Faculty Working Papers

Effect of Response Errors on Parameter

Estimates of Models of Savings Behavior

Robert Ferber and Lucy Chao Lee

University of Illinois

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana- Champaign

June 18, 1971

Effect of Response Errors on Parameter Estimates of Models of Savings Behavior

Robert Ferber and Lucy Chao Lee

University of Illinois

No. 17

Effect of Response Errors on Parameter Estimates of Models of Savings Behavior

Robert Ferber and Lucy Chao Lee

A considerable body of evidence has accumulated indicating that substantial errors exist in the reporting of asset and debt holdings in consumer financial surveys. The characteristics of these errors vary from one asset or debt to another, being more pronounced for more sensitive holdings, such as savings accounts, common stock and personal debt, and being of lesser importance for holdings such as life insurance, real estate and installment credit.[1] Overall, however, the evidence indicates that nonreporting of ownership may be substantial, that very small holdings may be overstated and very large holdings understated, and that those who refuse information on their holdings are likely to hold more of that asset or debt than would be expected on the basis of the usual averaging process.

Although current research may lead to means of detecting and correcting such errors, the fact remains that all of our past and current consumer financial survey data are subject to these errors. In view of the

---

[1]For example, see Lansing, J. B., Ginsburg, G. P., and Braaten, K., An Investigation of Response Error, University of Illinois, Bureau of Economic and Business Research, Studies in Consumer Savings, No. 2, 1961; Ferber, R., The Reliability of Consumer Reports of Financial Assets and Debts, University of Illinois, Bureau of Economic and Business Research, Studies in Consumer Savings, No. 6, 1966; Ferber, R., Forsythe, J., Guthrie, H. W., Maynes, E. S., "Validation of Consumer Financial Characteristics: Common Stocks," Journal of the American Statistical Association, June 1969, pp. 415-22; _____, "Validation of a National Survey of Consumer Financial Characteristics: Savings Accounts, "Review of Economics and Statistics, November 1969, pp. 436-44.

widespread interest in ascertaining and measuring the determinants of consumer savings behavior, it would, therefore, seem of critical importance to evaluate the effects of these errors on analytical studies of this type.

These effects are explored in this paper. More specifically, its objective is to assess the effects of errors in asset and debt variables on the estimates of parameters of models of consumer portfolios. This is done in a two-stage process. First, estimates are made of the magnitudes of the bias in estimates of the parameters of consumer portfolio models, using data from a validation study permitting relatively accurate determinations of the magnitude of response and nonresponse errors in the variables.

This first stage involves initially the formulation of alternative hypotheses on the determinants of consumer portfolios. These hypotheses are transformed into structural relations as a basis for the estimation of parameters. The parameters of the models are then estimated in two ways, one way by using the data on consumer portfolios as reported in the surveys and the other way after adjusting these variables for response and nonresponse errors in the data, based on the validation information. The latter adjustment is a rather tricky one, because the validation data provide only partial information on the errors in the variables, so that additional inferences of the nature of the error in the nonvalidated component of the variables have to be made. To obtain some idea of the sensitivity of the estimates of the parameters to these inferences, these estimates are made under alternative assumptions of these errors.

This first, econometric, approach yields rather narrow results, providing estimates of the effect of these errors on a particular type of sample. To obtain a more general idea of the nature of these effects, a simulation approach is used next. This approach involves five distinct steps. First, as before, certain structural hypotheses and corresponding portfolio functions are formulated. Second, based on the results obtained with the prior econometric approach, assumptions are made of the true values of the parameters of the models. Third, error properties are attributed to the portfolio variables based on the validation information from the sample used in the prior econometric approach as well as from previous validation studies of consumer financial behavior. Fourth, the same validation sample used in the first stage is run through the error process, and sets of observations are generated making use of the error properties postulated at the prior stage. These sets of observations are generated by Monte Carlo methods 150 different times, to yield some idea of the range of variables obtained with these error properties.

Finally, estimates are obtained of the parameters in each of these simulations and compared with the true values postulated in the second stage. The distributions of the parameter estimates around the true values provide a fairly comprehensive picture of the effect of these response and nonresponse errors, at least for the types of models postulated in this study.

The theoretical aspects of the effect of errors in variables has been well covered in the literature, in addition to a few empirical studies

of this question. A brief summary of this literature is provided in the following section, which serves as a framework for the present paper. Section 3 presents alternative formulations of models of savings behavior and describes the data used in this study. The results of the econometric approach are presented in Section 4 and that of the simulation approach in Section 5. The concluding section then summarizes the results and discusses their implications both for the study of savings behavior and with regard to the methodological issue of the effect of errors and variables on parameter estimates.

2. Review of Relevant Literature

The great majority of studies of consumer savings have focused on the flow aspects rather than on the stock of savings, which is not surprising in view of the much greater availability of data of the former type. Nevertheless, an increasing amount of data has begun to be available in recent years on household financial assets, and these data have served as a basis for a number of studies on the determinants of these asset holdings. Although the cross-section studies of this question are of primary interest, it seems desirable to refer briefly to some of the more recent time series studies because of their relevance to one of the principal aspects of a model of consumer portfolios. This is the question whether income or assets, or both, are most relevant to the determination of holdings of a financial asset.

The bulk of the evidence appears to point rather strongly toward some measure of wealth (usually net worth) rather than income as a more likely

primary determinant. Thus, in a study of factors influencing liquid asset holdings in England, Lydall found that new worth was far more important than the level of income.[2] Similar results were obtained by Meltzer as well as by Bronfenbrenner and Mayer.[3] In a still more recent study, Hamburger found that wealth was consistently more important than income in a number of single equation models of the influence of various factors on the demand for four financial assets -- marketable bonds, time and savings deposits at commercial banks, life insurance reserves, and savings accounts at credit unions, savings and loan associations and mutual savings banks.[4]

A study by Feige might also be cited in which he found that demand deposits, as well as time deposits of commercial banks, and savings and loan associations, were strongly influenced by an estimate of "permanent personal income."[5] However, the income variable reflected a weighted

---

[2]Harold Lydall, "The Life Cycle in Income, Saving and Asset Ownership," Econometrica, Vol. 23, April 1955, Pages 131-50.

[3]A. H. Meltzer, "The Demand for Money: The Evidence from the Time Series," Journal of Political Economy, Vol. 61, June 1963, Pages 219-246; M. Bronfenbrenner and T. Mayer, "Liquidity Functions in the American Economy," Econometrica, Vol. 28, October, 1960, Pages 810-834.

[4]M. J. Hamburger, "Household Demand for Financial Assets," Econometrica, Vol. 36, January, 1968, Pages 97-118.

[5]E. L. Feige, The Demand for Liquid Assets: A Temporal-Cross Section Analysis, Englewood Cliffs, New Jersey: Prentice Hall, 1964.

average of present and past values of personal income and is, therefore, closer in concept to a net worth variable than to a current income variable. Partly for the latter reason, no measure of wealth was used in this study.

The cross-section studies have been relatively few and of a somewhat varied nature because they necessarily had to be molded to fit the particular set of data. Thus, Watts and Tobin ran a series of multiple regressions on the 1950 BLS consumer expenditures data using as dependent various stock variables (mortgage debt, installment debt, cash balances and insurance) and also corresponding flow variables. They found a variety of socio-economic and demographic characteristics to influence these dependent variables, among which was income.[6] However, net worth was not available and therefore was not included in the study. It might be noted that housing level, a variable that was included and that might be considered as a proxy for permanent income and for net wealth, was highly significant in almost all cases.

In another cross-section study, using data from the Consumer Savings Project, Claycamp found that wealth, in the form of total assets, dominated income as a determinant of the proportion of assets held in liquid form as well as the proportion held in variable-dollar form (meaning assets whose value fluctuates with changes in prices).[7]

[6]H. W. Watts and James Tobin, "Consumer Expenditures and the Capital Account," in Irwin Friend and Robert Jones, Editors, Proceedings of the Conference on Consumption and Saving, Vol. 2, Philadelphia: University of Pennsylvania Press, 1960, Pages 1-48.

[7]H. J. Claycamp, Jr. The Composition of Consumer Savings Portfolios. Urbana, Illinois: University of Illinois Bureau of Economic and Business Research, Studies in Consumer Savings, No. 3, 1963.

In another study, Crockett and Friend found that regressions of different asset items on net worth gave about the same result as regressions of these holdings on disposable income, using data from the 1962 Federal Reserve Survey of Financial Characteristics of Consumers.[8] Later in the same study, results from a University of Michigan Survey Research Center panel of consumers for 1959-61 indicated that both income and net worth were significant in determining the flow into, and stocks of, particular assets, varying with the asset. As in the Watts and Tobin study, however, the income term was invariably a measure of long run, or "normal," income, which again might be construed as a proxy for net worth. Also, the bulk of this analysis focused on asset flows rather than on stocks.

Dorothy Projector found that equity in an asset at the beginning of the period, and also occasionally a net worth variable, were more important than disposable income in determining saving in the form of a publicly traded stock, checking accounts, savings accounts, and investment assets.[9] Invariably, equity in the particular asset at the beginning of the period was the dominant variable.

Also pertinent are various studies made by Kreinin with Survey Research Center Data on the factors influencing ownership of liquid assets,

---

[8]Jean Crockett and Irwin Friend, "Consumer Investment Behavior," in Robert Ferber, Editor, Determinants of Investment Behavior. New York: National Bureau of Research, 1967, pages 15-128.

[9]Dorothy S. Projector, Survey of Changes in Family Finances, Washington, D. C.: Board of Governors of the Federal Reserve System, 1968.

life insurance and common stock. A variety of socio-economic factors were found relevant, and stock ownership was found also to be influenced by liquid assets.[10]

3. Alternative Models

Based on the previous review of the literature, three alternative models were formulated for use in this experiment. In all three cases, assets are subdivided into three categories in accordance with the validation information that is available. These categories are savings accounts and savings certificates (S), common stock (C), and all other assets and debt (L).

The three models represent alternative hypotheses on the determinants of these three forms of asset holdings. The models are as follows:

Model A

This model assumes that each of the three asset holdings is dependent on the other asset holdings in accordance with the following hypothesis. Savings accounts and common stock are jointly dependent on each other and on total assets (T), while other asset holdings are dependent on savings accounts and on common stock. In addition, all three categories of assets are influenced by a set of family characteristics (Z), which are treated as exogenous. The model is formulated

---

[10] M. E. Kreinin, J. B. Lansing and J. N. Morgan, "Analysis of Life Insurance Premiums," Vol. 39, Feb., 1957, Pages 46-54; "Factors Associated with Stock Ownership," Review of Economics and Statistics, February, 1961, Vol. 43, Pages 76-80.

in linear arithmetic terms as follows:

1. $S_i = a_o + a_1C_1 + a_2T_i + a_3Z_i + u_i$

2. $C_i = b_o + b_1S_1 + b_2T_i + b_3Z_i + v_i$

3. $L_i = c_o + c_1S_i + c_2C_i + c_3Z_i + w_i$

4. $S_i + C_i + L_i \equiv T_i$

Model B

The second model is partially recursive in that it assumes savings accounts are determined first as a function of total asset holdings and of family characteristics. Holdings of common stock and of other assets are then assumed to be determined by savings accounts holdings and by total assets as well as by family characteristics. The common stock function is hence the same as in Model A. This hypothesis is in line with the general advice given by personal finance and money management people, that families just starting out should try to build up assets in the form of savings accounts (as well as life insurance) for reserves before accumulating other assets.

The exact form of the model is as follows:

1. $S_i = a_o + a_2T_i + a_3Z_i + u_i$

2. $C_i = b_o + b_1S_i + b_2T_i + b_3Z_i + v_i$

3. $L_i = c_o + c_1S_i + c_2T_i + c_3Z_i + w_i$

4. $S_i + C_i + L_i \equiv T_i$

Model C

This model differs primarily from the preceding models in treating as dependent the proportion of assets in a particular form rather than the absolute amount. In other words, the endogenous variables are the proportion of total assets in savings accounts (S/T), the proportion of total assets in common stock (C/T) and the proportion of total assets in other forms (L/T).

The basic hypothesis is similar to the preceding model, namely, that a family initially determines what proportion of its total assets should be in savings accounts, based on its total assets and its socio-economic characteristics. It then determines what proportion of its total assets should be in the form of common stock as a function of the prior determined proportion of its assets in the form of savings accounts, its total assets and its socioeconomic characteristics. The proportion of assets in other forms is obtained as a residual.

The exact form of the model is:

1. $(S/T)_i = a_o + a_2T_i + a_3Z_i + u_i$

2. $(C/T)_i = b_o + b_1 (S/T)_i + b_2T_i + b_3Z_i + v_i$

3. $(L/T)_i = 1 - (S/T)_i - (C/T)_i$

The Data

The empirical part of this study is based on a combination of two surveys carried out to validate reports of savings accounts and of common stock holdings in the Federal Reserve 1963 Survey of Financial

Characteristics (SFC). The SFC itself was a nationwide probability survey designed to obtain information on the complete financial position of families in this country, with over-representation from the high-income groups. The two validation studies were carried out toward the end of the SFC, using identical questionnaires and interview procedures and with the same interviewing organization (the U. S. Bureau of the Census). Unlike the SFC, however, the validation surveys were restricted geographically and, by their nature, contained only owners of that particular asset.[11]

It is clear, therefore, that the data used in this study do not represent a cross section of the U.S. population. Rather, these data constitute a very special sample for part of which savings account holdings are known and for the other part, stockholdings are known. In each case, however, the nature of the validation process precludes complete knowledge of either savings accounts or common stock for a particular family, so that adjustments have to be made for that part of the asset holding which was not validated and possibly not reported correctly. These adjustments, to be described shortly, introduce an additional source of error in the data. However, judging from the validation results presented in previous studies, there is little doubt that the resulting

---

[11]For a more complete description of these studies and for summaries of the results, see Robert Ferber, John Forsythe, Harold Guthrie, E. Scott Maynes, "Validation of a National Survey of Consumer Financial Characteristics: Common Stock," Journal of the American Statistical Association, June, 1969, Pages 415-32; -------, "Validation of a National Survey of Consumer Financial Characteristics: Savings Accounts," The Review of Economics and Statistics, November, 1969, Pages 436-444.

data are far more accurate with regard to family holdings of savings accounts and of common stock than of any other set of data that might be used for this purpose.

The types of adjustments made in these data and the manner in which they were made may be summarized as follows:

1. To adjust for errors in reported savings account balances or of stockholdings in the appropriate validation sample, the following rule was applied:

   a. If $V_R > 0$, and $T_R > V_R$:

   $$T_A = \frac{V_I}{V_R} \times (T_R - V_R) + V_I$$

   b. If $V_R > 0$, and $T_R = V_R$:

   $$T_A = \left| \frac{V_I - V_R}{N_V} \times (N_T - N_V) \right| + V_I$$

   c. If $V_R = 0$, and $T_R > 0$:

   $$T_A = (T_R + V_I) + V_I$$

   where T_A = adjusted total stocks or savings accounts

   T_R = reported total stocks or savings accounts

   V_I = institution record (amount) of validated stocks or savings accounts

   V_R = reported part of stocks or savings accounts for validation (amount)

   N_T = reported total number of batches of all stocks or all savings accounts

   N_V = reported number of batches of stocks or savings accounts for validation

2. To adjust for nonreporting of the particular holding in each of the validation samples, regressions were run using the adjusted total holding as the dependent variable (after adjusting the data from the preceding step) with a number of socioeconomic characteristics as independent. Separate regressions were obtained for common stock and for savings accounts. One of the independent variables in each case was the validated part of that holding. In both cases the fit was quite good, R^2 being .43 for the common stock regression and .37 for the savings account regression, with a large number of independent variables significant (particularly occupation, race, family size, income, and the validated part of that asset).

   For each of the sample members, its characteristics were substituted into the regression and an estimate obtained of the adjusted total holdings of that asset. The equation estimate was accepted except if the estimate was less than the validated figure for that sample member, in which case the validated figure was used as the total.

3. A further adjustment was made to spot nonreporting owners of savings accounts in the common stock sample, and nonreporting owners of common stock in the savings account sample. This was done by estimating the relative frequency of nonreporting of each asset by each income class. The same frequency of nonreporting was attributed to the comparable income class in the other

sample, which produced an estimate of the number of nonreporters in each income class in that validation sample. There were 28 estimated nonreporters for common stock and 13 estimated nonreporters of savings accounts.

To identify the specific sample members considered to be nonreporting owners in each validation sample, discriminant functions were derived for reporting of savings accounts and of stock for each sample, in each case the dependent variable being a 1-0 (reporter-nonreporter) variable. The independent variables were a variety of family characteristics including income, occupation, marital status, education, size of city, race and family size. A number of these variables were statistically significant at the .05 level, as was the coefficient of determination, but the overall goodness of fit was not high, namely, an R^2 of .10 for the stockholding function and .05 for the savings account function. Nevertheless, these functions were used on the nonvalidation sample to pinpoint nonreporters, namely, as those people with the highest estimated values of the function in each case on the presumption that these people were owners and should have reported their ownership.[12]

---

[12]It might be argued that the nonreporters should have been selected from the opposite end of the distribution, on the basis that low values of the dependent variables denoted nonreporters (but note that these people are also more likely to be nonowners). In any event, this approach was tested empirically and led to imputed ownership amounts for the nonvalidation sample which, when aggregated, yielded average amounts not reported far below the amounts indicated by the validation sample.

Once these nonreporting owners were pinpointed, estimates had to be made of the amounts the presumed nonowners owned in those particular assets. This was done by obtaining two additional least squares regressions, one regressing for the stock validation sample of the total savings accounts balances reported by the reporters as a function of socioeconomic characteristics; and the other for the savings account sample, the amount of stock holdings reported by the reporters as a function also of socioeconomic characteristics. In these instances, the goodness of fit was much better, namely, R^2 of .19 for savings accounts and .31 for stock holdings. A number of independent variables were significant at the .05 level, particularly income class, age of head, race and family size. Estimated amounts were accordingly obtained by substituting the characteristics of the presumed nonreporting owners into the appropriate function one at a time. It should be noted, however, that no adjustment could be made for reporting errors in this sample.

4. No adjustments were made for reporting or nonreporting in other assets, as there was no basis for doing so.

Effect of the Adjustments

A general idea of the effect of the adjustments in the asset variables is provided in Table 1. Not surprisingly, the table indicates that the average holdings per sample household were increased

substantially, by roughly 50 percent for savings accounts, by 22 percent for common stock holdings and hardly at all for other assets. As a result the average total asset holdings per sample household increased by 12 percent. Correspondingly, the proportion of total assets held in the form of savings accounts rises from 8.4 percent to 11.1 percent, while the proportion of total assets held in the form of common and preferred stock rises from 31.5 percent to 34.2 percent.

These increases are to be expected in view of the fact that the validation findings had indicated substantial nonreporting as well as reporting errors in the direction of underestimation for savings accounts, somewhat lesser reporting errors for common stocks, while the other validation studies had indicated low reporting errors for other assets. Since the adjustment procedures were designed to correct the data for these errors, as described in the previous section, substantial increases in asset holdings were only to be expected.

More or less paralleling these increases in average holding are the increases in the variances of these holdings. Because of the highly skewed nature of these holdings, the standard deviations exceed the means substantially for each type of asset. Not surprisingly, the adjustments serve to increase the standard deviations somewhat more though the coefficient of variation goes down for every variable but other assets.

4. Results of the Econometric Approach

The parameters of the three models shown on pages 8-9 were estimated both by ordinary least squares and by three-stage least squares. The

[illegible]

[illegible]

[illegible]

Table 1

Effect of Adjustments in Asset Variables on
Their Means and Standard Deviation

Variable	Mean		Standard Deviation	
	Unadj.	Adj.	Unadj.	Adj.
Savings accounts	$ 11,298	$ 16,917	$ 18,991	$ 24,107
Stocks	42,699	51,808	145,271	153,661
Other Assets	83,271	83,485	299,268	304,956
Total Assets	137,268	152,210	395,476	407,315
Savings accounts/total assets	8.4%	11.1%	4.8%	5.9%
Stocks/total assets	31.5%	34.2%	36.6%	38.6%
Number of observations	1,182	1,135	1,182	1,135

latter estimation procedure is much more efficient than ordinary least squares but may be more sensitive to specification error. The two sets of results taken together should provide some idea, however, of the extent to which differences in estimates of the parameters caused by errors in the data may be affected further by the estimation procedures.

The ordinary least squares estimates of the beta coefficients of the equations of Model A before and after adjustment for errors in the asset variables are presented in Table 2. The socioeconomic characteristics, the vector Z, used in all three equations are the same, namely, the variables listed in the table following the three asset variables at the top. The selection of these variables was governed partly by data availability and partly by the findings of previous studies regarding what characteristics appear to be related to household savings in one form or another.

As is evident from this table, differences between the two sets of estimates are more of degree than of anything else. For the savings function, the goodness of fit declines after the asset variables have been adjusted,[13] although five variables are significant at the .05 level after the adjustment as compared to three prior to the adjustment. Most significant perhaps is the change in the sign of the coefficient of the common

---

[13]For comparability with the 3SLS estimates, A & U statistics are given in addition to R^2, though the main reference in Tables 2-4 is to R^2. A is defined as the average absolute value of the residuals while U, Theil's measure of forecast accuracy is:

$$U = \frac{\sqrt{\Sigma_i (y_i - \hat{y}_i)^2/N}}{\sqrt{\Sigma_i y_i^2/N + \Sigma_i \hat{y}_i^2 / N}}$$

where $\hat{y}_i$ is the function estimate for the ith observation and $\hat{y}$ is the corresponding observed value.

Table 2

OLS Estimates of Beta Coefficients of Equations of Model A
Before and After Adjustment for Errors in Asset Variables

Variable	Savings accounts		Common stock		Other assets	
	Before	After	Before	After	Before	After
Total assets	.203**	.213**	.717**	.729**	----	----
Savings accounts	----	----	.038	-.033	.066*	.048
Common stock	.076	-.076	----	----	.442**	.484**
Married, spouse present	-.067	-.107*	-.047	-.055	.054	.060
Separated or widowed	.009	-.021	.007	.042	.012	-.009
Self-employed	.052	.024	-.092*	-.083*	.198**	.195**
Salaried professional	-.057	-.106	-.083	-.082	.064	.065
Clerical or sales	-.030	-.058	-.046	-.046	.030	.031
Craftsman, kindred work	-.053	-.069	-.053	-.054	.021	.021
Service worker	-.068	-.084*	.014	.008	-.004	-.005
Laborer	-.011	-.014	-.022	-.008	.003	-.007
Retired	-.068	-.055	.008	.043	-.017	-.044
City size[1]	.027	.033	.015	.037	-.157**	-.175**
City size[2]	.056	-.005	-.010	.009	-.106*	-.117*
City size[3]	.008	.055	.039	.060	-.134*	-.157**
City size[5]	.004	-.017	.009	.006	-.042	-.038
Male head	.143**	.163**	.035	.072*	-.005	-.024
Race white	.019	.006	-.112**	-.136**	.017	.040
Race nonwhite	-.013	-.022	-.039	-.052*	.001	.012
Education of head	.046	.012	.068**	.101**	-.005	-.029
Age of head	.162**	.184**	.042	.065*	-.004	-.016
Family size	-.057	-.064	-.045	-.076	-.005	.017
No. of children under 18	-.041	-.026	.022	.049	.031	.001
R^2 (adj.)	.138	.107	.575	.607	.274	.288
A	10.62	13.74	39.34	43.81	80.99	85.94
U	.490	.466	.347	.325	.514	.506

*Significant at .05 level
**Significant at .01 level

stock variable from a significant positive value to a nonsignificant negative value almost of the same magnitude (and which would have been judged significant at the .09 level). At the same time, the adjusted data indicate significance for marital status and for service workers and assign appreciably greater importance to the effect of the presence of a male head and of older people on increasing savings account balances.

The effect of the adjusted data seems to be more pronounced on the estimates of the parameters of the common stock function, but less so on the estimates of the parameters of the function for other assets. In the case of the common stock function, the adjustments serve to increase the number of coefficients significant at the .05 level from four to seven. In particular, the adjustments highlight the significance of nonwhites as a factor reducing common stock ownership, and of the presence of a male head and an older head for increasing stock ownership. The function also ascribes much greater importance to education in affecting (positively) the amount of stock owned.

The effects of the adjustment seem least pronounced on the estimates of the parameters of the function for other assets. In the case of this function, the goodness of fit is increased hardly at all while the number of variables significant at the .05 level declines from six to five.

The results for Model B are very similar to those obtained for Model A, as can be seen from Table 3. This is not especially surprising since the models are similar to each other (the common stock equation is in fact the same). Elimination of the common stock variable from the savings account function seem to have virtually no effect, as would be

expected from the relatively little importance attached to the coefficient of this variable in Model A. Nevertheless, the adjusted data do seem to have brought about much larger differences in the estimates of the parameters of the savings functions of Model B than of Model A. Thus, the effect of total assets is reduced substantially. On the other hand, the effects of the other significant variables are much more pronounced, particularly the now significant effect of marital status and of service occupations.

Highlighting the changed effect of the total assets variable is that the elasticity of savings account balances relative to total assets declines from 3.1 before the data adjustments to 1.4 after the adjustments. Though still elastic, the effect of total assets on savings accounts has been reduced by more than half.[14]

In the case of the function for other assets, the estimates of the parameters seem to have been affected dramatically not only by the adjustments in the data but also by the addition of the variable for total assets. The latter variable clearly dominates the relationship, as is evidenced by the increase in the goodness of fit of this function from an R^2 of approximately .30 before the inclusion of this variable to an R^2 of about .90 with its inclusion. The error adjustments in the asset variables increase the goodness of fit only slightly but increase substantially the number of variables significant at the .05 level--from 5 to 8. Presence of a male head and older age of head now have significant negative effect on holdings of other assets while nonwhite race has a negative effect.

---

[14]The corresponding elasticities for the savings function of Model A are 2.5 before adjustment of the data and 1.9 after the adjustments.

Table 3

Estimated Beta Coefficients of Equations of Model B
Before and After Adjustment for Errors in Asset Variables

Variable	Savings accounts		Other assets	
	Before	After	Before	After
Total Assets	.258**	.016**	.973**	.968**
Savings accounts	----	----	-.082**	-.062**
Married, spouse present	-.071	-.103*	.023	.028
Separated or widowed	.009	-.025	-.003	-.021
Self employed	.045	.030	.044*	.042*
Salaried professional	-.064	-.100	.040	.041
Clerical or sales	-.034	-.055	.022	.023
Craftsmen, kindred worker	-.057	-.065	.025	.027
Service worker	-.067	-.085*	-.007	-.004
Laborer	-.013	-.013	.011	.004
Retired	-.067	-.058	-.004	-.022
City size[1]	.028	.030	-.007	-.019
City size[2]	.055	-.005	.005	-.005
City size[3]	.011	.051	-.019	-.030
City size[5]	.005	-.018	-.004	-.003
Male head	.146**	.158**	-.017	-.036*
Race white	.010	.017	.054**	.068**
Race nonwhite	-.016	-.018	.019	.026*
Education of head	.051	.004	-.033**	-.051**
Age of head	.166**	.179**	-.020	-.033*
Family size	-.060	-.059	.022	.038
No. of children under 18	-.040	-.030	-.011	-.025
R^2 (adj.)	.137	.106	.900	.900
A	10.64	13.77	39.34	43.81
U	.491	.467	.155	.154

#Common stock function is the same as in Model A (Table 2).

*Significant at .05 level
**Significant at .01 level

Perhaps the most important changes brought about by the adjustment for the errors in the asset variables are apparent in the estimates of the parameters of the two equations of Model C, as shown in Table 4. In both cases, the goodness of fit is increased substantially, R^2 rising from .17 to .22 for the savings account function and from .24 to .39 for the common stock function. The number of variables significant at the .05 level, however, is hardly changed. The effect of total assets on the savings account ratio is substantially more negative, as is also true for marital status and education of family head.

In the case of the common stock function, the error adjustments produce nonsignificance of the coefficient of the total assets variable while increasing substantially the negative importance of the savings account ratio variable, of race and of family size. The adjustment also serves to remove the significance of the service and laboring occupations of heads of families, while at the same time highlighting strong positive effects due to age of head and to the number of children under 18. In this case it seems clear that any inferences regarding the effects of the other assets variable as well as of socioeconomic characteristic on the proportion of a family's total holdings in the form of common stock would be substantially different depending on which set of data were used.

It might be noted that the substitution of income for total assets, which was tested in a couple of cases, yielded much poorer goodness of fit. Thus, in the case of the common stock function of Model A, substitution of income for assets yielded an overall value of .251 for R^2 as compared to .575 when total assets was used.

Table 4

Estimated Beta Coefficients of Equations of Model C
Before and After Adjustment for Errors in Asset Variables

Variable	Savings accts./total assets		Common stock/total assets	
	Before	After	Before	After
Total assets	-.090**	-.131**	.120**	.041
Savings accts./total assets	----	----	-.273**	-.440**
Married, spouse present	-.183**	-.209**	-.099*	-.174
Separated or widowed	-.102*	-.104*	.024	-.032
Self employed	.015	-.039	-.175**	-.157**
Salaried professional	.027	-.048	-.183**	-.196**
Clerical or sales	.015	-.046	-.111*	-.139**
Craftsman, kindred worker	.084	.057	-.177**	-.187**
Service worker	.075*	.029**	-.094**	-.094
Laborer	.055	.094	-.089**	-.030
Retired	.065	-.006	-.061	-.024
City size[1]	.201**	.200**	.011	-.009
City size[2]	.069	.003	-.086	-.110**
City size[3]	.072	.094	-.041	-.049
City size[5]	.010	-.027	-.002	-.054
Male head	.026	.053	.037	.051
Race white	-.002	.012	-.040	-.132**
Race nonwhite	-.030	-.049	-.029	-.084**
Education of head	-.131**	-.197**	.177**	.180**
Age of head	.527	.074	.216**	.262**
Family size	-.228**	-.168**	-.118*	-.254**
No. of children under 18	.062	.053	-.012	.112*
R^2 (adj.)	.166	.220	.238	.393
A	.158	.188	.165	.166
$\dot{U}$	.429	.374	.416	.343

*Significant at .05 level
** Significant at .01 level

To summarize, the principal effects of the adjustment for the errors in the asset variables would seem to be some improvement in the goodness of fit and, more important, realignment of the importance of the asset variables in relation to each other and increased importance of a number of socioeconomic variables previously not judged significant. In view of these adjustments, the question naturally arises of the extent to which the particular adjustment process employed has predetermined these results.

There is little question that such predetermination is inherent in the adjustment process. At the same time, there is also little question that nonreporting of the financial data as well as reporting errors are related to socioeconomic characteristics. Thus, nonreporting of savings accounts tends to be higher among older people and among those with higher incomes.[15] Under the circumstances, adjustments to correct for these errors serve in effect to eliminate part of the noise in the data and to restore regularities which should have been there in the first place.

To be sure, there is always the possibility that the adjustments may have gone too far and have introduced irregularities which are not really present. Such a possibility cannot be eliminated simply from these results alone. It is worth noting, however, that the principal effect of these adjustments seems to have been on the common stock and the other-assets functions, both assets for which the amount and nature of the

---

[15]Robert Ferber, op. cit., Chapter 4.

adjustments were less than for savings accounts. More information about the effect of these adjustments is provided in the following section, which describes the simulation undertaken to explore in further detail the effects of such errors on the estimates of the parameters of these models.

In broad outline the results of the three-stage least squares estimates are the same as those just reported for the ordinary least squares estimates although appreciable differences are apparent in some of the individual functions. Thus, as is evident from the 3SLS of Model A in Table 5, the goodness of fit is improved primarily for the common stock function after adjustment for errors in the asset variables. The number of coefficients significant after adjustment is the same as the number before adjustment for two of the functions and is slightly higher for the savings account function. Also, the adjustment process serves to highlight the importance of common stock as a determining variable both in the savings accout function and in the function for other assets and, more generally, seems to accentuate the influence of the variables that are significant even before adjustment.

Rather surprisingly, however, total assets does not seem to be important at all in the common stock function in the 3SLS estimates whereas it is by far the most important variable in the OLS functions.

In the case of Model B, all three functions exhibit improvement in goodness of fit at least in terms of U, after adjustment for errors in the asset variables. At the same time, the common stock and other assets functions of this model show a very substantial improvement in goodness of

Table 5

Estimated 3SLS Beta Coefficients of Equations of Model A
Before and After Adjustment for Errors in Asset Variables

Variable	Savings accounts		Common stock		Other assets	
	Before	After	Before	After	Before	After
Total Assets	.500**	.373**	-.024	.468	---	---
Savings accounts			2.859	1.591	3.106**	3.142**
Common stock	-.306**	-.276**	----	----	.169	.557**
Married	-.087	-.115*	.152	.108	.254*	.398*
Separated or widowed	----	----	.003	.073	----	----
Self-employed	.086**	.087**	-.316*	-.170	-.231*	-.311*
Salaried professional	-.001	-.033	-.039	.039	-.047	.098
Clerical or sales	----	----	-.028	.008	.024	-.014
Draftsman, kindred worker	----	----	-.016	.009	.045	-.007
Service worker	-.037	-.042	.169	.126	.156	.110
Laborer	-.009	.009	----	----	.046	-.039
Retired	-.012	-.015	.115	.115	.088	-.066
City size1	----	----	----	----	.170**	.133**
City size2	----	----	----	----	.265**	-.016
City size3	----	----	----	----	.124**	.220**
City size5	----	----	----	----	.039**	-.065**
Male head	.142**	.174**	-.355	-.193	-.446**	-.577**
Race white	----	----	-.150**	-.176**	----	----
Race nonwhite	----	----	-.036	-.035	----	----
Education of head	----	----	.019	.100**	----	----
Age of head	.174**	.200**	-.436	-.251	-.532**	-.657**
Family size	-.105**	-.096**	.240	.102	.329**	.322*
A	10.88	13.89	234.9	154.1	541.0	576.2
U	.503	.468	.706	.593	.733	.742

*Significant at .05 level
**Significant at .01 level

fit in terms of both U and A compared to the corresponding functions of Model A.[16] Besides the fact that the total assets variable is highly significant in all three functions, both before and after error adjustment, more variables are significant after adjustment in both the common stock and other assets functions.

As compared to the OLS estimates, the main differences are that now many different variables are significant. In the far majority of cases, however, the direction of the relationship is the same.

The effect of the adjustment for errors on Model C appears mixed (Table 7) because on one basis, the statistic A, the goodness of fit worsens, whereas on another basis, Theil's U, the goodness of fit improves. However, since Model C entails a pronounced change in the unit of the dependent variable as compared to the previous models, it would not seem unreasonable to select U as the better basis for comparison, especially with the other models. On this basis, the goodness of fit for the savings account function after adjustment seems to be the lowest of any of the savings accounts functions in any of the models and about the same as by the OLS method of fit. The goodness of fit of the common stock function is also appreciably lower after adjustment than before, and in this sense presents a very similar result to those obtained for the common stock function of Model B.

---

[16]It will be recalled that the same improvement was evident for the other assets function by OLS but not for the common stock function which, by definition, is the same in the two models. This would seem to be a very good example of the greater efficiency of the 3SLS estimating procedure.

Table 6

Estimated 3SLS Beta Coefficients of Equations of Model B
Before and After Adjustment for Errors in Asset Variables

Variable	Savings accounts		Common stock		Other assets	
	Before	After	Before	After	Before	After
Total assets	.259**	.155**	.891**	.788**	1.069**	1.014**
Savings accounts	----	----	-.611	-.403	-.445	-.369
Common stock	----	----	----	----	----	----
Married	-.065	-.101	-.093	-.100	----	----
Separated or divorced	.011	.018	.003	.021	----	----
Self-employed	.102**	.107**	.013	-.026	.101**	.104**
Salaried professional	----	----	-.046	-.062	.062**	.085**
Clerical or sales	----	----	-.024	-.037	.032	.047**
Craftsman or kindred	----	----	-.020	-.043	.044*	.057**
Service worker	-.048	-.047	----	----	-.018	-.004
Laborer	.005	.012	----	----	.010	.019
Retired	----	----	.050	.058	.021	.004
City size[1]	----	----	----	----	----	----
City size[2]	----	----	----	----	----	----
City size[3]	----	----	----	----	----	----
City size[5]	----	----	----	----	----	----
Male head	.139**	.178**	.114	.118	.030	.024
Race white	----	----	-.109**	-.136**	.050**	.057**
Race nonwhite	----	----	-.034	-.045*	.026*	.033**
Education of head	----	----	.050*	.101**	-.056**	-.053**
Age of head	.149**	.194**	.126	.132	.025	.034
Family size	-.096**	-.087**	-.091	-.070	-.021	-.003
A	10.65	13.85	67.31	56.97	77.79	74.38
U	.494	.470	.434	.359	.219	.205

*Significant at .05 level
**Significant at .01 level

The number of significant coefficients is appreciably higher for both functions of Model C after error adjustment. Indeed, for both equations in this model the large majority of coefficients were statistically significant after adjustment, the only model or method of fit for which this was true.

All things considered, therefore, the results of the 3SLS estimation procedure supports that of the OLS estimates in improvement in the goodness of fit and in highlighting relationships previously judged not significant.

Results of the Simulation Study

The simulation study was carried out using only a fifth of the original sample because of the larger size of that sample (1,135) and because it was felt more important to obtain more simulations on a smaller sample than fewer simulations on a larger sample. Accordingly, after arranging the observations in numerical order every fifth observation was selected, yielding a sample of 226 observations. With this sample, 150 simulations were planned (50 simulations for each of the three models), a figure that could be accomodated with the available computer resources and which was felt to be large enough to yield reasonably stable results.

Using the adjusted data for this smaller sample, as explained in the section on methodology, estimates were obtained of the parameters of the equations of each of the three models using alternately single equation least squares and three-stage least squares. For the purposes of the simulation, these estimates serve as the "true"population values.

Table 7

Estimated 3SLS Beta Coefficients of Equations of Model C
Before and After Adjustment for Errors in Asset Variables

Variable	Savings assets/total assets		Common stocks/total assets	
	Before	After	Before	After
Total assets	-.094**	-.132**	-.137**	-.078*
Savings accts/total assets	----	----	-.089	-.182
Married	-.179**	-.211**	-.086	-.127*
Separation or divorced	-.096**	-.103**	----	----
Self employed	-.077	-.082**	-.172**	-.145**
Salaried professional	-.066	-.094**	-.182**	-.183**
Clerical or sales	-.057	-.086**	-.111*	-.127**
Craftsman or kindred	----	----	-.189**	-.201**
Service worker	----	----	-.105**	-.100**
Laborer	.022	.077**	-.100**	-.050
Retired	-.010	-.045	-.067	-.009
City size 1	.142**	.204**	----	----
City size 2	----	----	-.066*	-.065*
City size 3	-.003	.091**	----	----
City size 5	----	----	.009	-.029
Male head	.029	.057	.020	.033
Race white	----	----	-.040	-.132**
Race nonwhite	-.024	-.052	-.023	.070*
Education of head	-.133**	-.097**	.195**	.234**
Age of head	.032	.059	.214**	.221**
Family size	-.178**	-.123**	-.094	-.128**
A	.158	.189	.168	.171
U	.431	.375	.431	.369

*Significant at .05 level
**Significant at .01 level

Next, the simulation itself was carried out on the unadjusted data for the same 226 observations. The procedure for simulating the unadjusted asset information for each sample was as follows:

1. An estimate had already been obtained of the proportion of nonreporting owners of common stock, and of savings accounts, as described in steps two and three of the section on data adjustment. This yielded point estimates of 31.3 percent of the families not reporting common stock and 38.5 percent of the families not reporting savings accounts.
2. Adjustments for nonreporting of assets other than savings accounts and common stock were made on the basis of what is known about nonreporting errors of these other assets.[17] Based on those results, the average number of nonreporters of these other assets was taken as half of the number of nonreporters of common stock.
3. These estimated proportions were assumed to be normally distributed with the mean being the point estimate and with the variance that computed using that point estimate. Using this

---

[17]The main source for such information is the summary results of the validation studies conducted as part of the Consumer Savings Project of the Inter-University Committee for Research on Consumer Behavior. These results pertain to demand deposits, personal loans, auto loans, and farm assets. The results are summarized in Robert Ferber, The Reliability of Consumer Reports of Financial Assets and Debts, University of Illinois Bureau of Economic and Business Research, Studies in Consumer Savings, No. 6, 1966.

assumption, the numbers of nonreporters of each of these three assets for each of the 150 samples were generated with the aid of standardized random normal variants.

4. The specific nonreporters of each asset in each of the 150 samples was generated according to a uniform distribution, in the absence of any other information.[18]

5. To simulate errors in the asset holdings that were reported, three log normal distributions were used -- one for the ratio of reported total savings account balances to adjusted total balances in savings accounts, one for the ratio of reported holdings of common stock to adjusted total holdings of common stock, and one for the ratio of reported other assets to adjusted other assets (assuming a mean of 1.0 and a variance of .3, based on findings of other studies noted elsewhere in this paper). A log normal distribution was used because this was felt to be typical of most economic variables. The holdings of all the sample members were adjusted for each of the 150 samples by adjusting this ratio with the aid of standardized log normal random variables.

6. Using the 150 samples generated in this manner, estimates of the parameters of the equations of each of the three models were

---

[18]Contrary to what might be anticipated, the limited work that has been done provides no support for hypothesizing that nonreporters of one asset are more likely than reporters to be nonreporters of another asset.

generated by three-stage least squares. These estimates were compared with the "true" values of the same parameters by computing for each parameter estimate the mean of the estimates, the variance of the estimates and the mean square error. In addition, for each fitted equation two measures of the adequacy of the fit were computed, namely, A, and Theil's measure of forecast accuracy, U, as noted previously.

The results of the simulation are quite surprising and are exemplified to a large extent by the summary statistics in Table 8 pertaining to the OLS estimate for Model A. The summary statistics shown in this table for each parameter estimate of each of the three equations are the value of the true parameter (Column 2), the average of the 50 estimates obtained from the simulation (Column 3), the ratio of the latter to the former (Column 4), the proportion of times the 95% symmetrical confidence interval of the estimate contains the true parameter (Column 5), the average width (range) of this confidence interval (Column 6), the average lower bound (Column 7) and the average upper bound (Column 8) of this interval, and the coefficient of variation of the parameter estimate (Column 9).

The surprising nature of the results is perhaps best highlighted by the following capsule overview:

1. The average of the parameter estimates, even after 50 replications, does not come very close to the parameter, with some exceptions (Column 4). Only one quarter of these average of the estimates were within 10 percent, while nearly half of the

averages (21 of 47) deviated from the parameter by more than 20 percent; 9 of these 47 averages were in error by over 50 percent.

2. Nevertheless, the 95 percent confidence interval contained the true parameter almost invariably, with a few notable exceptions (Column 5). Thus this interval contained the parameter value more than 90 percent of the time for 40 of the 47 parameters in the model. (Note, however, that most of the key financial variables are among the exceptions -- the confidence intervals for five of the six financial variables excluded the true parameter value anywhere from 36 percent to 64 percent of the time.)

3. Why could the parameter estimates differ so appreciably from the parameter estimates and still invariably encompass the true values in their 95 percent confidence intervals? The answer is in Columns 6-8, where we see that with rare exceptions the width of the confidence interval is so large as to be virtually meaningless. Not only is the range as a rule many times the size of the parameter estimate but it tends to cover both negative and positive values. Hence, not only is there hardly any indication of the magnitude of the parameter but the significance of the variable, and the direction of any such effect, is in considerable doubt; this is true of 41 of the 47 variables.

4. In other words, the variances of the parameter estimates tend to be tremendous relative to the estimates themselves.

Table 8

OLS Simulation Estimates of Parameters and of Their Reliability, Model A

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				95% conference interval				
Variable	b_i	$\bar{\hat{b}}_i$	$\bar{\hat{b}}_i/b_i$	Pct. of time b_i included	Ave. range	Ave. lower bound	Ave. upper bound	$\bar{\hat{\sigma}}_{b_i}/\bar{\hat{b}}$
				Savings Accounts				
Constant	-1.87	-2.11	1.12	100%	40.3	22.2	18.1	490%
Total assets	.0259	.043	1.68	64	.047	.020	.067	28
Common stock	- .0094	-.040	4.30	58	.080	-.080	- .0003	51
Married	-8.62	- 8.10	.94	98	24.7	-20.4	4.24	78
Self-employed	-3.56	- 4.73	1.33	98	27.3	-16.4	6.90	126
Salaried	-5.23	-5.09	.97	100	18.6	-14.4	4.21	93
Services	-14.9	-12.4	.83	100	44.7	-34.8	9.90	91
Laborer	7.67	6.08	.79	98	39.2	-13.5	25.7	165
Retired	.953	1.60	1.68	100	24.2	-10.5	13.7	392
Male head	8.42	8.18	.97	92	26.9	- 5.29	21.7	84
Age of head	3.34	2.77	.83	98	7.13	-.801	6.34	66
Family size	.758	.629	.83	100	6.32	-2.53	3.79	259

Table 8 (Cont.)

OLS Simulation Estimates of Parameters and of Their Reliability, Model A

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				95% conference interval				
Variable	b_i	$\bar{\hat{b}}_i$	$\bar{\hat{b}}_i/b_i$	Pct. of time b_i included	Ave. range	Ave. lower bound	Ave. upper bound	$\bar{\hat{\sigma}}_{b_i}/\bar{\hat{b}}$
				Common Stock				
Constant	40.0	26.5	.66	86%	245	-95.8	148.9	236
Total assets	.489	.468	.96	36	-.099	.418	.517	5
Savings accts.	-.076	-.306	4.02	60	.933	-.773	.160	78
Married	-2.49	-4.18	1.68	100	93.6	-51.0	42.6	572
Separated	31.7	28.1	.89	96	97.6	-20.7	76.9	89
Self-employed	-78.1	-70.4	.90	98	115	-128	-12.8	24
Salaried	-36.3	-34.7	.96	100	108	-88.4	19.1	80
Clerical	-26.3	-22.9	.87	100	106	-75.9	30.1	118
Kindred workers	-25.8	-23.9	.93	100	104	-76.1	28.2	111
Services	2.74	-3.68	1.34	98	171	-88.9	81.6	1,185
Retired	.364	-1.29	3.55	100	107	-55.0	52.4	2,062
Male head	10.7	14.6	1.36	100	96.0	-33.4	62.6	167
White	-71.9	-54.8	.76	64	136	-123	13.3	64
Nonwhite	-67.1	-52.8	.79	100	385	-245	140	187
Age of head	.087	1.04	11.90	100	25.4	-11.7	13.7	625
Family size	-2.40	-1.95	.81	100	21.6	-12.8	8.87	283
Educ. of head	3.50	2.89	.83	94	7.84	-1.03	6.81	69

Table 8 (Cont.)

Simulation Estimates of Parameters and of Their Reliability, Model A

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				95% confidence interval				
Variable	b_i	$\hat{b}_i$	$\hat{b}_i/b_i$	Frequency b_i included	Ave. range	Ave. lower bound	Ave. upper bound	$\hat{\sigma}_{b_i}/\hat{b}$
				Other assets				
Constant	15.8	22.8	1.44	100	318	-136	182	357%
Savings accts	.288	.217	.76	92	1.55	-.556	.991	183
Common stock	.356	.297	.83	56	.274	.160	.434	24
Married	18.5	15.2	.82	100	144	-56.9	87.4	242
Self-Employed	193	186	.96	98	227	72.3	299	31
Salaried	50.4	46.1	.91	100	212	-60.0	152	117
Clerical	44.8	39.1	.87	100	216	-68.9	147	142
Kindred	30.9	24.6	.80	100	222	-86.5	136	231
Services	40.3	47.1	1.17	100	311	-108	203	169
Laborer	-9.06	-16.1	1.78	100	292	-162	130	464
Retired	.167	2.52	15.1	100	216	-105	110	2,190
Loc 1	-87.4	-85.0	.97	96	177	-174	3.62	53
Loc 2	-52.7	-50.2	.95	100	190	-145	44.6	97
Loc 3	-28.4	-22.5	.79	98	170	-107	62.4	193
Loc 5	-9.14	-11.0	1.20	100	317	-169	147	736
Male head	1.28	1.91	1.49	100	158	-77.3	81.1	2,026
Age of head	5.22	5.61	1.07	100	41.3	-15.0	26.2	189
Family size	-6.30	-7.51	1.19	100	36.7	-25.9	10.8	124

This is reflected in the very high coefficients of variation of these estimates (Column 9). Only one of the 47 coefficients of variation is less than 10 percent, only four more are less than 50 percent, while 29 of 47 exceed 100 percent. Note that four of the six financial variables have coefficients of variation of only about 50 percent or less, and for these variables the 95 percent confidence intervals are as likely as not to miss the true parameter.

5. Highly unstable estimates are most likely to characterize the "other assets" equation and least likely the savings accounts equation. Thus, though the bases are small, the proportion of coefficients of variation exceeding 100 percent is 78 percent for the former equation, 59 percent for the common stock equation, and only 24 percent for the savings accounts equation.

6. By specific variables no clear pattern is apparent in the reliability of the estimates except the scale phenomenon that coefficients with higher absolute values tend to have lower coefficients of variation. In other words, a variable estimate with a very large confidence interval in one equation may or may not have a very large confidence interval in another equation, but it is difficult to generalize on the basis of such a small sample.

All things considered, then, the picture is one of extreme instability of the parameter estimates brought about by the errors introduced into the data. As a result, the 95 percent symmetrical confidence

intervals have the rather odd property of high reliability in the sense of including the true parameter and at the same time of being meaningless because the intervals are too large to have any substantive value.

The extent to which the same results are borne out by the other models and by the 3SLS estimates is shown in Table 9. On the whole, the results are much the same as before. The parameter estimates for Model C by three stage least squares appear to be somewhat closer to the true values than the estimates obtained from the other models, but the gain in accuracy is not substantial. Even for this model, nearly one-third of the average values of the estimates after 50 simulations deviate from the true figure by 20 percent or more.

Also, as before, for more than 80 percent of the parameters and for each of the models, the 95 percent symmetrical confidence interval includes the parameter 90 percent of the time or more. At the same time, between 70 and 90 percent of the confidence intervals cover both plus and minus values (the exact percentage verying with the model and the method of fit) so that there is as a rule no reliability as to either the significance or the direction of the effect of a particular variable.

Table 9

Summary Statistics on Reliability of Simulation Model Estimates

Statistic	Value	OLS estimates			3 SLS estimates		
		Model A	Model B	Model C	Model A	Model B	Model C
Frequency $\bar{\hat{p}}_i$ within given percent of p_i	Within 10%	12	6	9	11	10	13
	10-19%	14	12	10	9	6	11
	20-49%	12	16	10	8	15	6
	50% or more	9	8	6	19	11	5
	Total	47	42	35	47	42	35
Pct. of parameters for which 95% confidence interval includes parameter 90% of the time		85%	79%	100%	75%	81%	100%
Pct. of average 95% confidence intervals covering both plus and minus values		87%	86%	72%	85%	90%	74%
Size distribution of average coefficients of variation	Under 10%	1	2	0	0	0	0
	10-49%	4	5	11	6	4	8
	50-99%	13	9	14	9	9	14
	100% or more	29	26	10	32	29	13
	Total	47	42	35	47	42	35

The reason is again brought out when we consider the size distribution of the average coefficients of variation, shown at the bottom of Table 9. For Models A and B, regardless of the method of fit, the far majority of the standard deviations of the regression coefficients exceed the estimates of the coefficients themselves. Here too, Model C turns in a better performance, by either method of fit, with most of the standard deviations of the regression coefficients being less than their standard errors. Indeed, in the case of the least squares estimates for this model nearly one-third of the standard deviations of the coefficients are less than half the size of the coefficients themselves, the best showing of all the models in the methods of fit.*

The fact remains, however, that even with Model C one would hesitate to impute much reliability to the results. The inescapable conclusion is that the effect of the noise (reporting errors) in the financial variables are such as to render highly questionable any estimates of parameters that might be obtained with such data.

---

*It might be noted also that Model C also contains the best acting equations both in terms of stability of the coefficients of variation and in terms of closeness of approximation of the parameter estimates to the parameters themselves. In the latter sense, 15 of the 35 average parameter estimates are within 20 percent of the true value by the 3SLS method of fit and 13 by the OLS method. The standard deviations of the parameter estimates of this equation were less than the estimates themselves 16 of 19 times for the OLS method of fit and 14 of 19 times for the 3SLS method of fit.

Conclusions

The results of this study might appear at first to be contradictory. Thus, the first part of the study, the econometric approach, suggests that adjustment for reporting errors in the financial variables, at least in so far as possible, brings about some improvement in the goodness of fit of different equations of a model and the highlights the significance of variables not otherwise significant.

At the same time, the results of the simulation approach suggest that introduction of errors into what are taken to be a relatively error-free set of data leads to parameter estimates that differ substantially from the "true" parameters, and to confidence intervals that are meaningless for all practical purposes. In other words, the data produce a very high degree of instability in the parameter estimates.

Are these two sets of results inconsistent? Not at all. This becomes apparent if we compare the percentage deviation of the parameter estimates from the econometric approach after adjustment for errors in the variables with the "before" estimates, taking the "after" estimates as the supposedly true values. Dividing the "before" estimates by the "after" estimates yield a set of ratios comparable to those shown in Column 4 of Table 8 for the simulation estimates. As an example, reproduced here are the ratios of the "before" to the "after" OLS estimates of Model A (Table 2) in conjunction with the ratios for the same variables from the simulation for the same equation from Table 8.

Variable	Econometric model estimates: "Before"/"After"	Simulation estimates: $\hat{b}_i/b_i$
Constant	11.29	1.12
Total assets	.95	1.68
Common stock	1.00	4.30
Married	.63	.94
Self-employed	2.16	1.33
Salaried	.54	.97
Services	.81	.83
Laborer	.79	.79
Retired	1.24	1.68
Male head	.88	.97
Age of head	.88	.83
Family size	.89	.83

It is rather striking that the two sets of ratios are of the same order of general magnitude, except that the ratios of the parameter estimates from the econometric approach appear to be much more volatile than those from the simulation approach. This is only to be expected because it should be recalled that the parameter estimates from the econometric approach are single estimates, while from the simulation approach they represent ratios of an average over 50 simulations. Even so, many of the ratios are very similar and, with only one exception, are also in the same direction.

This tabulation, plus others prepared from the other equations and other models, indicate strongly that the results of the two approaches tend to complement rather than conflict with each other.

It therefore seems clear that response errors of the magnitudes that appear to exist in financial data can distort seriously not only the means and variances of the corresponding univariate distributions but in addition estimates of the parameters of econometric models not only of these variables but of many other variables included in the model. The problem is clearly a most serious one. Where response errors of these magnitudes exist, parameters estimates have to be treated with a great deal of caution. Indeed, there may be no substitute to putting a great deal of additional effort into getting better data, partly through better data collection methods and partly through the use of such supplementary methods as validation techniques. Fortunately, most types of economic data do not seem to contain the

large response errors characteristic of consumer saving data, so possibly parameter estimates for models of other types of consumer behavior are less subject to distortion from this source. This is a question which remains to be investigated.