

NBER WORKING PAPERS SERIES

THE JOINT CONSUMPTION/ASSET DEMAND DECISION:  
A CASE STUDY IN ROBUST ESTIMATION

Marjorie Flavin

Working Paper No. 3802

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
August 1991

Comments welcome. The data utilized in this paper were made available by the Inter-University Consortium for Political and Social Research. The data were originally collected by the Survey Research Center of the University of Michigan. Neither the original source or collectors of the data nor the Consortium bear any responsibility for the analyses or interpretations presented here. Research support from the NSF is gratefully acknowledged. This paper is part of NBER's research program in Economic Fluctuations. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

THE JOINT CONSUMPTION/ASSET DEMAND DECISION:  
A CASE STUDY IN ROBUST ESTIMATION

**ABSTRACT**

The paper uses a previously unexploited data set -- the Michigan Survey of Consumer Finances -- to ask whether the finding that consumption tracks current income more closely than is consistent with the permanent income hypothesis can be attributed solely or partially to borrowing constraints. Using household data on income and asset stocks, the paper studies the saving side of the consumption/saving decision, and thus provides inferences on a comprehensive concept of consumption. To limit the influence of outliers, the paper uses a robust instrumental variables estimator, and argues that achieving robustness with respect to leverage points is actually simpler, both conceptually and computationally, in an instrumental variables context than in the OLS context.

The results indicate that households do use asset stocks to smooth their consumption, although this smoothing is far from complete. However, there is no evidence that the excess sensitivity of consumption to current income is caused by borrowing constraints. Compared to the conventional results, the robust instrumental variables estimates are more stable across different subsamples, more consistent with the theoretical specification of the model, and indicate that some of the most striking findings in the conventional results were attributable to a single, highly unusual observation.

Marjorie Flavin  
Department of Economics  
Rouss Hall  
University of Virginia  
Charlottesville, VA 22901  
and NBER

The paper starts by posing a straightforward economic question to a previously unexploited data set on household asset holdings. When the data fail to respond with an equally straightforward answer, the discussion embarks on a methodological detour concerning the impact, interpretation, and treatment of outliers. While the paper does eventually arrive at its intended destination in the sense that the original economic question gets a coherent answer, to many readers the detour -- basically a case study in robust estimation in an instrumental variables context -- will be the most interesting aspect of the paper. On the methodological issues, the paper argues that the robust instrumental variables estimator is an intuitively appealing method of dealing with the outliers common in household data sets, that the estimator is easy to implement computationally, and that the use of robust methods can make a very big difference to the economic conclusions drawn from the data.

In this case study, the economic application is the rational expectations permanent income hypothesis (RE-PIH). In tests based on aggregate data, there has emerged a high degree of consensus that consumption tracks current income more closely than is consistent with the RE-PIH. It is of obvious interest to determine whether the apparent "excess sensitivity" of consumption to current income found in the aggregate data also characterizes consumption behavior in micro data sets. If the finding is confirmed in data on households, it would also be useful to know whether the excess sensitivity arises because some households would like to smooth their consumption and are prevented from doing so by borrowing constraints, or because households fail to smooth their consumption even in the absence of borrowing constraints.

A number of authors, including Hall and Mishkin (1982), Altonji and

Siow (1987), Shapiro (1984), Zeldes (1989a), and Runkle (1991), have already examined these questions with panel data. Considering the fact that these authors ask closely related questions, use similar econometric methodology, and employ the same data set (the Panel Study of Income Dynamics), surprisingly little agreement about the substantive issues emerges from these studies.<sup>1</sup> For consumption applications, the PSID data have two important limitations. First, the only consumption data available in the survey are data on food consumption. While food consumption is a large fraction of total consumption for most households, a priori considerations suggest that the intertemporal behavior of food consumption may not be representative of other components of consumption.<sup>2</sup> Second, several of these authors have observed that the time series variation in the food consumption data for a given household is much too large to plausibly be attributed to revisions in the lifetime budget constraint or to movements in rates of return and have concluded that most of the variation in the data reflects some form of noise rather than the systematic consumption behavior addressed by the model.<sup>3</sup>

This study seeks to avoid the two major limitations of the previous work.

---

<sup>1</sup>Using the PSID food consumption data, Shapiro (1984) concludes that the RE-PIH is clearly rejected, Altonji and Siow (1987) find that the RE-PIH is generally supported and that there is suggestive but inconclusive evidence of borrowing constraints, Zeldes (1989a) finds that the RE-PIH is rejected and that the rejection can be attributed to borrowing constraints, and Runkle (1991) finds that the RE-PIH is strongly supported, with no evidence of borrowing constraints.

<sup>2</sup>For example, it seems likely that food consumption is subject to a larger degree of habit persistence and a smaller elasticity of intertemporal substitution than other types of consumption.

<sup>3</sup>Shapiro (1982, 1984) calculates that 95% of the variance of the year to year change in measured food consumption is attributable to measurement error. Similarly, of the variance of the change in food consumption not explained by variation in real interest rates, Runkle (1991) estimates that 76% is due to measurement error.

Instead of using the PSID data on food consumption, the data source is the Survey of Consumer Finances, a three year panel which provides data on the income and a comprehensive array of assets -- savings accounts, checking accounts, stocks, bonds, debt, and durable goods -- for 1600 households. While the SCF does not have a direct measure of consumption, inferences about the consumption of nondurables and services can be drawn from the observable behavior of income and asset stocks. Instead of explicitly constructing a consumption series from the data on income and changes in asset stocks, the econometric specifications focus on the saving side of the consumption decision. In addition to avoiding the need to construct an explicit consumption series, the focus on the saving side is more informative in the sense that the total saving flows of the household can be disaggregated by asset. Thus to the extent that the preliminary results indicate that households manage their total asset holdings in order to smooth their consumption in the face of fluctuations in income, one can also determine which particular assets are used most extensively in this role. Note that the macro implications of the results would be quite different depending on whether households accomplish their consumption smoothing by building up durable goods stocks, as opposed to using financial assets as the buffer.

Any household data set on consumption or asset holdings is likely to contain a substantial amount of noise, whether the noise is interpreted as the effect of preference shocks or as pure measurement error, and it is not claimed that the SCF data set is inherently less noisy than the PSID food consumption data. Instead, the paper uses a robust estimation approach designed to limit the influence exerted by outlying observations on the estimates.

While most of the statistics literature on limited influence estimation deals with robust versions of the OLS estimator, the application in this paper requires a robust version of an instrumental variables estimator. Expanding on

an observation made by Krasker, Kuh, and Welsch (1983), the paper argues that there is a sense in which achieving robustness with respect to observations which are outliers in  $X$ -space, or leverage points, is actually simpler, both conceptually and computationally, in an instrumental variables context than in the OLS context. Using a simplified version of the empirical problem estimated in the paper, the intuition for the robust instrumental variables estimator is presented graphically.

The robust estimates of the model are most interesting when studied in comparison with the conventional (non-robust) instrumental variables estimates. The conventional results, which are presented first, are highly unstable across different subsamples, and, in addition, the pattern of the conventional estimates is inconsistent with the theoretical specification of the model. The robust results are much more stable across different subsamples, conform to the theoretical specification, and indicate that some of the most striking findings in the conventional results were attributable to a single, highly unusual observation.

The empirical results indicate that households do use asset stocks to smooth their consumption in the face of income fluctuations to some extent, although this smoothing is far from complete. That is, 20% to 50% of a predictable increase in current income is buffered by asset stocks, with the remainder implicitly devoted to consumption of nondurables and services. Further, the results do not provide any evidence that the excess sensitivity of consumption to current income is caused by borrowing constraints.

Section 1: The model

The optimization problem of the household consists of maximizing the expected value of lifetime utility, subject to several constraints: (1) the lifetime budget constraint, (2) a non-negativity constraint on consumption, and (3) (potentially) a constraint on borrowing.<sup>4</sup> The utility function is assumed to be time-separable between consumption and leisure. The objective function is:

$$(1) \quad \max E_t \sum_{\tau=0}^T \left( \frac{1}{1+\delta_i} \right)^\tau U(c_{i,t+\tau}, \theta_{i,t+\tau})$$

where  $E_t$  denotes the expectation operator, conditional on information available at time  $t$

$\delta_i$  denotes household  $i$ 's rate of time preference

$c_{i,t+\tau}$  represents household  $i$ 's consumption during period  $t+\tau$

$\theta_{i,t}$  is a preference shock, or shift variable in the utility function

$U(c_{i,t+\tau}, \theta_{i,t+\tau})$  denotes the one period utility function

$T$  represents the number of periods (years) in the planning horizon.

The one period utility function is assumed to exhibit constant relative risk aversion (CRRA):

$$(2) \quad U(c_{i,t}, \theta_{i,t}) = \left( c_{i,t}^{1-\alpha} / (1-\alpha) \right) e^{\theta_{i,t}}$$

where  $\alpha$  is the coefficient of relative risk aversion. Because the CRRA utility function has the property that the marginal utility of consumption is infinite at zero consumption, it is not necessary to explicitly impose the non-negativity constraint on consumption.

The lifetime budget constraint is represented by an equation for the

---

<sup>4</sup>The specification of the household's optimization problem follows fairly closely the specification used by Zeldes (1989a).

evolution of non-human wealth over time and the terminal condition that wealth be non-negative at end of the planning horizon:

$$(3) \quad A_{i,t+1} = A_{i,t} \left( \sum_{j=1}^J w_{i,t}^j (1 + r_{i,t}^j) \right) + Y_{i,t} - c_{i,t}$$

$$A_{i,T} \geq 0$$

where  $A_{i,t}$  denotes non-human wealth of household  $i$  on January 1 of year  $t$   
 $w_{i,t}^j$  denotes the share of non-human wealth held in asset  $j$  in year  $t$   
 $r_{i,t}^j$  denotes the realized real return to asset  $j$  during year  $t$   
 $Y_{i,t}$  denotes real, after-tax labor income in year  $t$ .

Finally, households may face a constraint on the extent to which their non-human wealth can go below zero:

$$(4) \quad A_{i,t+r} \geq D_i$$

As a special case of the borrowing constraint, one could specify that the lower bound on non-human wealth be zero ( $D_i = 0$ ) and prohibit net borrowing altogether. More generally, though, the borrowing constraint imposes a limit (which may be individual specific) on the extent to which families can borrow.

When the borrowing constraint (equation (4)) is not currently binding, optimal consumption behavior is characterized using the Euler equation approach developed by Hall [1978, 1982] and by Hansen and Singleton [1982, 1983]. For the CRRA utility function, the Euler equation is:

$$(5) \quad E_t \left\{ \left[ 1 + r_{i,t}^j \right] c_{i,t+1}^{-\alpha} e^{\theta_{i,t+1}} \right\} = (1 + \delta_i) c_{i,t}^{-\alpha} e^{\theta_{i,t}}$$

As stressed by Zeldes (1989a, 1989b), the fact that the borrowing constraint is (or has some probability of being) binding in a future period should not cause a violation of the first order condition between  $c_{i,t}$  and  $c_{i,t+1}$ . As long as the household is not up against the borrowing constraint during period  $t$ , it is possible to reallocate consumption between periods  $t$  and  $t+1$  in order to satisfy



the marginal condition stated in equation (5).

In characterizing the distribution of the real rates of return and the marginal utility of consumption, I use the assumptions suggested by Breeden (1977). The conditional distributions of  $\ln(1 + r_{i,t}^j)$ ,  $\ln c_{i,t+1}^{-\alpha}$ , and  $\theta_{i,t+1}$  are each assumed to be normal. In general,  $\ln c_{i,t+1}^{-\alpha}$  and  $\theta_{i,t+1}$  will be correlated. To economize on notation, I specify the distributional assumptions concerning the sum,  $\ln c_{i,t+1}^{-\alpha} + \theta_{i,t+1}$ , instead of specifying the distributions of  $\ln c_{i,t+1}^{-\alpha}$  and  $\theta_{i,t+1}$  individually. That is, the distributions of  $\ln(1 + r_{i,t}^j)$  and  $\ln c_{i,t+1}^{-\alpha} + \theta_{i,t+1}$ , conditional on information available in period  $t$ , are given by:

$$(6) \quad \ln(1 + r_{i,t}^j) \sim N(m_{r_j}, v_{r_j})$$

$$\ln c_{i,t+1}^{-\alpha} + \theta_{i,t+1} \sim N(m_u, v_u)$$

and  $\text{cov}(\ln[1 + r_{i,t}^j], \ln c_{i,t+1}^{-\alpha} + \theta_{i,t+1}) = \text{cov}_{r_j u}$ .

While econometric identification will require some restrictions on the moments of the distributions of asset returns and the marginal utility of consumption, the theoretical model does not require these moments to be constant either across households or across time. For the theoretical discussion, think of the moments  $(m_{r_j}, v_{r_j}, m_u, v_u, \text{and } \text{cov}_{r_j u})$  as having both household ( $i$ ) and time ( $t$ ) subscripts, although these subscripts will be suppressed for notational convenience.

Replacing the left hand side of (5) with  $\exp(m_{r_j} + m_u + (v_{r_j} + v_u)/2 + \text{cov}_{r_j u})$ , taking logs, and rearranging, one obtains:

$$(7) \quad E_t \ln c_{i,t+1} - \ln c_{i,t} = \alpha^{-1} \left[ (v_{r_j} + v_u)/2 + \text{cov}_{r_j u} - \ln(1 + \delta_i) \right]$$

$$+ \alpha^{-1} E_t \ln(1 + r_{i,t}^j) + \alpha^{-1} (E_t \theta_{i,t+1} - \theta_{i,t}).$$

Since the expected growth rate of consumption is not observable, it is

necessary to add to both sides of the equation the forecast error in predicting  $\ln c_{i,t+1}$ . Define

$$(8) \quad \bar{\eta}_{i,t+1} = \ln c_{i,t+1} - E_t \ln c_{i,t+1}$$

Given the distributional assumptions made above, the random variable  $\bar{\eta}_{i,t+1}$  is normally distributed with mean zero and is uncorrelated with information available in period  $t$ . To sum up, the potentially testable implication of the optimal consumption model when the borrowing constraint is not currently binding can be expressed as:<sup>6</sup>

$$(9) \quad \ln(c_{i,t+1}/c_{i,t}) = \alpha^{-1} \left[ (v_{r_j} + v_u)/2 + \text{cov}_{r_j u} - \ln(1+\delta_1) \right] \\ + \alpha^{-1} E_t \ln(1 + r_{i,t}^j) + \alpha^{-1} (E_t \theta_{i,t+1} - \theta_{i,t}) + \bar{\eta}_{i,t+1}$$

As Zeldes (1989a) has pointed out, if a borrowing constraint is currently binding, the growth rate of consumption will exceed the optimal, unconstrained growth rate expressed on the RHS of equation (9). That is, in a situation in which the borrowing constraint is binding, the family would like to reallocate some consumption from  $t+1$  toward  $t$ , but is unable to do so. Since  $c_{i,t}$  is somewhat lower and  $c_{i,t+1}$  somewhat higher than the unconstrained optimal values, the actual growth rate of consumption exceeds the optimal, unconstrained growth rate when the borrowing constraint is currently binding. The presence of a borrowing constraint creates a one-sided violation of the Euler equation; because the household always has the option of saving more -- reallocating consumption from  $t$  toward  $t+1$ , a borrowing constraint never forces the expected growth rate of consumption to be lower than the unconstrained optimal growth rate.

---

<sup>6</sup>Note that the Euler equation (9) should hold for all assets which the family holds in nonzero amounts; the Euler equation for a given household should hold simultaneously for two or more assets with different expected returns provided the different assets have correspondingly different covariance properties with the marginal utility of consumption.

Following Zeldes, the effects of a borrowing constraint are incorporated into the empirical analysis by adding an additional disturbance  $z_{i,t}$ :

$$(10) \quad \ln(c_{i,t+1}/c_{i,t}) = \alpha^{-1} \left[ (v_r + v_u)/2 + \text{cov}_{ru} - \ln(1+\delta_i) \right] \\ + \alpha^{-1} E_t \ln(1 + r_{i,t}) + \alpha^{-1} (E_t \theta_{i,t+1} - \theta_{i,t}) + \bar{\eta}_{i,t+1} + z_{i,t}.$$

The disturbance  $z_{i,t}$  takes on a value of zero for agents not currently constrained and takes on positive values for agents for whom the borrowing constraint is binding. In general  $z_{i,t}$  will be correlated with current income and expected future income. (Note that to simplify the notation, I have suppressed the asset index  $j$ .)

At this point, one could impute consumption from the data on income and changes in asset levels, and test the Euler equation on the basis of this imputed consumption series. However, since the underlying data series consist of data on savings flows, it seems natural to test the model's implications for the savings side of the consumption/savings decision. Further, to the extent that we are interested in knowing whether households optimally smooth their consumption by using asset stocks to buffer income fluctuations, it will be interesting to see which assets in the portfolio - bank deposits, bonds, debt, or durable goods are used to absorb income fluctuations. To obtain the implications for savings flows, write the consumption growth rate as the sum of the first order Taylor series approximation and the associated remainder term:

$$(11) \quad \ln c_{t+1} - \ln c_t = \frac{\Delta c_{t+1}}{c_0} - \frac{1}{2} \left[ \left( \frac{c_{t+1} - c_0}{c_{t+1}^*} \right)^2 - \left( \frac{c_t - c_0}{c_t^*} \right)^2 \right]$$

where  $c_0$  is the point around which the linearization is taken,  $c_{t+1}^*$  is a point which lies between  $c_0$  and  $c_{t+1}$ , and  $c_t^*$  lies between  $c_0$  and  $c_t$ . Instead of setting  $c_0 = c_t$ , linearize the function  $\Delta \ln c_{t+1}$  around the point  $c_0 = y_t$ , giving:

$$(12) \quad \ln c_{t+1} - \ln c_t = \frac{\Delta c_{t+1}}{y_t} - \frac{1}{2} \left[ \left( \frac{c_{t+1} - y_t}{c_{t+1}^*} \right)^2 - \left( \frac{c_t - y_t}{c_t^*} \right)^2 \right]$$

To evaluate the likely magnitude of the remainder term, consider a household whose consumption both in  $t$  and in  $t+1$  is within 90% to 110% of the household's disposable income in period  $t$ . For such a household, the squared terms on the RHS of (12) would each be positive and smaller than 1%. Further, the deviations of  $c_{t+1}$  and  $c_t$  from  $y_t$  are likely to be positively correlated. Since the remainder amounts to half of the difference between two small and positively correlated terms, it seems sufficiently small to be safely ignored.

Using just the linear term of the approximation, then, the optimal consumption model, with or without a binding borrowing constraint, is given by

$$(13) \quad \frac{\Delta c_{i,t+1}}{y_{i,t}} = \alpha^{-1} \left[ (v_r + v_u)/2 + \text{cov}_{ru} - \ln(1+\delta_i) \right] \\ + \alpha^{-1} E_t \ln(1 + r_{i,t}) + \alpha^{-1} (E_t \theta_{i,t+1} - \theta_{i,t}) + \bar{\eta}_{i,t+1} + z_{i,t}$$

To obtain the implications for savings flows, both sides of equation (13) are subtracted from the growth rate of disposable income:

$$(14) \quad \frac{\Delta S_{i,t+1}}{y_{i,t}} = -\alpha^{-1} \left[ (v_r + v_u)/2 + \text{cov}_{ru} - \ln(1+\delta_i) \right] - \alpha^{-1} E_t \ln(1 + r_{i,t}) \\ + \frac{\Delta y_{i,t+1}}{y_{i,t}} - \alpha^{-1} (E_t \theta_{i,t+1} - \theta_{i,t}) - \bar{\eta}_{i,t+1} - z_{i,t}$$

where  $S_{i,t}$  denotes the total flow of saving of household  $i$  in year  $t$ .

Equation (14) is the flip side of consumption orthogonality. For households for whom the borrowing constraint is not currently binding ( $z_{i,t} = 0$ ), the growth in income between periods  $t$  and  $t+1$ , to the extent that it was predictable in  $t$ , should go entirely into saving under the permanent income hypothesis. In the empirical work, total saving of household  $i$ ,  $S_{i,t}$  is defined as the sum of the changes in the various asset stocks during year  $t$ :

$$(15) \quad S_{i,t} = \Delta \text{Savings}_t + \Delta \text{Checking}_t + \Delta \text{Bonds}_t + \Delta \text{Stocks}_t \\ - \text{Debt}_t + \text{Debt}_{t-1} + \text{Durables expenditure}_t$$

Durable goods have some consumption aspects and some investment aspects; in this paper durable goods are treated as assets rather than as consumption. Thus the complete array of assets modeled in the empirical work consists of savings accounts, checking accounts, bonds, stocks, durable goods, and debt.

An obvious testable implication of the model is the restriction that the coefficient of the growth rate of income should equal unity. If this restriction is rejected, the estimated coefficient of the income growth rate is easily interpreted if the alternative model of saving arises from a crude "Keynesian" consumption function:

$$(16) \quad c_t = \kappa + \beta y_t$$

In terms of saving behavior, the Keynesian consumption function implies:

$$(17) \quad \frac{\Delta S_{i,t+1}}{y_{i,t}} = (1-\beta) \frac{\Delta y_{i,t+1}}{y_{i,t}}$$

which is conveniently nested in equation (14).

## Section 2: Empirical specification

The discussion of the model can be summarized in the following equation for the change in savings flows:

$$(18) \quad \frac{\Delta S_{i,t+1}}{y_{i,t}} = \alpha^{-1} \left[ (v_r + v_u)/2 + \text{cov}_{ru} - \ln(1+\delta_i) \right] - \alpha^{-1} E_t \ln(1 + r_{i,t}) \\ + \gamma \frac{\Delta y_{i,t+1}}{y_{i,t}} - \alpha^{-1} (E_t \theta_{i,t+1} - \theta_{i,t}) - \bar{\eta}_{i,t+1} - z_{i,t}$$

where  $\gamma = 1$  under the null hypothesis and  $1-\gamma = \beta$ . The data for the study, described in more detail in the next section, consist of observations on income and assets for a panel of about 1600 households interviewed in 1967, 1968, and 1969. Because the specification addresses the saving, or the change in asset

levels between two points in time, the three year panel provides a pure cross-section of the change in the savings flows between 1968 and 1969.

The fact that the data is a pure cross-section introduces several considerations. First, the intercept term in equation (18) in principle varies both across households and across time. With only one observations per household it is obviously not possible to identify both the parameter of primary interest,  $\gamma$ , and a separate intercept term for each family. If one considers the households grouped according to the occupation of the primary earner, and assumes that the rate of time preference as well as the moments of the distribution of the marginal utility of consumption are constant across households within a given occupation, then some heterogeneity in the intercept term across households may be incorporated by estimating a set of occupation dummies instead of a single intercept term. However, in preliminary empirical work, the inclusion of occupation-specific intercept terms did not reveal significant variation in the intercepts across occupations, and did not have a noticeable effect on the estimates of  $\gamma$ . In the absence of evidence of significant variation in the moments across occupations, the specification was then simplified by assuming a common set of moments for all households.

Second, in a single cross-section, the expectational revision term,  $\bar{\eta}_{i,t+1}$ , need not have mean zero. This feature of the sample is easily accommodated by thinking of the households' expectational revision disturbances as the sum of an aggregate shock and an idiosyncratic shock:

$$(19) \quad \bar{\eta}_{i,t+1} = \eta_{t+1} + \eta_{i,t+1}$$

where the idiosyncratic shock,  $\eta_{i,t+1}$ , has mean zero and is uncorrelated across households. Third, the data do not provide time series variation in the rate of return. Taking the expected rate of return as constant across households, and

adding it to the constant term, along with the aggregate shock,  $\eta_{t+1}$ , gives<sup>6</sup>:

$$(20) \quad \frac{\Delta S_{i,t+1}}{y_{i,t}} = \mu + \gamma \frac{\Delta y_{i,t+1}}{y_{i,t}} - \alpha^{-1} (E_t \theta_{i,t+1} - \theta_{i,t}) - \eta_{i,t+1} - z_{i,t}$$

With the additional assumptions, the intercept term

$$\mu = -\eta_{t+1} - \alpha^{-1} \left[ (v_r + v_u)/2 + \text{cov}_{ru} - \ln(1+\delta) + E_t \ln(1+r_t) \right]$$

is constant across households. Note that unlike tests of the model based on aggregate time series data, the specification does not impose the restriction that the moments be time-invariant.

With a time series or panel data set, the obvious way to identify equation (20) would be with the use of lagged variables. The hypothesis would then be tested by asking whether predictable changes in income go entirely into saving, with "predictable" being implicitly defined in terms of time series predictions.<sup>7</sup>

Instead of identifying the model by relying on time series predictions, the paper uses data on explicit statements made by households concerning their income expectations. More concretely, the 1967 survey asked the following questions:

"Will your family income for this year be higher or lower than last year?

If higher (lower), do you think it will be a lot higher (lower), or just a little higher (lower)?" Allowable answers were: a lot higher, a little higher, same, a little lower, or a lot lower.

"Thinking ahead about four years, would you say that your family income will

---

<sup>6</sup>The survey does contain information on the interest rates which households received on their savings accounts, and on their expectations of inflation, both of which exhibited considerable cross sectional variation. While the intertemporal substitution parameter is in principle identified with the cross sectional variation in nominal asset returns and expectations of inflation, preliminary estimates of equation (18) gave small and imprecise estimates of the intertemporal substitution parameter, and, more importantly, the inclusion or exclusion of the expected rate of return made no material difference to the estimates of  $\gamma$ , the parameter of primary interest.

<sup>7</sup>For a different but related problem, Deaton (1990) discusses an approach which uses cross sectional sample moments for household data to make inferences about the parameters of a time series process.

be much higher, a little higher, the same, or smaller than it is now?"

The use of the stated expectations variables offers two advantages. First, the extent to which households smooth consumption in the face of predictable income fluctuations can be estimated without requiring an assumption concerning the way expectations are formed. Second, in forming their expectations of future income, households undoubtedly have access to useful information not available to the econometrician. Since this "private" information will be reflected in the household's explicit statement of income expectations, the expectational variables may be more strongly correlated with the household's income growth rate and therefore provide more precise estimates.

For the moment, consider a sample for which the borrowing constraint is not binding, so that  $z_{i,t} = 0$  for each household in the subsample. The expectational variables are obviously uncorrelated with the idiosyncratic part of the expectational revision term,  $\eta_{i,t+1}$ . However, assuming that the expectational variables are uncorrelated with the term representing the expected change in the preference shifter,  $E_t \theta_{i,t+1} - \theta_{i,t}$ , is less obvious. Changes in family composition will in general cause simultaneous changes in consumption and in income; further, the changes in consumption and income associated with changes in family size are likely to be forecastable, at least in part. For this reason, it seems likely that the instrumental variables (stated expectations of future income) would be correlated with the expected change in the preference shifter for families experiencing a change in the family composition. To avoid this correlation, any household which experienced a change in household composition over the three year period was excluded from the sample, reducing the sample size by about 20%. For households with stable family composition, it seems reasonable to assume that the expected change in the preference shifter is uncorrelated with the expectational variables used as instruments.



For samples in which the borrowing constraint is binding for some households, the shadow price of the constraint,  $z_{i,t}$ , may be positive. Since the households subject to a binding constraint ( $z_{i,t} > 0$ ) are likely to be those which expect positive growth rates of income, the expectational variables will be correlated with the disturbance for these observations.

The discussion suggests the following strategy for distinguishing between the effects of borrowing constraints, on the one hand, and "Keynesian" consumption behavior, on the other. Following Zeldes (1989a) and Runkle (1991), the sample is split into two subsamples, one of which contains only households for which the  $z_{i,t}$  term is assumed to be zero on the basis of the households' observed asset holdings, and the other of which contains households for which  $z_{i,t} \geq 0$ . In the unconstrained sample, the fact that  $z_{i,t}$  is identically zero for all observations implies that the estimation of equation (20), using the expectational variables as instruments, provides a consistent estimate of  $\gamma$ . For this subsample, the null hypothesis is embodied in the restriction  $\gamma = 1$ , and, if the null hypothesis is rejected,  $1 - \gamma = \beta$  provides an estimate of the Keynesian marginal propensity to consume under the alternative. Since we would expect, on a priori grounds, the expectational variables to be positively correlated with the shadow price of the constraint, the model predicts that the presence of binding borrowing constraints will result in a downward biased estimate of  $\gamma$  for the subsample of households with low levels of assets. Thus testing the restriction that the estimates of  $\gamma$  are the same across the two subsamples provides a test of the hypothesis that borrowing constraints are not binding, even for households with low levels of assets.

Data

The data for this study are from the Survey of Consumer Finances, conducted by the Survey Research Center of the University of Michigan, for the years 1967, 1968, and 1969. In recent years, the Survey of Consumer Finances has been a pure cross-section survey; that is, a completely different group of families is sampled each year. However, of the roughly 3,000 households interviewed in 1967, half were designated as a panel and reinterviewed in 1968 and 1969. For each of the three years, the survey has data on total disposable income of the household, and expenditure on certain components of consumption, such as housing, additions and repairs, cars, and "other durable goods". For each of the major categories of expenditure on durable goods, the survey also has data on any debt incurred with the purchase of the durable good. In 1968 and 1969, the survey requested information not only on the level of the respondent's holdings of various financial assets (checking accounts, savings accounts, bonds, and stocks) but also the change in the respondent's holdings of each of these assets over the past year.

While a consumption series is not explicitly constructed, one can, of course, use the results to make inferences about consumption behavior, since nondurable consumption can be thought of as the equation which has been dropped from a singular system. In order to interpret the implications of the empirical work for consumption, it is useful to note the categories of consumption expenditure which are included in the implicit consumption series. In particular, payments for housing services are included in the implicit consumption series. For households which own their home, the interest portion of the mortgage payments would be reflected in the consumption measure but the principal payments would not, since the debt variable includes mortgage debt. Similarly, for

renters, the implicit consumption series would reflect rent payments.

Several aspects of the data are worth noting:

$y_t$ , disposable household income, is the sum of earned income, mixed labor/capital income, capital income, and transfer payments, minus total income tax. The data on capital income is based on flows of income from capital such as dividends and interest payments and does not include capital gains. Transfer payments include "help from relatives" as well as unemployment compensation and welfare payments.

$\Delta$ Saving<sub>t</sub>,  $\Delta$ Checking<sub>t</sub>, and  $\Delta$ Bonds<sub>t</sub> are based on responses to questions of the form: "Thinking back to this time last year, has the amount in all your family unit's savings accounts gone up or down? About how much has it gone (up/down) since this time last year?"

$\Delta$ Stocks<sub>t</sub> reflects the household's cashflow into or out of stocks, explicitly excluding unrealized capital gains. The respondent is first asked whether the household purchased, sold, or both purchased and sold stocks in the past year. If the household sold or purchased stocks (but not both) the amount of the sale or purchase is recorded. If the household both sold and purchased stocks, the respondent is asked: "Disregarding changes in stock prices, 'on balance' did you put new money into stocks or take money out during the last twelve months? About how much was this?"

Durables Expenditure<sub>t</sub> is the sum of expenditure on additions and repairs, expenditure on all cars purchased in year t (net of value of any cars traded-in), and expenditure on other durable goods (net of any trade-ins).

Debt<sub>t</sub> is total debt remaining at the end of year t.

Section 3: Conventional results

Empirical results are presented for four different samples. The criteria for inclusion in a given sample and the number of observations in the sample are described below.

<u>descriptive label</u>	<u>sample size</u>	<u>criterion for inclusion in sample</u>
whole sample	774	Of the 1,568 households in the panel, 371 experienced a change in household composition during the three year period. Because such a change in household composition is likely to lead to correlation between the preference shift part of the error term and the instruments (stated expectations of future income), these households were excluded, reducing the sample size to 1,197 households. Deleting households whose records contained missing observations for any of the required variables further reduced the sample size to 774.
poor	424	subset of whole sample for which the sum of the household's wealth in the form of checking accounts, savings accounts, bonds, and stocks at the end of year t was less than \$1000 in 1968 dollars
rich	350	subset of whole sample for which the sum of the household's wealth in the form of checking accounts, savings accounts, bonds, and stocks at the end of year t was greater than or equal to \$1000 in 1968 dollars
truly wealthy	77	subset of whole sample for which the household held strictly positive levels of <u>each</u> of the following assets: bank deposits (i.e., savings and/or checking accounts), bonds, and stocks

Equation (20) was estimated by instrumental variables using, in addition to the expectational variables described earlier, lagged disposable income,  $y_{i,t}$ , as an instrument. The two categorical variables on income expectations were each expanded into sets of dummy variables, with one dummy variable for each response. The results reported in Table 1 are anomalous in several respects. From the discussion of the empirical specification, recall that for those samples which

Table 1

Instrumental variables estimates of  $\gamma$

sample	estimates of:		s.e.e.	number of observations
	intercept	$\gamma$		
whole sample	.032 (.029)	.53 (.24)	.583	774
poor	.034 (.044)	.62 (.34)	.628	424
rich	.041 (.034)	.27 (.26)	.524	350
truly wealthy	-.023 (.049)	1.33 (.22)	.383	77

Note: Standard errors are in parentheses.

contain households with low levels of nonhuman wealth, the disturbance term,  $z_{i,t}$ , which reflects the shadow price of the borrowing constraint, is included in the error term. Because households which expect their income to rise are more likely than average to have a binding borrowing constraint ( $z_{i,t} > 0$ ), the estimates of  $\gamma$  will be inconsistent for these samples. Thus only the rich and truly wealthy subsamples provide consistent estimates of  $\gamma$ .

While the slope coefficients for the whole sample and the poor subsample do not provide consistent estimates of  $\gamma$ , they do provide a rough check on the specification of the model. If the expectation of income growth is positively correlated with  $z_{i,t}$ , and the  $z_{i,t}$  term enters the saving equation with a negative sign, the presence of the borrowing constraint should cause the estimates of  $\gamma$  to be downward biased in the poor subsample and in the whole sample. Thus under the interpretation of the disturbance term developed above, one would expect the estimate of  $\gamma$  for the rich subsample to exceed the estimate for the poor subsample. The result that the estimate of  $\gamma$  for the poor is .62 and statistically significant, while the estimate of  $\gamma$  for the rich is .27 and statistically insignificant does not inspire confidence in the specification and the assumptions about the disturbance term.

The estimate of  $\gamma$  for the truly rich is troubling for two reasons. First, the point estimate exceeds unity, although not by a statistically significant margin. Second, the estimate of  $\gamma$  from the truly wealthy subsample, 1.33, differs dramatically from the estimate of .27 from the rich subsample, a finding which again fails to support the premise that both of these subsamples should provide consistent estimates of  $\gamma$ .

Despite its anomalous results, Table 1 does contain the striking finding that for the subsample designated as "truly wealthy", essentially all of the anticipated growth in income goes directly into assets. For this subsample, it

seems natural to follow up the analysis by disaggregating the change in total savings flows,  $\Delta S_{i,t+1}$ , into the changes in savings flows associated with the various assets. That is, for the truly wealthy sample, which consists of households which hold strictly positive amounts of bank deposits (the sum of savings accounts and checking accounts), bonds, and stocks, one could estimate the system:

$$(21) \quad \begin{aligned} \Delta S_{i,t+1}^{LA}/y_{i,t} &= \mu_1 + \varphi_1 \gamma (\Delta y_{i,t+1}/y_{i,t}) - e_{i,t+1}^{LA} \\ \Delta S_{i,t+1}^{St}/y_{i,t} &= \mu_2 + \varphi_2 \gamma (\Delta y_{i,t+1}/y_{i,t}) - e_{i,t+1}^{St} \\ \Delta S_{i,t+1}^B/y_{i,t} &= \mu_3 + \varphi_3 \gamma (\Delta y_{i,t+1}/y_{i,t}) - e_{i,t+1}^B \\ \Delta S_{i,t+1}^{Debt}/y_{i,t} &= \mu_4 + \varphi_4 \gamma (\Delta y_{i,t+1}/y_{i,t}) - e_{i,t+1}^{Debt} \\ \Delta S_{i,t+1}^{Dur}/y_{i,t} &= \mu_5 + \varphi_5 \gamma (\Delta y_{i,t+1}/y_{i,t}) - e_{i,t+1}^{Dur} \end{aligned}$$

where

$$\begin{aligned} \Delta S_{i,t+1}^{LA} &= (\Delta \text{Savings}_{t+1} - \Delta \text{Savings}_t) + (\Delta \text{Checking}_{t+1} - \Delta \text{Checking}_t) \\ \Delta S_{i,t+1}^{St} &= \Delta \text{Stocks}_{t+1} - \Delta \text{Stocks}_t \\ \Delta S_{i,t+1}^B &= \Delta \text{Bonds}_{t+1} - \Delta \text{Bonds}_t \\ \Delta S_{i,t+1}^{Debt} &= \Delta \text{Debt}_{t+1} - \Delta \text{Debt}_t \\ \Delta S_{i,t+1}^{Dur} &= \text{Durable expenditure}_{t+1} - \text{Durable expenditure}_t \end{aligned}$$

Table 2 presents the results of estimating system (21) for the truly wealthy subsample, using the same instruments as before. Again, the results are both striking and unsettling. One would expect that the assets used most extensively to buffer income fluctuations would be liquid assets, debt, and durable goods. While the results indicate that these households do, to some extent, use durable goods stocks and debt to smooth their consumption, the results also state that most of the consumption smoothing is accomplished by the purchase or sale of stocks. The coefficient of the income growth rate is .759 for the stock equation, with a standard error of .151. While the corresponding coefficients in

Table 2

Instrumental variables estimates  
of  $\varphi_i \gamma$  for disaggregated asset stocks

truly wealthy subsample

dependent variable	estimate of:		s. e. e.
	intercept	$\varphi_i \gamma$	
$\Delta S_{i,t+1}^{LA} / Y_{i,t}$	.026 (.031)	.030 (.139)	.240
$\Delta S_{i,t+1}^{St} / Y_{i,t}$	-.008 (.033)	.759 (.151)	.260
$\Delta S_{i,t+1}^B / Y_{i,t}$	-.009 (.003)	.010 (.015)	.026
$\Delta S_{i,t+1}^{Debt} / Y_{i,t}$	.004 (.014)	-.220 (.065)	.111
$\Delta S_{i,t+1}^{Dur} / Y_{i,t}$	-.029 (.029)	.306 (.130)	.224

Note: Standard errors are in parentheses. Number of observations was 77.



the equations for debt and durables are plausible in magnitude, statistically significant, and of the correct sign (-.220 for debt, and .306 for durables), their role in the explanation of consumption smoothing is upstaged by the equation for stocks. Surprisingly, liquid assets and bonds play virtually no role in buffering income fluctuations.

Scanning the data series on the magnitudes of stock sales or purchases reveals that the majority of households who own stocks make no net transactions in stocks in a given year. A plot of the dependent variable,  $\Delta\text{Stock}_{t+1}/y_t$ , against the growth rate of income,  $\Delta y_{t+1}/y_t$ , resolves the contradiction between the results of Table 2, which asserts that stocks are the primary vehicle for consumption smoothing, and the observation that even among the subsample of households which own stocks, few families actively buy and sell stocks in a typical year. Figures 1a and 1b reflect the following facts: 1) most of the households had no net transactions in stocks, 2) for 76 out of the 77 households the line which best fits the scatter is a horizontal one, and 3) the entire result that stocks are used extensively to smooth consumption is attributable to a single observation.

According to the data plotted in Figure 1a, the outlier observation experienced a quadrupling of his income between 1968 and 1969, and increased his net purchases of stocks by an amount equal to over 400% of his 1968 income. To determine whether the outlier was a legitimate, albeit highly unusual, observation or the result of a coding error, the raw data on the observation contained in the extract is compiled in the appendix. The data records substantiate the huge growth in income by revealing that the head of household of the observation (hereafter identified as household 676) was a student in an advanced degree program in 1967, received a professional degree, possibly an M.D., worked only part of the year in 1968, and was employed at the professional

Figure 1a

*scatter plot for stocks*

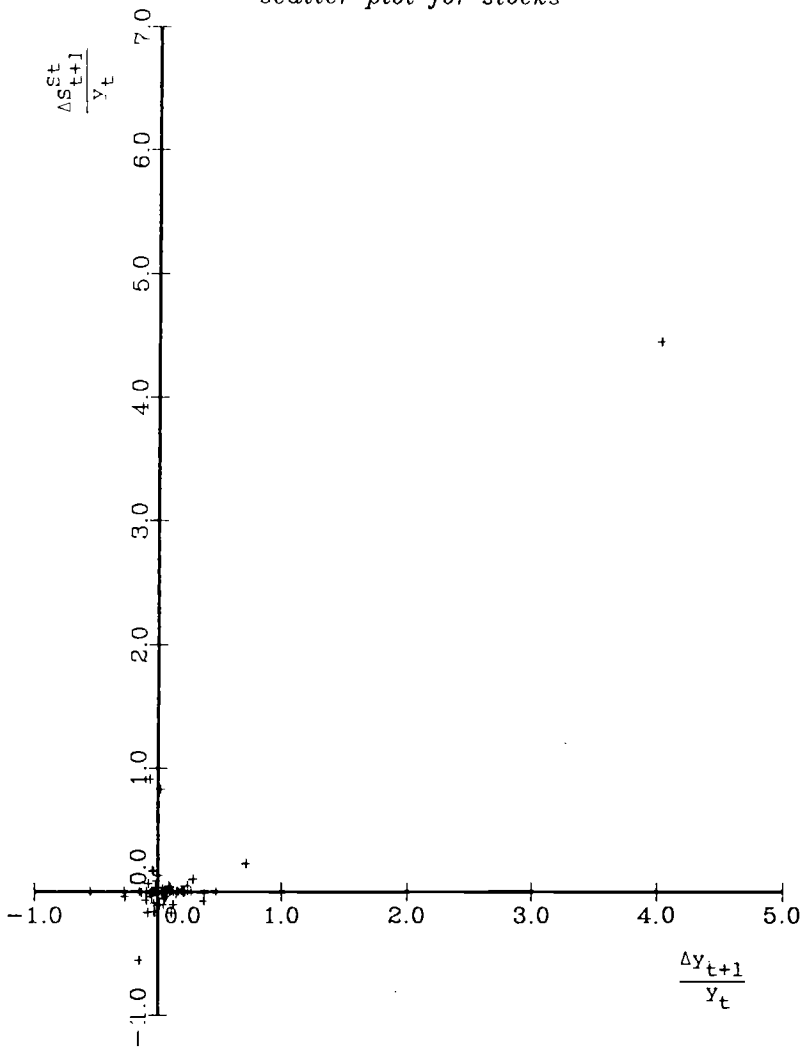
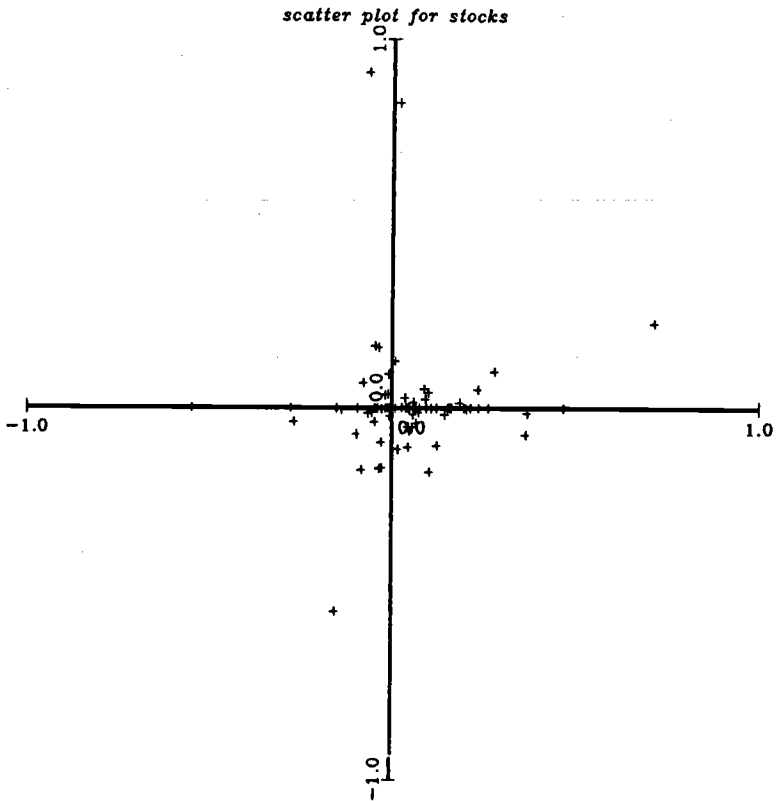


Figure 1b

(enlarged scale version of Figure 1a  
for observations other than #676)



level in the health field for all of 1969. With a disposable income of \$22,573 (in 1969 dollars), the respondent stated that a net purchase of \$20,000 of stocks was made in 1969.

Based on the data on household 676's expenditures on durables, a car, additions and repairs, and mortgage payments it is not possible that the \$22,573 income figure and the \$20,000 figure for net stock purchases are both accurate. Underreporting of the household's total income is likely, given that for the student years, the respondent reported total income levels of \$100 (1967) and \$4,400 (1968) for the support of a six person household. Further, the \$5,500 house which the household owned and lived in in 1967 was retained as a real estate investment in 1968 and 1969 after the head of household received his degree and bought a new residence. Presumably the first house was generating rental income in 1968 and 1969, and this income flow does not seem to appear in the recorded data on the family's capital income.

While there are clearly some discrepancies or omissions in the data on household 676, the basic attribute which endowed the observation with so much leverage on the empirical results -- that the household experienced a huge increase in income and saved a large part of it in the form of stock purchases -- appears to be legitimate. Even if the recorded data underrepresents the household's income by \$5,000 and if the \$20,000 stock purchase number has been "rounded up" by a comparable magnitude, the "corrected data", if available, would have had a similar effect of the empirical results, in the sense that the behavior of the single outlier would determine the results.

Household 676 appears to be an extremely unusual, but basically legitimate observation. While the inordinate amount of influence exerted by household 676 could be crudely limited by removing the observation from the sample and reestimating the specifications for the rest of the subsample, it seems likely

that the sample contains other influential observations. Rather than engage in a subjective process of eliminating unusual observations, the paper deals with outliers by employing a robust instrumental variables estimation procedure.<sup>8</sup>

#### Section 4: Approach to robust estimation

While limited influence estimation has been an active area of research in the statistics literature for the past twenty years, robust estimation techniques have not been fully exploited in applied econometric work, despite the widespread recognition that economic data sets, and particularly household data sets, are very likely to contain influential observations, such as household 676. An econometrician delving into the statistical literature on robust estimation is deterred by two aspects of the literature. One barrier is a reflection of the origins of robust estimation techniques in the statistics, rather than econometrics, literature. While an array of limited influence analogs of OLS have been proposed, evaluated, and programmed, research on the theory and performance of limited influence estimators suitable for simultaneous equations systems (i.e., analogs of instrumental variables estimators) is less well developed.

Second, the limited influence literature is split into two major strands. One strand develops estimation techniques which are efficient under failures of the assumption that the disturbances are distributed normally, i.e., techniques

---

<sup>8</sup>Among several factors which could account for the opposing conclusions reached by Runkle (1991) and Zeldes (1989), Runkle notes that he used more stringent criterion for inclusion in the sample. That is, Runkle excluded observations for which measured food consumption increased by more than 300% or declined by more than 75% from one year to the next, whereas Zeldes did not exclude observations with extreme values. Altonji and Siow (1987) exclude observations for which food consumption increased by more than 400%, decreased by more than 75%, as well as observations for which real wages or family income increased by more than 500% or decreased by more than 80%.

which are robust to outliers in epsilon-space. A second strand develops techniques which limit the influence of leverage points, or observations which are outliers in X-space.

Modifying an estimator to make it robust with respect to outliers in epsilon-space is straightforward, both conceptually and computationally. However, in the general case in which influential data points may be outliers in X-space, both the derivation and computational implementation of robust estimators is much more complicated. That robustness with respect to outliers in X-space is more difficult to achieve than robustness with respect to outliers in epsilon-space is a direct consequence of the fact that, for each observation,  $x$  is vector-valued, whereas epsilon is scalar-valued. While an outlier in epsilon-space is easily identified as an observation for which the absolute value of the residual is large, identifying and limiting the influence of outliers in the vector of observations on the right hand side variables is nontrivial.<sup>9</sup>

Presumably econometricians have been deterred from using robust estimation techniques on the assumption that the instrumental variables version of an estimator which is robust to outliers in X-space as well as outliers in epsilon-space will be even more cumbersome than the corresponding OLS version. Surprisingly, though, there is a sense in which achieving robustness with respect to outliers in X-space, or leverage points, is much simpler, both conceptually and computationally, in an instrumental variables context than in the OLS context.

To develop this point, this section first reviews the Huber [1973,1977]

---

<sup>9</sup>One might be tempted to identify outliers in X-space by calculating the norm of the  $x$  vector as the "obvious" generalization of the absolute value of the epsilon. However, by this method, the classification of an observation as an outlier would not be invariant to the choice of units of the data. Identification of outliers in a way which satisfies basic criteria of reasonable estimators, such as invariance to the choice of units, is the complicating factor in developing estimators robust with respect to outliers in X-space.

estimator, basically a modification of OLS which limits the influence of outliers in epsilon-space. Next, the assertion that the generalization of the framework from an OLS to an instrumental variables context actually simplifies the treatment of leverage points is developed. Based on this discussion, the section describes a simple algorithm for computing instrumental variables estimates which are robust to outliers in both X-space and epsilon-space. The section concludes with some plots which provide the visual intuition for the estimator.

To establish notation, assume

$$(22) \quad y_i = x_i \beta + \epsilon_i$$

where  $x_i$  is a  $1 \times k$  vector of explanatory variables

$\beta$  is a  $k \times 1$  vector of parameters

and  $\epsilon_i$  is a disturbance from a symmetric distribution with scale parameter  $\sigma$ .

Consider the class of estimators which can be written as the solution to a minimization problem in which the objective function is a function of the residuals:

$$(23) \quad \text{objective function} = \min \sum_{i=1}^n \rho((y_i - x_i \beta) / \sigma)$$

for some function  $\rho : \mathbb{R} \rightarrow \mathbb{R}^+$

Defining  $r_i = y_i - x_i \beta$

$$\text{and} \quad \psi(r_i) = \frac{\partial \rho(r_i)}{\partial r_i} .$$

the solution to the minimization problem will satisfy:

$$(24) \quad \sum_{i=1}^n \psi((y_i - x_i \beta) / \sigma) x_i = 0 .$$

OLS, of course, is the estimator which uses the quadratic objective function

$$(25) \quad \rho\left(\frac{r_i}{\sigma}\right) = \frac{1}{2} \left[ \frac{r_i}{\sigma} \right]^2, \quad \text{which implies} \quad \psi(r_i) = \frac{\partial \rho}{\partial r} = \frac{r_i}{\sigma^2} .$$

Outliers in epsilon-space

While OLS is efficient when the errors are distributed normally, the sensitivity to large residuals inherent in the quadratic specification of the objective function leads to inefficiency when the disturbances are distributed non-normally. Compared to OLS, the Huber estimator downweights observations with large residuals by using as the  $\psi(r_i)$  function:

$$(26) \quad \psi(r_i) = \frac{r_i}{\sigma} \left[ \min \left\{ 1, \frac{c}{\left| \frac{r_i}{\sigma} \right|} \right\} \right]$$

where the parameter  $c$  determines the bound such that observations with residuals exceeding  $c\sigma$  are downweighted. Thus the Huber estimator chooses the estimate vector  $\beta$  which satisfies:

$$(27) \quad \sum_{i=1}^n \min \left\{ 1, \frac{c}{\left| \frac{r_i}{\sigma} \right|} \right\} \cdot \left( \frac{y_i - x_i \beta}{\sigma} \right) x_i = 0$$

From equation (27), the Huber estimator can be interpreted as a weighted least squares estimator in which "well-behaved" observations with residuals smaller in absolute value than  $c\sigma$  receive a weight of unity and outliers, defined as observations with residuals larger than  $c\sigma$ , receive a weight of  $\frac{c\sigma}{|r_i|}$ . Intuitively, the Huber estimator transforms the data by "moving" observations with large residuals to the  $c\sigma$  bound around the regression line. Computationally, the Huber estimates are obtained by iterating on the weighted version of the normal equations (equation (27)).

Outliers in X-space

While the Huber estimator provides a simple, and intuitively appealing, method of reducing the sensitivity of the estimates to outliers in epsilon-space, it does not provide protection against leverage points, or outliers in X-space. If the data actually conform to the assumptions of the parametric model, limiting



the influence of outliers in X-space is undesirable because the leverage exerted by these outliers increases the precision of the estimates. However, if we view the parametric model not as a specification of "the truth", but instead as an approximation designed to succinctly characterize the bulk of the data, the disproportionate influence exerted by outliers in X-space is less desirable. For example, if the true relationship is nonlinear while the model that approximates it is linear, one would like the estimates of the slope parameters to reflect the linear relationship which best approximates the bulk of the observations. If the true model is nonlinear, a single observation with an extreme value of x can result in an estimated linear approximation which differs substantially from the linear relationship which characterizes the bulk of the data.

In the present application, the specification of the parametric model embodies the assumption that all of the data points are generated by the same structure. Even for the subsample of households which have substantial asset holdings, one would suspect that the data may be generated by more than one structure, with some households' behavior consistent with the optimal consumption model and other households failing to conform to the model. Even if the possibility of heterogeneous behavior is not an important a priori consideration, confrontation with the data (e.g. Figure 1a) can reveal the implausibility of the single structure assumption. Given the data on stock accumulations vs. income growth plotted in Figure 1a, the choice between conventional and robust estimation techniques amounts to a choice between characterizing the behavior of household 676 or characterizing the behavior of the other 76 households in the sample.

Surprisingly, limiting the influence of outliers in X-space is simpler, both conceptually and computationally, in an instrumental variables context than in an OLS context. In OLS the vector of estimates is chosen to satisfy the normal

equations,

$$(28) \quad 0 = \sum_{i=1}^n (y_i - x_i b) x_{ji} \quad \text{for } j = 1 \text{ to } k$$

where  $x_i$  = ith row of X

and  $x_{ji}$  = ith observation on jth RHS variable.

Note that the contribution to this sum, and therefore the influence on the estimates, of observation i is a function of the product  $\hat{\epsilon}_i x_{ji}$ . If the observations on the x variables are well-balanced, highly influential observations will arise only in the instance of outliers in epsilon-space and therefore can be easily downweighted with the Huber estimator. However, if the x's are not "balanced" in the sense that there are one or more observations which contain extreme values of the x variable, such as household 676, these observations will be highly influential ones, even if the associated epsilon is well-behaved. Since observations which act as leverage points can exert a large degree of influence on the estimates without being "tagged" with a large residual, the Huber estimator will not identify and downweight them. In the OLS case, achieving robustness with respect to outliers in X-space requires 1) finding the appropriately defined norm for the k-element x-vector for each observation, 2) downweighting individual observations which are outliers according to this norm, and 3) computing weighted least squares estimates on the basis of the reweighted observations. Obtaining the appropriate norm of the x-vector for each observation is the burdensome step in the algorithm.

To think about the effect of outlying x-values in the instrumental variables context, consider the instrumental variables analog of the normal equations:

$$(29) \quad 0 = \sum_{i=1}^n (y_i - x_i b) z_{ji} \quad \text{for } j = 1 \text{ to } k$$

where  $z_{ji}$  = ith observation on the jth instrumental variable.

The influence of the ith observation on the estimates depends on the product

$\hat{\epsilon}_i z_{ji}$ , and not on  $x_{ji}$ . While the  $x$ -values appear in the normal equations, and obviously help to determine the estimates, Krasker, Kuh, and Welsch (1983) point out that in the instrumental variables context the  $x_i$  vector affects the estimates only via its effect on the  $i$ th residual, not directly as in the OLS case. Thus even if the data on the  $x$ -variables contain gross outliers, as long as the instruments are balanced the initial, nonrobust instrumental variables estimates will result in large residuals being associated with the leverage point observations. A Huber procedure in which observations with large residuals are iteratively downweighted can then be used to limit the influence of the outlier observations.

For this approach to work, one needs a specification in which the instruments are well-balanced. In some cases, the instruments suggested by theoretical considerations will be well-balanced just by the nature of the variable. For example, in the model estimated in this paper, the natural choice of instruments consists of expectations of the future, stated in qualitative terms such as household income being expected to rise by "a lot" or "a little", a variable which is obviously not contaminated with extreme value outliers. In other applications with micro data sets, variables commonly used as instruments include demographic data such as age, gender, household size, and education. These demographic variables also are by nature well-balanced, at least in comparison with economic variables such as income or wealth.

More generally, however, there is a sense in which the projection of the  $x$  variables on the primitive instruments inherent in the instrumental variables estimator tends to eliminate leverage points, even if both the observations on  $x$  and the observations on the primitive instruments, considered separately, are not balanced. To see this, note that the instruments can be thought of as fitted values of the  $x$  variables from projections on the list of primitive instruments,

and that as a consequence of this projection, some of the observations with extreme values of  $x$  will nevertheless have non-extreme values of fitted  $x$ .

Suppose that the structure generating  $X$  (the matrix of observations on the RHS variables) and  $Z$  (the matrix of observations on  $\ell$  primitive instrumental variables) is of the form:

$$(30) \quad \begin{array}{l} X = V\delta + \nu \quad \text{where } \ell \geq k \\ \begin{matrix} (nxk) & (nx\ell)(\ell xk) & (nxk) \end{matrix} \\ \\ Z = V + \eta \\ \begin{matrix} (nx\ell) & (nx\ell) & (nx\ell) \end{matrix} \end{array}$$

and  $\nu$ ,  $\eta$ , and  $V$  are mutually uncorrelated.

Whether the number of primitive instruments is equal to, or exceeds, the number of right hand side variables, equation (29) can be written as:

$$(31) \quad 0 = \sum_{i=1}^n (y_i - x_i b) \hat{x}_{ji}$$

where  $\hat{x}_{ji} = Z(Z'Z)^{-1}Z'x_j$

Note that the  $x$  variables can be outliers either because they contain extreme values of  $v$  or extreme values of  $\nu$ . For those outliers created because of extreme values of  $\nu$ , the projection on the  $z$ 's purges this component and leaves a non-outlier value of  $\hat{x}$ . Similarly, the primitive instruments may contain observations with extreme values, but as long as the outliers reflect large  $\eta$ 's rather than  $v$ 's, the fitted  $x$ 's will be well-balanced.

It would be too strong a statement to say that the first stage projection in the instrumental variables estimator guarantees that the fitted  $x$ 's will be devoid of leverage points. However, to the extent that the outliers in both  $x$  and  $z$  are the result of pure measurement error, uncorrelated across variables, this measurement error will be purged in the projection. Further, the fitted  $x$ 's are obviously observable; in an application in which the  $x$  variables are unbalanced, one can always check to see whether the outliers in the  $x$  variables and in the  $z$  variables were attributable to extreme values of  $v$  and therefore led

to extreme values in  $\hat{x}$ .

The algorithm for robust instrumental variables estimates, described in more detail below, obtains a first vector of residuals based on the initial, non-robust instrumental variables estimates. A robust estimate of the standard deviation of the residuals is calculated by scaling up the median absolute residual. Using the estimated standard deviation of the residuals and a choice of the parameter  $c$ , the Huber weights for each observation are computed. The Huber-weighted observations are then used to compute a new set of parameter estimates and thus a new set of residuals and observations weights. Iteration continues until the parameter estimates converge.

Algorithm for robust instrumental variables

1. Form instruments and obtain initial, nonrobust, estimates and residuals:

$$b_0 = (\hat{X}'X)^{-1}\hat{X}'y \quad \hat{X} = Z(Z'Z)^{-1}Z'X \quad r = y - Xb_0$$

2. Calculate estimate of scale parameter,  $\hat{\sigma}$ :

$$\hat{\sigma} = 1.483 \text{ median}(|r_i|)$$

This is a robust method of estimating the scale parameter. If the residuals are normally distributed,  $\hat{\sigma}$  is a consistent estimate of the standard deviation; however, if the residuals contain outliers, these observations will not affect the estimate.

3. Compute observation weights based on residuals,  $\hat{\sigma}$ , and  $c$ .

$$w_i = \min\left(1, \frac{c\hat{\sigma}}{|r_i|}\right)$$

4. Noting that  $0 = \sum_i w_i (y_i - x_i b) x_i = \sum_i w_i (y_i - x_i b) \sqrt{w_i} x_i$ , calculate the square roots of the weights,  $\sqrt{w_i}$ , and transform the data by multiplying each element in the  $i$ th row of  $y$ ,  $X$ , and  $Z$  by  $\sqrt{w_i}$ :

$$\bar{y}_i = \sqrt{w_i} y_i \quad \bar{x}_i = \sqrt{w_i} x_i \quad \bar{z}_i = \sqrt{w_i} z_i$$

5. Form new IV estimates with the transformed data:

$$\hat{\bar{X}} = \bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'\bar{X} \qquad b_{\text{new}} = (\hat{\bar{X}}'\hat{\bar{X}})^{-1}\hat{\bar{X}}'\hat{\bar{y}}$$

6. Form a new vector of residuals:

$$r_{\text{new}} = y - Xb_{\text{new}}$$

Note that the residuals are calculated using the new parameter estimates, but the untransformed data.

7. Go to step 2, and use the new vector of residuals to calculate a new iteration of values of  $\hat{\sigma}$ , the weights, and the parameter estimates. Continue until parameter estimates converge.

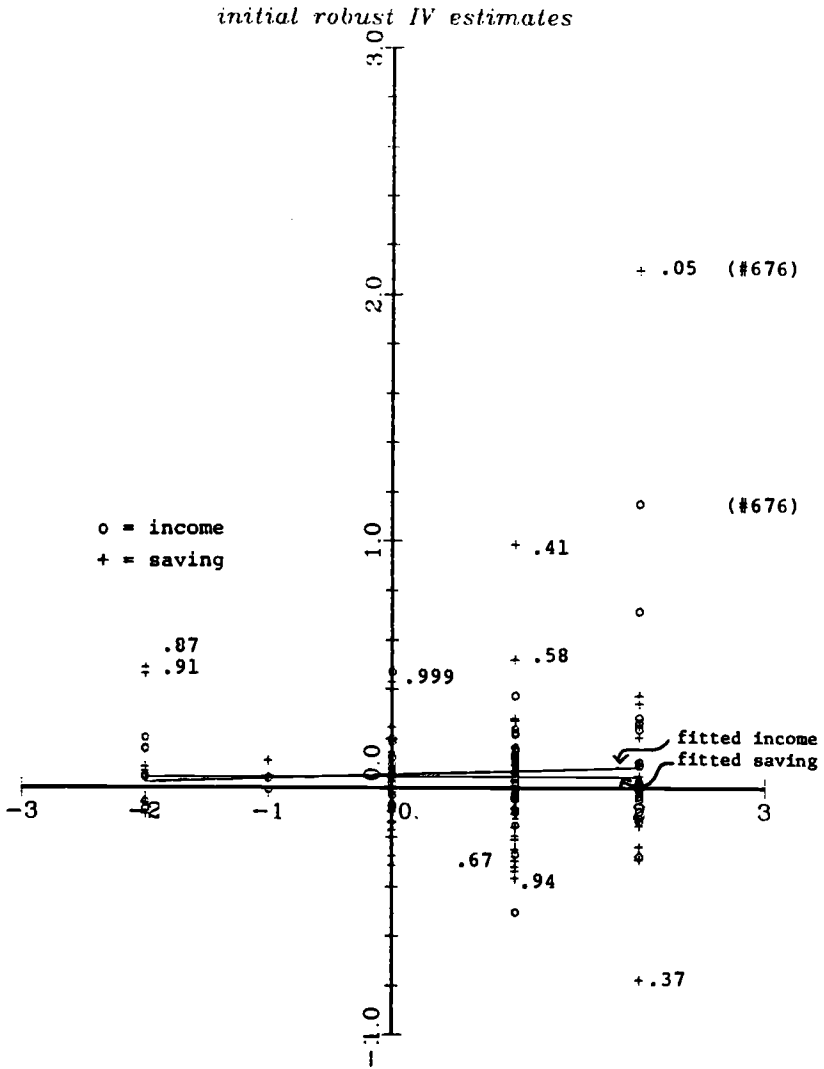
The key point exploited in the algorithm -- that as long as the instruments are balanced, outliers in X-space will be iteratively downweighted by the instrumental variables version of the Huber estimator -- is illustrated in Figures 2 and 3. These figures plot the data for the instrumental variables estimates of equation (20) for the truly wealthy subsample.

The empirical work in the paper estimates equations of the form of (20) with 8 instruments (in addition to the constant), 7 of which are dummy variables constructed by expanding responses to the two categorical expectational variables into sets of dummy variables. Figures 2 and 3 provide the visual intuition for the robust instrumental variables estimator used in the empirical work, the only simplification being the use of a single instrumental variable in place of the complete set of 8 instrumental variables. The single instrument was the response to the question, "Will your family income this year be higher or lower than last year?". In the plots, the response "a lot lower" was assigned the value -2, "a little lower" - -1, the same - 0, "a little higher" = 1, and "a lot higher" = 2. Obviously the instrument is "well-balanced" in the sense required for the estimator to limit the influence of outliers in X-space.

Having simplified the problem to one in which the model consists of a single RHS variable with a single instrumental variable, the instrumental variables



Figure 3





estimate of the slope parameter can be obtained by taking the ratio of 1) the estimated slope in the OLS regression of the dependent variable ( $\Delta S_t/y_{t-1}$ ) on the instrument, to 2) the estimated slope in the OLS regression of the RHS variable ( $\Delta y_t/y_{t-1}$ ) on the instrument. Figure 2 contains plots of both the savings data ( $\Delta S_t/y_{t-1}$ ) against the instrument (represented by plus signs), and the income data ( $\Delta y_t/y_{t-1}$ ) against the instrument (represented by circles). Also plotted in Figure 2 are the fitted lines from the OLS regressions of the savings variable on the instrument (estimated slope = .0709), and of the income data on the instrument (estimated slope = .0541).

Anticipating his graduation from medical school, household 676 appropriately stated that he expected his income in the coming year to be "a lot higher"; hence both the data on savings and income growth for this household show up at  $z = 2$  as gross outliers in the vertical dimension. As an outlier, household 676 is a highly influential observation in both of the OLS regressions illustrated in Figure 2, and therefore a highly influential observation in determining the nonrobust instrumental variables estimate of  $\gamma$  of 1.31.

Note that although household 676 is a gross outlier in X-space, by the choice of instruments, he is not an outlier in instrument-space. Thus while he exerts a high degree of influence on the nonrobust instrumental variables estimates, his extreme value of the x variable exerts its influence on the estimates not directly, as it would in an OLS context, but indirectly through the large residuals in the two projections. Since the influential effect of the observation works through the creation of an outlier in epsilon-space, the Huber approach is able to identify and downweight the outlying observation.

Based on the nonrobust IV estimates of equation (20), 5 observations had residuals which exceed  $2\sigma$  in absolute value. The Huber weights for these

observations,  $\frac{2\sigma}{|r_i|}$ , which will be used in the next iteration, are recorded on Figure 2 next to the point representing the value of the observation's LHS variable. Household 676 will receive a weight of .29 in the next iteration; four other observations receive weights less than unity, but are not downweighted as severely as observation 676.

The initial iteration of the robust estimates is plotted in a similar way in Figure 3 (note change in vertical scale). Using the Huber weights generated by the estimates plotted in Figure 2, Figure 3 plots the weighted observations and the two OLS regressions computed on the basis of the weighted observations. Since the downweighting of household 676, in particular, "moves" this observation toward the scatter of the other observations, household 676 now exerts less leverage on the estimates. With the slope of the income regression equal to .015, and the slope of the saving equation equal to -.00003, the estimate of  $\gamma$  is -.002 in the initial robust IV regression. Based on the residuals of this regression, new weights for each observation are computed for the next iteration. For the second iteration, the weight received by household 676 has fallen to .05. Seven other observations are downweighted, receiving weights between .37 and .999. Note that one iteration of the estimator results in a change in the economic interpretation of the data from  $\gamma = 1$  (asset stocks provide complete buffering of predictable fluctuations in income) to  $\gamma = 0$  (essentially no buffering by asset stocks).

Section 5: Robust results

Robust instrumental variables estimates were computed using the complete set of instruments and the algorithm described in the previous section. Table 3a reports the robust estimates of  $\gamma$  for the four samples.

The anomalous results which plagued the conventional estimates are gone. An embarrassing feature of the conventional results -- the fact that the inconsistency due to the nonzero value of the shadow price of the borrowing constraint should create downward bias in the estimate of  $\gamma$  for the poor subsample, whereas the estimate of  $\gamma$  for the poor exceeded the estimate of  $\gamma$  for the rich -- has been eliminated. While the point estimate of  $\gamma$  for the poor (.22) is slightly larger than the corresponding estimate for the rich (.18), this difference is not significant, either statistically or in terms of economic magnitude. Since the estimate of  $\gamma$  is consistent for the rich subsample and inconsistent for the poor subsample if borrowing constraints are binding for some of the households in the poor subsample, one could conduct a specification test for the presence of binding borrowing constraints (strictly positive values of the  $z_{i,t}$ ) by testing the equality of the estimates of  $\gamma$  across the two subsamples. Given the very similar estimates of  $\gamma$  across the two subsamples, a formal test is not required; the empirical results based on the sample split between rich and poor households reveal no evidence of borrowing constraints.

For the truly wealthy subsample, the robust point estimate of  $\gamma$  is .33, compared to the point estimate of 1.33 from the conventional estimator. The dramatic drop in the estimate of  $\gamma$  is in large part a result of the severe downweighting of household 676, which received a final weight of .06 in these estimates. Table 3b records the distribution of the final weights assigned to observations for the rich and the truly wealthy subsamples. While 40 out of the 350 observations in the rich subsample, and 10 out of the 77 observations in the

Table 3a

Robust instrumental variables estimates of  $\gamma$

sample	estimates of:		s.e.e.	number of observations	
	intercept	$\gamma$		total	downweighted
whole sample	.022 (.010)	.23 (.11)	.206	774	96
poor	.020 (.013)	.22 (.12)	.191	424	55
rich	.026 (.015)	.18 (.16)	.230	350	40
truly wealthy	.002 (.025)	.33 (.26)	.181	77	10

Note: Standard errors are in parentheses.<sup>12</sup>

Table 3b

Distribution of final weights for rich and truly wealthy subsamples

weight (or range)	number of observations	
	<u>rich subsample</u>	<u>truly wealthy subsample</u>
1	310	67
.9 to 1	7	4
.8 to .9	3	0
.7 to .8	5	1
.6 to .7	7	0
.5 to .6	7	1
.4 to .5	6	2
.3 to .4	0	1
.2 to .3	1	0
.1 to .2	3	0
0 to .1	1	1

<sup>12</sup>The standard errors were taken from the estimated covariance matrix:

$$\hat{\sigma}^2 (\hat{X}'\hat{X})^{-1} \hat{X}'\hat{X}(\hat{X}'\hat{X})^{-1}$$
 where  $\hat{\sigma}$  is the estimate of the standard deviation of the disturbances based on the median absolute deviation of the untransformed residuals.

truly wealthy subsample, were downweighted, household 676 was an outlier even among outliers. That is, for these two subsamples, fewer than 3% of the observations received final weights less than .40, and household 676 was the only observation with a weight less than .10.

For the truly wealthy subsample and the rich subsample, Tables 4a and 4b present the robust estimates of the savings equations for the disaggregated assets. In this exercise, saving in the form of liquid assets, stocks, and bonds has been aggregated into "financial assets", and the disaggregation is between financial assets, debt, and durable goods.<sup>10</sup> According to Table 4a, which pertains to the truly wealthy subsample, 30% of an anticipated increase in income is saved in the form of financial assets, 6% in the form of purchases of durable goods, and another 20% is saved in the form of reductions in debt. Note that compared to the earlier results for saving aggregated across these assets, the disaggregated results indicate that a larger fraction of predictable income

---

<sup>10</sup>For the truly wealthy subsample, it is possible to disaggregate further, and calculate the robust instrumental variables estimates for separate equations for liquid assets, stocks, and bonds. The robust results for the disaggregated financial assets had the following attributes. First, the estimate of the slope parameter,  $\phi_1\gamma$ , was zero for liquid assets and for bonds, and was close to zero (.04) and statistically insignificant for stocks. Second, for the disaggregated financial asset equations, the robust estimator downweighted almost a third of the observations (18 for liquid assets, 24 for stocks, and 26 for bonds). My interpretation of these results is that even for households which hold strictly positive levels of liquid assets, stocks, and bonds, in the presence of transactions costs a household which adjusts its total holdings of financial assets will tend to lump the adjustment into one or two of the disaggregated assets rather than adjust all three. If, at a high level of disaggregation, the bulk of the observations for an individual asset indicate no net transaction, the estimator may be downweighting any (or most) of the observations which reflect a nonzero transaction in the asset. Because transactions costs imply a large number of observations with a zero net transaction for highly disaggregated assets, liquid assets, stocks and bonds are aggregated into "financial assets" for the robust results.

Table 4a

Instrumental variables estimates  
of  $\varphi_1 \gamma$  for disaggregated asset stocks

truly wealthy subsample

dependent variable	estimate of: intercept	$\varphi_1 \gamma$	s.e.e.	# of observations downweighted
$\Delta S_{i,t+1}^{\text{Fin}} / Y_{i,t}$	.004 (.016)	.291 (.165)	.114	13
$\Delta S_{i,t+1}^{\text{Debt}} / Y_{i,t}$	.009 (.011)	-.218 (.047)	.082	10
$\Delta S_{i,t+1}^{\text{Dur}} / Y_{i,t}$	-.019 (.022)	.059 (.194)	.166	7

Note: Standard errors are in parentheses. Number of observations was 77.

Table 4b

Instrumental variables estimates  
of  $\varphi_1 \gamma$  for disaggregated asset stocks

rich subsample

dependent variable	estimate of: intercept	$\varphi_1 \gamma$	s.e.e.	# of observations downweighted
$\Delta S_{i,t+1}^{\text{Fin}} / Y_{i,t}$	.006 (.008)	.098 (.071)	.115	65
$\Delta S_{i,t+1}^{\text{Debt}} / Y_{i,t}$	.009 (.004)	-.060 (.043)	.060	84
$\Delta S_{i,t+1}^{\text{Dur}} / Y_{i,t}$	.015 (.009)	-.018 (.095)	.138	54

Note: Standard errors are in parentheses. Number of observations was 350.

fluctuations goes into saving.<sup>11</sup> For the rich subsample, the breakdown is: 10% of a predictable increase in income is saved in the form of financial assets, 6% in the form of a reduction in debt, and durables are not used as a mechanism for consumption smoothing.

The implausibility of the assumption that all of the data are generated by a single structure was one motivation for the use of the robust estimation approach. For the truly wealthy subsample, the estimator found observation 676 -- whose asset accumulation behavior seems to follow the optimal consumption model with a vengeance -- basically inconsistent with the incomplete smoothing behavior of the bulk of the observations. It is interesting to note that for all of the samples, the robust estimates of  $\gamma$  are uniformly lower than the corresponding conventional estimates. One interpretation of this pattern is that all of the samples contain a mixture of households, some of which display the incomplete smoothing behavior which is summarized by the robust results, and others of which conform more closely to the optimal consumption model.

### Conclusions

The paper uses a previously unexploited data set -- the Survey of Consumer Finances -- to ask whether the finding that consumption tracks current income more closely than is consistent with the permanent income hypothesis can be attributed partially or solely to the presence of borrowing constraints. Previous empirical studies of this question using the PSID data on food

---

<sup>11</sup>In comparing the disaggregated results with the aggregated results, note that in the final iteration the data are not identical in each of these equations in the sense that a given observation may well receive different weights in the different equations; an observation which is an outlier with respect to accumulation of financial assets may not be an outlier with respect to durable purchases or with respect to total saving.

consumption are inconclusive for several reasons: 1) the intertemporal behavior of food consumption may be unrepresentative of other components of consumption, 2) the data appear to be dominated by some form of noise, presumably the effects of preference shocks or of measurement error, and 3) the studies reach conflicting conclusions.

Using the SCF data on the income and asset stocks of households, the paper studies the saving side of the consumption/saving decision, and thus provides inferences on a comprehensive concept of consumption. Because the SCF data, like the PSID data, is apt to contain a substantial amount of noise, and, in addition, because it is more realistic to regard the empirical specification as a highly stylized approximation than as a complete and final description of the truth, the paper uses a robust instrumental variables estimator in order to limit the influence of outliers.

A comparison of the conventional and robust instrumental variables estimates indicates that some of the most striking aspects of the conventional results are driven by a single, extremely unusual observation. This highly influential data point appears to be basically a legitimate observation rather than a keypunch error. Further, unlike the conventional results, the pattern of the robust estimates across different subsamples is consistent with the assumptions which underlie the specification. One implication of those assumptions is the notion that, to the extent that borrowing constraints play an important role in determining the consumption of low-asset households, the estimated coefficient of the growth rate of income will be biased downward for the sample of low-asset households. Since the estimates of this parameter are essentially the same in two subsamples split according to household asset-holding, the results do not provide evidence of an important role for borrowing constraints. For the bulk of the observations, the results indicate that households exhibit incomplete



smoothing of consumption, with 20% to 50% of predictable movements in income being buffered by asset stocks. Finally, the estimated model summarizing the behavior of the majority of households is not an adequate model for all households; while a systematic treatment of the outlier observations is not provided, it seems likely that at least some of the heterogeneity in the data reflects the presence of households which conform more closely to the optimal consumption model.

References

- Altonji, Joseph G., and Aloysius Siow, "Testing the Response of Consumption to Income Changes with (Noisy) Panel Data", Quarterly Journal of Economics, 102, May 1987.
- Bernanke, Ben, "Permanent Income, Liquidity, and Expenditure on Automobiles: Evidence from Panel Data," Quarterly Journal of Economics, 1984.
- Bewley, Truman, "The Permanent Income Hypothesis: A Theoretical Formulation," Journal of Economic Theory, 16, 1977.
- Breeden, Douglas T., Ph.D. dissertation, Graduate School of Business, Stanford University, 1977.
- Breeden, Douglas T., "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," Journal of Financial Economics, vol. 7, pp. 265-296, 1979.
- Deaton, Angus, "Saving and Income Smoothing in the Cote d'Ivoire", working paper, Princeton University, 1990.
- Flavin, Marjorie A., "The Adjustment of Consumption to Changing Expectations about Future Income," Journal of Political Economy, 89, October 1981 974-1009.
- Flavin, Marjorie A., "The Excessive Sensitivity of Consumption to Current Income - Liquidity Constraints or Myopia?", Canadian Journal of Economics, Feb. 1985.
- Friedman, Milton, A Theory of the Consumption Function, Princeton University Press, Princeton, N.J., 1957.
- Garber, Peter M. and Robert G. King, "Deep Structural Excavation? A Critique of Euler Equation Methods," Technical Working Paper 31, NBER, Nov. 1983.
- Hall, Robert E., "Stochastic Implications of the Life Cycle -- Permanent Income Hypothesis: Theory and Evidence," Journal of Political Economy, 86, December, 1978, 971-987.
- Hall, Robert E., "Intertemporal Substitution in Consumption," Journal of Political Economy, 1988.
- Hall, Robert E. and Frederic S. Mishkin, "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households." Econometrica, 50, March 1982, 461-481.

- Hampel, Frank R., Elvezio M. Ronchetti, Peter Rousseeuw, and Werner A. Stahel, Robust Statistics: The Approach Based on Influence Functions, Wiley, 1986.
- Hansen, Lars Peter and Kenneth J. Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," Econometrica, 50, September 1982, 1269-1286.
- Hansen, Lars Peter and Kenneth J. Singleton, "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," Journal of Political Economy, April 1983, 249-266.
- Hayashi, Fumio, "The Effect of Liquidity Constraints on Consumption: A Cross-Sectional Analysis," Quarterly Journal of Economics, 1985.
- Hayashi, Fumio, "The Permanent Income Hypothesis: Estimation and Testing by Instrumental Variables," Journal of Political Economy, 90, 1982, 895-916.
- Hayashi, Fumio, "The Permanent Income Hypothesis and Consumption Durability: Analysis Based on Japanese Panel Data," Quarterly Journal of Economics, 100, November 1985.
- Huber, P.J., "Robust Regression: Asymptotics, Conjectures, and Monte Carlo", Annals of Statistics, 1, 1973.
- Huber, P.J., "Robust Methods of Estimation of Regression Coefficients", Math. Operationsforschung Statist. Ser. Statist., 8, 1977.
- Krasker, William S., Edwin Kuh, and Roy E. Welsch, "Estimation for Dirty Data and Flawed Models", Handbook of Econometrics, Vol. 1, Griliches and Intriligator, eds, North Holland, 1983.
- Runkle, David E., "Liquidity Constraints and the Permanent Income Hypothesis: Evidence from Panel Data," Journal of Monetary Economics, 27, February 1991.
- Shapiro, Matthew D., "A Note on Tests of the Permanent Income Hypothesis in Panel Data", unpublished, May 1982.
- Shapiro, Matthew D., "The Permanent Income Hypothesis and the Real Interest Rate: Some Evidence From Panel Data," Economic Letters, 14, 1984, 93-100.
- Zeldes, Stephen, "Consumption and Liquidity Constraints: An Empirical Investigation," Journal of Political Economy, 97, April 1989(a).
- Zeldes, Stephen, "Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence," Quarterly Journal of Economics, 104, May 1989(b).

Appendix

Raw data on household 676

<u>question</u>	<u>1967</u>	<u>1968</u>	<u>1969</u>
(Interviews were conducted in Jan - March of each year)			
panel status		panel, same respondent as in 1967	panel, same respondent as in 1968
sex of head of household	male	male	male
age of head of household	30	31	32
# of adults 18 years or older	2	2	2
# of children under 18	4	4	4
how long married (years)	6	7	9
education of head	college, bachelors	college, advanced or professional degree	college, advanced or professional degree
Are you (head) working now?	student	working now	working now
What kind of business?	inapp.	Government Medical, Health and Education Services	Services
occupation	student	professional technical and kindred workers	professional technical and kindred workers
weeks unemployed last year	none; inapp.	none, inapp.	21
earned income, family unit	0	3,600	27,000
mixed labor/capital income	0	0	0
total capital income	100	800	100
total transfer payments	0	0	0
total income	100	4,400	27,100
total income tax	0	94	4,527

<u>question</u>	<u>1967</u>	<u>1968</u>	<u>1969</u>
value of any money, stocks, bonds, property inherited since last year	0		0
present value or cost of house	5,500		17,000
monthly mortgage payments or rent			125
net equity in house	5,500	0	2,000
Do you own any real estate such as a lot, summer home, apartment building, or business property (other than this place here)? What is the value of property?		7,000	15,000
Do you owe any money on the property?		0	
total price of car bought in 66/67	1,500		
amount borrowed if car bought in 66/67	0		
Did you trade in or sell a car when you bought that one?	no		
total value of the car stock	2,170		
net outlay on all cars bought in year preceding survey	1500	0	3,300
expenditures on additions and repairs	460	0	2,400
amount borrowed on additions and repairs			0
total outlay on all durables (does not include cars)	115	255	3,250
amount spent on vacations	400	200	500
expenditures on hobbies and recreation	25		0
If you plan to buy a car in the coming year, how much do you expect to spend?		3,000	
amount you expect to spend on durables		1,000	

<u>question</u>	<u>1967</u>	<u>1968</u>	<u>1969</u>
Thinking ahead to next year, do expect you family income to be higher or lower?	higher, a lot		
Thinking ahead about four years, do you expect your family income to be higher or lower?	much higher		
Do you expect prices to go up, down, or stay the same?	will go up		will go up
What do you expect the inflation rate to be?	2% or less		5%
total amount in checking		100	1000
size of increase in checking acct		0	0
Has the amount in savings account gone up or down since last year?		no change	0
savings account net of CD's		0	0
value of CD's		0	0
total liquid assets (checking accounts, savings accounts, CD's)		600	
increase in liquid assets since last year		0	
value of stocks and shares		1,000	25,000
amount owed on stocks		0	0
amount of purchase (or sale) of stocks		0	20,000
face value of bonds		500	0
how much more in bonds than a year ago		0	0
total remaining debt incurred in all years (medical and dental included)	0	4,000	4,000

Note: Blanks in this table should not be interpreted as zeros. A blank indicates that the question was not asked in that survey year or I did not pull the variable in the extract.