SM-88-9

A Series of Occasional Papers in Draft Form Prepared by Members of the Research Department for Review and Comment

STAFF MEMORANDA

IMPERFECT INFORMATION AND THE PERMANENT INCOME HYPOTHESIS

Abhijit V. Banerjee and Kenneth N. Kuttner

FEDERAL RESERVE BANK OF CHICAGO

Imperfect Information and the Permanent Income Hypothesis

Abhijit V. Banerjee Kenneth N. Kuttner [1]

August 4, 1988

[1]The authors are grateful to Gary Chamberlain, Benjamin Friedman, Zvi Griliches, Greg Mankiw, Knut Mork, Philippe Weil, and to participants in the NBER Consumption Group for comments; and to Iain Cockburn, Cheri Minton, and Andy Mitrusi for valuable technical assistance.

Abstract

The purpose of this paper is to explore the nature of the information set used by consumers in making their counsumption decisions. Specifically, it re-examines the evidence for the Permanent Income view of consumption under the assumption that consumers may not always be able to distinguish transitory income shocks from permanent shocks. For these 'indistinguishable' shocks, we assume that consumers use an optimal linear forecast to calculate the annuity value of the shock. Because this implies that the consumer treats some portion of each temporary shock as if it were permanent, the resulting response of consumption would appear, in the 'distinguishable shocks' context, to be excessive. This hypothesis offers an explanation for the excess sensitivity puzzle reported by earlier econometric studies of consumption.

The contribution of this paper lies in its attempt to estimate a parameter describing the 'amount' of information utilized by consumers, i.e., the degree to which income shocks can be discerned as either transitory or permanent. This methodology is also relevant for the evaluation of business-cycle theories which rely on agents' confusion between kinds of shocks to generate output fluctuations.

The first section of the paper discusses a simple example, demonstrating that incorrectly attributing to consumers the ability to distinguish transitory from lifetime shocks can lead to erroneous conclusions, such as the spurious finding of excess sensitivity. The next section proposes a general model, incorporating both distinguishable and indistinguishable shocks, and discusses its identification. The next section covers the data issues, and the minimum-distance estimation method. The final two sections summarize our empirical findings, and draw some general conclusions about the role of information in modeling consumer behavior. The empirical results we present are inconclusive, but do appear to offer some weak evidence against the Perfect Information view embodied in much of the literature.

KEYWORDS: Permanent Income Hypothesis, dynamic factor models, minimum-distance estimation.

1 Introduction

The responsiveness of consumer spending to changes in income is of vital importance for policy questions, particularly those relating to demand management. Empirically assessing this responsiveness is also important in a larger sense, in that one can ask to what extent it is 'rational,' or consistent with the efficient exploitation of all available information at the disposal of consumers.

Studies which test the Permanent Income Hypothesis have usually distinguished between two aspects of this responsiveness. The first is the responsiveness of consumption to anticipated changes in income; this, according to the rational expectations version of the Permanent Income Hypothesis, should always equal zero.[1] The second concerns the *magnitude* of the response of consumption to income 'surprises'; studies examining this second angle address themselves to the question of whether the sensitivity of consumption to these income innovations is appropriate, given plausible values of the prevailing interest rate. Empirical work attempting to assess this sensitivity tends to find a level of sensitivity far in excess of what can be justified by Permanent Income theory.[2]

Our purpose in this paper is to reexamine both of these issues, with particular emphasis on the second, 'excess sensitivity' issue, taking into account alternative assumptions regarding the consumer's information set as it affects the measurement of the key sensitivity parameter.

The issue of how much information to attribute to agents in a rational expectations model is a thorny one; consumers are assumed to respond rationally to all available information, but what to include in that information set is usually unclear, and left unspecified by the theory. Nor is it usually apparent how to determine from the data the answer to this question. Yet the correct choice is essential in empirical work; if we model behavior as the response to unanticipated 'surprises,' the econometric specification of that surprise will

[1]Hall (1978) was the first to emphasize this orthogonality property. Flavin (1981) re-interprets the orthogonality property in terms of the response to anticipated income shocks.

[2]Hall and Mishkin (1982) is the first study to uncover this excess sensitivity phenomenon.

certainly affect our interpretation of the response.

This issue is particularly germane to empirical work in consumption. Those studies which purport to estimate the response of consumption to income innovations have so far paid scant attention to this problem and its implications. The original work by Hall and Mishkin in this area, for example, and its successors[3] assert that consumers possess enough information to discern two distinct kinds of shocks to their income: high persistence (lifetime), and low persistence (transitory).

While it is easy to think of a few examples of unmistakably transitory income shocks (e.g., lottery payoffs, temporary tax surcharges), and a few shocks with an identifiably longer persistence (e.g., the Tax Reform Act of 1986), the majority of changes to households' incomes would seem very difficult to classify as either lifetime or transitory. Changes in real income due to movements in the price level, indirect taxes, and 'temporary' layoffs are good examples of what could be called 'indistinguishable' shocks.[4]

One of the goals of this paper is to examine explicitly the nature of the information set available to consumers when they plan their consumption expenditures. In particular, it seeks a way of empirically discerning what part of (the variance of) consumption is due to 'distinguishable' shocks, as opposed to 'indistinguishable' shocks.[5]

The answer to this question has important implications for policy, and for a proper appraisal of the Permanent Income Hypothesis. One of the implications of our work, for example, is that it can provide an explanation for the excess sensitivity puzzle reported by Hall and Mishkin. Specifically, we show that estimating the permanent income consumption model under the Perfect Information assumption of Hall and Mishkin, when the Imperfect Information model is true, will deliver an estimate of the senstivity parameter (representing

[3]Other work incorporating the same the same kinds of informational assumptions includes papers by Mork and Smith (1986), and by Altonji, Martins and Siow (1988).

[4]Even income movements which are the direct result of announced 'permanent' or 'temporary' policy measures might be subject to problems of dynamic inconsistency, and therefore be placed, to some degree, in the 'indistinguishable' class. After all, who knows what a Democrat in the White House might do with the Tax Reform Act of 1986?

[5]Of course, a finding in favor of indistinguishable shocks need not imply that *no* distinction can be made regarding the persistence of shocks — simply that most of the observed variation in income is of the indistinguishable type.

the annuity value of a transitory income shock) which is inconsistent and biased upwards, spuriously indicating excess sensitivity.

Moreover, if consumers are uninformed about the sources of income shocks, policy measures that affect income will produce changes in consumption whose dynamics are quite different from those we would see if consumers could actually discern the nature of their income shocks, and could use this information to divine their true persistence. In particular, evidence in favor of the 'indistinguishability' hypothesis could be interpreted as indirect evidence in favor of the sort of price-level versus relative-price confusion which is key to much of modern business cycle theory, as in Lucas (1972).

In order to examine these issues, we construct several models of consumption which include features of the Rational Expectations - Permanent Income Hypothesis. Alternative specifications we consider include the 'Perfect Information' version, which embodies the distinguishability hypothesis implicit in Hall and Mishkin, and an alternative 'Imperfect Information' hypothesis, which drops the distinguishability assumption. We also consider nested specifications which include both of these models as special cases, and include, in the tradition of earlier studies, a portion of 'rule-of-thumb' consumers, whose consumption tracks income one-for-one. We use an Optimal Minimum Distance (OMD) technique to fit these models on family-level food expenditure and labor income data from a subset of the Panel Study of Income Dynamics (PSID), for the 1978-1984 survey years. This OMD method has been shown to have advantages over maximum-likelihood methods in the presence of non-normal disturbances and conditional heteroskedasticity. We estimate the model both using unweighted data, and using data weighted by each household's mean income level to correct for heteroskedasticity.

Our empirical results confirm the Hall-Mishkin finding of excess sensitivity — but only when we impose the Perfect Information restrictions. In contrast, when we impose the restrictions implied by the Imperfect Information hypothesis, this excess sensitivity result vanishes; consumption appears to respond more or less as predicted by the Permanent Income Hypothesis, although large standard errors make a precise assessment of this sensitivity difficult.

Actually nesting the Perfect and Imperfect Information specifications proves to be somewhat problematic, however. Although the respective hypotheses imply distinct covariance patterns between changes in consumption and subsequent changes in income, the differences are subtle, and the data are not particularly sympathetic to either specification. The result is that the two key parameters β and ϕ, the sensitivity of consumption to income innovations and the proportion of 'perfect information' consumption, are not well identified separately.

The orthogonality restrictions implied by the Permanent Income Hypothesis are accepted in the unweighted data, but rejected in the weighted data. Including a fraction of consumption attributed to rule-of-thumb consumers further muddies the empirical waters, as the rule-of-thumb component of consumption can also account for patterns of covariance between changes in income and lagged changes in consumption which are similar to those implied by both types of Permanent Income consumption.

The general picture to emerge from these results is that the data are slightly more consistent with Imperfect Information versions of the Permanent Income model than they are to Perfect Information versions. While the data do not appear to be rich enough to enable us to effectively discriminate between the two hypotheses, the Imperfect Information version, at least, yields an estimate of the sensitivity parameter which is more consistent with the implications of the Permanent Income Hypothesis than the one delivered by the Perfect Information version. None of the parametric models of consumption we try, including those which include rule-of-thumb behavior, appears to fit the data very well, however, as indicated by their χ^2 statistics.

2 An Illustrative Model

A simple model of life-cycle consumption will illustrate the consequences of inappropriately specifying the consumer's information set by attributing to the consumer knowledge about the source of each shock. Specifically, we show how estimating the sensitivity parameter, β, from the moment restrictions implied by the Perfect Information model (when the Imperfect Information model is true) will deliver an estimate of the senstivity parameter which is inconsistent and biased upwards. For the sake of illustration, we will discuss the case with a zero rate of time preference, a zero interest rate, and serially uncorrelated transitory income shocks. In Section 3, we will drop these assumptions, and cover a somewhat more realistic case with a constant interest rate (assumed to equal to the rate of time preference), and serially correlated transitory income shocks.

As in the Hall and Mishkin paper, there are two kinds of shocks to consumers' income: *lifetime* and *transitory*. Lifetime income shocks are assumed to exhibit infinite persistence, while the transitory shocks decay over time. In other words, a lifetime income shock *permanently* alters an individual's earnings prospects, while a transitory income shock reflects temporary 'blips' to earnings. The simplest stochastic specification of such a latent variable process is to model the lifetime income shocks as innovations in a random walk process, while the transitory component is simply white noise:

$$x_t = x_{t-1} + \epsilon_t \tag{1}$$

$$y_t = x_t + \eta_t. \tag{2}$$

Here, x is lifetime income, ϵ the shocks to lifetime income, y observed income, and η the transitory income shocks. The two components are assumed to be serially uncorrelated (for the time being), and uncorrelated with one another.

Obviously, good permanent-income consumers would, if they were able to discern the two kinds of shocks, consume the full amount of the lifetime income shock. On the other hand, rather than consume the full amount of the transitory income shock, rational consumers would clearly want to consume only the annuity value of the amount of the shock (at some appropriate interest rate), thereby spreading the windfall over the duration of their lifetimes.

The assumption that these shocks are distinguishable to the consumer is, as we have argued, inappropriate for many of the income changes we observe. The question we intend to explore is what the consequences would be of attributing to consumers more information on the nature of these shocks than they actually have, and estimating the consumption model *as if* consumers could separately discern the two components.[6]

In order to say something specific about the joint behavior of consumption and income, we need to specify a model of permanent-income consumption. For the purposes of this example, we will make use of the simplest model imaginable. We assume throughout that consumers maximize an additively separable, quadratic utility function in discrete time, in which the consumer knows his lifetime with certainty. For the time being, we also take both the rate of time preference and the rate of interest to be equal to zero. The maximization problem is therefore:

$$\max_{\{c\}} \sum_{i=0}^{T} E_t u(c_{t+i}),$$

where

$$u(c) = d_0 + d_1 c + d_2 c^2,$$

subject to the budget constraint:

$$\sum_{i=0}^{T} c_{t+i} = \overline{W}$$

where $\overline{W}$ is the sum of current assets plus the present value all future income. The budget constraint is assumed to hold *ex post*.[7]

With a zero interest rate, the following consumption rule solves the above maximization problem:

$$c_t = \frac{1}{T+1}\left(A_t + y_t + \sum_{i=1}^{T} E_t y_{t+i}\right)$$

where A_t is the value of the consumer's assets at the beginning of period t, and the expression within parentheses is the expected value of lifetime wealth. First differencing and using the

[6] In the vocabulary of dynamic factor models, of which this is an example, the issue is whether a one-factor model is more appropriate than a two-factor version.

[7] For the purposes of this model, we overlook the complications introduced by allowing the budget constraint to hold in an expected value sense, enabling the consumer to die with negative net worth. See Ritter (1988) for a discussion of this issue.

law of motion for A_t,

$$A_t = A_{t-1} + y_t - c_t,$$

yields an expression for Δc which, because it is a function entirely of the revision in the consumer's expectations about future income between period t and period $t+1$, embodies the random walk principle of Hall (1978) so long as those expectations are formed rationally:

$$\Delta c_{t+1} = \frac{1}{T}\left(\sum_{i=1}^{T} E_{t+1}y_{t+i} - \sum_{i=1}^{T} E_t y_{t+i}\right).$$

Because lifetime income is a random walk and transitory income is serially uncorrelated, the consumer's best forecast for income j periods hence is exactly the same as his one-step-ahead forecast of income:

$$E_t y_{t+j} = E_t y_{t+1}.$$

Using this fact in the expression for Δc yields a simplified expression in terms of revisions in expectations:

$$\Delta c_{t+1} = \frac{1}{T}\left(y_{t+1} + (T-1)E_{t+1}y_{t+2} - TE_t y_{t+1}\right). \tag{3}$$

This is the point at which the assumption about the nature of the information set available to the consumer becomes crucial. If the consumer can, in fact, distinguish the two shocks (thereby observing his own lifetime income, x), $E_{t+1}y_{t+2}$ is simply equal to x_{t+1}, and $E_t y_{t+1}$ is just x_t. In this Perfect Information case, the expression for Δc_{t+1} simplifies to:

$$\Delta c_{t+1} = \epsilon_t + \frac{1}{T}\eta_t. \tag{4}$$

These forecasts are clearly infeasible for consumers who are unable to distinguish one kind of shock from the other. In a sense, an uninformed consumer is suffering from an errors-in-variables problem similar to that experienced by econometricians trying to estimate permanent income models. Because he is unable to use the unobservable lifetime income to forecast his future income, the consumer must come up with a forecast based only on those elements in his information set. While each individual clearly has a large amount of idiosyncratic information on which he can base his forecast of future income (e.g., education, promotion prospects, etc.), we will model his prediction problem as if he had only the information in his earnings history to go on.

The construction of an optimal forecast rule for this restricted information set is simplified by the observation that the second moments of a latent variable time series process such as Equations 1 and 2 are equal to the second moments of an alternative ARMA process. In other words, to someone who could not discern the underlying latent variables, Δy would 'look' just like an ARMA process. An uninformed consumer, who could not separately discern the latent variables in Equations 1 and 2 could, therefore, construct an optimal linear forecast of his earnings based on this corresponding ARMA process. In our case, where lifetime income follows a random walk and transitory income is serially uncorrelated, Δy can be written as:

$$\Delta y_t = \epsilon_t + \eta_t - \eta_{t-1},$$

with autocovariances:

$$\begin{aligned} E(\Delta y_t^2) &= \sigma_\epsilon^2 + 2\sigma_\eta^2 \\ E(\Delta y_t \Delta y_{t-1}) &= -\sigma_\eta^2 \\ E(\Delta y_t \Delta y_{t-k}) &= 0 \text{ for } k \geq 2. \end{aligned}$$

Because the autocovariances of Δy are zero beyond the first, one can find some MA(1) process which will generate exactly the same set of autocovariances as those generated by our latent variable model. If b is the moving-average parameter of the corresponding MA(1) process, then Δy can be written as:

$$\Delta y_t = (1 + b)e_t$$

with autocovariances:

$$\begin{aligned} E(\Delta y_t^2) &= (1+b^2)\sigma_e^2 \\ E(\Delta y_t \Delta y_{t-1}) &= b\sigma_e^2 \\ E(\Delta y_t \Delta y_{t-k}) &= 0 \text{ for } k > 1. \end{aligned}$$

Equating the two sets of autocovariances and solving for b as a function of σ_η^2 and σ_ϵ^2 yields the following expression for b:[8]

$$b = -1 - \frac{\sigma_\epsilon^2}{2\sigma_\eta^2} + \frac{\sigma_\epsilon}{\sigma_\eta}\sqrt{1 + \frac{\sigma_\epsilon^2}{4\sigma_\eta^2}}.$$

Using the standard forecast rule for MA(1) processes,

$$E_t y_{t+1} = (1+b)\sum_{i=0}^{\infty}(-b)^i y_{t-i},$$

to substitute for the expectations in Equation 3 yields an error-learning equation for Δc:

$$\Delta c_{t+1} = \left(1 + b\frac{T-1}{T}\right)(y_{t+1} - E_t y_{t+1}) \tag{5}$$

where the MA forecast error, in terms of the latent variables, is:

$$y_{t+1} - E_t y_{t+1} \equiv e_{t+1} = \epsilon_{t+1} + \eta_{t+1} + \sum_{i=0}^{\infty}(-b)^{i+1}\epsilon_{t-i} - (1+b)\sum_{i=0}^{\infty}(-b)^i \eta_{t-i}.$$

While obviously not orthogonal to lagged ϵ or η individually, this forecast error (and the corresponding change in consumption) will be uncorrelated with all elements in the *uninformed* consumer's information set: that is, all lagged changes in his *observed* income. This orthogonality condition places a testable restriction on the covariance matrix between

[8]This is identical to the expression derived by Muth (1960) by explicitly minimizing the mean-square error of a linear forecast of a random walk with measurement error.

Δc and lagged Δy:

$$E(\Delta c_t \Delta y_{t-k}) = 0 \text{ for } k \geq 1.$$

In the Imperfect Information case, multiplying the error-learning equation for Δc (Equation 5) by (leads of) the expression for Δy and taking expectations yields the other restrictions on the elements of the covariance matrix:

$$E(\Delta c_t \Delta y_t) = \left(1 + b\frac{T-1}{T}\right)\left(\sigma_\epsilon^2 + (2+b)\sigma_\eta^2\right) \tag{6}$$

$$E(\Delta c_t \Delta y_{t+1}) = -\left(1 + b\frac{T-1}{T}\right)\sigma_\eta^2. \tag{7}$$

On the other hand, multiplying the Perfect Information consumption rule in Equation 4 by Δy and taking expectations yields a different set of covariance restrictions:

$$E(\Delta c_t \Delta y_t) = \sigma_\epsilon^2 + \frac{1}{T}\sigma_\eta^2 \tag{8}$$

$$E(\Delta c_t \Delta y_{t+1}) = -\frac{1}{T}\sigma_\eta^2. \tag{9}$$

Regardless of the assumptions made about the consumer's information set, the autocovariances of the income process are:

$$E(\Delta y_t^2) = \sigma_\epsilon^2 + 2\sigma_\eta^2 \tag{10}$$

$$E(\Delta y_t \Delta y_{t-1}) = -\sigma_\eta^2. \tag{11}$$

Under either assumption about the nature of the consumer's information, these four equations can be used to identify, from the estimated covariances of Δc and Δy, the parameters of the income process and the consumption model. The task is to show how changing the informational assumption alters the mapping from the structural parameters of the consumption model to the moments of the joint distribution of Δc and Δy in such a way as to lead to biased and inconsistent estimates of the parameters of the consumption model.

Our concern here is the responsiveness of consumption to income innovations, which we parameterize with β. In the context of the Perfect Information model, β can be thought of as the annuity value of a transitory shock, which reflects a combination of the interest rate used to discount future income, the length of the consumer's planning horizon, and the persistence of transitory income shocks. In this illustrative model, β is a relatively uninteresting quantity, simply equal to $\frac{1}{T}$. (In the extended model, β will depend not only on the length of the horizon, but also on the prevailing interest rate, and the persistence of the transitory shocks.)

The definition of the β is exactly the same in the pure Imperfect Information case as in the Perfect Information case. With no distinction between transitory and permanent shocks, however, its interpretion is somewhat less obvious. The correct interpretion of β parameter is still as an index of the sensitivity of consumption to income innovations. Now, β measures the degree to which this sensitivity exceeds the 'baseline' (infinite horizon) sensitivity, $(1+b)$. Alternatively, the Imperfect Information β can be thought of as responsiveness to transitory shocks consistent with the observed response to indistinguishable shocks.

Our plan is to explore the sensitivity issue by estimating the consumption model, comparing the estimate of β with what could be thought of as reasonable values for that parameter. In the Perfect Information case, one can identify β through the covariance of Δc_t with Δy_{t+1} (Equation 8), and the first autocovariance of Δy (Equation 11), since, in the perfect information case,

$$\beta = \frac{E(\Delta c_t \Delta y_{t+1})}{E(\Delta y_t \Delta y_{t-1})}.$$

Forming the ratio of the sample analogs of these moments should, if the model is correctly specified, deliver a consistent estimate of β.

If, on the other hand, the sample covariances are generated by the consumption of individuals who are unable to distinguish between the two kinds of shocks, then using this ratio to identify β yields an inconsistent $\hat{\beta}$, a linear combination of the true β and 1 with weights $-b$ and $1+b$ (where b is negative):

$$\hat{\beta} = (1+b) - b\beta.$$

Thus, in this simple example, the $\hat{\beta}$ obtained from estimating the model under the incorrect assumption of perfect information would be subject to a potentially serious inconsistency problem, leading to an overstatement of the response of consumption to income.[9] Such a problem could at least partially account for the Hall-Mishkin finding of excess sensitivity in the response of consumption to innovations in transitory income.

3 The Extended Model

In this section, we extend the basic model of consumption outlined in the preceding section to include two additional features: serially correlated transitory income shocks, and a nonzero (but constant) interest rate. We also expand the specification to allow for advance information about changes in income, and construct a specification which includes both distinguishable and indistinguishable shocks, thereby nesting the Hall-Mishkin restrictions within a more general model. Finally, because the PSID study covers only food expenditures (rather than total consumption) for most years, we modify the model to describe the behavior of food consumption.

3.1 Serially Correlated Transitory Income Shocks

In order to approximate the dynamics of the changes households' earnings, we model the transitory income component as following an AR(1) process, while the lifetime income component continues to be a random walk with uncorrelated errors. In the example above, with white noise transitory income shocks, the time series process which replicated the autocovariances of the latent variable process for Δy was an ARMA(0,1); here, the equivalent time series process is an ARMA(1,1). We will briefly sketch the mapping between the two

[9] The β parameter is not the only one for which bias may be a problem when the information set is misspecified. Because the PSID data set reports only food expenditures, it is necessary to jointly estimate both β and α, the slope of the Engel curve for food. A similar argument can be made that the incorrect specification will lead to an inflated estimate of α, if it is also estimated using sample covariances.

representations. In terms of the latent variables, the income model is:

$$(1-L)x_t = \epsilon_t \quad (12)$$

$$y_t = x_{t-1} + \eta_t \quad (13)$$

$$(1-\phi L)\eta_t = \nu_t \quad (14)$$

where η and ν are white noise. L denotes the lag operator.

As above, if consumers are unable to make out the lifetime and transitory components separately, this latent variable process will look to them just like some ARMA process. Specifically, Equations 12 through 14 can be rewritten in terms of Δy as:

$$(1-\phi L)\Delta y_t = \Delta \nu_t + (1-\phi L)\epsilon_t,$$

which is recognizable as an ARMA(1,1) in Δy, with a composite error term consisting of terms in ν and ϵ.

Calling the autoregressive parameter a and the moving-average parameter b, the corresponding ARMA process can be written as:

$$(1-aL)\Delta y_t = (1+bL)e_t.$$

Equating the autocovariances of the two representations and solving for a, b, and σ_e^2 in terms of ϕ, σ_ϵ^2, and σ_ν^2 yields expressions for the parameters of the ARMA(1,1) representation in terms of ϕ and the ratio of the variances of the lifetime and the transitory shocks:

$$a = \phi$$

$$b = \frac{-1 - \frac{1}{2}(\phi^2+1)\frac{\sigma_\epsilon^2}{\sigma_\nu^2} + (1-\phi)\frac{\sigma_\epsilon}{\sigma_\nu}\sqrt{\frac{1}{4}(1+\phi)^2\frac{\sigma_\epsilon^2}{\sigma_\nu^2}+1}}{\phi\frac{\sigma_\epsilon^2}{\sigma_\nu^2}+1}.$$

Having ascertained the parameters of the appropriate ARMA process, all that remains is to insert the forecast rules:

$$E_t\Delta y_{t+1} = (\phi+b)\sum_{i=0}^{\infty}(-b)^i\Delta y_{t-i}$$

$$E_t\Delta y_{t+k} = \phi^k E_t\Delta y_{t+1} = \phi^{k-1}(\phi+b)\sum_{i=0}^{\infty}(-b)^i\Delta y_{t-i}$$

into the error-learning consumption equation of our uninformed consumer, and use the result to generate the corresponding moment restrictions.

3.2 Nonzero Rate of Interest

While the introduction of a nonzero interest rate and serially correlated transitory income complicates matters somewhat, the consumer's decision rule retains the error-learning structure it had in the simpler version of the model. The main difference is in the definition of the sensitivity parameter, β, which now is a function of the interest rate and the serial correlation parameter, as well as the length of the consumer's horizon. Specifically, the β which describes the 'correct' degree of sensitivity to income surprises takes the form:

$$\beta \equiv \omega/\mu, \text{ where}$$

$$\mu \equiv \frac{1-(\frac{1}{1+r})^{T+1}}{1-(\frac{1}{1+r})}, \text{ and } \omega \equiv \frac{1-(\frac{a}{1+r})^{T+1}}{1-(\frac{a}{1+r})}.$$

As T increases, the β parameter, defined in this way, approximates the interest rate, r; in the limit as $T \to \infty$, $\beta \to r/(1+r-a)$. Both the Imperfect Information and the Perfect Information versions of the consumption rule can now be rewritten in terms of this β parameter.

At this point, it is convenient to introduce the modification required to estimate the model on food expenditure alone. Going on the assumption that the Engel curve for food is approximately linear, with slope equal to α and a nonzero intercept, the only modification required is straightforward, and simply involves inserting this α parameter into the equations describing the consumption rule. Incorporating these changes, and reinterpreting Δc as referring to food consumption only, yields the following generic expression for Δc:

$$\Delta c_t = \frac{\alpha}{\mu}\sum_{i=1}^{T}\left(\frac{1}{1+r}\right)^i (E_t y_{t+i} - E_{t-1}y_{t+i}).$$

To derive the imperfect information version, we combine this expression with that for the ARMA(1,1) forecast rule, yielding an error-learning equation for Δc as a function of the period t forecast error:

$$\Delta c_t = \alpha\left(\frac{1+b}{1-\phi} - \frac{\phi+b}{1-\phi}\beta\right)(y_t - E_{t-1}y_t). \tag{15}$$

The perfect information model, on the other hand, implies a very different consumption rule, specifying a separate response to each component:

$$\Delta c_t = \alpha\epsilon_t + \alpha\beta\nu_t. \tag{16}$$

As in the simple example of the preceding section, this error-learning rule for consumption implies a specific set of restrictions on the elements of the covariance matrix of Δc and the leads and lags of Δy. Also, as before, it implies the orthogonality condition between Δc and all lagged elements of the information set, including lagged changes in y.

The other restrictions on the covariance matrix implied by the imperfect information model can be found by substituting the ARMA forecast error (in terms of the latent variables) into the error learning rule, multiplying by Δy, and taking expectations. They are:

$$E(\Delta c_t \Delta y_t) = \alpha\left(\frac{1+b}{1-\phi} - \frac{\phi+b}{1-\phi}\beta\right)\left(\sigma_\epsilon^2 + \frac{2+b-\phi}{1+\phi b}\sigma_\nu^2\right) \tag{17}$$

$$E(\Delta c_t \Delta y_{t+1}) = -\alpha\left(\frac{1+b}{1-\phi} - \frac{\phi+b}{1-\phi}\beta\right)\frac{(1-\phi)^2}{1+\phi b}\sigma_\nu^2 \tag{18}$$

where b is defined as above. On the other hand, the perfect information model implies a distinct pattern of covariance between Δc and Δy:

$$E(\Delta c_t \Delta y_t) = \alpha\sigma_\epsilon^2 + \alpha\beta\sigma_\nu^2 \tag{19}$$

$$E(\Delta c_t \Delta y_{t+1}) = -\alpha\beta(1-\phi)\sigma_\nu^2. \tag{20}$$

As before, the pattern of autocovariances of Δy is independent of the specification of the consumer's information set:

$$E(\Delta y_t^2) = \frac{2}{\phi+1}\sigma_\nu^2 + \sigma_\epsilon^2 \tag{21}$$

$$E(\Delta y_t \Delta y_{t-1}) = \frac{\phi-1}{\phi+1}\sigma_\nu^2 \tag{22}$$

$$E(\Delta y_t \Delta y_{t-2}) = \phi\frac{\phi-1}{\phi+1}\sigma_\nu^2. \tag{23}$$

Comparing Equations 15 and 16, the consumption rules for the Imperfect and the Perfect Information cases, it is clear that the competing hypotheses imply qualitatively different reactions in response to a shock to income. One way to see this is to compare the expressions for the covariance between Δy_{t+1} and Δc_t.

Consider first the perfect information case, and consider a period in which an individual receives a positive transitory income 'blip.' If consumption behaves according to Equation 16, Δy_{t+1} and Δc_t will be negatively correlated for the following reason: In the current period, our consumer will adjust his consumption upwards. In the subsequent period, income will fall — but the consumer *knew* it was going to, so consumption won't change. The correlation between this period's increase in consumption and next period's decrease in income generates the negative correlation. Furthermore, as β approaches zero, this negative covariance in the Perfect Information case shrinks to zero, since the accompanying change in consumption also shrinks to zero.

On the other hand, under the Imperfect Information assumption, even as β goes to zero, we will see a negative covariance between Δy_{t+1} and Δc_t as successive observations of y yield additional information on whether the initial change in income was a lifetime or transitory income shock. Thus, it is the covariance of Δc_t with Δy_{t+1}, relative to $Cov(\Delta c_t \Delta y_t)$ and $Var(\Delta y_t)$, which will be key in attempting to distinguish the competing hypotheses.

Before we estimate the model we will first discuss two additional ways in which we augment the basic indistinguishable shocks model of Section 2, and then discuss the question of identification in that more general model.

3.3 Allowing For Advance Information

A striking feature of the household consumption data is that the correlation between the current change in consumption and the change in income one period hence is significantly *positive*, rather than negative as implied by the Permanent Income Hypothesis under either informational assumption. The standard explanation of this phenomenon is that consumers, when they make their 'current' consumption decision, already have some advance information about the subsequent period's innovations in income. The timing of the PSID survey corroborates this interpretation. The survey, which is administered in March of every year, contains income questions which refer to the *previous* year's earnings. The consumption questions, on the other hand, are usually interpreted as pertaining to *current* consumption. Therefore, it is only natural that consumption in March of 1984 should already have responded to some of the earnings news for 1984.

To compensate for this timing problem, we assume, following Hall and Mishkin, that the 'true' current change in consumption is a convex combination of the theoretical 'current' change in consumption and 'next period's' change in consumption. If γ is the 'proportion with no advance information,' then:

$$\begin{aligned} \Delta c_t &= \gamma\alpha\left(\frac{1+b}{1-\phi} - \frac{\phi+b}{1-\phi}\beta\right)(y_t - E_{t-1}y_t) \\ &\quad + (1-\gamma)\alpha\left(\frac{1+b}{1-\phi} - \frac{\phi+b}{1-\phi}\beta\right)(y_{t+1} - E_t y_{t+1}). \end{aligned}$$

One justification for this specification is that it corresponds to a model of information propagation in which there is a γ probability every period that the next period's innovation in income will be known before this period's consumption decision is taken.

3.4 Nesting The Two Models

While it seems plausible that consumers often cannot distinguish between temporary and permanent shocks to their incomes, it is unlikely that they could never make this distinction. To allow for the possibility that consumers face a situation of *partial* information, we will assume that the consumer faces *two* sets of income shocks: one distinguishable set, and one indistinguishable set. Here, we will describe a fully general (but underidentified) specification, and will later discuss the restrictions on the model required to achieve identification. The required restrictions will condense this general specification into a convex combination of the Hall-Mishkin model and our imperfect information model. We will start with an unrestricted latent variable specification for the Δy process:

$$\Delta y = \epsilon_{1,t} + \eta_{1,t} - \eta_{1,t-1} + \epsilon_{2,t} + \eta_{2,t} - \eta_{2,t-1}$$

where $\epsilon_{1,t}$, $\epsilon_{2,t}$ $\eta_{1,t}$, and $\eta_{2,t}$ are independent of each other, and are described by the same stochastic processes as ϵ and η above (i.e., $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are white noise with variances $\sigma^2_{1,\epsilon}$ and $\sigma^2_{2,\epsilon}$, while $\eta_{1,t}$ and $\eta_{2,t}$ follow the processes $(1-\phi_1 L)\eta_{1,t} = \nu_{1,t}$ and $(1-\phi_2 L)\eta_{2,t} = \nu_{1,t}$, where the variance of $\nu_{1,t}$ and $\nu_{2,t}$ are $\sigma^2_{1,\nu}$ and $\sigma^2_{2,\nu}$).

We now assume that the consumer cannot separately observe $\epsilon_{1,t}$ and $\eta_{1,t}$, but *can* separately observe $\epsilon_{2,t}$ and $\eta_{2,t}$. This makes our augmented consumption model a sum of two parts: one which corresponds to the Imperfect Information model (Equation 15), and one which corresponds to the Perfect Information model (Equation 16). Ignoring, for the time being, the advance information complication, this extension implies the following equation for Δc:

$$\begin{aligned}
\Delta c_t &= \alpha \left(\frac{1+b}{1-\phi} - \frac{\phi+b}{1-\phi}\beta \right) e_t + \alpha \left(\epsilon_{2,t} + \nu_{2,t}\beta' \right) \\
e_t &= \epsilon_{1,t} + \nu_{1,t} - (\phi+b)\sum_{i=0}^{\infty}(-b)^{i+1}\epsilon_{1,t-1-i} - (1+b)\sum_{i=0}^{\infty}(-b)^{i}\eta_{1,t-1-i}
\end{aligned}$$

and β and β' are defined exactly as before for the two subprocesses.

This formulation nests the Perfect Information specification of Hall and Mishkin with the Imperfect Information specification, as special cases of a more general model. The two special cases correspond to setting $\epsilon_{1,t}$ and $\nu_{1,t}$, or $\epsilon_{2,t}$ and $\nu_{2,t}$ identically equal to zero. However, as we will argue in the next section, this most general formulation does not impose enough restrictions to achieve identification. In order to construct an estimable specification, we must impose a number of restrictions on the variances of the observed and unobserved components discussed above.

The system of equations we use to identify the parameters appears in full in the appendix. What follows is a brief discussion of some of the problems we face in identifying the nested versions of the model. There are essentially six variances and covariances which provide independent information about the parameters of this model. They are: $E(\Delta c_t \Delta y_t)$, $E(\Delta c_t \Delta y_{t+1})$, $E(\Delta c_t \Delta y_{t+2})$, $E(\Delta y_t^2)$, $E(\Delta y_t \Delta y_{t-1})$, and $E(\Delta y_t \Delta y_{t-2})$. As the model is specified above (incorporating both the advance information extension and the two classes of income innovations), there are nine parameters to identify: γ, β, β', ϕ_1, ϕ_2, $\sigma^2_{1,\epsilon}$, $\sigma^2_{2,\epsilon}$, $\sigma^2_{1,\nu}$, and $\sigma^2_{2,\nu}$.

With such an excess of parameters relative to the number of moments, the model as it stands is clearly underidentified. To identify the model, we proceed by assuming that the distinguishable and indistinguishable shocks come from populations with identical autoregressive parameters, and equal variances, up to a constant of proportionality:

$$\begin{aligned}
\phi_1 &= \phi_2 \\
\frac{\sigma^2_{1,\epsilon}}{1-\rho} &= \frac{\sigma^2_{2,\epsilon}}{\rho} \equiv \sigma^2_\epsilon \\
\frac{\sigma^2_{1,\nu}}{1-\rho} &= \frac{\sigma^2_{2,\nu}}{\rho} \equiv \sigma^2_\nu .
\end{aligned}$$

In addition to eliminating a ϕ and two σs, these assumptions also imply that $\beta = \beta'$. Introducing the constant of proportionality between variances, ρ, adds a single parameter, so that we are left with exactly six coefficients to estimate, not counting α, which will be estimated separately.

These assumptions essentially say that the two pairs of innovations to the income process are identical, except to the extent that one pair may have a higher variance than the other pair. One way to interpret this restricted model is the following: consumers receive only two kinds of shocks, ϵ and ν. Each year, some proportion of consumers, ρ, receive information on the source of their shocks, while the rest, $1-\rho$, receive no information on the source of their shocks.

This is a non-trivial assumption, since one can think of cases where the distinguishable shocks come from one kind of population (say, changes in direct taxes) while the indistinguishable shocks come from a different kind of population (say, changes in indirect taxes); however, it is unavoidable if we are to achieve identification. In this restricted form, our model nests both the Perfect and Imperfect Information models as special cases corresponding to $\rho = 1$ and $\rho = 0$, respectively.

3.5 Estimating α

One final identification problem remains. The α parameter, which appears in each of the $\Delta c - \Delta y$ elements of the covariance matrix, is not identifiable from covariances alone, unless the Perfect Information assumption is maintained. In this case, the fact that both the transitory and the lifetime factors have their own coefficients, $\alpha\beta$ and α, respectively, allows α to be identified from the covariances. This accords with intuition, which suggests that we can discern α from the response to a lifetime shock. Then, knowing α and observing the response to a transitory shock, we can 'back out' an estimate of β.

In the Imperfect Information situation, this is no longer the case: α and a term containing β always enter the covariance restrictions as a product, meaning that the two cannot be disentangled from this information alone. It is possible to verify from the system of equations in the appendix that even with some non-zero fraction of Perfect Information consumption, the presence of the ρ nesting parameter makes it impossible to determine α from the covariances.

We choose an alternate method of estimating the slope of the Engel curve, involving an auxiliary regression of consumption on income — in effect, using information from the levels of consumption and income, rather than the differences, to achieve identification.[10]

This is just the kind of regression which is susceptible to the errors-in-variables problem identified by Friedman (1957). The problem is that measured income is the sum of transitory and permanent components; a naïve regression of consumption on income will yield an attenuated estimate of the marginal propensity to consume food out of permanent income.

We remedy this problem by estimating the Engel curve on time averages of each household's data — that is, we regress the average consumption level for household i on the average income of that household in a regression of the form:

$$\bar{c}_i = \kappa + \alpha \bar{y}_i.$$

The idea behind this 'between' estimator of α is that the measurement errors induced by transitory income tend to cancel each other out over time, so that $\bar{y}$ is a relatively noise-free estimate of permanent income. Such a procedure will not completely eliminate measurement error, but will at least reduce it.

Table 1 below shows the equation we use to obtain our estimate of α, in which we also control for the number of household members. The point estimate from the linear version is 0.08; this is the value we will use in subsequent estimations. The nonlinear specification in Table 1 shows only a very small amount of curvature in the Engel curve for our sample, a fact we attribute to the homogeneous composition of the sample we selected for analysis. The next section describes this sample in greater detail.

[10] An additional benefit of this method is that it is free of the specification bias which, as we argued above, could contaminate the $\hat{\alpha}$ and $\hat{\beta}$ estimated from the covariances under the incorrect informational assumption.

Table 1: The 'Between' Estimate of α

	Independent Variable				$\overline{R}^2$
	Intercept	Y	Y^2	HHSIZE	
1	259.89 (30.50)	0.0803 (0.0026)	—	168.47 (7.79)	0.5122
2	230.35 (51.07)	0.0867 (0.0093)	$-2.7E-7$ $(3.7E-7)$	167.42 (7.92)	0.5121

Dependent Variable: Real food consumption
Y = Real disposable income
HHSIZE = Number of household members
Data are 1978-84 averages, in 1967 dollars.
Standard errors are in parentheses.

4 Estimation

4.1 The Data

The data we use come from the University of Michigan's Panel Study of Income Dynamics (PSID), from survey years 1978 through 1984.[11] We use only family-level data throughout, taking the household, as defined by PSID conventions, as being the appropriate level of aggregation for the analysis of consumption decisions. Bearing in mind the timing complication described above, we take the responses from each year's survey to refer to the previous year's values, but allow for consumption decisions to be made with advance information.

For our income variable, we use the sum of the labor incomes of the Head and the Wife (or 'Wife'), adjusted for federal income taxes and FICA payroll taxes.[12] For consumption, we use the sum of the household's expenditures on food at home and in restaurants. We deflate the income variable with the consumer price index to put it in terms of constant 1967 dollars. Similarly, using the food price component of the consumer price index, we express food consumption in terms of 1967 dollars.

Rather than using the entire PSID panel of 6,918 families (as of Wave 17), we choose to focus our analysis on a carefully selected subset of the panel. We first drop a number of observations which appear to be 'bad' data, outliers, or have some other (observable) problem. Second, in an attempt to avoid the problems involved in modelling non-Permanent-Income behavior, we omit observations on families which are likely to be constrained in one way or another.

The data we drop in the first round are the following:

1. Families which report zero labor income for both Head and Wife

[11] Technical constraints imposed by the computer software forced us to use only seven years of data (six sets of differences). However, because households are being continually added to the survey and because our method requires a balanced sample, including fewer years increases the number of observations available for estimation.

[12] Ideally, we would also want to adjust for state taxes in arriving a measure of changes in disposible income. However, because the PSID does not contain any state tax data, this would involve either adjusting by some representative marginal state tax rate, or combining the PSID information on the household's state of residence with statutory tax rate data to estimate each household's state taxes.

2. Families which report zero food expenditure
3. Very wealthy families (inflation adjusted needs ratio greater than 20)
4. Observations with a major assignment to food or income data
5. Families in which both the Head and the Wife were institutionalized, students, or non-participants in the labor force for any other reason
6. Observations in which the reported labor income was truncated by the number of digits on the PSID tape[13]
7. Families which reported a change in real income greater than 50 per cent in absolute value, relative to the previous year.

Second, in order to focus on the informational issue discussed above, we wish to concentrate our analysis on those families who are most likely to behave according to the Permanent Income Hypothesis. Accordingly, we drop those observations corresponding to very poor families whose behavior is likely to be liquidity constrained. Specifically, we discard observations of families whose inflation adjusted needs ratio was less than unity, and those of families which received food stamps during the sample period.

Finally, on the grounds that estimating a dynamic earnings structure as in Equations 12 and 13 makes little sense for retired people who generally have little or no labor income, we eliminate retirees from the sample. The result of this series of cuts is a relatively homogeneous balanced sample of 1,978 observations.

4.2 Estimation Method

In the most general terms, dynamic factor models of consumption, like those described above, act to place sets of restrictions on the covariance matrix of (the leads and lags of) Δy and Δc. Therefore, an estimation method which allows us to impose those restrictions

[13]For 1978-83, this amount was $99,999; for 1984, it was $999,999.

directly on the elements of the covariance matrix is better adapted to fitting these models than one which imposes restrictions on the ratios of the off-diagonal elements to the diagonal elements, as does the regression method.

The Minimum Distance estimation method is very well adapted to estimating a model with such a structure.[14] The idea is to minimize a quadratic criterion function of the form:

$$\min_{\theta} \, [g(\theta) - z]' \, A \, [g(\theta) - z]$$

where θ is an m-dimensional vector of parameters to be estimated, z is an n-dimensional vector of unconstrained estimates, and g is the mapping from the constrained parameter space to the unconstrained parameter space which incorporates the restrictions on the moments implied by the behavioral model. The minimum distance estimator, $\hat{\theta}$, is the θ which solves the first-order condition:

$$Dg(\theta)'A \, [g(\theta) - z] = 0.$$

Gauss-Newton iterations can be used to numerically solve this equation and obtain a value for $\hat{\theta}$.

It can be shown that, under very general conditions, the Minimum Distance estimator is consistent and asymptotically normal — regardless of the choice of weighting matrix, A, used in the minimization. If the inverse of the covariance matrix of z is used as the weighting matrix, then the minimum-distance method delivers the 'Optimal' Minimum Distance estimator, yielding the most efficient (relative to other choices of A) estimate of θ.[15] The method of Maximum Likelihood is analogous to using the inverse of the matrix of fourth moments implied by the normal distribution for A; for a non-normal distribution of disturbances, the ML method would therefore utilize a sub-optimal weighting matrix.

[14]Other applications of OMD estimation of covariance structures include Abowd and Card (1986), and Altonji, Martins and Siow (1987).

[15]See Chamberlain (1982) and (1984) for a complete presentation of the Minimum Distance method, and its optimality properties.

Another benefit of using $Cov(z)^{-1}$ as the weighting matrix is that the minimized criterion function, multiplied by the number of observations, is distributed asymptotically as a χ^2, with degrees of freedom equal to the number of restrictions placed on the model, i.e., $n - m$; in Table 3 we use this fact to perform χ^2 tests of various restrictions. In the results that follow, we use a feasible version of the Optimal Minimum Distance estimator, in which the A matrix is replaced by the inverse of the estimated covariance matrix of z, the matrix of sample fourth moments.

In our application, the θ vector consists of the parameters of the income process and the consumption model discussed above. The g function maps these parameters into our z vector, which is comprised of the unique elements of the sample covariance matrix, e.g.:

$$\frac{1}{N}\sum_{i=1}^{N}\Delta c^*_{i,t}\Delta y^*_{i,t\pm s},\ \frac{1}{N}\sum_{i=1}^{N}\Delta y^*_{i,t}\Delta y^*_{i,t-s} \text{ and } \frac{1}{N}\sum_{i=1}^{N}\Delta c^*_{i,t}\Delta c^*_{i,t-s},\ \text{for } s \leq 5,$$

where $\Delta c^*_{i,t} \equiv \Delta c_{i,t} - \overline{\Delta c_t}$ and $\Delta y^*_{i,t} \equiv \Delta y_{i,t} - \overline{\Delta y_t}$. Even though our models place no restrictions on the autocovariances of Δc, we include these moments in the z vector, but impose only the stationarity restrictions on that subvector.

5 The Results

The results we get are mixed. As it should be evident from Tables 2 and 3, the data resist our attempts to impose stationarity. This should not surprise the reader; most studies based on this data find the same rejection.[16] Orthogonality restrictions imposed by our theory are rejected in the weighted data but not in the unweighted data; the implications of this will be examined in the next section. The income process we impose is, however, accepted even at the 10% level which suggests that our ARMA(1,1) specification for Δy is not obviously worse than the ARMA(0,2) specification adopted by Hall and Mishkin.[17]

[16]See for example Altonji, Martins and Siow(1987).

[17]MaCurdy (1982) also concludes that the data are indifferent between the ARMA(1,1) and the ARMA(0,2) specifications.

The interpretation of our results should therefore be qualified by the fact that the basic structure we impose on the data to make estimation possible is not entirely supported by the data. A further problem arises because the relation between ρ and β in our model is highly nonlinear and, as a result, in our estimation of the nested model, yields two sets of values for ρ and β which generate the same values of the objective function. To make matters worse, one set of these values is typically outside the legitimate range. (Both ρ and β get negative point estimates although in general one cannot reject the hypothesis that they are both zero.) Our emphasis in interpreting the results will therefore be in the direction of comparing χ^2 values for different alternative hypotheses rather than looking at point estimates.

The strongest support for the Imperfect Information view comes from looking at the weighted data. The numbers in Tables 2 and 3 show that imposing the pure Imperfect Information restriction ($\rho = 0$) on the nested model causes the χ^2 value to change very slightly. On the other hand, imposing the Perfect Information restriction ($\rho = 1$) is rejected quite emphatically by the data. The results in the unweighted data are less clear cut. The pure Imperfect Information restriction has a somewhat lower χ^2 value than the Perfect Information hypothesis, but neither set of restrictions on the nested model can be rejected even at the 10% level. (At the 15% level we can reject the Perfect Information hypothesis but we cannot really insist that this is a very meaningful rejection).

The point estimates reported in Tables 4 and 5 correspond to the non-negative set of roots. As discussed above, there is some reason to be sceptical of the information value of these estimates. The one which seems most reasonable is the estimate for γ which comes out to be 0.51 in the unweighted case and 0.62 in the weighted case. At least the value from the weighted model, suggesting 38% advance information, is not inconsistent with Hall and Mishkin's claim that there is a lag of one quarter between the income and the consumption data. Also, the value of γ corresponding to the negative roots is not very different from these values, which suggests a degree of robustness in the estimate of this parameter.

The estimates of β and ρ reported in the tables are hard to interpret. When we estimate

Table 2: Goodness-of-Fit Statistics

Model	Restrictions	k	χ^2 Statistics	
			Unweighted	Weighted
0	Unrestricted	78	0.00	0.00
1	Stationarity Constrained	23	100.00	114.51
2	(1) Plus Orthogonality	18	106.51	126.95
3	(2) Plus Income Process	15	111.44	130.52
	Without Rule-of-Thumb Consumption			
4	Nested Perfect & Imperfect Info	12	124.21	140.60
5	Perfect Information Only	11	126.31	144.54
6	Imperfect Information Only	11	125.30	141.55
	With Rule-of-Thumb Consumption			
7	Nested Perfect & Imperfect Info	13	122.70	138.41
8	Perfect Information Only	12	122.88	139.59
9	Imperfect Information Only	12	123.60	138.53

Table 3: Tests of Alternative Specifications

Model		DF	Unweighted Data		Weighted Data	
H_A	H_0		χ^2 Stat	P-Value	χ^2 Stat	P-Value
0	1	55	100.00	0.00	114.51	0.00
1	2	5	6.51	0.26	12.44	0.03
2	3	3	4.93	0.18	3.57	0.31
3	4	3	12.77	0.01	10.08	0.02
4	5	1	2.10	0.15	3.96	0.05
4	6	1	1.09	0.30	0.95	0.33
7	8	1	0.18	0.67	1.18	0.28
7	9	1	0.90	0.34	0.12	0.73
7	4	1	1.51	0.22	2.19	0.14
7	5	2	3.61	0.16	6.13	0.05
7	6	2	2.60	0.27	3.14	0.21
8	5	1	1.43	0.23	4.95	0.03
9	6	1	1.70	0.19	3.02	0.08

Table 4: Parameter Estimates, Unweighted Data

Model	Parameters						
	σ_ϵ^2	σ_ν^2	ϕ	β	γ	ρ	ζ
3	1.19	0.96	0.31	—	—	—	—
	(0.10)	(0.09)	(0.05)				
4	1.21	0.94	0.33	0.47	0.51	0.69	0.00
	(0.11)	(0.09)	(0.05)	(0.14)	(0.06)	(0.24)	—
5	1.15	0.98	0.30	0.53	0.48	1.00	0.00
	(0.09)	(0.08)	(0.05)	(0.11)	(0.06)	—	—
6	1.21	0.94	0.33	0.08	0.52	0.00	0.00
	(0.11)	(0.09)	(0.05)	(0.13)	(0.06)	—	—
7	1.24	0.92	0.34	0.46	0.54	0.85	0.24
	(0.11)	(0.09)	(0.05)	(0.17)	(0.06)	(0.34)	(0.19)
8	1.24	0.92	0.33	0.47	0.54	1.00	0.29
	(0.11)	(0.09)	(0.05)	(0.16)	(0.06)	—	(0.14)
9	1.25	0.92	0.34	0.03	0.56	0.00	0.26
	(0.11)	(0.09)	(0.05)	(0.20)	(0.05)	—	(0.18)

Standard Errors are in parentheses.

Table 5: Parameter Estimates, Weighted Data

Model	Parameters						
	σ_ϵ^2	σ_ν^2	ϕ	β	γ	ρ	ζ
3	15.18	12.85	0.22	—	—	—	—
	(1.14)	(1.05)	(0.04)				
4	14.93	12.89	0.23	0.44	0.62	0.60	0.00
	(1.14)	(1.05)	(0.04)	(0.17)	(0.08)	(0.22)	—
5	14.11	13.44	0.20	0.42	0.63	1.0	0.00
	(1.04)	(0.98)	(0.04)	(0.14)	(0.10)	—	—
6	14.89	12.90	0.23	0.06	0.64	0.0	0.00
	(1.13)	(1.04)	(0.04)	(0.13)	(0.18)	—	—
7	15.41	12.57	0.24	0.34	0.66	0.61	0.29
	(1.19)	(1.07)	(0.04)	(0.29)	(0.07)	(0.37)	(0.19)
8	15.21	12.69	0.24	0.32	0.69	1.0	0.39
	(1.16)	(1.06)	(0.04)	(0.25)	(0.08)	—	(0.16)
9	15.42	12.56	0.24	-0.01	0.66	0.0	0.32
	(1.18)	(1.07)	(0.04)	(0.19)	(0.05)	—	(0.17)

Standard Errors are in parentheses.

ρ freely, the estimate comes out to be quite high (between 0.6 and 0.7), and it is accompanied by high β values (not less than 0.4).[18] The same kind of β values are generated if we impose the Perfect Information restrictions. On the other hand, imposing the pure Imperfect Information restrictions yields a β value of 0.06, which is perfectly consistent with the interest rates and lifetimes we consider reasonable, and the decline in the fit, as we saw above, is very slight. The fact that the deterioration in fit is so slight, in apparent contradiction to the high point estimates of ρ, is due to the non-linearity of the model, and the presence of the second root corresponding to a small, negative value of ρ. Small changes in the specification or the selection of an alternative subset of the PSID panel may make this second set of roots positive, in which case a stronger conclusion in favor of our hypothesis would be warranted.

[18] In constrast, with a rate of interest equal to 4% and a horizon of 25 years, the 'correct' value of β is approximately 0.05.

6 Rule-of-Thumb Consumption?

The most unsatisfactory aspect of our current results is the strong violation of the Euler equation restrictions in the weighted case. In combination with the fact that the Euler equation restrictions are not rejected in the unweighted case, this suggests that the violation may come from the behavior of the members of the sample with relatively low incomes. Intuitively one expects their influence on the results to be greater in the weighted case.

This suggests that the source of the trouble may be liquidity constraints on low income families. The trouble is, as Zeldes (1985) points out, that there is no simple rule for predicting how liquidity constrained agents would behave.[19]

The same problem also arises with the second candidate for an alternative hypothesis, namely, that at least some agents are not rational and use rules of thumb to decide their consumption. There is no obvious candidate for such a rule of thumb. Current fashion favors the so-called Keynesian consumption function, which simply says that $\Delta c = \Delta y$, but there seems to be no reason to prefer this over $\Delta c = k\Delta y$ where k is less than 1. Keynes himself preferred the latter and called it the 'fundamental psychological law.'[20] And as long as k is positive, the version with $k < 1$ is *a priori* no worse in explaining violations of the Euler equation.

In a future extension of this paper we will consider the question of the best specification for rule-of-thumb behavior. In this paper we limit ourselves to examining what happens if we add to our model the 'Keynesian' alternative mentioned above. Our motive for doing so is for comparability with other studies like Hall and Mishkin and Campbell and Mankiw (1988) which make this assumption.

The method we use for incorporating 'Keynesian' behavior follows Hall and Mishkin (1982). We assume that a fraction ζ of consumption is determined by the rule of thumb, $\Delta c = \Delta y$, and the rest is determined by the permanent income model, as above. We

[19] See Hall (1987) and Hayashi (1987) for surveys of the evidence on liquidity constraints.

[20] See Keynes (1936).

therefore write:

$$\Delta c = \zeta \Delta y + (1 - \zeta)\Delta c^p$$

where c^p is the consumption predicted by the permanent income model introduced above.

The method of incorporating rule-of-thumb consumption suffers from the defect that it actually implies that all agents sometimes follow a rule of thumb and sometimes follow the permanent income model. By contrast what we actually want is to be able to model the fact that some agents follow one of these models most of the time and that others follow the other model most of the time. As a result, Hall and Mishkin's interpretation of our ζ parameter as the fraction of rule-of-thumb consumers is not strictly correct.

We follow Hall and Mishkin in using the moments of consumption with lagged income to identify ζ. Our results again depend on which data set we use. With both sets we find that the restriction that $\zeta = 0$ cannot be rejected at the 5% level. The same is true if we compare the pure Imperfect Information model ($\zeta = 0,\ \rho = 0$) to the unrestricted model. The restriction of Perfect Information ($\zeta = 0,\ \rho = 1$) is rejected in the weighted data. If we allow ζ to be freely estimated, once again one cannot reject the restriction of pure Imperfect Information, but now one cannot reject Perfect information either, and in the unweighted data it actually performs slightly better. To a limited extent therefore, the introduction of rule-of-thumb consumption does make the Hall-Mishkin assumption perform better relative to the Imperfect Information hypothesis.

However, the conclusion of Hall and Mishkin that the inclusion of rule-of-thumb consumers is enough to generate reasonable point estimates is not confirmed by our study. In the Hall-Mishkin case ($\rho = 1$, unweighted data) the estimate of ζ we get, 0.29, is not so different from the estimate they report, 0.2, but the β we get, 0.47 is much larger than their estimate of 0.17 and cannot be reconciled with rational behavior. The only case where we obtain an estimate of β consistent with the theory is in the unrestricted case with weighted data. In this case, we get an estimate of 0.33 with a standard error of 0.28, but the finding that this is not inconsistent with the theory is mainly driven by the unusually large standard error. In all other cases the estimate of β is not really changed by the inclusion of rule-of-thumb consumers. The estimate of ρ is also surprisingly insensitive to the inclusion of rule-of-thumb consumers.

7 Conclusions

The most general conclusion to be drawn from this paper is that alternative specifications of the information set used in inferring the 'surprise' movements of a variable can make a substantive difference in the estimation of rational expectations models.

In the context of consumption, our specific result is that an alternative specification of consumers' information sets, which endows them with 'less' information than is customary, performs at least as well (in terms of fit) as the stronger Perfect Information specification. In addition, this Imperfect Information assumption is able to resolve the excess sensitivity puzzle found in other studies. In particular, when we impose the weaker informational restrictions, we obtain a point estimate of the sensitivity parameter which is justifiable in the presence of plausible interest rates and horizons, although a large standard error makes it difficult to do precise inference. In this sense, our results can be interpreted as being favorable to the Permanent Income Hypothesis, although some kind of rule-of-thumb behavior appears to be marginally relevant.

There are a number of reasons for caution in interpreting our results. First, the power of the covariances of food consumption with income to distinguish these competing hypotheses is very low. While the Imperfect Information versions appear to perform slightly better than the Perfect Information versions, the improvement in fit is marginal. It is therefore entirely plausible that the Perfect Information model is really the appropriate specification of consumer behavior, but that consumers simply respond with excess vigor to income shocks; there is not enough information in our data to reject one hypothesis in favor of the other. It may, however, be possible to extend our analysis to include additional 'indicator' variables, such as hours worked, asset income, or saving, which would yield information on households' expectations of future income, and by so doing, improve our ability to discern one informatonal hypothesis from the other. This remains a topic for future work.

A second caveat is due from the observation that the χ^2 statistics indicate that *all* of the consumption models we try fit the data rather poorly. Even the stationarity restrictions we impose at the outset fail spectacularly. While conclusions drawn from these misspecified models must be treated with caution, such models can serve as useful approximations for divining the structure in the data.

Finally, the problem of measurement error is one we were unable to properly address; although we made an attempt to discard outliers and data points to which major assignments were made, a proper measurement-error correction would require more degrees of freedom than we have at our disposal, using only income and consumption data.

Despite these caveats, we believe that the results presented here are a first step towards resolving some of the outstanding questions in the study of household-level consumer behavior.

A Appendix: The Nested Model

The following set of equations defines the mapping, g, from the parameters of the structural model of consumption and income to the covariances of Δc with Δy, and the autocovariances of Δy.

If lifetime income is a random walk, while transitory income is described by a stationary AR(1), then the autocovariances of Δy are:

$$\begin{aligned}
E(\Delta y_t^2) &= \frac{2}{\phi+1}\sigma_\nu^2 + \sigma_\epsilon^2 \\
E(\Delta y_t \Delta y_{t-1}) &= \frac{\phi-1}{\phi+1}\sigma_\nu^2 \\
E(\Delta y_t \Delta y_{t-2}) &= \phi\frac{\phi-1}{\phi+1}\sigma_\nu^2 \\
E(\Delta y_t \Delta y_{t-k}) &= \phi^{k-1}\frac{\phi-1}{\phi+1}\sigma_\nu^2.
\end{aligned}$$

If we define:

$$\begin{aligned}
k &= \left(\frac{1+b}{1-\phi} - \frac{\phi+b}{1-\phi}\beta\right), \text{ and} \\
b &= \frac{-1-\frac{1}{2}(\phi^2+1)\frac{\sigma_\epsilon^2}{\sigma_\nu^2} + (1-\phi)\frac{\sigma_\epsilon}{\sigma_\nu}\sqrt{\frac{1}{4}(1+\phi)^2\frac{\sigma_\epsilon^2}{\sigma_\nu^2}+1}}{\phi\frac{\sigma_\epsilon^2}{\sigma_\nu^2}+1},
\end{aligned}$$

then the restrictions placed on the covariances of Δc and Δy can be written:

$$\begin{aligned}
E(\Delta c_t \Delta y_{t-k}) &= 0 \text{ for } k \geq 1 \\
E(\Delta c_t \Delta y_t) &= \alpha\gamma\left[k(1-\rho)\left(\sigma_\epsilon^2 + \frac{2+b-\phi}{1+\phi b}\sigma_\nu^2\right) + \rho\left(\sigma_\epsilon^2 + \beta\sigma_\nu^2\right)\right]
\end{aligned}$$

$$E(\Delta c_t \Delta y_{t+1}) = \alpha k(1-\rho)\left[(1-\gamma)\left(\sigma_\epsilon^2 + \frac{2+b-\phi}{1+\phi b}\sigma_\nu^2\right) - \gamma\frac{(1-\phi)^2}{1+\phi b}\sigma_\nu^2\right]$$

$$+ \alpha\rho\left[(1-\gamma)\left(\sigma_\epsilon^2 + \beta\sigma_\nu^2\right) + \gamma\beta(1-\phi)\sigma_\nu^2\right]$$

$$E(\Delta c_t \Delta y_{t+2}) = -\alpha\left(\gamma\phi + (1-\gamma)\right)\left[k(1-\rho)\frac{(-\phi)^2}{1+\phi b} + \rho\beta(1-\phi)\right]\sigma_\nu^2$$

$$E(\Delta c_t \Delta y_{t+k}) = -\alpha\phi^{k-2}\left(\gamma\phi + (1-\gamma)\right)\left[k(1-\rho)\frac{(-\phi)^2}{1+\phi b} + \rho\beta(1-\phi)\right]\sigma_\nu^2 \text{ for } k \geq 3.$$

References

Abowd, J.M. and D. Card (1986), "On the Covariance Structure of Earnings and Hours Changes", NBER Working Paper #1832.

Altonji, J.G. and A. Siow (1987), "Testing the Response of Consumption to Income Changes with (Noisy) Panel Data", *The Quarterly Journal of Economics* 102, 293-328.

Altonji, J.G., A.P. Martins and A. Siow (1987), "Dynamic Factor Models of Consumption, Hours, and Income", NBER Working Paper #2155.

Altonji, J.G., A.P. Martins and A. Siow (1988), "Using Cross-Equation Restrictions between Asset Income and Non-Asset Income to Estimate the Permanent Income Hypothesis", Northwestern University.

Campbell, J.Y. and N.G. Mankiw (1987), "Permanent Income, Current Income, and Consumption", NBER Working Paper #2436.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data", *Journal of Econometrics* 18, 5-46.

Chamberlain, G., 1984, "Panel Data", in Griliches and Intriligator, eds: *Handbook of Econometrics* Volume II, Amsterdam: North-Holland.

Flavin, M.A. (1981), "The Adjustment of Consumption to Changing Expectations about Future Income", *Journal of Political Economy* 89, 974-1009.

Friedman, M., 1957, *A Theory of the Consumption Function.* Princeton, NJ: Princeton University Press.

Hall, R.E. (1978), "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence", *Journal of Political Economy* 86, 971-987.

Hall, R.E. and F. Mishkin (1982), "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households", *Econometrica* 50, 461-481.

Hall, R.E. (1987), "Consumption", NBER Working Paper #2265.

Hayashi, F., 1987, "Tests for Liquidity Constraints: A Critical Survey", in Bewley, ed.: *Advances in Econometrics, Fifth World Congress* Volume II, Cambridge: Cambridge University Press.

Keynes, J.M., 1936, *The General Theory of Employment, Interest, and Money.* London: Macmillan.

Lucas, R.E. (1972), "Expectations and the Neutrality of Money", *Journal of Economic Theory* 4, 103-124.

MaCurdy, T.E. (1982), "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis", *Journal of Econometrics* 18, 83-114.

Mork, K.A. and V.K. Smith (1986), "Testing the Life-Cycle Hypothesis on Panel Data Using Detailed Consumption Diaries and Income Based on Tax Records", Vanderbilt University.

Muth, J.F. (1960), "Optimal Properties of Exponentially Weighted Forecasts", *Journal of the American Statistical Association* 55, 299-306.

Ritter, J.A. (1988), "Endogenous Borrowing Constraints and Consumption", University of Texas.

Zeldes, S. (1985), "Consumption and Liquidity Constraints: An Empirical Investigation", Wharton School, University of Pennsylvania.