−897.73	−897.72	−897.71	−897.69
2	−889.69	−889.69	−889.69	−889.68	−889.68
3	−879.02	−879.02	−879.02	−879.02	−879.02

As is evident from Table 3, the restricted log-likelihood function is not very sensitive to the pre-specified choice of $\log(\lambda)$. Comparing the restricted likelihood values in Table 3 to the unrestricted likelihood values in Table 4, the evidence against the present-value-budget-balance restriction is very weak. In estimating the unconstrained ($N = 3$) model we were unable to find likelihood values that showed any appreciable improvement over the constrained value. Notice that for $N = 3$ the unconstrained likelihood value is only .10 higher than the constrained value. Thus we find that evidence against permanent-income-type models cannot be attributed to violation of present value budget balance.

Table 4 reports the fraction of variance explained by the first factor. Since these numbers are insensitive to the choice of $\log(\lambda)$, only numbers for $\log(\lambda) = -.015$ are reported.

Table 4: Fraction of Variation Explained by the First Factor for the Restricted Parameterization ($N = 3$)

N	net surplus	labor income
1	.16	.52
2	.22	.53
3	.19	.51

Not surprisingly these results are very similar to those reported in Table 2.

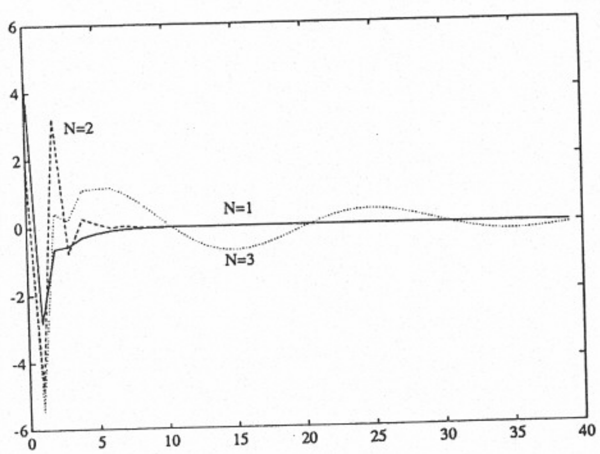

Figure 3. Impulse Responses of Net Surplus in the Restricted Models, N = 1, 2, 3.

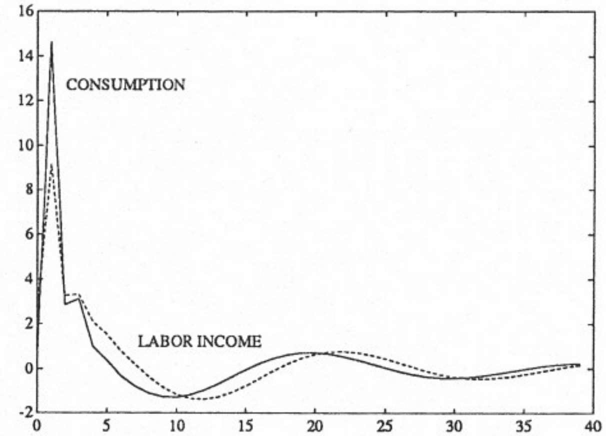

Figure 4. Impulse Responses of Consumption and Labor Income in the Restricted N = 3 Model.

Figure 3 gives the impulse response functions for net surplus and Figure 4 gives the impulse response functions for consumption and labor income for the restricted $[\log(\lambda) = -.015]$ version of the model with $N = 3$. These figures are the restricted counterparts to Figures 1 and 2 respectively. Figures 2 and 4 are remarkably similar for the reasons described previously.

6. Approximation Error

In this section we investigate the effect of model approximation error when testing restriction R6. For simplicity, we focus on the case in which $\{y_t\}$ is two-dimensional. Recall that the payout series is a linear combination of the net surplus series and the receipt series. In this section, we find it convenient to make $(1-L)p_t$ rather than $(1-L)s_t$ the first component of y_t. The moving-average representation for $\{y_t\}$ is denoted:

$$(6.1) \qquad y_t = \begin{bmatrix} p_t - p_{t-1} \\ r_t - r_{t-1} \end{bmatrix} = C(L)w_t^+$$

Consistent with A3 and R5 for $\gamma(z) = 1 - z$, we assume that $C(z)$ is lower triangular:

$$(6.2) \qquad C(z) = \begin{bmatrix} \pi_1(z) & 0 \\ \rho_1(z) & \rho_2(z) \end{bmatrix} .$$

We impose an additional assumption that accommodates a rich collection of moving-average representations:

Assumption A5: $(1/2\pi)\int_{-\pi}^{\pi}\left\{|\rho_1[\exp(-i\theta)]|^2/|\rho_2[\exp(-i\theta)]|^2\right\}d\theta < \infty.$

In contrast to the previous sections, we assume that present-value budget balance is violated for the first component of the shock process $\{w_t^+\}$:

Assumption A6: $\pi_1(\lambda) \neq \rho_1(\lambda)$.

Our goal is to construct a sequence of approximating models that converge, in some precise sense, to the true model such that each approximating model satisfies R6. If we can construct such an approximating sequence, it will seriously undermine the testability of R6 when (even small) approximation errors are present.

The approximating models we consider all have the following structure. The $C(z)$ matrix function is replaced by an approximating function that is also lower triangular:

$$C^a(z) = \begin{bmatrix} \pi_1(z)b(z) & 0 \\ \rho_1(z)b^*(z) & \rho_2(z) \end{bmatrix} \tag{6.3}$$

where

$$b(z) = (z-\zeta)/(1-z\zeta) \quad \text{and} \quad b^*(z) = (z-\zeta^*)/(1-z\zeta^*) \tag{6.4}$$

for real numbers ζ and ζ^* in the interval $(0,1)$. Consistent with the approximation criterion that is implicit in maximum likelihood estimation, we use an information measure of the magnitude of the approximation error. Let S^a be the spectral density of the approximating model and let S be the spectral density of the true model. Our measure of approximation error uses the population counterpart to the Gaussian log-likelihood function and has a frequency domain representation:

$$\begin{aligned} \eta(S^a,\, S) = {}& (1/2\pi)\int_{-\pi}^{\pi}\left\{\log\,\det[S(\theta)] - \log\,\det[S^a(\theta)]\right\}d\theta \\ & + (1/2\pi)\int_{-\pi}^{\pi}\, \mathrm{trace}\,[I - S(\theta)S^a(\theta)^{-1}]d\theta\,. \end{aligned} \tag{6.5}$$

Notice that $\eta(S,S)$ is zero. In addition, it can be shown that $\eta(S^a,S)$ is strictly positive unless $S^a(\theta)$ is equal to $S(\theta)$ for almost all θ.

Given an approximating model satisfying (6.3), representation (6.5) simplifies considerably. Recall that

$$\begin{aligned} S(\theta) &= C[\exp(-i\theta)]\; C[\exp(i\theta)]' \\ &= \begin{bmatrix} |\pi_1(z)|^2 & \rho_1(z^{-1})\pi_1(z) \\ \rho_1(z)\pi_1(z^{-1}) & |\rho_1(z)|^2 + |\rho_2(z)|^2 \end{bmatrix} \end{aligned} \tag{6.6}$$

for $z = \exp(-i\theta)$. Similarly,

$$\begin{aligned} S^a(\theta) &= C^a[\exp(-i\theta)]\; C^a[\exp(i\theta)]' \\ &= \begin{bmatrix} |\pi_1(z)|^2 & \rho_1(z^{-1})\pi_1(z)b^*(z^{-1})b(z) \\ \rho_1(z)\pi_1(z^{-1})b^*(z)b(z^{-1}) & |\rho_1(z)|^2 + |\rho_2(z)|^2 \end{bmatrix} \end{aligned} \tag{6.7}$$

since $|b(z)|^2 = |b^*(z)|^2 = 1$ for all $|z| = 1$. Note that the matrices $S(\theta)$ and $S^a(\theta)$ have the same determinants and that

$$\begin{aligned} &\mathrm{trace}\,[I - S(\theta)S^a(\theta)^{-1}] = \\ &\qquad |\rho_1(z)\pi_1(z^{-1})|^2\; [1 - b^*(z)b(z^{-1})]/\det[S(\theta)] \; + \\ &\qquad |\rho_1(z)\pi_1(z^{-1})|^2\; [1 - b^*(z^{-1})b(z)]/\det[S(\theta)] \\ &= |1 - b^*(z)b(z^{-1})|^2\; |\rho_1(z)|^2/|\rho_2(z)|^2 \end{aligned} \tag{6.8}$$

for $z = \exp(-i\theta)$. Substituting into (6.5), it follows that the magnitude of the approximation error is

$$\begin{aligned} &\eta(S^a,\, S) = \\ &(1/2\pi)\int_{-\pi}^{\pi}\Big(|1 - \left\{b^*[\exp(-i\theta)]\, b[\exp(i\theta)]\right\}|^2 \\ &\qquad |\rho_1[\exp(-i\theta)]|^2/|\rho_2[\exp(-i\theta)]|^2\Big)d\theta\,. \end{aligned} \tag{6.9}$$

Formula (6.9) is very similar in structure to one derived in Sims (1972a) for least squares estimation with strictly exogenous regressors. Notice that the magnitude of the approximation error is a weighted integral of the squared error in approximating 1 by the product $b^*\,[\exp(-i\theta)]$ $b[\exp(i\theta)]$. In light of Assumption A5, the density that weights this squared error induces a finite measure on $(-\pi,\, \pi]$. Among other things, this assumption guarantees that the approximation error is finite for members of our class of approximating models.

We now construct a sequence of approximating models $\{C_j\}$ for which $\{\eta(C_j, C)\}$ converges to zero. The matrix C_j is given by the right side of (6.3) with b_j replacing b and b_j^* replacing b^* where

$$b_j(z) = (z - \zeta_j)/(1 - z\zeta_j) \text{ and } b_j^*(z) = (z - \zeta_j^*)/(1 - z\zeta_j^*) \tag{6.10}$$

for sequences of real numbers $\{\zeta_j\}$ and $\{\zeta_j^*\}$ in the interval $(0,1)$. For each j the pair (ζ_j, ζ_j^*) is chosen so that the present-value budget balance restriction R6 is satisfied. We focus on the special case in which $\pi_1(\lambda)$ and $\rho_1(\lambda)$ are both different from zero. Arguments for the other cases proceed in a similar but slightly more complicated fashion.[19]

Let $\{\zeta_j\}$ be a sequence of real numbers in the interval $(0,1)$ that are distinct from λ but that converge to λ. Note that $\pi_1(\lambda)\, b_j(\lambda)$ is different from zero but that the sequence $\{\pi_1(\lambda)\, b_j(\lambda)\}$ converges to zero. Next choose ζ_j^* so that

$$\pi_1(\lambda)\, b_j(\lambda) = \rho_1(\lambda)\, b_j^*(\lambda) \tag{6.11}$$

It follows from the analysis in Section 4 that (6.11) implies that restriction R6 is satisfied for each of the approximating models. Viewing (6.11) as an equation in ζ_j^*, it follows that ζ_j^* must satisfy

$$\pi_1(\lambda)\, b_j(\lambda)\, (1 - \lambda\zeta_j^*) = \rho_1(\lambda)\, (\lambda - \zeta_j^*)\ . \tag{6.12}$$

For sufficiently large j, equation (6.12) determines ζ_j^* uniquely because $\{b_j(\lambda)\}$ converges to zero and $\rho_1(\lambda)$ is different from zero. Solving for ζ_j^* gives

$$\zeta_j^* = [\pi_1(\lambda)\, b_j(\lambda) - \rho_1(\lambda)\lambda]/[\lambda\pi_1(\lambda)\, b_j(\lambda) - \rho_1(\lambda)] \tag{6.13}$$

Since $\{b_j(\lambda)\}$ converges to zero, $\{\zeta_j^*\}$ converges to λ. By omitting some of the initial entries of the sequence of $\{\zeta_j\}$ and hence also of $\{\zeta_j^*\}$, we can guarantee that all entries of both sequences are in the interval $(0,1)$.

It remains to show that the approximation error $\eta(C_j, C)$ can be made arbitrarily small. Since $\{\zeta_j\}$ and $\{\zeta_j^*\}$ both converge to λ, $\{b_j^*(z)\, b_j(z^{-1})\}$ converges to one for all $|z| = 1$. Furthermore,

$$\begin{gathered} |b_j^*(z)|\ |b_j(z^{-1})| \le \\ [(1+\zeta_j)(1+\zeta_j^*)]/[(1-\zeta_j)(1-\zeta_j^*)] \quad \text{for} \quad |z| = 1\ . \end{gathered} \tag{6.14}$$

The inequality permits us to apply the Dominated Convergence Theorem and conclude that $\{\eta(S, S_j)\}$ converges to zero. We have established:

Proposition 8: Suppose Assumptions A1–A6 are satisfied. Then there exists a sequence $\{S_j\}$ of spectral density functions satisfying R6 such that $\{\eta(S_j, S)\}$ converges to zero.

In light of Proposition 8, the empirical content of Restriction R6 is tenuous. Even when the restriction is not satisfied, it is possible to construct a sequence of models, each of which satisfies R6, that approximate the original model arbitrarily well. These approximating models have the feature that the corresponding sequences $\{\rho_1^j\}$ and $\{\pi_1^j\}$ have zeros that are arbitrarily close to λ. To avoid the negative implications of Proposition 8, one must be able to rule out zeros of these functions in a neighborhood of λ.[20] Recall that in Section 5 we assumed that $\rho_1(z)$ and $\pi_1(z)$ are ratios of finite-order polynomials. If the orders of these polynomials are chosen to be too large, then a finite-parameter version of Proposition 7 holds. Of course it is very difficult in practice to determine the precise order of numerator and denominator polynomials. Furthermore, it is often the case that the parameterizations used in estimation are best thought of as approximations to models with more complicated time series correlation structure. Hence, it seems hard to circumvent the negative conclusion of Proposition 8.

7. Variable Real Interest Rates

In this section, we permit uncertainty in the real interest rate δ of (2.1). We rewrite (2.1) as

$$k_{t+1} = (1+\delta_t)\, k_t + s_t \quad \text{for} \quad t = 0, 1, 2, \dots\ . \tag{7.1}$$

Let J_{t-1} denote information available to agents at time $t-1$. Throughout this section, we require

$$E\,(\delta_t \mid J_{t-1}) = \delta\ , \tag{7.2}$$

so that the conditional expectation of interest rates is constant over time. Since $\{\delta_t\}$ is no longer deterministic, the martingale implications described in Sections 2 and 3 may no longer be applicable.

To deduce testable restrictions, we take expectations of (7.1) conditioned on time $t-1$ information. This yields

$$E(k_{t+1} \mid J_{t-1}) = (1+\delta)\, k_t + E(s_t \mid J_{t-1})\ ,\ t = 0, 1, \dots\ , \tag{7.3}$$

with k_0 given, since k_t is determined at time $t-1$. We impose assumption A1 and terminal condition

$$E \sum_{t=0}^{\infty} \lambda^t\, (k_t)^2 < +\infty\ ,$$

where $\lambda = 1/(1+\delta)$. Solving (7.3) forward gives

$$k_t = -\sum_{j=0}^{\infty} \lambda^{j+1} E(s_{t+j} \mid J_{t-1}) \quad , \quad t = 0, 1, \ldots . \tag{7.4}$$

For reasons similar to those given in Section 2, (7.4) imposes no testable restrictions on the joint process $\{(r_t, p_t)\}$ beyond those described in Section 2. However, (7.4) does have interesting implications for the question studied by Hamilton and Flavin (1986), namely the restrictions imposed by present value budget balance on the joint process $\{(k_{t+1}, r_t, p_t)\}$. In fact (7.4) is a version of the class of present value models investigated by Campbell and Shiller (1987).

We continue to assume that information available to agents has the structure described in Section 2. In particular, recall that s_t is the first component of an m-dimensional vector process y_t that has representation

$$y_t = C(L)\, w_t \,,$$

where w_t is the vector of innovations to agents' information. The first row of the above system of equations is

$$s_t = \sigma(L)\, w_t \,.$$

Using the above formula together with results of Hansen and Sargent (1980a), permits (7.4) to be represented as

$$k_{t+1} = \kappa(L) w_t \ \text{ where } \ \kappa(z) = -\left[\frac{\lambda\sigma(z) - \lambda\sigma(\lambda)}{(z-\lambda)}\right] . \tag{7.5}$$

Equation (7.5) translates directly into a restriction on the moving-average representation for $\{(k_{t+1}, y_t')\}$ in terms of $\{w_t\}$.

So long as the dimension n of w_t is greater than the dimension m of y_t, a composite process $\{(k_{t+1}, y_t')\}$ satisfying (7.5) can be stochastically nonsingular. Stochastic nonsingularity can emerge when k_{t+1} cannot be expressed as a linear function of current, past and future values of y_t. In this circumstance, k_{t+1} can reveal additional information about w_t. Indeed, $\{(k_{t+1}, y_t'), (k_t, y_{t-1}') \ldots\}$ could well generate a smaller information set than does $\{w_t, w_{t-1}, \ldots\}$.[21]

Hamilton and Flavin (1986) formulated and tested a version of (7.5) that emerges under the special assumption that s_t is a univariate process with an innovation that reveals w_t, so that s_t is Granger caused by no other variables in $\{(k_{t+1}, y_t')\}$. Under this special assumption,

restriction (7.5) implies that the $\{(k_{t+1}, y_t')\}$ process is stochastically singular and that the regression of k_{t+1} on current and lagged s_t fit by Hamilton and Flavin will have an R^2 of unity. If one were to drop the assumption that s_t is Granger caused by no other components of $\{(k_{t+1}, y_t')\}$, a serially correlated *Shiller error* would emerge in the relation fit by Hamilton and Flavin. The presence of that error would require resorting to estimators of the class described by Hansen and Sargent (1982) and Hayashi and Sims (1983) in order to obtain consistent parameter estimates and valid test statistics.

Further insights about efficient estimation and about the structure of the restrictions are gained by noting that the model for $\{(k_{t+1}, y_t')\}$ induced by (7.5) is a member of a class of *exact linear rational expectations models* studied by Hansen and Sargent in Chapter 3. The results in Chapter 3 imply that the restrictions given by (7.5) always apply to a Wold moving-average representation for $\{(k_{t+1}, y_t')\}$.[22] Furthermore, even if the moving-average representation in terms of w_t turns out to be a Wold representation, it is always possible to find some other moving-average representation with the same number of noises that satisfies the restrictions and is not a Wold representation. Thus, unlike the restrictions studied in Section 2, the restrictions given in (7.5) apply to both the Wold moving-average representation and to some other moving-average representations. Evidently, the restrictions (7.5) have a different structure from those studied in Section 2.

Finally, it is straightforward to extend the analysis in this section to accommodate unit roots in the payoff and receipt processes along the lines described in Section 2. The asset process $\{k_t : t = 0, 1, \ldots\}$ will inherit the nonstationarity; however, (7.5) implies that the process $\{k_{t+1}, r_t, p_t) : t = 0, 1, \ldots\}$ will be co-integrated in the sense of Engle and Granger (1987) (e.g. see Chapter 3 and Campbell and Shiller 1987).

8. Conclusions

In Section 2, we supplemented the hypothesis of present-value budget balance with two more hypotheses, the existence of a time invariant moving-average representation for the process $\{(r_t, p_t)\}$ and constancy of the net rate of interest. We found that present-value budget balance is so weak an hypothesis that it imposes literally no observable restrictions on the process $\{(r_t, p_t)\}$. This weakness illustrates how difficult it is to verify a present-value budget constraint that restricts the entire (infinite-dimensional) future of the $\{(r_t, p_t)\}$ process. The restriction acquires content only if the hypothesis of a time-invariant linear representation is supplemented with additional hypotheses that reduce the parameter space in ways that render inadmissible the transformations that we exhibited in Section 2. We showed that the present-value budget balance restriction is delicate in the sense that it cannot hold for a Wold moving-average representation, which is the representation that is typically recovered using vector autoregression methods.

In Sections 3 and 4, we showed that the present value budget balance restriction acquires additional content when it is part of a class of linear-quadratic models of optimal consumption (or optimal tax collections). In those models, present value budget balance imposes an additional restriction on the $\{(r_t, p_t)\}$ process above and beyond those imposed by martingale characterizations stemming from Euler equations. A sense in which even these restrictions are tenuous was described in Section 6.

Section 7 described the restrictions on the joint $\{(k_{t+1}, p_t, r_t)\}$ process imposed by the assumption that $E_{t-1}\delta_t = \delta$ for all t. It turns out that those restrictions deliver an *exact linear rational expectations model*. In the next paper in this volume (Chapter 6) Roberds estimates such a model for post war U.S. time series on government debt and deficits.

Appendix A

In this appendix we show that $D(z)$ given in (2.20) can be used to build a Wold representation for the composite process $\{x_t\}$. Use the Wold representation for $\{y_t\}$, namely,

$$y_t = C^*(L)w_t^* \tag{A.1}$$

to form a two-sided moving-average representation for $\{x_t\}$ by adding a row

$$k_t = \kappa^*(L)w_t^* , \qquad \text{where} \quad \kappa^*(z) \equiv -\lambda z\sigma^*(z)/(z-\lambda) , \tag{A.2}$$

and $\sigma^*(z)$ is the first row of $C^*(z)$. Use $D(z)$ given in (2.20) to form an alternative moving-average representation for $\{x_t\}$:

$$y_t = C^+(L)w_t^+ , \quad k_t = \kappa^+(L)w_t^+ \tag{A.3}$$

where

$$C^+(z) \equiv C^*(z)D(z) \quad \text{and} \quad \kappa^+(z) \equiv \kappa^*(z)D(z) . \tag{A.4}$$

Note that

$$\begin{aligned} \kappa^+(z) &= -\lambda z\sigma^*(z)D(z)/(z-\lambda) \\ &= -\lambda z\sigma^*(z)\, Q \begin{bmatrix} 1/(1-\lambda z) & 0 \\ 0 & [1/(z-\lambda)]I \end{bmatrix} . \end{aligned} \tag{A.5}$$

Since $\sigma^*(\lambda)Q$ has zeroes in all entries except the first, $\kappa^+(z)$ has a removable singularity at $z=\lambda$, implying that $\kappa^+(L)$ is one-sided. Therefore, $(A.3)$ gives a one-sided moving-average representation for $\{x_t\}$.

It remains to show that representation $(A.3)$ is a Wold representation. Suppose to the contrary that the rank of the matrix $\begin{bmatrix} C^+(\lambda) \\ \kappa^+(\lambda) \end{bmatrix}$ is m-1. In this case there exists an orthogonal matrix Q^+ such that the first column of $\begin{bmatrix} C^+(\lambda) \\ \kappa^+(\lambda) \end{bmatrix} Q^+$ contains all zeroes. Hence the matrix function

$$\begin{bmatrix} \tilde{C}(z) \\ \tilde{\kappa}(z) \end{bmatrix} \equiv \begin{bmatrix} C^+(\lambda) \\ \kappa^+(\lambda) \end{bmatrix} Q^+ \begin{bmatrix} (1-\lambda z)/(z-\lambda) & 0 \\ 0 & I \end{bmatrix} \tag{A.6}$$

has a removable singularity at $z=\lambda$, and the composite process $\{x_t\}$ has a one-sided moving-average representation:

$$y_t = \tilde{C}(L)\tilde{w}_t , \quad k_t = \tilde{\kappa}(L)\tilde{w}_t \tag{A.7}$$

for some vector white noise $\{\tilde{w}_t\}$. Consequently, the first row of $\tilde{C}(\lambda)$ satisfies R1. However,

$$\tilde{C}(z) = C^*(z)Q \begin{bmatrix} (z-\lambda)/(1-\lambda z) & 0 \\ 0 & I \end{bmatrix} Q^+ \begin{bmatrix} (1-\lambda z)/(z-\lambda) & 0 \\ 0 & I \end{bmatrix} \tag{A.8}$$

implying that $\det[C^*(z)] = \det[\tilde{C}(z)]$. Thus $\tilde{C}(z)$ also satisfies R2 which contradicts Proposition 2. This proves that $(A.3)$ is a Wold representation for $\{x_t\}$.

Notes

1. The recent paper of Trehan and Walsh (1988) studied these questions as well as tax smoothing.

2. More formally, the notation J_t will be used both to denote the sigma algebra generated by w_t, w_{t-1}, ... and the space of all random variables with finite second moments that are measurable with respect to this sigma algebra.

3. For definitions of concepts from the theory of complex variables such as Laurent series, Taylor series, poles and removable singularities see Churchill, Brown and Verhey (1974). For applications of these concepts in macroeconomics, see Hansen and Sargent (1980a), Whiteman (1983) and Sargent (1987b).

4. Sargent (1987a) did not show the sense in which the restriction is vacuous and did not fully explore the implication that the restriction does not apply to a Wold representation.

5. Recall that the white noise sequence in a Wold representation is not necessarily a martingale difference sequence but instead satisfies the weaker requirement of being serially uncorrelated. We do not investigate the implications of the stronger restriction that the vector white noise be a martingale difference sequence.

6. While A2 is essential to the proof of Proposition 2, it is adopted only as a matter of convenience for Proposition 3. We leave it to the reader to show that Proposition 3 holds when only A1 is maintained. Also, several examples in Sargent (1987b, Ch. XIII) purport to show instances in which restriction R1 is testable. These examples all hinge on maintaining a low-dimensional parameterization of C. The reader can verify that these restrictions vanish when the dimensionality of the parameterizations is expanded.

7. The language *receipt* and *payouts* is inspired by the permanent income model of consumption, but is perverse when applied to Barro's tax smoothing model. (See (3.1)–(3.2) in the text.) In Barro's model, what we call *receipts* corresponds to government expenditures, while what we call *payouts* corresponds to total tax collections.

8. In the spirit of this paper, issues that occur when $\beta(z)$ is approximated by ratios of polynomials potentially are relevant to our analysis. Sims (1972a) demonstrated that objects such as sums of lag coefficients are particularly sensitive to errors in approximating $\beta(z)$. The same observation, however, does not apply to approximating $\beta(z)$ for $|z| < 1$.

9. As pointed out to us by Chi Wa Yuen, if λ is not known, then R4 can be used to help identify λ. In general, R4 can only be used to identify λ locally and not globally because the function $(1 - z)\beta(z) - \gamma(z)$ can have multiple zeros in (0,1). A finding of no zeroes in this region constitutes evidence against R4. However, the testable implications of R4 are considerably weaker and the resulting statistical tests harder to implement when λ is not known *a priori*.

10. In Barro's (1979) model, the martingale implication might be construed as applying to tax rates instead of taxes. Let τ_t denote the tax rate. In this case the present value budget balance restriction could be tested by regressing the surplus process onto the current and past values of $(\tau_t - \tau_{t-1})/\tau_{t-1}$ and checking the discounted sum of coefficients.

11. Restriction R4 is absent in the permanent income model studied by Campbell (1987) because k_{t+1} in Campbell's model is not restricted to be in J_t. Instead k_{t+1} also depends on a shock labeled *unanticipated capital gains* that is only observed by economic agents as of time $t + 1$.

12. West (1988) correctly pointed out that obtaining a more powerful test may not be essential in the case of the permanent income model of consumption because evidence using his test is sufficient to challenge the validity of the model.

13. Hansen (1987) showed that specifications in which there are more services than goods can often be converted into specifications in which the number of goods and services are the same without changing the optimal decision rules for consumption. Hence the restriction in A4 that $B(z)$ be a square matrix can often be relaxed. The further restriction that $B(z)$ be nonsingular rules out cases in which *stable* service processes may require *unstable* consumption processes as in, for example, the rational addiction model of Becker and Murphy (1988).

14. Heaton (1989) deduced a continuous time counterpart in a model with a single consumption good but a general specification of the temporal nonseparabilities in preferences.

15. Wilcox (1989) noted that one of the difficulties in NIPA measures of consumption is splitting retail sales into durable and nondurable components. Our model avoids having to make this split.

16. To illustrate that $\psi(z)$ can be zero inside the unit circle of the complex plane, consider the following setup. Suppose that

$$B(z) = \begin{bmatrix} 1 & bz \\ 0 & 1 \end{bmatrix}$$

for $b > 0$ and $\mu' = [1, 1]$. Then

$$\psi(z) = [1 + (1 - bz)(1 - b\lambda)]/[1 + (1 - b\lambda)^2] .$$

It is easy to verify that b can be chosen so that $\psi(z)$ is zero for some $|z| < 1$. For instance, note that

$$\psi(-\lambda) = [2 - (b\lambda)^2]/[1 + (1 - b\lambda)^2] .$$

Then $\psi(-\lambda) = 0$ for $b = \sqrt{2}/\lambda$. An objection to this specification is that in a deterministic steady state, the marginal utility for the first service is negative. This defect can be overcome by premultiplying $B(z)$ by an appropriately chosen orthogonal matrix. This transformation of $B(z)$ does not alter the indirect preferences for consumption goods. In addition, there are initial conditions for the capital stock and a constant endowment sequence for which both consumption goods are strictly positive in a steady state. Such a specification should tolerate at least small amounts of uncertainty in the endowment process and still have the vector of services, consumption goods, and marginal utilities be positive with high probability.

17. An alternative approach suggested by Andy Atkeson is to assume that the indirect preferences for consumption are separable in the first good. This is equivalent to restricting $B(z)$ to have all zeros in the first row and column except for the (1,1) entry. If the first consumption good is measured without error, then no other processes adapted to the information sets of economic agents should Granger-cause the process for the first consumption good. Furthermore, the univariate innovation in this process should reveal a component of the information set of economic agents to which the present value budget balance restriction applies. To test this restriction still requires that the surplus process also be measured without error. However, it is important to remember that if other consumption goods are also measured without error, the model described in Section 4 implies a stochastic singularity for the observed time series tht is likely to be counter-factual.

18. Recall that the time series are scaled by an estimated growth factor as part of the estimation. The discount factor λ applies to the scaled time series. The estimates of the growth rate $\log(\mu)$ were typically around .0075.

19. For instance, if $\pi_1(\lambda)$ has single zero at λ, then the function $b_j(z)$ should have an additional factor of the form $[(z - \lambda_j)(1 - \lambda z)]/[(1 - \lambda_j z)(z - \lambda)]$ where $\{\lambda_j\}$ converges to λ. Among other things, this extra factor is designed so that $\pi_1(z)b_j(z)$ has a removable singularity at $z = \lambda$.

20. For the (low order) parameterizations adopted in Section 5, approximation error does not seem to be the source of our nonrejection because $\pi_1(\lambda)$ and $\rho_1(\lambda)$ both seem not to be zero. For both the unrestricted and the restricted estimates, we computed the zeros of $a_{11}(z)$ and $a_{21}(z)$. The only zeros that were similar in magnitude for $a_{11}(z)$ and $a_{22}(z)$ were complex and had absolute values exceeding unity. There were no common zeros in the vicinity of plausible values of λ.

21. This point is related to those that we discuss under the topic of the first of the two difficulties in our paper "Two Difficulties in Interpreting Vector Autoregressions" (Chapter 4). That paper treats a class of examples related to (7.5) in which it can be taken that $n = m$, and in which current and lagged values of the process (k_{t+1}, y_t) spans a smaller information set than does $\{w_t, w_{t-1}, \ldots\}$.

22. Among other things, Campbell and Shiller (1987) exploit this observation and derive the restrictions implied on a finite-order vector autoregression.