NBER WORKING PAPER SERIES

ROBUST ESTIMATION OF THE JOINT CONSUMPTION/ASSET DEMAND DECISION

Marjorie Flavin

Working Paper 7011
http://www.nber.org/papers/w7011

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1999

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 source or collectors of the data nor the Consortium bear any responsibility for the analyses or interpretations presented here. For helpful comments, I thank Angus Deaton, James Hamilton, and Roger Koenker, as well as seminar participants at Rice University, U. of Houston, U. of Texas, Texas A&M, U. of Montreal, Berkeley, and University of Washington. Research support from the NSF is gratefully acknowledged.The views expressed in this paper are those of the authors and do not reflect those of the National Bureau of Economic Research.

Robust Estimation of the Joint Consumption/
Asset Demand Decision
Marjorie Flavin
NBER Working Paper No. 7011
March 1999
JEL No. C20, F21

ABSTRACT

The paper proposes an instrumental variables version of the Huber estimator as an alternative to the IV-Krasker Welsch estimator. The IV-Huber estimator is analytically and computationally much simpler than IV-Krasker Welsch. In the context of an empirical study of the importance of borrowing constraints on consumption, the paper reports the results for the following estimators: 1) conventional (non-robust) IV, 2) conventional IV after the subjective rejection of outliers, 3) conventional IV after trimming, 4) IV-Huber, and 5) IV-Krasker-Welsch. In the presence of a heavy-tailed error distribution, both the IV-Krasker Welsch and the IV-Huber estimators provide substantial improvements in efficiency over conventional IV. Further, the informal robust procedure of using conventional IV after trimming does not match the efficiency gains of the formal robust methods.

The empirical results indicate that households exhibit incomplete smoothing of consumption, with about 20-50% of predictable movements in income being buffered by asset stocks. When saving is disaggregated by type of asset, the results provide some evidence of borrowing constraints: households which are not subject to a liquidity constraint use financial assets as their primary means of buffering income fluctuations, while constrained households use purchases of durable goods almost exclusively as the vehicle for consumption smoothing.

Marjorie Flavin
Economics Department, 0508
UCSD
9500 Gilman Drive
La Jolla, CA 92093-0508
and NBER
mflavin@ucsd.edu

While robust estimation has been an active area of research in the statistics literature for the past twenty years, robust estimators have not been fully exploited in applied econometric work, despite the widespread recognition that economic data sets are very likely to contain highly influential observations or outright data errors. One barrier to transfer of robust estimation techniques from the statistics literature into econometric practice is the fact that research on the theory and performance of robust estimators suitable for simultaneous equations systems (i.e., analogs of instrumental variables estimators) is considerably less well developed than the large volume of research on robust analogs of OLS.

A notable exception is Krasker and Welsch, Econometrica (l985), which provides an instrumental variables version of the earlier Krasker-Welsch estimator (JASA (1982)). While the Krasker-Welsch estimator is preferable to non-robust methods, or informal and ad hoc "robust" methods, the estimator has a number of limitations, and is moderately difficult to calculate. Whether for these or other reasons, the Krasker-Welsch estimator does not enjoy widespread or even significant usage in applied econometric work; in this sense it has not passed the basic "market test".

This paper proposes an instrumental variables version of the Huber estimator, and argues that the IV-Huber estimator avoids some of the undesirable features of the IV-Krasker-Welsch estimator. Perhaps more importantly, the proposed IV-Huber estimator is much simpler, both analytically and computationally, than IV-Krasker-Welsch. In addition to proposing the IV-Huber estimator and stating its asymptotic properties, the paper reports the results of Monte Carlo experiments to assess the relative efficiency of various alternative estimators (non-robust, non-robust after subjective rejection of outliers, IV-Huber, and IV-Krasker-Welsch), estimate the magnitude of small sample bias, and verify the asymptotic errors. The Monte Carlo experiments show that the improvement in efficiency under heavy-tailed error distributions is large enough to be of practical importance to applied econometricians.

In addition to proposing the IV-Huber estimator, a second purpose of the paper is to promote the use of robust methods in applied econometric work, by arguing that robust estimation is a) more necessary, and b) easier to implement, than is commonly believed. To demonstrate that robust estimators can be crucial in determining the economic conclusions which emerge from an empirical study, the paper compares the estimates generated by conventional, nonrobust instrumental variables and by the two robust instrumental variables procedures in the context of a fully specified empirical problem. The economic question addressed is: Can the finding that consumption tracks income more closely than is consistent with general specifications of the optimal consumption model be attributed solely or partially to the effects of borrowing constraints? For this application, it turns out that the use of a robust estimator completely reverses the economic conclusions which are obtained with the nonrobust estimator. The empirical work uses a previously unexploited household data set —the University of Michigan Survey of Consumer Finances — and is further distinguished from previous studies by framing the empirical specification in terms of the implications for saving rather than consumption. By using data on income and a comprehensive array of asset stocks, the empirical work presented in the paper provides inferences on a comprehensive concept of consumption, rather than a limited category such as food consumption. Further, to the extent that households do use asset stocks to smooth their consumption path, one can use the asset data, disaggregated by type of asset, to determine which assets play the major role in buffering income fluctuations.

The paper uses a generous definition of saving, in which expenditures on durable goods are considered saving rather than consumption. The empirical results indicate that households exhibit incomplete smoothing of consumption, with 20 - 50% of predictable movements in income being buffered by asset stocks. The results based on total saving flows fail to reveal any evidence of borrowing constraints. When saving is disaggregated by type of asset, however, the results do provide some evidence of borrowing constraints: households which are not subject to liquidity constraints use financial assets as their primary means of buffering income fluctuations, while constrained households use purchases of durable goods almost exclusively as the vehicle for consumption smoothing.

Section 1 derives the consumption model and discusses the data used in the empirical work; Section 2 states the IV version of the Huber estimator and compares the IV-Huber estimator with the IV-Krasker-Welsch estimator; Section 3 reports the estimation of the model under the alternative estimators and the results of the Monte Carlo experiments.

Section 1: The model

The specification of the household's optimization problem follows Zeldes (l989a). Households maximize expected lifetime utility, subject to several constraints: (1) a lifetime budget constraint, (2) a non-negativity constraint on consumption, and (3) (potentially) a constraint on borrowing. The objective function is:

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

where E_t denotes the conditional expectation operator, δ_i denotes household i's rate of time preference; $c_{i,t}$ represents household i's consumption during period t; and $\theta_{i,t}$ is a preference shock.

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

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

where α is the coefficient of relative risk aversion.[1] The lifetime budget constraint is given by:

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

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

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; and $Y_{i,t}$ denotes real, after-tax labor income in year t.

Finally, households may face a constraint on the extent to which their assets can go negative:

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

When the borrowing constraint (equation (4)) is not currently binding, optimal consumption behavior is characterized using the Euler equation approach developed by Hall (l978, l982) and by Hansen and Singleton (l982, l983). For CRRA utility, the Euler equation is:

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

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). Using the distributional assumptions suggested by Breeden (l977), assume that the rate of return, the marginal utility of consumption, and the preference shock are jointly log normally distributed. With these distributional assumptions, the Euler equation implies that

$$(6) \quad \ln\left(\frac{c_{i,t+1}}{c_{i,t}}\right) = \frac{1}{\alpha}\left[E_t \ln\left(1+r_{i,t}^j\right) + \frac{v_{r_j} + v_u}{2} + \mathrm{cov}_{r_j u} - \ln(1+\delta_i)\right] + \frac{1}{\alpha}\left(E_t\theta_{i,t+1} - \theta_{i,t}\right) + \tilde{\eta}_{i,t+1}$$

In equation (6), $\tilde{\eta}_{i,t+1}$ is the error in forecasting ln $c_{i,t+1}$, and v_{r_j}, v_u, and $\mathrm{cov}_{r_j u}$ are moments of the joint distribution of the asset return, the marginal utility of consumption, and the preference shock:[2]

$$(7) \quad \ln\left(1+r_{i,t}^j\right) \sim N\left(m_{r_j}, v_{r_j}\right)$$

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

$$\mathrm{cov}\left(\ln\left(1+r_{i,t}^j\right), \ln c_{i,t+1}^{-\alpha} + \theta_{i,t+1}\right) = \mathrm{cov}_{r_j u}$$

Under these distributional assumptions, $\tilde{\eta}_{i,t+1}$ is normally distributed with mean zero.

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 (6); that is, the household would like to reallocate some consumption from t+1 toward t, but is unable to do so. 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 to t+1 — a borrowing constraint never forces the expected growth rate of consumption to be lower than the unconstrained optimal growth rate. Following Zeldes, the effects of a borrowing constraint are incorporated into the empirical analysis by adding an additional disturbance $z_{i,t}$:

$$(8) \quad \ln\left(\frac{c_{i,t+1}}{c_{i,t}}\right) = \frac{1}{\alpha}\left[E_t \ln\left(1+r_{i,t}^j\right) + \frac{v_{r_j} + v_u}{2} + \mathrm{cov}_{r_j u} - \ln(1+\delta_i)\right] + \frac{1}{\alpha}\left(E_t\theta_{i,t+1} - \theta_{i,t}\right) + \tilde{\eta}_{i,t+1} + z_{i,t}$$

where 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. To obtain the implications for savings flows, rewrite the consumption growth rate using Taylor's Theorem:

$$(9) \qquad \ln c_{t+1} - \ln c_t = \frac{\Delta c_{t+1}}{c_0} - \frac{1}{2}\left[\left(\frac{c_{t+1} - c_0}{c^*_{t+1}}\right)^2 - \left(\frac{c_t - c_0}{c^*_t}\right)^2\right]$$

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. Letting $c_0 = y_t$ be the point of linearization, (9) becomes

$$(10) \qquad \ln c_{t+1} - \ln c_t = \frac{\Delta c_{t+1}}{y_t} - \frac{1}{2}\left[\left(\frac{c_{t+1} - y_t}{c^*_{t+1}}\right)^2 - \left(\frac{c_t - y_t}{c^*_t}\right)^2\right]$$

Using just the linear term, the optimal consumption model can be approximated by:[3]

$$(11) \qquad \frac{\Delta c_{i,t+1}}{y_{i,t}} = \frac{1}{\alpha}\left[E_t \ln\left(1+r^j_{i,t}\right) + \frac{v_{r_j} + v_u}{2} + \mathrm{cov}_{r_j u} - \ln(1+\delta_i)\right] + \frac{1}{\alpha}\left(E_t\theta_{i,t+1} - \theta_{i,t}\right) + \tilde{\eta}_{i,t+1} + z_{i,t}$$

Subtracting from the growth rate of income to obtain the implications for savings flows:

$$(12) \frac{\Delta S_{i,t+1}}{y_{i,t}} = \frac{\Delta y_{i,t+1}}{y_{i,t}} - \frac{1}{\alpha}\left[E_t \ln\left(1 + r^j_{i,t}\right) + \frac{v_{r_j} + v_u}{2} + \mathrm{cov}_{r_j u} - \ln(1+\delta_i)\right] - \frac{1}{\alpha}\left(E_t\theta_{i,t+1} - \theta_{i,t}\right) - \tilde{\eta}_{i,t+1} - z_{i,t}$$

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

Equation (12) is the flip side of the usual consumption orthogonality condition. 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 consumption smoothing model.

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:

$$(13) \qquad S_{i,t} = \Delta\text{Saving}_t + \Delta\text{Checking}_t + \Delta\text{Bonds}_t + \Delta\text{Stocks}_t - \text{Debt}_t + \text{Debt}_{t-1} + \text{Durables expenditure}_t \ .$$

Durable goods have some consumption aspects and some investment aspects; in this paper durable goods are treated as assets rather than as consumption.

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:

(14) $c_t = \kappa + \beta y_t$.

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

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

which is conveniently nested in equation (12).

Thus the task is to estimate the following model for the change in saving:

(16) $$\frac{\Delta S_{i,t+1}}{y_{i,t}} = \gamma\frac{\Delta y_{i,t+1}}{y_{i,t}} - \frac{1}{\alpha}\left[E_t \ln\left(1+r^j_{i,t}\right) + \frac{v_{r_j} + v_u}{2} + \mathrm{cov}_{r_j u} - \ln(1+\delta_i)\right] - \frac{1}{\alpha}\left(E_t\Delta\theta_{i,t+1}\right) - \tilde{\eta}_{i,t+1} - z_{i,t}$$

where γ=1 under the null hypothesis and 1-γ =β. 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 l967, l968, and l969. 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 l968 and l969.

The fact that the data is a pure cross-section introduces several considerations.[4] First, in a single cross-section, the expectational revision term, $\tilde{\eta}_{i,t+1}$, 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:

(17) $$\tilde{\eta}_{i,t+1} = \eta_{t+1} + \eta_{i,t+1}$$

where the idiosyncratic shock, $\eta_{i,t+1}$, has mean zero and is uncorrelated across households. Second, 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, η_{t+1} gives[5]

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

With these additional assumptions, the intercept term

$$\mu = -\eta_{t+1} - \frac{1}{\alpha}\left[E_t \ln\left(1 + r_{i,t}^{j}\right) + \frac{v_{r_j} + v_u}{2} + cov_{r_j u} - \ln(1+\delta_i) \right] \quad .$$

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 (18) 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.[6] 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 l967 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)?" "Thinking ahead about four years, would you say that your family income will 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 model 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 and therefore provide more precise estimates.

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 expectation 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 shift for families experiencing a change in 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 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 household's 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 (18), using the expectational variables as instruments, provides a consistent estimate of γ. 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 binding borrowing constraints will result in a downward biased estimate of γ for the subsample of households with low levels of assets. Thus testing the restriction that the estimates of γ 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.

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 l967, 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 payment 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:

1. 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 included capital gains. Transfer payments include "help from relatives", unemployment compensation, and welfare payments.
2. ΔSaving$_t$, ΔChecking$_t$, and ΔBonds$_t$ 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?"
3. ΔStocks$_t$ 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?"
4. Durables Expenditure$_t$ 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).
5. Debt$_t$ is total debt remaining at the end of year t.

Section 2: An alternative robust estimator

This paper proposes an instrumental variables version of the Huber estimator, and argues that the IV-Huber estimator avoids some of the undesirable features of the IV-Krasker-Welsch estimator. Perhaps more importantly, the proposed IV-Huber estimator is much simpler, both analytically and computationally, than IV-Krasker-Welsch. While the two estimators are derived from different optimization problems, the first order conditions which define the estimates are closely related, and in many contexts the two estimators will generate similar estimates. In the ordinary regression context, the optimization problem motivating the original Krasker-Welsch estimator was the minimization of the asymptotic covariance matrix subject to a bound on the sensitivity (or maximum influence of any single observation on the estimates) of the estimator.[7] In the instrumental variables context, Krasker and Welsch derive their estimator by starting with the standard instrumental variables estimator and modifying the estimator only by downweighting any observations which would otherwise violate the sensitivity bound.

To establish notation, assume

(19) $$y_i = x_i\beta + \varepsilon_i$$

where $x_i = 1\times k$ vector of explanatory variables
$\beta = k\times 1$ vector of parameters
$\varepsilon_i =$ disturbance from a symmetric distribution with scale parameter σ
$\hat{x}_i = 1\times k$ vector of instruments.

Denoting the sensitivity of the estimator as δ, and the bound on the sensitivity as $a>0$, the IV-Krasker-Welsch estimator for sensitivity bound $\delta<a$ is defined as the solution to:

(20a) $$\sum_{i=1}^{n} \min\left(1, \frac{a}{\left|\frac{y_i - x_i\beta}{\sigma}\right| \left\{\hat{x}_i A^{-1} \hat{x}_i{}'\right\}^{1/2}}\right)\left(\frac{y_i - x_i\beta}{\sigma}\right)\hat{x}_i = 0$$

(20b) $$A = \frac{1}{n}\sum_{i=1}^{n} E\left[\min\left(\eta^2, \frac{a^2}{\hat{x}_i A^{-1} \hat{x}_i{}'}\right)\right]\hat{x}_i{}'\hat{x}_i$$

where η ~ N(0,1) and A is the kxk matrix which satisfies (20b).

In addition to satisfying the heuristic criterion of being "as close as possible" to the standard instrumental variables estimator, the IV-Krasker-Welsch estimator has the property that when equation (20) is evaluated for the special case of OLS (i.e., when $\hat{x}_i = x_i$), the estimator coincides with the original Krasker-Welsch estimator for the ordinary regression case, which was shown to have a certain efficiency property[8] under the assumptions of the central model (i.e., under the assumption that the disturbances are distributed normally).

The Huber estimator is based on a different optimization problem. Huber (l964) considered the following two-person zero-sum game: Nature chooses the distribution of the disturbances within a neighborhood of the central model, and the statistician chooses an estimator from the class of M-estimators. Nature's objective is to maximize the asymptotic variance of the estimator, while the statistician's objective is to minimize the asymptotic variance. One class of distributions considered by Huber is the gross-error model, in which a known fraction, ε, of the data come from a symmetric, but otherwise unknown, distribution and the remainder, 1-ε, come from the central model (i.e., normal) distribution. For this class of distributions, the error distribution, F, is given by:

$$F = (1-\varepsilon)\Phi + \varepsilon H, \tag{21}$$

where Φ denotes the cumulative standard normal distribution and H denotes the contaminating distribution.

For the gross error model, the saddle point of the game is: Nature chooses the distribution, f_0, which is normal in the middle and exponential in the tails:

$$\begin{aligned} f_0(t) &= (1-\varepsilon)(2\pi)^{-1/2}\exp\left(-\frac{1}{2}t^2\right) && \text{for } |t| < c \\ &= (1-\varepsilon)(2\pi)^{-1/2}\exp\left(-c|t| + \frac{1}{2}c^2\right) && \text{for } |t| \ge c \end{aligned} \tag{22}$$

and the econometrician chooses the Huber estimator, which for the ordinary regression case is given by:

$$\sum_{i=1}^{n}\min\left(1, \frac{c}{\left|\frac{y_i - x_i\beta}{\sigma}\right|}\right)\left(\frac{y_i - x_i\beta}{\sigma}\right)x_i = 0 \tag{23}$$

where the "tuning constant", c, depends on the fraction of gross errors, ε .

Huber's approach is aptly described as a "minimax strategy" in the sense that the objective is to minimize the asymptotic variance for the worst case distribution in a neighborhood of the central model. The Huber estimator is robust in the sense that while the estimator is slightly less efficient than OLS at the central (normal) model, the asymptotic variance does not deteriorate as rapidly as OLS under distributions in a neighborhood of the normal.

From equation (23), 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 $\frac{c\sigma}{|y_i - x_i\beta|}$. Intuitively, the Huber estimator transforms the data by "moving" observations with large residuals to the $c\sigma$ bound around the regression line.

The dependence of the tuning constant, c, on the fraction of gross errors, ε, is given by:

(24) $$(1-\varepsilon)^{-1} = \int_{-c}^{+c} \varphi(t)dt + \frac{2\varphi(c)}{c}$$ where $\varphi(t)$ is the standard normal density.

In the limiting case of no contamination ($\varepsilon = 0$ and $c = \infty$), the estimator coincides with OLS; in the opposing limiting case of 100% contamination ($\varepsilon \to 1$ and $c \to 0$), it coincides with least absolute deviations.

Restricting our attention to the ordinary regression case for the moment, note that despite the fact that the Krasker-Welsch and Huber estimators were derived from different optimization problems, the resulting first-order conditions are very closely related. The ordinary regression version of Krasker-Welsch is given by:

(25) $$\sum_{i=1}^{n} \min\left(1\,,\, \frac{a}{\left|\frac{y_i - x_i\beta}{\sigma}\right| \left\{x_i A^{-1} x_i'\right\}^{1/2}}\right)\left(\frac{y_i - x_i\beta}{\sigma}\right) x_i = 0$$

$$A = \frac{1}{n}\sum_{i=1}^{n} E\left[\min\left(\eta^2, \frac{a^2}{x_i A^{-1} x_i'}\right)\right] x_i' x_i$$

and the Huber estimator is given by (23). In terms of the first-order conditions, the difference between the two estimators is that the Huber estimator downweights an observation only if its residual exceeds a bound, while the Krasker-Welsch estimator downweights an observation if the observation would otherwise exceed the sensitivity bound on the basis of the size of its residual and its value of x_i, combined.

Both the Huber estimator and the Krasker-Welsch estimator have limitations. A major limitation of the Huber estimator is that while the estimator limits the influence of outliers in ε-space, the influence of outliers in x-space, or leverage points, is not bounded. That is, the Huber estimates can be inordinately affected by a single observation, or small group of observations, if those observations have extreme values of the x-variables and are unrepresentative of the bulk of the sample.

The Krasker-Welsch estimator was designed precisely to overcome this lack of robustness with respect to outliers in x-space. However, placing a bound on the overall influence of any observation, whether the source of the influence is the observation's ε or its position in x-space, comes at the price of additional complexity relative to the Huber estimator. That is, since the first order condition (25) for the Krasker-Welsch estimator requires the robust distances (robust distance = $d_i = (x_i A^{-1} x_i')^{1/2}$) of each of the observations, one must first solve for the fixed point of the kxk matrix A in (25). Aside from the additional programming and computation required, the introduction of the robust distances introduces some limitations in addition to the desired robustness with respect to outliers in x-space. First, both the statement of the optimization problem and the implementation of the Krasker-Welsch estimator invoke the assumption of normally distributed errors. The optimization problem is the minimization of the asymptotic variance of the estimates under the central model (i.e., normally distributed errors), subject to the sensitivity bound. Since the fixed point for A, and therefore the robust distances, are computed under the assumption of normally distributed errors, the behavior of the Krasker-Welsch estimator under non-normally distributed errors is not known.

Second, in order to implement the Krasker-Welsch estimator, one needs to specify the value of the sensitivity bound, a. One approach, suggested by Krasker and Welsch, is to specify the desired efficiency of the K-W estimator relative to OLS under the ideal conditions of a nonstochastic X matrix and normally distributed errors, and iterate on a to find the value of the sensitivity bound which achieves the desired relative efficiency. Alternatively, Krasker and Welsch suggest that the ad hoc rule of setting $a = 1.8\sqrt{k}$ works well in practice, in the sense that this choice generally results in plausible percentages of observations being downweighted. However, regardless of whether one sets the sensitivity bound to achieve a given relative efficiency or uses an ad hoc rule, the specification of the bound amounts to picking a particular point on the trade-off between the tightness of the sensitivity bound and efficiency. It's difficult to know how this trade-off should be made. For the Huber estimator, it is, of course, unrealistic to assume that one knows a priori the value of ε, and therefore the optimal choice of the tuning constant, c. However, because the Huber estimator downweights outliers smoothly rather than trimming observations, the estimator is fairly insensitive to the value of c. Huber shows that for the true fraction of gross errors within the range $0 \le \varepsilon \le .2$, the asymptotic variance of the estimates computed for any choice of c in the range $1 \le c \le 2$ is only slightly larger than the minimum variance estimates (i.e., the estimates computed with the optimal choice of c).

In the ordinary regression context, many economic applications, particularly those which use household data, will exhibit observations that are outliers in x-space. In these instances, a strong case can be made for the Krasker-Welsch estimator over the Huber estimator, despite the additional complexity of the estimator and its dependence on the assumption of normality.

Each of the estimators has a straightforward instrumental variables version, equation (20) for IV-Krasker-Welsch (IV-KW), and equation (26) for the IV-Huber (IV-H):

$$(26) \qquad \sum_{i=1}^{n} \min\left(1\,, \frac{c}{\left|\frac{y_i - x_i\beta}{\sigma}\right|}\right)\left(\frac{y_i - x_i\beta}{\sigma}\right)\hat{x}_i = 0$$

As a member of the class of GMM estimators, the IV-Huber estimator is asymptotically normal. One might assume that, parallel to the OLS case, the presence of extreme values or errors in the x variables would make a strong case for preferring IV-KW over IV-Huber. Contrary to this presumption, I argue that the shift from ordinary regression to instrumental variables itself provides some protection against outliers in x-space, so that in many cases IV-H will provide estimates that are robust with respect to outliers in both ε-space and x-space, without requiring the additional complexity and limitations of the IV-KW estimator.

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

$$(27) \qquad 0 = \sum_{i=1}^{n}\left(y_i - x_i b\right)\hat{x}_i$$

The influence of the ith observation on the estimates depends on the product $\hat{\varepsilon}_i \hat{x}_i$, and not on x_i. While the x-values appear in the normal equation, and obviously help to determine the estimates, Krasker, Kuh, and Welsch (l983) point out that in the instrumental variables context the x_i vector affects the estimates only via its effect on the ith 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. To make the point another way, consider the formula for the IV-KW estimator (equation (20)). The only difference between IV-H and IV-KW results from the use in IV-KW of the robust distances. In the IV context, the distances that matter are the robust distances of the instruments, not the x-variables themselves. If the instrumental variables are well-balanced (i.e., devoid of extreme values) the robust distances will be reasonably uniform across observations, and the vectors of observation weights used by the two estimators will be fairly similar.

For this approach to work, one needs a specification in which the instruments are well-balanced. For this paper, the natural choice of instruments consists of the household's stated expectations of the future, expressed in qualitative terms such as household income being expected to rise by "a lot" or "a little", etc., a variable that by construction contains no extreme values. 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 on the primitive instruments, considered separately, contain outliers.

Section 3: Empirical results

Empirical results are presented for the following four samples:

descriptive label	sample size	criteria for inclusion in sample
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 to 1,197 households. Deleting households with missing data further reduced the sample to 774.
poor	424	subset of the whole sample for which the sum of the household's financial

		wealth (checking accounts, savings accounts, bonds, and stocks) at the end of l967 was less than $1000 in l968 dollars
rich	350	subset of the whole sample for which household financial wealth at the end of 1967 was greater than or equal to $1000 in l968 dollars
truly wealthy	77	subset of the rich subsample for which the household held strictly positive levels of each of the following assets: bank deposits (i.e., savings and/or checking accounts), bonds, and stocks

For comparison, the model was estimated using conventional (nonrobust) IV, IV-Krasker-Welsch, and two informal robust methods, in addition to IV-Huber. The most common method of dealing with outliers consists of excluding (or trimming) from the sample observations for which the value of the dependent variable falls in the extreme tails of its distribution. Since this approach involves a zero/one decision in which some observations are excluded completely, the cut-off points are usually set so that only a few percent of the data is lost.[9] The second informal method of dealing with outliers is simply the subjective exclusion of individual data points which appear to have an inordinate effect on the estimates in preliminary results.

For the Survey of Consumer Finances data, preliminary estimates reflected the presence of (at least) one highly unusual and highly influential observation. Unlike most of the households in the data set, which exhibited incomplete consumption smoothing, household 676 followed the consumption smoothing model with a vengeance: the household experienced a quadrupling of its income between 1968 and 1969, and increased its net purchases of stocks by an amount equal to over 400% of its 1968 income.[10] To represent the informal robust estimation procedure based on subjective exclusion of highly influential observations, the model was estimated by conventional IV after exclusion of household 676.

Table 1: Estimates of γ

	conventional	**delete 676**	**trimmed**	**IV-Huber c=2.0**	**IV-Huber c=1.4**	**IV-KW a=1.8√k**
poor:						
estimate	.61	.61	.08	.22	.23	.30
s.e.	(.43)	(.43)	(.15)	(.20)	(.23)	(.19)
% weighted				13.0%	22.2%	17.5%
rich:						
estimate	.27	.003	.15	.18	.15	.14
s.e.	(.34)	(.26)	(.25)	(.33)	(.35)	(.52)
% weighted				11.4%	22.3%	17.4%
truly wealthy:						

estimate	1.33	.40	.40	.33	.28	.18
s.e.	(.38)	(.47)	(.47)	(.42)	(.44)	(.35)
% weighted				13.0%	22.1%	20.8%
pooled:						
estimate	.53	.44	.18	.23	.21	.24
s.e.	(.32)	(.32)	(.16)	(.18)	(.19)	(.18)
% weighted				12.4%	22.2%	18.7%

Note: Huber-White standard errors[11] are reported in parentheses.

Table 1 reports the estimates of γ from the following estimators: conventional (non-robust) IV, the two informal robust procedures,[12] IV-Huber with c=2.0 (the optimal value of c for ε=.01), IV-Huber with c=1.4 (optimal for ε=.05), and IV-Krasker-Welsch. The estimates of γ based on conventional IV are anomalous in several respects. According to the analytical model, the presence of non-zero values of the shadow price of the borrowing constraint will lead to an inconsistent (downward biased) estimate of γ for the poor subsample, while the rich subsample should yield consistent estimates of γ. According to the conventional estimates, the estimate of γ is .61 (and insignificantly different from its value of unity under the null hypothesis) for the poor, while the estimate of γ is .27 (and significantly different from unity) for the rich. The estimate of γ for the truly wealthy is troubling for two reasons. First, the point estimate actually exceeds unity, although not by a statistically significant margin. Second, the analytical model implies that both the rich and the truly wealthy subsamples should yield consistent estimates of γ. Instead, the estimate of γ from the truly wealthy subsample, 1.33, differs dramatically from the estimate of .27 for the rich subsample.

Next consider the empirical results from the three formal robust estimators: the anomalous results that plagued the conventional estimates are gone. While the IV-Huber point estimate of γ 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 γ is consistent for the rich subsample and inconsistent for the poor subsample if borrowing constraints are binding, one could conduct a specification test for the presence of binding borrowing constraints by testing the equality of the estimates of γ across the two subsamples. Given the very similar estimates of γ 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 rich subsample, the IV-Huber estimates indicate that one can reject a value of unity for the marginal propensity to save, while a value of zero cannot be rejected; for this subsample, the conventional results are not grossly different. For the poor subsample, however, the IV-Huber estimates provide a decisive rejection of the null hypothesis that $\gamma=1$, while the conventional estimates cannot reject a value of unity.

A comparison of the observation weights assigned by IV-Huber (for c=1.4) with the weights assigned by IV-Krasker-Welsch is provided by Figures 1a and 1b. Note that both estimators downweight observation 676 severely: IV-Huber assigns a weight of .06 while Krasker-Welsch assigns a weight of .02. Note that the poor subsample (which does not contain household 676) has a number of severe outliers and that the rich subsample has several extremely influential observations in addition to household 676.

Each of the three formal robust estimators delivers the same answer to the basic economic question posed by the paper: When one divides the sample into constrained and unconstrained subsamples, both subsamples yield an estimated marginal propensity to save of about 20%. For both subsamples, the estimated marginal propensity to save is statistically significantly different from its value of unity under the null hypothesis. While the estimates are not as precise as one would like, they are sufficiently precise to conclude that the optimal consumption model is rejected, and, further, the rejection cannot be attributed simply to borrowing constraints.

To what extent do the informal robust estimators succeed in removing the anomalous effects of outliers exhibited by the conventional results? The subjective exclusion of the single most spectacular outlier results in a decline in the estimated marginal propensity to save for the truly wealthy from 1.33 to .4, but actually worsens the anomalous relationship between the estimates for the poor and rich subsamples. The ad hoc robust procedure of trimming observations results in estimates which are reasonably consistent with the formal robust estimates.

Figure 1a
Comparison of Weights: IV-Huber vs. IV-Krasker Welsch
subsample: rich

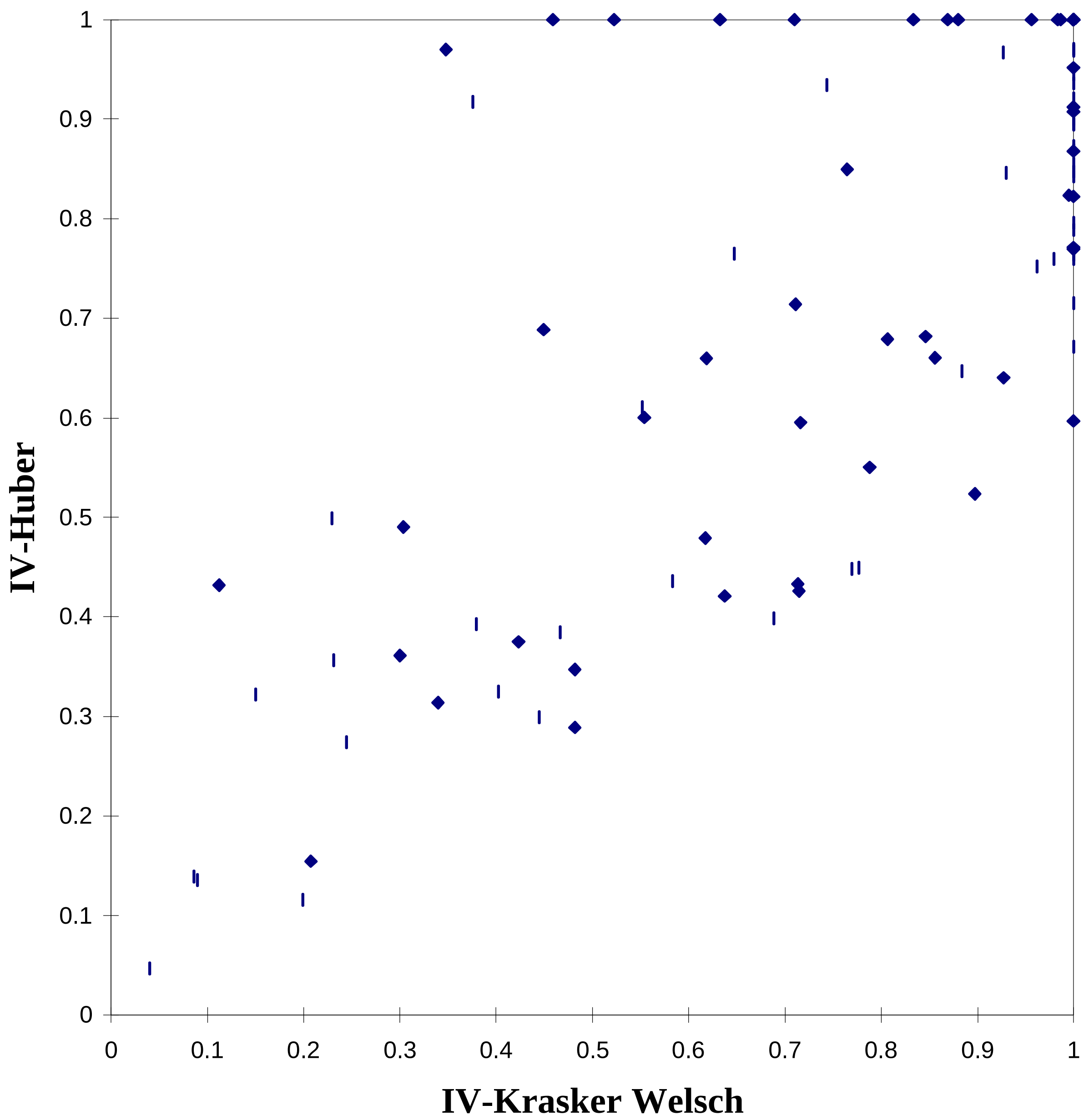
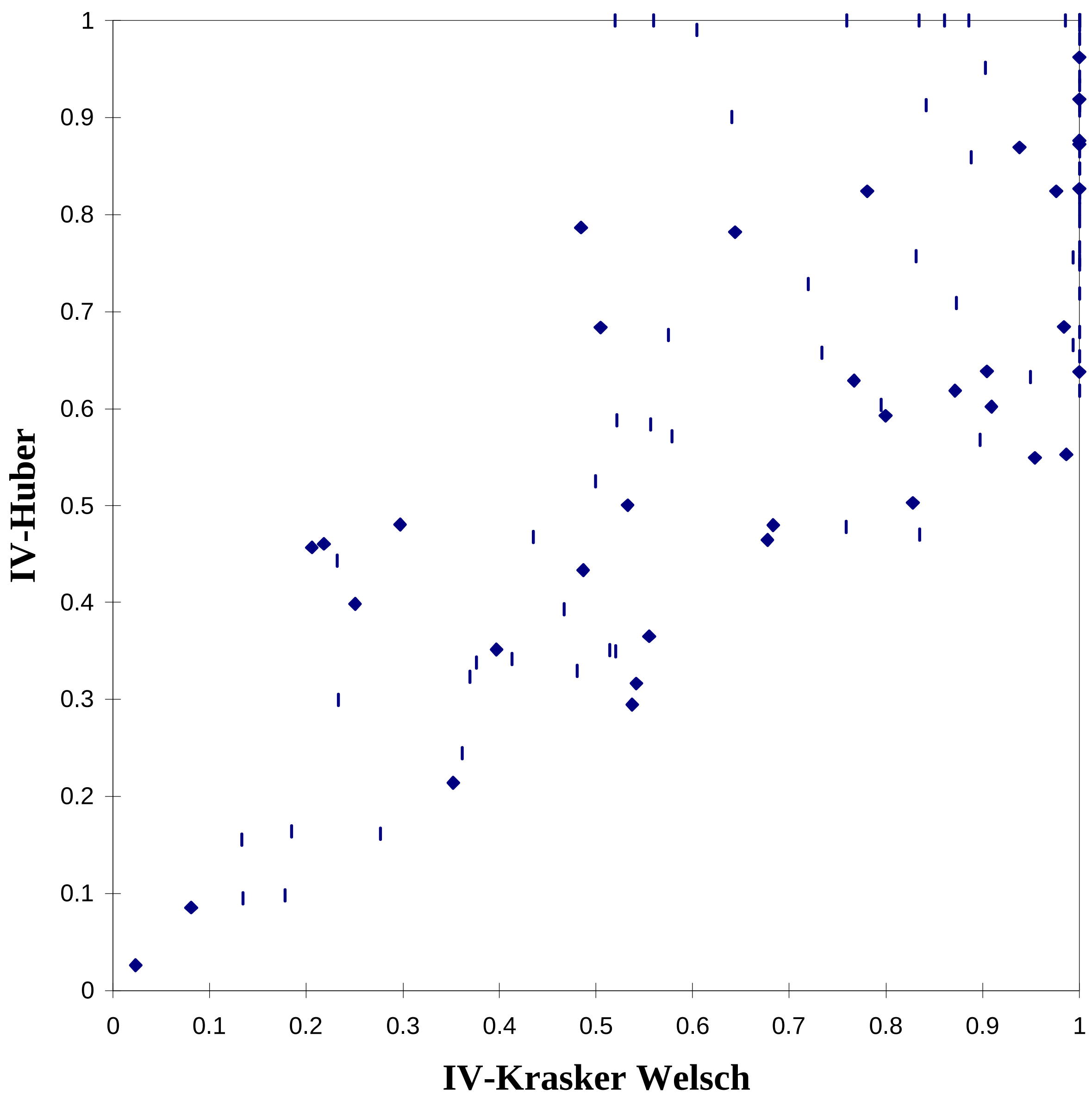


Figure 1b
Comparison of weights: IV-Huber vs. IV-Krasker-Welsch
subsample: poor



Monte Carlo experiments were conducted in order to assess the relative efficiency of the various estimators, estimate the magnitude of small sample bias, and verify the asymptotic standard errors. The Monte Carlo study was based on the model:

$$\text{(28)} \qquad \begin{aligned} X &= Z\Pi + \varepsilon_x \\ y &= X\gamma + \varepsilon_y \end{aligned}$$

where γ is the vector of structural parameters of interest, and Z is the matrix containing the instruments (assumed uncorrelated with both ε_y and ε_x) for X. Data on X and y were generated using the reduced form of the model:

$$\text{(29)} \qquad \begin{aligned} X &= Z\hat{\Pi} + \nu_x \\ y &= Z\hat{\Pi}\gamma + \nu_y \end{aligned}$$

Estimates of the parameter matrix, $\hat{\Pi}$, and the covariance matrix of the reduced form disturbances were calculated using the 350 observations in the rich sample, since data on the unconstrained households alone provide consistent estimates of the parameters. The Monte Carlo simulations were generated using the actual data matrix for the instruments (Z), the estimated $\hat{\Pi}$ matrix and an assumed value for the vector of structural parameters, γ (intercept of .026 and slope coefficient of .18). To capture the correlation structure of the disturbances, ν_x and ν_y were generated as linear combinations of two independent unit-variance disturbances, e_1 and e_2:

$$\text{(30)} \qquad \begin{aligned} \nu_x &= w_a e_1 \\ \nu_y &= w_b e_1 + w_c e_2 \end{aligned} \qquad \text{for } w_a=.46405811,\ w_b=.050450198,\ w_c=.53302651.$$

To reflect the presence of leverage points in the X data, the e_1 disturbance was generated as a mixture of normals; with probability .9 a variate was drawn from the standard normal distribution and with probability .1 a variate was drawn from a normal distribution with standard deviation equal to 10, then the resulting vector of variates was rescaled to have unit variance. The Monte Carlo estimates are reported under two alternative cases for the distribution of the error e_2. In one set of simulations, the e_2 variate is mixed normal, with probability .8 of being drawn from the standard normal distribution and with probability .2 from a normal distribution with standard deviation 10 (then rescaled to have unit variance). Since the “cost” incurred in the use of robust estimators comes in the form of reduced efficiency when the errors are normally distributed, the estimators are compared in another set of simulations (labeled "normal") in which e_2 is simply a standard normal variate.

Monte Carlo results are presented in Table 2 for the rich and truly wealthy subsamples. In each case, the actual data on the instrument matrix Z for the particular subsample was used (thus the simulations reflect a sample size of 350 for the rich and 77 for the truly wealthy). Simulated data on X and y was then generated using a common set of assumptions, described above, on the values of $\hat{\Pi}$, γ, and the joint distribution of ν_x and ν_y.

The magnitude of the small sample bias to the estimate of the marginal propensity to save was calculated by comparing the average estimate of γ, over 1000 simulations, to the assumed value of .18. The estimated bias was negative and fairly small (about -.017) for the sample of 350 observations. For the truly wealthy, the bias is more noticeable (around -.05), but small relative to sampling error.

Table 2: Monte Carlo results (1000 draws)

	conventional	**trimmed**	**IV-Huber c=2.0**	**IV-Huber c=1.4**	**IV-KW a=1.8√k**
			mixed normal distribution		
rich:					
estimated bias	-.017	-.036	-.018	-.017	-.015
RMSE	.299	.241	.198	.183	.137
truly wealthy:					
estimated bias	-.059	-.062	-.051	-.049	-.030
RMSE	.435	.363	.359	.345	.289
			normal distribution		
rich:					
estimated bias	-.023	-.023	-.021	-.021	-.019
RMSE	.297	.297	.307	.314	.383
truly wealthy:					
estimated bias	-.055	-.053	-.056	-.057	-.046
RMSE	.449	.446	.498	.530	.647

Improved efficiency under heavy-tailed error distributions is the theoretical rationale for the formal robust estimators. The Monte Carlo experiments show that the improvement in efficiency is large enough to be of practical importance to applied econometricians: For the rich subsample, the root mean square error of the IV-Huber estimators is about two-thirds that of the conventional estimator. IV-Krasker-Welsch does even better, with RMSE less than half that of conventional IV. For the truly wealthy subsample, the IV-Huber estimators have RMSE approximately 80%, and IV-Krasker-Welsch approximately 66%, of conventional IV. Since most applied econometricians use some ad hoc form of rejection of outliers, the comparison of the formal robust methods to conventional IV is one of primarily theoretical interest. The issue of practical interest concerns the comparison of ad hoc robust procedures such as trimming observations to the formal robust methods. Based on the rich sample, the trimmed IV estimator achieves about half the reduction in RMSE of the IV-Huber estimator, and a third the reduction of IV-Krasker-Welsch.

As expected, the robust estimators have larger RMSE than conventional IV when e_2 is normally distributed. Note, however, that based on the simulations for the rich sample, the RMSEs for IV-Huber are only 3% to 6% higher than the RMSE for conventional IV. That is, the efficiency loss under normality is an order of magnitude smaller than the efficiency gain with the mixed normal distribution. For IV-Krasker-Welsch, the efficiency loss under normal errors is more substantial: the RMSE is 30% larger than the RMSE for conventional IV.

Based on the Monte Carlo simulations with the mixed normal distribution, the RMSE for all of the robust estimators are somewhat smaller than the corresponding estimated standard errors. For example, for the rich sample, the standard error for the IV-Huber estimator (c=2.0) was .33, compared to a RMSE of .198 in the simulation. If the assumptions of the Monte Carlo simulations accurately reflect the actual data generation process, this would suggest that inferences based on the reported standard errors are conservative.

Tables 3a, 3b, and 3c present the conventional and robust estimates of disaggregated savings equations. In this exercise, the dependent variable is defined as savings flows in the form of financial assets, debt, or durable goods purchases separately. “Financial assets” are defined as savings accounts, checking accounts, bonds, and stocks. According to the IV-Huber estimates reported in Table 3a, which pertains to the rich sample, only 10% of an expected income fluctuation is buffered via changes in financial asset stocks. With a standard error of .15, one cannot rule out the hypothesis that the MPS in the form of financial assets is actually zero. However, note that the MPS in the form of financial assets is estimated with sufficient precision to rule out a wide range of plausible economic behavior. While other assets, such as debt and durable goods can be used, to some extent, to buffer

Table 3: Comparison of estimates of marginal propensity to save for disaggregated asset stocks

3a: rich subsample

dependent variable	conventional IV	IV-Huber, c=2.0	IV-Huber, c=1.4	IV-KW, a=1.8√k
$\Delta S^{Fin}_{i,t+1} / Y_{i,t}$				
estimate	.16	.10	.09	.01
s.e.	(.22)	(.15)	(.15)	(.12)
% weighted		18.6%	27.4%	23.7%
$\Delta S^{Debt}_{i,t+1} / Y_{i,t}$				
estimate	-.11	-.06	-.04	-.03
s.e.	(.08)	(.09)	(.07)	(.07)
% weighted		24.0%	36.6%	33.7%
$\Delta S^{Dur}_{i,t+1} / Y_{i,t}$				
estimate	.01	-.02	.03	.06
s.e.	(.15)	(.15)	(.13)	(.15)
% weighted		15.4%	29.7%	23.1%

3b: poor subsample

dependent variable	conventional IV	IV-Huber, c=2.0	IV-Huber, c=1.4	IV-KW, a=1.8√k
$\Delta S^{Fin}_{i,t+1} / Y_{i,t}$				
estimate	.35	-.003	-.001	-.005
s.e.	(.38)	(.006)	(.003)	(.007)
% weighted		39.6%	47.1%	39.2%
$\Delta S^{Debt}_{i,t+1} / Y_{i,t}$				
estimate	-.18	-.02	-.03	-.04
s.e.	(.16)	(.09)	(.09)	(.08)
% weighted		12.7%	23.6%	17.0%
$\Delta S^{Dur}_{i,t+1} / Y_{i,t}$				
estimate	.08	.12	.11	.17
s.e.	(.13)	(.20)	(.18)	(.22)
% weighted		21.0%	29.2%	28.5%

3c: truly wealthy subsample

dependent variable	conventional IV	IV-Huber, c=2.0	IV-Huber, c=1.4	IV-KW, a=1.8√k
$\Delta S^{Fin}_{i,t+1} / Y_{i,t}$				
estimate	.80	.29	.29	.23
s.e.	(.24)	(.29)	(.24)	(.20)
% weighted		16.9%	27.3%	23.4%
$\Delta S^{Debt}_{i,t+1} / Y_{i,t}$				
estimate	-.22	-.22	-.22	-.08
s.e.	(.06)	(.06)	(.05)	(.11)
% weighted		13.0%	22.1%	19.5%
$\Delta S^{Dur}_{i,t+1} / Y_{i,t}$				
estimate	.31	.06	.08	.05
s.e.	(.16)	(.34)	(.30)	(.27)
% weighted		9.1%	26.0%	14.3%

income fluctuations, one would expect financial assets to play the major role. Based on the robust estimates, one can statistically reject the hypothesis that the MPS in the form of financial assets is any greater than 40%. The estimates confirm, with greater statistical precision, a basic implication of the estimates based on the aggregated definition of saving: the null hypothesis of optimal consumption smoothing is rejected, even for the unconstrained subsample. Further, for the poor subsample, the only disaggregated component of savings with a quantitatively nontrivial role in consumption smoothing is expenditure on durable goods, with an estimated savings propensity of .12. With a standard error of .2, one cannot reject other economically plausible values (e.g. 0 or .2). However, note that for the poor sample, the estimated MPS in the form of financial assets is almost exactly zero, and extremely precisely estimated. With any of the robust estimates, one can reject the hypothesis that the MPS in the form of financial assets is as much as 1%.

While the robust estimates of the overall propensity to save out of expected income growth are very similar for the constrained and the unconstrained subsamples, the disaggregated results suggest that purchases of durable goods are the primary (perhaps sole) vehicle for consumption smoothing for the poor, while accumulation of financial assets and reductions in debt are the major means of consumption smoothing for the rich. The idea that purchases of durable goods are actually less strongly correlated with expected income growth for the rich than for the poor makes sense intuitively; for a household with a substantial level of financial assets, one would expect income fluctuations to be buffered initially by changes in these holdings. For an unconstrained household, purchases of durable goods represent a reallocation of the household's portfolio between financial assets and durable goods, but the timing of durable goods expenditures would presumably be determined by factors other than the time path of income. In contrast, for a household with little or no financial assets, the borrowing constraint tends to force the household to defer acquisition of durable goods until the additional income is available, causing the timing of durable goods acquisitions to coincide more closely with the timing of income.

The truly wealthy sample is a small (77 observations) subsample of the rich sample. None of the estimates based on this small subsample are required for the principal economic conclusions of the paper: the optimal smoothing model can be statistically rejected, and this rejection cannot be attributed to the presence of borrowing constraints. Further, the Monte Carlo experiments revealed substantial small sample bias (downward bias of .05

for an assumed parameter of .18) for the truly wealthy sample size. With these caveats in mind, note that the estimated MPS in the form of financial assets is .29, with a standard error of .24. Unlike the larger rich sample, the data cannot reject the notion that financial assets are used by this elite group of households to buffer a substantial part of expected income fluctuations. The estimated MPS in the form of reductions in debt is large (22%) and precisely estimated (s.e. of .05). The estimated MPS in the form of durable goods acquisitions is closer to zero, both in terms of magnitude and in statistical significance, for the truly wealthy than for the poor.

The results based on the disaggregated asset stocks put an important twist on the earlier result that constrained and unconstrained households have essentially the same saving propensities. The results for total saving used a generous definition of saving which counted expenditure on durable goods as saving rather than consumption. When saving is disaggregated by type of asset, however, the results suggest that households in the rich, or unconstrained, sample use financial assets as their primary means of consumption smoothing, while households in the constrained sample save almost exclusively in the form of purchases of durable goods. Note that empirical studies of borrowing constraints which rely on consumption data usually use nondurable consumption as the dependent variable. Conceptually, using nondurable consumption as the measure of consumption is equivalent to using total saving, inclusive of durable goods expenditures, as the measure of saving. Thus empirical studies relying on nondurable consumption would miss the evidence of borrowing constraints which appears in the disaggregated results.

Conclusions

The paper proposes an IV version of the Huber estimator as an alternative to the IV-Krasker Welsch estimator. The IV-Huber estimator performs comparably to IV-Krasker Welsch in an empirical application to household data on saving flows, and in a related set of Monte Carlo experiments, but is considerably simpler, both analytically and computationally, than IV-Krasker Welsch. Assessing the various estimators according to their efficiency under heavy tailed error distributions, informal robust procedures such as the subjective rejection of outliers and trimming fall about halfway between the nonrobust estimator and the formal robust estimators;[13] compared to the informal robust procedures both IV-Krasker Welsch and IV-Huber provide efficiency gains which are large enough to be of practical importance to applied econometricians.

[1]Because the CRRA utility function has the property that the marginal utility of consumption is infinite at zero consumption, it is not necessary to impose the non-negativity constraint on consumption explicitly.
[2]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 $(m_{r_j}, v_{r_j}, m_u, v_u, cov_{r_j u})$ as having both household (i) and time (t) subscripts, although these subscripts will be suppressed for notational convenience.
[3]To evaluate the likely magnitude of the remainder term, consider a household whose consumption both in t and in t+1 is within 90% of the household's disposable income in period t. For such a household, the squared terms on the RHS of (10) 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.
[4]The intercept term in equation (16) in principle varies both across households and across time. With only one observation per household it is obviously not possible to identify both the parameter of primary interest, γ, 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 γ. 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.
[5]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 (16) gave small and imprecise estimates of the intertemporal substitution parameter. More importantly, the inclusion or exclusion of the expected rate of return made no material difference to the estimates of γ, the parameter of primary interest.
[6]For a different but related problem, Deaton (l990) discusses an approach which uses cross-sectional sample moments for household data to make inferences about the parameters of the time series process.
[7]Krasker and Welsch define the sensitivity of the estimator, δ, as:

$$\delta = \sup_{y,x} \sup_{\lambda} \frac{|\lambda'\Omega(y,x)|}{(\lambda' V\lambda)^{1/2}}$$

. The influence function, Ω(y,x), gives the effect of the ith observation on the vector of estimates, $\hat{\beta}$. Formally, the influence function is a function Ω: $RxR^k \rightarrow R^k$ which satisfies:

$$\sqrt{n}\left[\hat{\beta} - \beta - \frac{1}{n}\sum_{i=1}^{n}\Omega(y,x)\right] \rightarrow 0 \quad \text{in probability as } n \rightarrow \infty \quad .$$

λ is a kx1 vector

$V = E[\Omega(y,x)\Omega(y,x)']$, the asymptotic covariance matrix of the estimates.

[8]In the ordinary regression case, Krasker and Welsch were not able to show that the estimator was strongly efficient. Instead, they showed that if a strongly efficient estimator existed, it would be of the form of the Krasker-Welsch estimator. Subsequently, Ruppert (l985) provided an example in which no strongly efficient estimator exists. In Ruppert's example, while other estimators can provide more efficient estimates of particular parameters, the Krasker-Welsch estimator does not perform substantially worse. Thus the practical significance of Ruppert's example is unclear.
[9]For example, the papers using the PSID food consumption data exclude observations for which the (absolute value of the) change in log consumption exceeds a specified value. 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. Zeldes excluded observations for which food consumption increased by more than 200% or decreased by more than two-thirds. Altonji and Siow (l987) exclude observations for which food consumption increased by more than 400%, or 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%.

[10]Based on a more detailed examination of the data record, household 676 appears to be a legitimate observation and not a keypunch error. The head of household was a student in an advanced degree program in l967, received a professional degree, possibly an M.D., worked only part of the year in l968, and was employed at the professional level in the health field for all of l969. With a disposable income of $22,573 (in l969 dollars), the respondent stated that a net purchase of $20,000 of stocks was made in l969.

[11]The Huber-White standard errors are calculated as:

$$\hat{\text{var}}(b) = (\hat{X}' DX)^{-1} \hat{X}' V\hat{X}(X' D\hat{X})^{-1}$$

where D= (nxn) diagonal matrix with diagonal elements d_i such that d_i=1 if w_i=1 and d_i=0 if w_i<1 and V=(nxn) diagonal matrix with diagonal elements $v_i=w_i^2 r_i^2$. In the special case in which all the weights are fixed at unity (the conventional IV estimator), $\hat{\text{var}}(b)$ would be the usual White heteroskedasticity-consistent covariance matrix. The effect of the endogenous downweighting of observations is captured by the D matrix.

[12]For the "trimmed" estimator, observations for which the dependent variable was less than -.67 or greater than 2 were excluded from the sample.

[13]On this point the paper reinforces the finding of Relles and Rogers (l977), who conduct a Monte Carlo experiment pitting actual statisticians using subjective rejection of outliers against formal robust methods. In Relles and Rogers, the statistical problem is simply the estimation of a location parameter. In the context of estimation of the mean, Relles and Rogers find that the formal robust methods substantially outperform the informal methods, even when implemented by trained statisticians.

Appendix: Algorithm for IV-Huber estimator

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

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

2. Calculate estimate of scale parameter, σ:
$\hat{\sigma} = 1.483\ \mathrm{median}(|r_i|)$
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 the residuals, $\hat{\sigma}$, and c.

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

4. Noting that $0 = \sum_{i=1}^{n} w_i\left(y_i - x_i b\right)\hat{x}_i = \sum_{i=1}^{n}\sqrt{w_i}\left(y_i - x_i b\right)\sqrt{w_i}\,\hat{x}_i$, calculate the square roots of the weights, $\sqrt{w_i}$, and transform the data by multiplying each element of the ith row of y, X, and Z by $\sqrt{w_i}$:

$$\tilde{y}_i = \sqrt{w_i}\,y_i \qquad \tilde{x}_i = \sqrt{w_i}\,x_i \qquad \tilde{z}_i = \sqrt{w_i}\,z_i$$

5. Form new IV estimates with the transformed data:

$$\hat{\tilde{X}} = \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\tilde{X} \qquad b_{new} = \left(\hat{\tilde{X}}'\tilde{X}\right)^{-1}\hat{\tilde{X}}'\tilde{y}$$

6. Form a new vector of residuals:

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

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.

References

Altonji, Joseph G., and Aloysius Siow, "Testing the Response of Consumption Income Changes with (Noisy) Panel Data", Quarterly Journal of Economics, 102, May l987.

Bassett, G.W., and Koenker, R.W., "The Asymptotic Distribution of the Least Absolute Error Estimator", JASA, 73, l978, 618-622.

Bernanke, Ben, "Permanent Income, Liquidity, and Expenditure on Automobiles: Evidence from Panel Data," Quarterly Journal of Economics, l984.

Breeden, Douglas T., Ph.D. dissertation, Graduate School of Business, Stanford University, l977.

Breeden, Douglas T., "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," Journal of Financial Economics, vol. 7, l979, 265-296.

Carroll, Christopher D., "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence," Brookings Papers of Economic Activity, l992:2(b), 61-156.

Deaton, Angus, "Saving and Liquidity Constraints," Econometrica, 59, l991, 1221-48.

Deaton, Angus, Understanding Consumption, Oxford University Press, l992.

Deaton, Angus, "Saving and Income Smoothing in the Cote d'Ivoire", Journal of African Economies, 1, 1992, 1-24.

Flavin, Marjorie, "The Adjustment of Consumption to Changing Expectations about Future Income", Journal of Political Economy, 89, October l981 974-1009.

Flavin, Marjorie, "The Excess Smoothness of Consumption: Identification and Interpretation", Review of Economic Studies, 60, July l993, 651-666.

Hall, Robert E., "Stochastic Implications of the Life Cycle -- Permanent Income Hypothesis: Theory and Evidence," Journal of Political Economy, 86, December, l978, 971-987.

Hall, Robert E., "Intertemporal Substitution in Consumption," Journal of Political Economy, l988.

Hall, Robert E. and Frederic S. Mishkin, "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households." Econometrica, 50, March l982, 461-481.

Hampel, Frank R., Elvezio M. Ronchetti, Peter Rousseeuw, and Werner A. Stahe, Robust Statistics: The Approach Based on Influence Functions, Wiley l986.

Hansen, Lars Peter and Kenneth J. Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," Econometrica, 50, September l982, 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 l983, 249-266.

Hayashi, Fumio, "The Effect of Liquidity Constraints on Consumption: A Cross-Sectional Analysis," Quarterly Journal of Economics, l985(a).

Hoaglin, D.C., and Welsch, R.E., "The Hat Matrix in Regression and ANOVA", The American Statistician, 32, l978, 17-22.

Huber, P.J., "Robust Estimation of a Location Parameter", Annals of Mathematical Statistics, 35, l964, 73-101.

Huber, P.J., "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions", Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, l967, 221-233.

Huber, P.J., "Robust Regression: Asymptotics, Conjectures, and Monte Carlo" Annals of Statistics, 1, l973.

Huber, P.J., "Robust Methods of Estimation of Regression Coefficients", Math Operationsforschung Statist. Ser. Statist., 8, l977.

Huber, P.J., Robust Statistics, New York, John Wiley, l981.

Huber, P.J., "Minimax Aspects of Bounded-Influence Regression", JASA, March l983, Vol 78, No. 381.

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, l983.

Krasker, William S., "Estimation in Linear Regression Models With Disparate Data Points", Econometrica, l980, 48, 1333-1346.

Krasker, William S., and Roy E. Welsch, "Efficient Bounded-Influence Regression Estimation, JASA, September l982, Vol. 77, No. 379.

Krasker, William S., and Roy E. Welsch, "Resistant Estimation for Simultaneous-Equations Models Using Weighted Instrumental Variables" Econometrica, Vol. 53, No.6, l985.

Powell, J.L., "The Asymptotic Normality of Two-Stage Least Absolute Deviation Estimators", Econometrica, 51, l983.

Relles, Daniel A., and William H. Rogers, “Statisticians are Fairly Robust Estimators of Location”, JASA, March 1977, Vol. 72, 107-111.

Runkle, David E., "Liquidity Constraints and the Permanent Income Hypothesis Evidence from Panel Data," Journal of Monetary Economics, 27, February l991.

Ruppert, David, "On the Bounded-Influence Regression Estimator of Krasker and Welsch," Journal of the American Statistical Association, 80, March 1985, 205-208.

Shapiro, Matthew D., "A Note on Tests of the Permanent Income Hypothesis in Panel Data", unpublished, May l982.

Shapiro, Matthew D., "The Permanent Income Hypothesis and the Real Interest Rate: Some Evidence From Panel Data," Economic Letters, 14, l984, 93-100.

Zeldes, Stephen, "Consumption and Liquidity Constraints: An Empirical Investigation," Journal of Political Economy, 97, April l989(a).

Zeldes, Stephen, "Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence," Quarterly Journal of Economics, 104, May 1989(b).