

NBER WORKING PAPER SERIES

IS CONSUMPTION TOO SMOOTH?

John Campbell

Angus Deaton

Working Paper No. 2134

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1987

The research reported here is part of the NBER's research program in Economic Fluctuations. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Is Consumption too Smooth?

ABSTRACT

For thirty years, it has been accepted that consumption is smooth because permanent income is smoother than measured income. This paper considers the evidence for the contrary position, that permanent income is in fact *less* smooth than measured income, so that the smoothness of consumption cannot be straightforwardly explained by permanent income theory. Quarterly first differences of labor income in the United States are well described by an AR(1) with a positive autoregressive parameter. Innovations to such a process are "more than permanent;" there is no deterministic trend to which the series must eventually return, and good or bad fortune in one period can be expected to be at least partially repeated in the next. Changes to permanent income should therefore be greater than the innovations to measured income, and changes in consumption should be more variable than innovations to measured income. In fact, changes in consumption are much less variable than are income innovations. We consider two possible explanations for this paradox, first, that innovations to labor income are in reality much less persistent than appears from an AR(1), and second, that consumers have more information than do econometricians, so that only a fraction of the estimated innovations are actually unexpected by consumers. The univariate time series results are less than decisive, but the balance of the evidence, whether from fitting ARMA models or from examining the spectral density, is more favorable to the view that innovations are persistent than to the opposite view, that there is slow reversion to trend. The information question is taken up within a bivariate model of income and savings that can accommodate the feedback from saving to income that is predicted by the permanent income theory if consumers have superior information. Nevertheless, our results are the same; changes in consumption are typically smaller than those warranted by the change in permanent income. We show that our finding of "excess smoothness" is consistent with the earlier findings of "excess sensitivity" of consumption to income. Our analysis is conducted within a "logarithmic" version of the permanent income hypothesis, a formulation that recognizes that rates of growth of income and saving ratios have greater claim to stationarity than do changes in income and saving flows.

John Campbell
NBER
1050 Massachusetts Avenue
Cambridge, MA 02138

Angus Deaton
Woodrow Wilson School
Princeton University
Princeton, NJ 08544

0. Introduction

For thirty years, it has been accepted that consumption is smooth because permanent income is smoother than measured income. Indeed, the smoothness of consumption is a principal *raison d'être* for the permanent income theory, as every macro textbook carefully explains. This paper considers the evidence for the contrary position, that permanent income is in fact *less* smooth than measured income, so that the smoothness of consumption cannot be straightforwardly explained by permanent income theory. The relationship between permanent and measured income depends on the long-run properties of the stochastic process generating income, and it is always difficult to make inferences about the long-run from typical macroeconomic time-series. Even so, we feel that the weight of the evidence we consider is against the permanent income story as an adequate explanation for the smoothness of consumption. Several different calculations suggest the same conclusion, that innovations in labor income are typically "more than" permanent, in the sense that the expected present discounted value of an innovation in income is greater than the innovation itself.

The paradox that we consider, smooth consumption versus noisy permanent income, was first raised as a possibility by Deaton (1986). In the current paper, we attempt to do two things. The first is to thoroughly examine the univariate time-series properties of labor income to see whether there is a paradox to be explained, or whether the conventional wisdom can be supported. Since the answers are not transparent from the data, we employ a range of different techniques so as to allay fears that our results are an artefact of a particular methodology or choice of functional form. Our second aim is to relate the "excess smoothness" result to the literature on "excess

sensitivity" of consumption to income, a literature that dates from Flavin (1981). Flavin found that, contrary to theoretical predictions, consumption responds to lagged or anticipated changes in labor income. Flavin's results are not in fact inconsistent with the findings of this paper, as we shall show. Following the work of Campbell (1987), we formulate the permanent income theory of consumption so that a single set of restrictions will guarantee both that changes in consumption are orthogonal to previously known information, and that consumption changes are exactly those warranted by the revision to permanent income as perceived by consumers. Taken together, the restrictions can be rejected against the data, as in Campbell's earlier work, and consistently with other results, including those of Flavin. Moreover, we show that the reaction of changes in consumption to *anticipated* changes in income is also the root cause of the failure of consumption to respond sufficiently to *innovations* in income. Since these tests are conducted within a bivariate model of consumption and income, we also have a check that the result on the persistence of income innovations carries through to the bivariate case.

Section 1 is a preliminary one that establishes the context in which the smoothness paradox is examined. Our formulation of the permanent income model is a standard one, but we introduce a log-linearization of the model that in many respects is easier to handle. Most of the relevant time series are more easily transformed to stationarity in logarithmic form, and most atheoretical specifications of consumption functions have found that logarithmic forms tend to fit the data better. The linearization given here is designed to reap the benefits of logarithms without sacrificing the inherently additive structure of the permanent income model. The first section

also gives simple forms of the smoothness result, for both linear and logarithmic formulations. Section 2 is concerned with the univariate time-series analysis of the disposable labor income series created by Blinder and Deaton (1985). We estimate a range of ARIMA processes for income and use them to calculate permanent income. A variance components model suggested by Watson (1986) is also discussed, and we pay a good deal of attention to non-parametric estimation of the long-run properties of the series, measured by the (normalized) spectral density at zero frequency. We argue that the smoothness of consumption cannot plausibly be explained by the presence of slow trend reversion in the univariate income process. Section 3 considers models in which saving and income are represented within a bivariate system that allows for the possibility that consumers have more information about future income than is contained in the history of the income process. We show that consumption is excessively smooth in the bivariate system, just as it is in the univariate case, so that the smoothness cannot readily be attributed to consumers' superior information. This section contains the discussion of the relationship between excess smoothness and excess sensitivity. It also contains a brief discussion of the relationship between our results and those in the literature, as well as concluding remarks.

1. Consumption, permanent income, and innovations

We shall take the following equation as representing the permanent income theory:

$$c_t = \frac{r}{1+r} \left[A_t + \sum_{i=0}^{\infty} (1+r)^{-i} E_t y_{t+i} \right] \quad (1)$$

where c_t is consumption at time t , r is the real rate of interest, assumed to be a constant, A_t is non-human wealth at the end of period t , so that $rA_t/(1+r)$ is capital income, E_t is the expectation operator for expectations formed at t , and y_t is labor income received at time t . Point expectations, a constant real interest rate, and the infinite horizon are all adopted to enable us to illustrate as simply as possible the issues with which we are concerned. The evolution of assets over time is governed by

$$A_{t+1} = (1+r)(A_t + y_t - c_t). \quad (2)$$

The first difference of equation (1) can be written, using (2), in the form

$$\Delta c_{t+1} = r \sum_{i=0}^{\infty} (1+r)^{-i} (E_{t+1} - E_t) y_{t+i} \quad (3)$$

so that changes in consumption are driven by innovations in labor income. More precisely, in this infinite horizon model, the change in consumption is simply the annuity value of the present discounted value of change in the expected value of future labor incomes. As in more general models of consumption under uncertainty, the change in consumption depends on neither the past history of nor previously anticipated changes in labor income. We shall also make use of an alternative but equivalent expression first derived in Campbell (1987). This explains saving, s_t , defined by

$$s_t = rA_t/(1+r) + y_t - c_t \quad (4)$$

by the "saving for a rainy day" equation

$$s_t = -\sum_{i=0}^{\infty} (1+r)^{-i} E_t \Delta y_{t+i}. \quad (5)$$

whereby saving is the discounted present value of expected future declines in income. (The equivalence of (5) and (1) can readily be seen by using (4) to substitute for s_t , and then "unscrambling" the changes in labor income.)

Equations like (3) and (5) can be given empirical content by specifying some mechanism for forming expectations. We follow standard practice by specifying a time-series process for labor income, and using the estimated parameter values to calculate predictions that can be revised as new information becomes available. For example, if y_t follows the ARMA process

$$y_t = \sum \phi_k^* y_{t-k} + \epsilon_t + \sum \theta_k^* \epsilon_{t-k} \quad (6)$$

which may or may not contain unit roots, then we have, see Flavin (1981) or Hansen and Sargent (1981),

$$\Delta c_t = \frac{r[1 + \sum (1+r)^{-i} \theta_i^*]}{(1+r)[1 - \sum (1+r)^{-i} \phi_i^*]} \epsilon_t \quad (7)$$

In cases where there is a unit root in the autoregressive component of (6), it is convenient to consider changes in expectations about changes rather than levels of income. Since y_{t+i} is y_t plus the sum of Δy_{t+j} from $j=1$ to $j=i$, we can substitute in the right hand side of (3) and rearrange to obtain

$$\Delta c_{t+1} = \sum_{i=0}^{\infty} (1+r)^{-i} (E_{t+1} - E_t) \Delta y_{t+1+i} = \frac{1 + \sum (1+r)^{-i} \theta_i}{1 - \sum (1+r)^{-i} \phi_i} \quad (8)$$

where θ_i and ϕ_i are the autoregressive and moving average coefficients in the ARMA representation of the *first-difference* of income.

To take a simple example that illustrates the algebra as well as the phenomenon with which we shall be concerned, the following simple auto-regression in first differences gives a good fit to the labor income series. Our data period runs from 1953,2 to 1984,4, so that with differencing and lags there are 125 observations: by OLS

$$\Delta y_t = \underset{(3.2)}{8.2} + \underset{(5.5)}{0.442} \Delta y_{t-1} \quad \sigma_\epsilon = 25.2 \quad (9)$$

so that, in terms of the ARMA representation, the θ 's are zero, as are the ϕ 's except for $\phi_1=0.442$, so that (8) is

$$\Delta c_t = \frac{(1+r)}{0.558+r} \epsilon_t \quad (10)$$

The multiplier on the right hand side of (10) is 1.79 when r is zero, and decreases only slowly with r , for example to 1.76 when r is 10% per annum, so that (9) predicts that the standard deviation of changes in consumption should be at least 1.76 times larger than the standard deviation of the innovation of labor income, estimated by (9) to be 25.2 (\$ 1972 per capita per annum). In fact, the standard deviation of the change in consumption is 27.3, and even this is an overestimate since *purchases* rather than consumption of durables are included in the total. The standard deviation of changes in consumption of non-durables and services is 12.4, and scaling this by the ratio of the mean of total consumption to the mean of consumption of non-durables and services gives a figure of only 15.8. The problem arises because the stochastic process (9) has the implication that shocks to labor income are indefinitely persistent, so that equation (10) predicts that consumption should be noisier than income. If the income

increase in one quarter is larger than expected, not only will that bonus never have to be repaid, but it is reasonable to expect the good fortune to be repeated, at least partially, in subsequent periods. Of course, this is only an example, and we are still very far from having established that consumption is in fact excessively smooth.

In spite of the fact that autoregressions like (9) fit the data well, it is clear that the change in labor income is not a stationary time series. Much more reasonable is the supposition that the first difference of the *logarithm* of labor income is stationary. However, the permanent income model relates consumption and changes in consumption to the level and changes in the level of labor income, so that some reformulation is needed in order to work in logarithms. The basic idea is to work with the *ratios* of saving and consumption to labor income, and to relate them to expectations about the ratios of future to current income. Since we assume that the rate of growth of labor income is stationary with mean μ , say, expressions of the form $E_t(y_{t+j}/y_t)$ are readily decomposed into an expected growth component $\exp(j\mu)$, and a residual. And because this residual is likely to be small relative to the growth component, it is possible to adopt a convenient linearization that yields a logarithmic version of the model. The details are confined to a brief Appendix; here we report the loglinear forms for the key equations (3) and (5), and present the equations to be used in the empirical work.

The "rainy day" equation, (5), in which saving anticipates future declines in labor income, has a very similar form in logarithms, in which the saving *ratio* anticipates future *logarithmic* declines in income, viz.,

$$(s_t/y_t) \approx -\sum_1^{\infty} \rho^i E_t \Delta \log y_{t+i} - \kappa \quad (11)$$

where κ is a constant given in the Appendix. The discount factor in (11) is not $1/(1+r)$, but $(1+\mu)/(1+r)$, or approximately $1+\mu-r$. One of the costs of moving from linear to logarithmic specifications is the need to assume that $r > \mu$, that the real interest rate is larger than the rate of growth of real labor income. The discounted present value of linear growth will exist provided only that the discount rate is positive, but this is clearly not the case for proportional growth. We believe that the stronger assumption is warranted by the increase in realism of the logarithmic model for incomes.

Changes in consumption can also be related to changes in expectations about rates of income growth. The logarithmic counterpart to equation (3) takes the form

$$\frac{\Delta c_{t+1}}{y_t} \approx \frac{r}{r-\mu} \sum_{i=1}^{\infty} \rho^i (E_{t+1} - E_t) \Delta \log y_{t+i} \quad (12)$$

so that the ratio of the change in consumption to labor income is proportional to the change in the present value of future rates of growth of income, where once again, the discount rate is the excess of the real interest rate over the rate of growth.

Equations (11) and (12) make somewhat different approximations, and since we shall use both in the analysis, it is important that they be reconciled. The derivation of equation (11) requires that the saving ratio be small, and in the Appendix we show that this will only be true if both r and μ/r are small. If so, $r/(r-\mu)$ in (12) is approximately equal to unity, while if the lagged value of (11) is divided by ρ and subtracted from (11), we have

$$\frac{s_t}{y_t} - \Delta \log y_t - \frac{s_{t-1}}{\rho y_{t-1}} \approx - \sum_{i=0}^{\infty} \rho^i (E_t - E_{t-1}) \Delta \log y_{t+i} \approx - \frac{\Delta c_t}{y_{t-1}} \quad (13)$$

The approximate equivalence of the first and last expressions may be checked directly, and is satisfied in our data. Our concern will be to test whether either of the outermost expressions is equal to that on the inside.

The simple autoregressive scheme that works for first differences in (9) also works for rates of growth of labor income, viz., and again by OLS

$$\Delta \log y_t = 0.00263 + 0.443 \Delta \log y_{t-1} \quad \sigma_\epsilon = 0.00792 \quad (14)$$

(3.3) (5.5)

so that, if ϵ_t is the innovation in the process, we have

$$\sum_{i=0}^{\infty} \rho^i (E_t - E_{t-1}) \Delta \log y_{t+i} = \sum_{i=0}^{\infty} (0.432\rho)^i \epsilon_t = \frac{\epsilon_t}{1-0.443(1+\mu+r)} \quad (15)$$

The sample average quarterly rate of growth of labor income, μ , is 0.00451, or 1.805% per annum. Hence, when r is zero, the multiplier for ϵ_t in (15) is 1.80, and is increasing in r . We therefore have "excess smoothness" in consumption if the standard deviation of either of the two outermost expressions in (13) is less than 1.80 times 0.00792, i.e. 0.0142 per quarter or 5.68% per annum. In fact, the standard deviation of the ratio of the change in consumption of non-durables and services to lagged income is 3.27% per annum, while that of the expression involving saving ratios is 3.57% per annum. Table 1 gives these and other results for a variety of different data series. The first panel shows means and standard deviations for rates of growth of both total (i.e. inclusive of capital) income z_t , and labor income y_t . The mean and standard deviation of the present value of innovations in labor income are calculated as above, but converted to an annual

basis. The second panel gives rates of growth of total consumption, and the two measures of consumption change. Apart from sign, the two are close to one another, and both have standard deviations much smaller than predicted from the AR(1) behavior of the changes in log income. The final panel calculates similar statistics for consumption of non-durables and services;

Table 1

Means, actual and predicted standard deviations

Variable	Mean (% per annum)	Standard Deviation (% p.a.)	
$\Delta \log z_t$	2.155	3.293	
$\Delta \log y_t$	1.805	3.516	
$\Sigma \rho^i \Delta E_t \Delta \log y_t$	0.000	5.539	
Total consumption			
$\Delta \log c_t$	2.204	3.141	
$\Delta c_t / y_{t-1}$	2.307	3.272	
$\Delta \eta_t$	-2.155	3.574	
Scaled consumption of non-durables and services			Scale factor λ
$\Delta \log c_{nt}$	1.991	1.950	-
$\lambda \Delta c_{nt} / y_{t-1}$	2.073	2.019	1.274
$\Delta \eta_t$	-2.155	3.574	1.274
$\lambda \Delta c_{nt} / y_{t-1}$	2.433	2.370	1.495
$\Delta \eta_t$	-1.194	2.488	1.495

Notes: A constant discount rate of 6% per annum is assumed. The calculations in the third line assume that $\Delta \log y_t$ follows an AR(1) process with coefficient $\phi=0.432$. $\Delta \eta_t$ is shorthand for $(s_t/y_t) - \Delta \log y_t - \rho^{-1}(s_{t-1}/y_{t-1})$.

this grows somewhat less rapidly than the total, and is very much smoother. Two scale factors are considered. The first, 1.274, is the ratio of the mean of total consumption to the mean of consumption excluding durables. Once again, the two change measures are close in both mean and variance, and the variances are much less than those predicted by the theory. The second scale factor, 1.495, comes from Campbell (1987), and is the reciprocal of the marginal propensity to consume estimated from a simple bivariate reg-

ression of consumption of non-durables and services on total income. This "cointegrating" factor is the number λ that will make stationary the saving series $z_t - \lambda c_t$. Under this definition of consumption, the two approximations diverge significantly, particularly in means. One of the difficulties here is that purchases of durable goods are becoming a steadily larger share of total consumption, so simple scaling is not an adequate substitute for calculating the consumption of durables.

Even so, these results clearly show that the excess smoothness result can be described in logarithms just as well as in levels. Even total consumption, changes in which are certainly more variable than in a true measure of consumption, is smoother than it ought to be if permanent income theory is true. More accurately, it is too smooth if the simple AR(1) model of changes in labor income adequately characterizes the long-run behavior of the series. Whether or not that is true is the topic of the next section.

2. Univariate time-series representations of income

Since changes in consumption are typically less variable than are changes in labor income, the question of whether or not consumption is *too* smooth rests on whether or not changes in permanent income are typically of smaller magnitude than changes in income. Permanent income will only change in response to new information about income itself, and the change in permanent income will be larger or smaller as the innovations in income are expected to be more or less persistent. In this section we study the persistence of univariate innovations to income, reserving to the next section the possibility that what here are labelled innovations may in reality be partly anticipated by consumers.

We investigate these questions under the assumption that $\Delta \log y_t$ is a stationary time series. It is important to note that this does not presume the answer and that the assumption is consistent with a wide range of processes. For example, if $\log y_t$ were the sum of a stationary ARMA process and a deterministic linear time trend, the first difference, $\Delta \log y_t$, would be a stationary ARMA process with a unit root in the moving average part. Consider the moving average representation of the series,

$$(\Delta \log y_t - \mu) = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} \quad (16)$$

where θ_0 is taken to be unity. Note that for the AR(1) scheme given by (14), $\theta_j = \phi^j$, whereas the presence of a deterministic trend would tend to show up in negative values for θ_j sometime after the first. For example, if $\log y_t$ were the sum of white noise and a linear trend (which it is not), we should have $\theta_1 = -1$, and $\theta_j = 0$ for $j > 1$. The moving average representation (16) is convenient, not only because it allows such unit roots, but also because we have

$$(E_t - E_{t-1}) \Delta \log y_{t+k} = \theta_k \epsilon_t \quad (17)$$

In consequence, the quantity we are interested in for evaluating excess smoothness in (13) is

$$\sum_{i=0}^{\infty} \rho^i (E_t - E_{t-1}) \Delta \log y_{t+i} = \left(\sum_{i=0}^{\infty} \rho^i \theta_i \right) \epsilon_t = \pi(\rho) \epsilon_t \quad (18)$$

Hence, if we can find the MA representation of $\Delta \log y_t$, the expression $\pi(\rho)$ tells us what should be the ratio of the standard deviation of $\Delta c_t / y_t$ to the standard deviation of the innovation in labor income. Note that $\pi(1)$ is the

persistence measure defined in Campbell and Mankiw (1986a) and calculated by them for real gross national product. It measures the ratio by which an innovation has to be multiplied to determine its final effect upon the level of the series. Since $\rho = 1 + \mu - r \approx 1$, the measure that we require here is likely to be numerically close to the simple persistence measure. Note finally that whether or not $\pi(\rho) > 1$, whether the series is persistent or whether permanent income is noisier than measured income, is not simply a matter of whether the series in levels does or does not possess a unit root. Consider for example the sum of a random walk and white noise; the series in levels has a unit root, but its persistence measure, which depends on the ratio of the variances of the white noise to the innovation variance of the random walk, is always less than unity. We shall consider a closely related example below.

Perhaps the most obvious way to estimate $\pi(\rho)$ is to fit ARMA models to the series $(\Delta \log y_t - \mu)$, and this is what we do first. For the period 1953,2 through 1984,4 we used the Kalman filtering algorithm, see Harvey (1981), to calculate exact maximum likelihood estimates for ARMA(p, q) models for all combinations of $p, q \leq 2$. A selection of the results is shown in Table 2.

Each vertical column in the table represents a particular specification, with a general increase in complexity as we move from left to right. The first horizontal panel shows the parameter estimates for the autoregressive part, while the second shows those for the moving average. The third panel contains the estimate of the innovation variance and the maximized value of twice the logarithm of the likelihood. The roots of both autoregressive and moving average parts are shown next, while the final panel gives the persistence measures that are our main focus of interest. The first three of

Table 2

ARMA models for changes in log labor income

	(1) white noise	(2) AR(1)	(3) ARMA(1,1)	(4) ARMA(1,2)	(5) ARMA(2,0)	(6) ARMA(2,1)
Autoregressive parameters						
ϕ_1	-	0.45(.09)	0.44(.18)	0.59(.39)	0.45(.10)	1.44(.09)
ϕ_2	-	-	-	-	-0.00(.10)	-0.45(.09)
Moving average parameters						
θ_1	-	-	0.01(.21)	-0.14(.42)	-	-1.00(.11)
θ_2	-	-	-	-0.09(.22)	-	-
Innovation variance x 10 ³ and likelihood						
σ_ϵ	8.76(0.5)	7.83(0.4)	7.83(0.4)	7.82(0.4)	7.83(0.4)	7.77(0.4)
2logL	836.42	864.32	864.32	864.81	864.32	865.42
Roots						
AR	0	0.45	0.44	0.59	0.44, 0.01	0.97, 0.47
MA	0	0	0.01	0.38, -0.24	0	-1.00
Persistence measures						
5 years	1	1.80(0.3)	1.80(0.3)	1.90(0.4)	1.81(0.4)	1.06(3.0)
20 years	1	1.80(0.3)	1.80(0.3)	1.90(0.4)	1.81(0.4)	0.17(6.2)
$\infty \pi(1)$	1	1.80(0.3)	1.80(0.3)	1.90(0.4)	1.81(0.4)	0.00(1706)
$\pi(0.9945)$	1	1.80(0.3)	1.80(0.3)	1.89(0.4)	1.80(0.4)	0.29(1429)
$\pi(0.9895)$	1	1.80(0.3)	1.79(0.3)	1.88(0.4)	1.79(0.4)	0.49(1242)
$\pi(0.9845)$	1	1.79(0.3)	1.78(0.3)	1.87(0.4)	1.79(0.4)	0.64(1096)

these show what fraction of an income innovation can be expected to remain in the level of (log) income after 5 years, 20 years, and in the limit. This limiting persistence measure, or $\pi(1)$, is simply the sum of the coefficients of the MA representation of the first difference, while the last three lines show the same sum with the successive coefficients discounted according to (18) above. The three figures shown correspond to real interest rates of 4%, 6% and 8% per annum respectively.

Column (1) of the Table shows the estimates under the hypothesis that the rate of change of income is (a constant plus) white noise, so that the logarithm of labor income is a random walk plus drift. Since the change in a random walk is immediately consolidated into the series, but has no predictive power for the future, all of the persistence measures are unity. The likelihood for the random walk model is significantly exceeded by that for the AR(1) model for first differences in column (2), which is essentially the same model that was discussed above. For this model, the persistence measure attains its final value of 1.80 quite rapidly, and is quite insensitive to variations in the discount rate within the relevant range of real interest rates. Improving the likelihood beyond the figure in column (2) is more difficult. MA(1) and MA(2) processes, not shown, fit the data very much worse than does the AR(1), while the extension of the AR(1) to an ARMA(1,1) in column (3) gives a very small MA coefficient and an almost imperceptible increase in the likelihood. The ARMA(1,2) does very little better, nor is there any evidence that adding AR coefficients with no MA coefficients improves the fit, see the AR(2) in column (5). All of the models in columns (2) through (5) have similar persistence characteristics, and for none does the discounting make much difference to the persistence measure.

The last column, the ARMA(2,1) model, tells a somewhat different story. All three parameters are individually significant, although there are also very high correlations between the estimates. The moving average part has a unit root, while the autoregressive part has roots of 0.97 and 0.47. Since the 0.97 root is close to unity, and not significantly different from it, there is an almost exact cancellation of roots, in which case we are back with the AR(1) model in column (2). Indeed, a standard likelihood ratio

test cannot reject the AR(1) as a specialization of the ARMA(2,1). Even so, the ARMA(2,1) has both estimated AR roots less than unity, (one only just), and contains a unit root in the moving average part, so that innovations will always be eliminated in the end, and the ultimate measure of persistence is zero. Note that even after five years, the persistence measure is greater than unity, so that the behavior of this model deviates from that of the others only a long time after the original shock, when there are a long string of small negative coefficients in the MA representation that eventually eliminate the early positive effects. Of course, since the roots almost cancel, there is very great uncertainty about the persistence estimates. Furthermore, the discounting now makes a considerable difference with higher interest rates favoring the early positive effects at the expense of the later negative ones.

It is instructive to compare the ARMA(2,1) with the components model suggested by Watson (1986). In this, $\log y_t$ is written as the sum of two components, one of which is a random walk with drift, while the other is a stationary AR(2) process, i.e.

$$\begin{aligned} \log y_t &= r_t^1 + r_t^2 \\ r_t^1 &= r_{t-1}^1 + \mu + \zeta_t^1 \\ (1 - \phi_1 L - \phi_2 L^2) r_t^2 &= \zeta_t^2 \end{aligned} \tag{19}$$

where ζ_t^1 and ζ_t^2 are independent white noise disturbances with variances $\sigma_{1\zeta}^2$ and $\sigma_{2\zeta}^2$. It can be shown that Watson's model can be represented as an ARMA(2,2) in the first differences, although the parameters are restricted. If the model is fitted to our data, and once again estimation is straightforward using the Kalman filter, we obtain an estimate of $\sigma_{1\zeta}^2$ of almost

exactly zero, in which case (19) is identical to the ARMA(2,1) in column (6). Note that in this case, the random walk in the second line of (19) becomes a deterministic trend, so that the model represented by column (6), as well as the version of the Watson model that best fits our data, is of a (just) stationary AR(2) process around a deterministic linear trend. Such a model has no long run persistence, because income will eventually return to trend, and the effects of innovations on permanent income are limited, though they can still be quite sizeable, especially if the long-run negative effects are heavily discounted.

There is no formal statistical basis for favoring one or other of these two representations. An external criterion like parsimony would lead to column (2), the simple AR(1) in first differences. But many economists have strong prior beliefs in models with deterministic trends, and they are unlikely to be persuaded by parsimony. Formal tests for unit root models are available from the work of Dickey and Fuller (1981) and Phillips and Perron (1986), and these cannot reject the AR(1) model in favor of the deterministic trend model in column (6). But that does not mean that the deterministic trend model is incorrect, but simply that it is not possible to discriminate on these data with these models.

Measurement of persistence depends on the long-run properties of the income series, and it is conceivable that a mere thirty years worth of data is insufficient for this task. However, another possibility is that the ARMA models in Table 2 are not the most efficient way of measuring persistence since they do so indirectly, estimating not persistence itself, but a set of parameters which are then used to calculate persistence. An alternative, more direct procedure, is to examine the representation of income in

the frequency domain. Long run properties of the series correspond to very low frequency components, so that, for example, as Watson points out, the model (18) has a spectral density at zero frequency of zero, as does the AR(2,1) model in column (6), while the spectral density of the AR(1) series in column (2) reaches its *maximum* at zero. The use of the spectral density at zero as a measure of persistence has also been advocated by Cochrane (1986), and its (close) relation to the measures used here is given by Campbell and Mankiw (1986a).

We denote the j th autocovariance of the log difference of labor income as α_j , and write $C(z)$ for the sum $\sum(\alpha_j z^j)$ with j running from $-\infty$ to $+\infty$. If σ^2 is the variance of $\Delta \log y_t$, its spectral density at zero can be written as $v=C(1)/\sigma^2$, which is the sum from minus to plus infinity of all autocorrelations. From Granger and Newbold (1977, p.31), we have $C(z)=\pi(z)\pi(z^{-1})\sigma_\epsilon^2$, where $\pi(\cdot)$ is defined in (18), and σ_ϵ^2 , as before, is the innovation variance of $\Delta \log y_t$, so that

$$v = C(1)/\sigma^2 = (\sigma_\epsilon^2/\sigma^2)[\pi(1)]^2. \quad (20)$$

Consequently, if $R^2=1-\sigma_\epsilon^2/\sigma^2$ is the fraction of the variance that can be predicted from the past history of the process, the relationship between the two persistence measures can be written

$$\pi(1) = \sqrt{[v/(1-R^2)]} \quad (21)$$

From the point of view of this paper, the main virtue of the measure v is that there exists a direct non-parametric estimator, see for example Priestly (1982). Write the sample autocorrelations $\hat{\gamma}_j$ as

$$\hat{\gamma}_j = \frac{T}{T-j} \frac{\sum_{t=j+1}^T (\Delta \log y_t - \mu)(\Delta \log y_{t-j} - \mu)}{\sum_{t=1}^T (\Delta \log y_t - \mu)^2}. \quad (22)$$

then the estimate of v based on a triangular (Bartlett) window of size k is

$$\hat{v}^k = 1 + 2 \sum_{j=0}^k [1-j/(k+1)] \hat{\gamma}_j \quad (23)$$

The window weights in (23) give linearly declining weights to higher-order autocorrelations up to and including the k th. Provided this window size is increased with the sample size, \hat{v}^k is a consistent estimator of v , and has an asymptotic t -value given by

$$t(\hat{v}^k) = \sqrt{[0.75T/(k+1)]} \quad (24)$$

The fact that v can be estimated directly is of great convenience, but a number of complications ought to be noted before turning to the results. First, we can only move from v to $\pi(1)$ with through (21), and that requires knowledge of extent to which the series can be predicted from its past. The required R^2 can be sensibly estimate from the ARMA models, but it cannot be directly estimated from the data without some parametric model. Even so, we know that R^2 is at least as large as γ_1^2 , the squared first order autocorrelation coefficient, so that $\sqrt{[v/(1-\gamma_1^2)]}$ provides a lower bound for $\pi(1)$, whatever the value of R^2 . Replacing γ_1^2 by its sample estimate provides an estimate of this lower bound, π^* . Second, it is not really $\pi(1)$ that we require, but $\pi(\rho)$, and there appears to be no way of allowing for the discount factor in the estimation. However, note from Table 2 that discounting makes little difference for most of the models, and it is clear what sort of patterns of persistence will cause this result not to be true.

Table 3 presents the estimates of the spectral density and the implied

Table 3

Non-parametric estimates of persistence

Labor income, $\Delta \log y_t$										
Autocorrelations at lag:										
1	2	3	4	5	6	7	8	9	10	
0.446	0.193	0.156	0.056	-0.152	-0.017	-0.010	-0.070	-0.024	0.123	
Persistence estimates:										
Window size		$\hat{\vartheta}^k$	s.e. ($\hat{\vartheta}^k$)		π^*					
10		2.212	0.755		1.663					
20		2.334	1.100		1.707					
30		2.523	1.445		1.775					
40		2.443	1.609		1.704					
50		1.986	1.459		1.575					
60		1.424	1.144		1.333					
Total income, $\Delta \log z_t$										
Autocorrelations at lag:										
1	2	3	4	5	6	7	8	9	10	
0.373	0.124	0.103	-0.039	-0.206	-0.074	-0.068	-0.120	-0.031	0.141	
Persistence estimates:										
Window size		$\hat{\vartheta}^k$	s.e. ($\hat{\vartheta}^k$)		π^*					
10		1.589	0.542		1.358					
20		1.518	0.716		1.328					
30		1.529	0.871		1.329					
40		1.438	0.947		1.292					
50		1.088	0.800		1.124					
60		0.793	0.637		0.960					

persistence measures. The autocorrelations are shown for lag lengths up to ten quarters, and it is interesting to note that after the first four, which, as predicted by the simple AR(1), are positive and geometrically declining, there is a sequence of *negative* coefficients, which are not predicted by the AR(1), and which may indicate some tendency for the effects of the innovations to be negated some considerable time after they have

occurred. Nevertheless, the estimate of ν is consistently above 2 for all reasonable window sizes; note that as the window size tends to the sample size, $\hat{\nu}$ tends to a mechanical and meaningless value of zero. However, it is also the case that the standard error increases throughout the same range, so that it would be difficult to persuade a determined believer that persistence is indeed greater than unity. For very small window sizes, say with $k < 10$, the estimates of ν exceed unity by more than two standard deviations, but given the pattern of the autocorrelations, this is hardly a fair test. Nevertheless, the lower bound estimates for $\pi(1)$ shown in the table are strikingly similar to those found in Table 2 for the AR(1) and related models. There is no evidence here of the zero long-run persistence that would be predicted by a model with a deterministic trend.

Since the series for labor income is a "manufactured" series, see Blinder and Deaton (1985) for details, we also present details of similar estimates for the more "official" series on total income. The pattern is somewhat similar, and although the persistence estimates are smaller there is certainly no suggestion that the results are extremely sensitive to the definition of income. A more serious question is about the reliability of the non-parametric estimator, particularly in samples of this size, and it is legitimate to be concerned whether or not these procedures can be expected to discriminate between genuinely persistent series and those, such as the series generated by the Watson model, that have ultimately zero persistence. Following Campbell and Mankiw (1986b), some relevant Monte Carlo results are presented in Table 4. The second panel reports Campbell and Mankiw's results for a time series that is a stationary AR(2) process around a deterministic trend so that the first difference is an ARMA(2,1) with a unit root

Table 4

Experimental results on persistence estimates

Window	Truth is AR(1) in first differences		Truth is AR(2) with deterministic trend	
	$\hat{\nu}^k$ (s.d.)	$\pi^*(1)$	$\hat{\nu}^k$ (s.d.)	$\pi^*(1)$
10	1.94 (0.59)	1.50 (0.27)	1.32 (0.38)	1.24(0.20)
20	1.87 (0.86)	1.45 (0.36)	0.83 (0.32)	0.97(0.20)
30	1.70 (0.99)	1.37 (0.42)	0.57 (0.27)	0.80(0.19)
40	1.55 (1.04)	1.29 (0.46)	0.43 (0.23)	0.69(0.19)
50	1.38 (1.06)	1.19 (0.48)	0.34 (0.21)	0.60(0.19)
60	1.21 (1.05)	1.09 (0.50)	0.27 (0.19)	0.53(0.19)

Notes: The AR(1) model has a parameter of 0.4, while the AR(2) parameters are 1.34 and -0.42. The results are based on 500 replications of using generated series of length 130. True ν and $\pi(1)$ are 2.33 and 1.67 for the AR(1) model and are zero for the AR(2) model. Standard deviations are standard deviations of the *empirical* distribution of the estimates.

in the moving average part; the parameters of the AR(2) are 1.34 and -0.42, corresponding to roots of 0.84 and 0.50. The first panel gives similar data for the AR(1) model in first differences. In each case, 500 artificial series of length 130 were generated, and persistence estimates calculated for various window sizes following exactly the procedures used in Table 3. Note first that the estimates for the AR(1) model are biased downward and that the bias worsens as the window size is enlarged. Similarly, the estimates for the AR(2) with deterministic trend are too large; the long-run return to trend can only be captured by window sizes that are larger than the sample size can bear. Even so, the difference between the left and right panels is very apparent, and is in the right direction; at all window sizes the AR(1) is estimated to be more persistent than the ARMA(2,1), and the difference is particularly marked when the window size is 20 or more.

Comparing Table 4 with our "real" results in Table 3, labor income generates even larger persistent estimates than does the theoretical AR(1), although the standard deviations in Table 4 suggest that our results are well within the sampling distribution. But it is very hard to believe that labor income is indeed described by the a deterministic trend plus the stationary AR(2) in Table 2; the estimates in Table 3 are extremely unlikely to have come from the population described by the right hand panel of Table 4.

In summary then, we tend to believe that the evidence is in favor of the persistent models, so that, if consumers predict income on the basis of a univariate process, consumption is much too smooth to be consistent with the permanent income model. Choice between the ARMA models in Table 2, and in particular between an ARIMA(1,1,0) and a stationary AR(2) around a deterministic trend, is very much a matter of taste, and it has become clear to us there are very different tastes in the profession. The non-parametric estimates of persistence do not resolve the question in favor of the model with the unit root, but they do suggest a good deal more persistence than is consistent with a relatively rapid return to a deterministic trend.

3 Bivariate models and the excess sensitivity question

In this section we work with a bivariate autoregressive representation of the saving ratio and the change in log of income. Such a representation allows a much richer representation of the permanent income theory, it is immune to a number of important criticisms of the univariate approach, and it will allow us to link our results with the literature. The simplest form that we shall consider is

$$\begin{pmatrix} \Delta \log y_t^{-\mu} \\ s_t/y_t^{-\sigma} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \Delta \log y_{t-1}^{-\mu} \\ s_{t-1}/y_{t-1}^{-\sigma} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (25)$$

where σ is the mean saving ratio. In matrix notation, (25) is

$$x_t = Ax_{t-1} + u_t. \quad (26)$$

Most of our analysis can be done with the simple first-order vector autoregression (25), but we shall occasionally incorporate additional lags. If there are M lags, x_t becomes a $2M$ vector containing current $\Delta \log y_t$, its first $(M-1)$ lags, followed by the equivalent M terms for the saving ratio. The matrix A is then a $2M$ square matrix of coefficients.

Equation (25) asserts that the saving ratio is stationary but satisfies what can be thought of as an error correction mechanism whereby deviations of the saving ratio from its (unconditional) mean exert a lagged influence that helps return the series to its equilibrium level. The change in income has the same autoregressive part as before, but the lagged saving ratio is also permitted to exert an influence. The existence of the "cross" effects not only permits rather general univariate time-series representations of the two series, but also permits a much more satisfactory treatment of the informational structure of the problem. It is reasonable to suppose that consumers make use of whatever information is available to them in making forecasts of future income, and that while current and past income levels are likely to be relevant, they are unlikely to be exclusively so. Consumers may be able to predict labor income much more accurately than do the simple univariate time series models of the previous section, so that what we class as "innovations" to income could have been largely predicted by consumers. If so, the smoothness of consumption may simply reflect the

lower innovation variance that is guaranteed by superior information, see in particular West (1986a), although West (1986b) has also constructed a test that shows that the data do not support this hypothesis.

In contrast to the univariate representations, bivariate models such as (25) are capable of recognizing that consumers may have superior information. If the permanent income theory is correct, saving (or the saving ratio) incorporates consumers' expectations about future income, so that saving behavior reveals those expectations to the observer. This is perhaps the most important reason for expecting saving to Granger-cause income. In period $t-1$, consumers may receive advance notice of an income change in t , for example through an innovation in money, stock prices, or whatever. Such information will be reflected in saving behavior in period $t-1$, but will show up in income only in period t , so that saving will Granger-cause income in a bivariate system such as (25). Because of these effects, and again provided that the permanent income model is correct, the econometrician's perceived innovation to permanent income must be the same as the innovation experienced by the consumer. Because saving is included in our information set, our prediction of permanent income must contain the consumer's prediction of permanent income, while any "advance notice" possessed by consumers will be reflected, not in the true innovation being smaller than the apparent innovation, but rather in the ability of lagged saving to predict income. The formal basis for these results is as stressed in Campbell (1987); the permanent income equations (11) and (13) remain valid when projected on to the econometrician's information set, provided that the savings ratio itself is included in the set.

We use the vector autoregression (25) to examine three related questions. First, we fit (25) to the data, and examine the corresponding univariate representation for income. This provides yet another way of looking at the persistence questions of section 2. Second, we follow Campbell (1987), and use the model to test the standard "orthogonality" condition, that changes in consumption are not predictable from lagged information. As in other studies, we find evidence of "excess sensitivity," that changes in consumption are predictable by *anticipated* changes in income. Thirdly, we repeat the tests for "excess smoothness" of consumption, and show that, once again, consumption does not respond sufficiently to *unanticipated* changes in income. We show that the excess smoothness and excess sensitivity results are consistent with one another, and that they stem from the same basic underlying feature in the data; they are essentially the same phenomenon.

The first-order vector autoregression (25) generates parameter estimates that are shown in the first panel of Table 5. The usual first-order autoregression in $\Delta \log y_t$ is again apparent, although there is a small but significant negative feedback from the lagged saving ratio to changes in income. The saving rate is also well described by an AR(1), especially when consumption is total consumption; when non-durable and service consumption is used, the "own" autoregressive parameter is close to unity, and there is a much larger feedback from lagged income changes.

Univariate representations for both series can readily be derived using standard techniques, see e.g. Granger and Newbold (1977, p.217). Both series can be represented as ARMA(2,1) processes, which suggests that there is no immediate contradiction between the first-order VAR and the best ARMA models in Table 2. Characteristics of the ARMA(2,1) for the rate of growth

Table 5

Univariate representations of income growth from bivariate VAR

	(1)		(2)		(3)	
VAR parameters						
a_{11}, a_{12}	0.443	-0.177	0.454	-0.067	0.458	-0.066
	(0.08)	(0.06)	(0.08)	(0.03)	(0.08)	(0.03)
a_{21}, a_{22}	0.072	0.803	0.212	0.972	0.266	0.956
	(0.07)	(0.05)	(0.07)	(0.03)	(0.08)	(0.03)
ARMA parameters and AR roots						
ϕ_1, ϕ_2	1.247	-0.369	1.425	-0.455	1.414	-0.455
	(0.09)	(0.07)	(0.08)	(0.08)	(0.08)	(0.03)
θ_1	-0.799	(0.05)	-0.954	(0.02)	-0.952	(0.02)
r_1, r_2	0.763	0.483	0.943	0.482	0.918	0.496
Persistence measures						
20 years	1.642	(0.25)	1.546	(0.81)	1.170	(0.39)
$\infty \pi(1)$	1.648		1.533		1.172	
$\pi(0.9945)$	1.646		1.560		1.216	
$\pi(0.9895)$	1.644		1.580		1.252	
$\pi(0.9845)$	1.642		1.596		1.283	

Notes: (1) refers to calculations in which consumption is taken to be total consumption, including purchases of durables. Panel (2) uses only non-durables and services consumption inflated by a factor of 1.274. Panel (3) is the same as panel (2) but with an inflation factor of 1.495.

of labor income are given in Table 3; because of the feedbacks between income and the saving ratio, the representation varies with the definition of consumption. However, all three representations have an autoregressive root that is between 0.48 and 0.50, while the second autoregressive root is always very close to cancelling with the moving average root. Hence, all three ARMA(2,1) processes shown are very close to being AR(1) processes with a positive autoregressive parameter of 0.48. However, all of the ARMA(2,1) processes shown in the Table have high persistence, with estimates of $\pi(\rho)$ that are always greater than unity, and that are largely insensitive to variations in the real interest rate. All of this is confirmation of the

results of the previous section, and to the extent that there are differences with the univariate processes, the difference is in favor of greater rather than less persistence.

We consider next the relationship between the estimated vector autoregression and the theoretical properties of income, saving and consumption that are presented in the introduction. The version of the permanent income model on which we focus is that given by equation (13) in section 1. In the notation of the VAR (25), this can be rewritten in the convenient form

$$\frac{s_t}{y_t} - \Delta \log y_t - \frac{s_{t-1}}{\rho y_{t-1}} = (e'_2 - e'_1)x_t - \rho^{-1} e'_2 x_{t-1}. \quad (27)$$

where e'_1 and e'_2 are the vectors (1,0) and (0,1) respectively. (Note that (27) also holds when the VAR contains more lags, provided e_1 and e_2 are appropriately redefined.) Substitution of the VAR (26) into the right hand side gives

$$\frac{s_t}{y_t} - \Delta \log y_t - \frac{s_{t-1}}{\rho y_{t-1}} = [(e'_2 - e'_1)A - \rho^{-1} e'_2]x_{t-1} + (e'_2 - e'_1)u_t \quad (28)$$

From (26), it is clear that $(E_t - E_{t-1})x_{t+i} = A^i u_t$, so that, since $\Delta \log y_t$ is the first element of x_t ,

$$\sum_{i=0}^{\infty} \rho^i (E_t - E_{t-1}) \Delta \log y_{t+i} = \sum_{i=0}^{\infty} e'_1 \rho^i A^i u_t \quad (29)$$

Now, by (13), the left hand sides of (28) and (29) should be equal, so that the permanent income theory requires that two sets of constraints be satisfied, viz.,

$$(e'_2 - e'_1)A - \rho^{-1}e'_2 = 0 \quad (30a)$$

$$\sum_{i=0}^{\infty} e'_1 \rho^i A^i = e'_2 - e'_1 \quad (30b)$$

The interpretation of these two sets of restrictions is an important part of our story. Equation (30a) guarantees that the left hand side of (28) is independent of either lagged income growth or the lagged saving ratio. It is therefore a test of the unpredictability of consumption, and includes a test of the absence of "excess sensitivity," of the lack of relationship between changes in consumption and lagged values of income. By contrast, the restrictions in (30b) ensure that the change in consumption is that which is warranted by the change in the present value of income. If (30b) is satisfied, there can be no "excess smoothness" of consumption.

The orthogonality condition (30a) implies that the A matrix can be written in the form

$$A = \begin{bmatrix} \alpha & \beta \\ \alpha & \beta + \rho^{-1} \end{bmatrix} \quad (31)$$

where α and β are not restricted. Since ρ is non-zero, the determinant of $(I - \rho A)$ will be non-zero if, and only if $\beta \neq 0$, i.e. if saving Granger-causes income. Given $\beta \neq 0$, the inverse $(I - \rho A)^{-1}$ exists, so that (30a) can be rearranged to yield

$$e'_1 (I - \rho A)^{-1} = e'_2 - e'_1 \quad (32)$$

which is identical to (30b). Hence, provided that lagged saving ratio has predictive power for the change in labor income, the orthogonality condition and the condition for smoothness are identical. If consumption changes can-

not be predicted by past changes in either consumption or income, then the consumption change must be equal to the change in permanent income.

If $\beta=0$ in (31), the orthogonality condition will not generally guarantee that the change in consumption equals the innovation in permanent income. In this case, the change in income is a simple AR(1) as in Section 2, so that the discounted innovation in (29) is $u_{1t}/(1-\rho\alpha)$. This will be equal to the innovation in $(s_t/y_t - \Delta \log y_t)$ if $u_{2t} - u_{1t}$ is equal to $u_{1t}/(1-\rho\alpha)$, which requires the existence of a linear dependency between the two innovations which there is no reason to expect to hold in the data. Further, we know from the results of the previous sections that consumption changes are not sufficiently variable to match discounted innovations if the first differences of income are indeed an AR(1). Lastly, we note that the results in Table 5 suggest that the saving ratio *does* Granger-cause changes in income, so that the case of $\beta \neq 0$ that is the one that is of practical importance.

Table 6 presents some of the relevant data, again for the three alternative definitions of consumption. The first column presents standard tests of orthogonality using VAR's with one lag of the two variables, and VAR's with five lags of each. All of these test statistics show rejections, with the possible exception of the five lag VAR for the larger scaling of consumption of non-durables and services. Even allowing for the asymptotic nature of these results, and their dependence on an estimated variance covariance matrix of the residuals, they are not favorable for the hypothesis and are in line with other findings using similar (or the same) data.

The remainder of the table compares theoretical and actual innovation variances, or standard deviations. For each case, we calculated the theoretical innovation variance of $(s_t/y_t - \Delta \log y_t)$, which is (approximately)

Table 6

Tests for excess smoothness and excess sensitivity

	Wald test (p-value)	predicted innovation variance	actual innovation variance	ratio (s.e.)
Total consumption				
VAR-1	19.6 (5.4x10 ⁻⁵)	5.044	3.309	0.656 (0.10)
VAR-5	45.9 (1.5x10 ⁻⁶)	4.768	3.017	0.633 (0.15)
Non-durables and services consumption x 1.274				
VAR-1	18.4 (1.0x10 ⁻⁴)	3.864	2.290	0.593 (0.21)
VAR-5	31.4 (5.0x10 ⁻⁴)	4.686	2.135	0.456 (0.20)
Non-durables and services consumption x 1.495				
VAR-1	10.0 (0.0066)	3.188	2.382	0.747 (0.16)
VAR-5	18.2 (0.0509)	4.022	2.257	0.561 (0.13)

Notes: The Wald test is the test of the estimated parameters in the VAR for conformity with the restrictions (30a). The predicted innovation is the standard deviation of the last term in (29), i.e. the square root of the quadratic form $e_1'(I-\rho A)^{-1}\Omega(I-\rho A')^{-1}e_1$ where Ω and A are estimated from the unrestricted VAR. The actual innovation is the standard deviation of $u_{2t} - u_{1t}$, or the square root of $(e_2' - e_1')\Omega(e_2 - e_1)$. All calculations use a ρ of 0.9895, which corresponds to a real interest rate of 6% per annum.

equal to the theoretical innovation in $\Delta c_t/y_{t-1}$, and is calculated from the last term in (29). For example, in the one-lag case, the VAR's are given in Table 5, the variance covariance matrices of the residuals are calculated in the usual way, and a value for the real interest rate of 6% per annum is assumed. The actual innovation in $(s_t/y_t - \Delta \log y_t)$ is, by (28), simply $u_{2t} - u_{1t}$, the standard deviation of which requires only the variance covariance matrix. In every case, the theoretical innovation variance is larger than the actual innovation variance, and in all but one case, is more than twice as large. Consumption is markedly smoother than it ought to be if the permanent income theory were correct.

Although Table 6 is a useful summary of our findings, rather more insight can be obtained from a slightly different approach. If we examine the three different A matrices shown in Table 5 and compare their structures with the structure required for orthogonality in (31), it is apparent that the restriction on the second column is approximately satisfied, so that the rejection of the model reflects the fact that $a_{11} \neq a_{21}$. This informal impression is easily confirmed by calculating the test statistics for each of the two hypotheses separately. The data can therefore be characterized straightforwardly by writing A , not as in (31), but as

$$A = \begin{pmatrix} \alpha & \beta \\ \alpha - \chi & \beta + \rho^{-1} \end{pmatrix} \quad (33)$$

where χ is the "excess sensitivity" parameter. Note that, if A satisfies (33), we have, from (28),

$$\frac{s_t}{y_t} - \Delta \log y_t - \frac{s_{t-1}}{\rho y_{t-1}} = -\chi \Delta \log y_{t-1} + (e'_2 - e'_1) u_t \approx -\frac{\Delta c_t}{y_{t-1}} \quad (34)$$

so that $\chi > 0$ reflects the fact that changes in consumption respond positively to last period's changes in income. It is this stylized fact of the data that has repeatedly caused the model to fail, as for example in Flavin's (1981) paper. Note also that the AR(1) structure for $\Delta \log y_t$ with a positive coefficient implies that $\Delta \log y_{t-1}$ predicts $\Delta \log y_t$, so that $\chi > 0$ can also be interpreted as a response of consumption to anticipated changes in current income, as would be generated by the existence of liquidity constraints.

The implications of the existence of excess sensitivity for forecasts of future income can most clearly be seen by evaluating (29) when A has the structure given by (33):

$$\sum_{i=0}^{\infty} \rho^i (E_t - E_{t-1}) \Delta \log y_{t+i} = \sum_{i=0}^{\infty} e_1' \rho^i A^i u_t = e_1' (I - \rho A)^{-1} u_t = (u_{2t} - u_{1t}) / (1 - \rho \chi) \quad (35)$$

The theoretical innovation is therefore the actual innovation multiplied by $(1 - \rho \chi)^{-1}$, which, since χ is a little less than 0.4, means that the predicted standard deviation will be about fifty percent too large, which is essentially what is shown in Table 6. In the bivariate framework used here, the untoward sensitivity of changes in consumption to *anticipated* changes in income inevitably implies that consumption will respond by less than is warranted given the innovation in income. There is no contradiction between excess sensitivity and excess smoothness; they are the same phenomenon.

It is worth explicitly reconciling these findings with those of Flavin, since Flavin's interpretation of her results is somewhat different from ours. Flavin defines "excess sensitivity" as existing when the response of consumption to current and lagged changes of income is larger than can be justified by the permanent income model. Since the model implies that changes in consumption should not be related to lagged income, part of her excess sensitivity test is clearly identical to the orthogonality tests presented here. Further, and in spite of our different econometric procedures (most importantly our use of logarithms, and Flavin's detrending of the data), the orthogonality conditions fail here as they do in Flavin's paper. However, Flavin also finds that consumption is excessively sensitive to *current* income changes and writes, for example, "Using either nondurables consumption or consumption of nondurables and services as the dependent variable, the hypothesis that consumption exhibits no excess sensitivity to current income can be rejected at the 0.5 percent level." (p1086). Since our principal finding is that consumption is too smooth, that it does not

respond enough to unanticipated changes in current labor income, a brief reconciliation seems in order.

Consider the following simple model adapted from Flavin but using our notation and our logarithmic form:

$$\frac{\Delta c_t}{y_{t-1}} = \sum_{i=0}^{\infty} \rho^i (E_t - E_{t-1}) \Delta \log y_{t+i} + \chi \Delta \log y_t + v_{2t} \quad (36)$$

$$\Delta \log y_t = \mu(1-\phi) + \phi \Delta \log y_{t-1} + v_{1t} \quad (37)$$

Following Flavin, we have added an error term v_{2t} to the consumption equation, and the term $\chi \Delta \log y_t$ represents the excess sensitivity of consumption to changes in current income. It is the quantity χ that Flavin finds to be positive and significantly different from zero. (Note that it is straightforward to add further lagged changes in income to (36) and no new issues of principle arise.) Equation (36) can be rewritten as

$$\Delta c_t / y_{t-1} = \chi \mu (1-\phi) + \phi \chi \Delta \log y_{t-1} + [\chi + (1-\rho\phi)^{-1}] v_{1t} + v_{2t} \quad (38)$$

so that (37) and (38) are the reduced form of the system. If we follow Flavin, and allow the covariance of v_{1t} and v_{2t} , ω_{12} , say, to be unrestricted, then (37) and (38) are an exactly identified two equation SUR system, and the indirect least squares (maximum likelihood) estimate of χ is given by

$$\bar{\chi} = \frac{\text{cov}(\Delta \log y_{t-1}, \Delta c_t / y_{t-1})}{\text{cov}(\Delta \log y_{t-1}, \Delta \log y_t)} \quad (39)$$

This expression does not involve any covariance between the change in consumption and the current change in income, and given that $\Delta \log y_t$ is first-

order autocorrelated, the probability limit of $\tilde{\chi}$ is zero if, and only if, $\Delta c_t/y_{t-1}$ is uncorrelated with *lagged* changes in income. Hence when Flavin estimates an excess response of consumption to changes in current income, she is measuring the same thing that we are measuring, the failure of changes in consumption to be orthogonal to lagged changes in income. Note that in a different world, where changes in consumption are orthogonal to lagged information, but there is an excess sensitivity of consumption to *innovations* in income, then the plim of $\tilde{\chi}$ would be zero and the response of consumption to income innovations would show up in a positive estimate of ω_{12} . In fact, the estimate of ω_{12} in Flavin's paper is large and *negative*, as it must be if the smoothness of consumption is to be captured by the model.

Our results are therefore consistent with Flavin's in the sense that both studies find that changes in consumption are correlated with lagged information about income. Of course, in other respects the formulation in this paper is quite different from Flavin's. Our inclusion of lagged saving in the income generation equation is required to justify our claims that excess smoothness cannot be explained by consumers' superior information. Nor is it open to us to work with consumption and income data that have been detrended. Mankiw and Shapiro (1985) have demonstrated that detrending can seriously compromise econometric tests for orthogonality, and it is important to note that the results in Section 3 are not subject to this difficulty. But more seriously from our point of view, the assumption that income is stationary around a deterministic trend would prejudge the issue with which we began, "Is consumption too smooth?"

List of works cited:

- Blinder, A.S., and A.S. Deaton, (1985), "The time series consumption function revisited," *Brookings Papers on Economic Activity*, 2, 465-511
- Campbell, J.Y., and N.G. Mankiw, (1986a), "Are output fluctuations transitory?" NBER, Cambridge, Mass., processed. (September)
- Campbell, J.Y., and N.G. Mankiw, (1986b), "Permanent and transitory components in macroeconomic fluctuations," NBER, Cambridge, Mass., processed. (December)
- Campbell, J.Y., (1987), "Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis," *Econometrica*, forthcoming.
- Cochrane, J.H., (1986), "How big is the random walk in GNP?" Department of Economics, University of Chicago, processed. (April)
- Deaton, A.S., (1986), "Life-cycle models of consumption: is the evidence consistent with the theory?" NBER Working Paper No. 1910.
- Dickey, D.A., and W.A. Fuller, (1981), "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, 49, 1057-72
- Flavin, M., (1981), "The adjustment of consumption to changing expectations about future income," *Journal of Political Economy*, 89, 974-1009
- Granger, C.W.J. and P. Newbold, (1977), *Forecasting economic time-series*, New York. Academic Press.
- Hansen, L.P., and T.J. Sargent, (1981), "A note on Wiener-Kolmogorov prediction formulas for rational expectations models," *Economics Letters*, 8, 255-60.
- Harvey, A.C., (1981), *Time series models*, Deddington. Philip Allan.
- Mankiw, N.G., and M.D. Shapiro, (1985), "Trends, random walks, and tests of the permanent income hypothesis," *Journal of Monetary Economics*, 16, 165-74.
- Phillips, P.C.B., and P. Perron, (1986), "Testing for a unit root in a time series regression," Cowles Foundation Discussion Paper No 795, Yale University, New Haven, processed.
- Priestly, M.R., (1981), *Spectral analysis and time series*, London. Academic Press.
- Watson, M.W., (1986), "Univariate detrending methods with stochastic trends," *Journal of Monetary Economics*, 18, 1-27

West, K.D., (1986a), "Dividend innovations and stock price variability," Cambridge, NBER Working Paper No 1833. (*Econometrica*, forthcoming)

West, K.D., (1986b), "The insensitivity of consumption to news about income," Woodrow Wilson School, processed. (December)

Appendix

A logarithmic version of the permanent income model

Start from the "rainy day" equation (5), i.e.

$$s_t = -\sum_{i=0}^{\infty} (1+r)^{-i} E_t \Delta y_{t+i}. \quad (A1)$$

Subtract capital income from both sides of (A1) to give

$$c_t - rA_t / (1+r) = y_t - s_t = \frac{r}{1+r} [y_t + \sum_{i=1}^{\infty} (1+r)^{-i} E_t y_{t+i}] \quad (A2)$$

Dividing through by current labor income y_t gives

$$1 - (s_t / y_t) = \frac{r}{1+r} [1 + \sum_{i=1}^{\infty} (1+r)^{-i} E_t (y_{t+i} / y_t)] \quad (A3)$$

Now, for all $j > 0$,

$$(y_{t+j} / y_t) = \exp[j\mu + \sum_{k=0}^j (\Delta \log y_{t+k} - \mu)] \approx e^{j\mu} [1 + \sum_{k=0}^j (\Delta \log y_{t+k} - \mu)] \quad (A4)$$

where $\mu = E(\Delta \log y_t)$ is the (unconditional) mean of the rate of growth of labor income. If the final term in (A4) is substituted in the right hand side of (A3) and terms rearranged, we reach

$$1 - (s_t/y_t) \approx \frac{r}{(1+r)(1-\rho)} \left[1 + \sum_{i=1}^{\infty} \rho^{-i} E_t(\Delta \log y_{t+i} - \mu) \right] \quad (\text{A5})$$

where $\rho = [(1+\mu)/(1+r)] \approx 1+\mu-r$. Provided the ratio of saving to labor income is sufficiently small, we can take logs of both sides and approximate once again to give:

$$(s_t/y_t) \approx -\sum_{i=1}^{\infty} \rho^i E_t \Delta \log y_{t+i} - \kappa \quad (\text{A6})$$

$$\kappa = \log(r/(1+r)) - \log(1-\rho) - \mu\rho/(1-\rho) \quad (\text{A7})$$

Equation (A6) is equation (11) in the main text. From (A5), we have

$$E(s_t/y_t) \approx 1 - \frac{r}{(1+r)(1-\rho)} \approx r - \frac{\mu}{r} \quad (\text{A8})$$

so that the saving ratio is small if r is small and if μ/r is small.