Philippine Review of Economics and Business
Volume XVIII, Nos. 1 & 2, March & June, 1981

TRUNCATION BIAS IN HOUSEHOLD MONEY DEMAND TESTS

By Manuel F. Montes*

1. Introduction

Because of continuing controversy on the effectiveness of government policy on aggregate economic conditions and because of its importance in forecasting models, the interest in research on the nature of the demand for money equation is undiminished. It is unfortunate that on both counts, the research results so far have been less than conclusive. Cooley and LeRoy (1982) discuss the weakness of the results of these studies on the basis of identification problems and research reporting bias. The "case of the missing money" (Hamburger, 1977) in which money equations overestimated money demand in the 1970s is a recent example of the forecasting problems that have recently been noted.

The interest of this paper is more in line with the first motivation. This paper reports on research undertaken at the household level with a focus on testing various hypotheses about the money demand function. These tests have typically been carried out using aggregate data because of convenience and the potential for forecasting application mentioned above. These tests are weak on both theoretical and empirical grounds.

One important issue in this field — the income (or wealth) elasticity of demand — is extremely difficult to test on aggregate money holdings, a portion of which is held by governments which cannot be easily thought of as having any "income", and by business

*Assistant Professor of Economics, University of the Philippines. This paper is partly based on Appendix B of my Ph.D. dissertation (Montes, 1981). I am indebted to Teresa Jayme Ho for introducing me to the sample selection literature (see also Ho, 1980). This paper has benefited from the suggestion of a colleague, Dr. Rolando A. Danao. The data utilized in this paper were made available in part by the Inter-University Consortium for Political and Social Research (ICPSR). Neither the original source or collectors of the data nor the consortium bear any responsibility for the analyses or interpretations presented here.

ıterprises, which on the basis of first principles ought not to anage its money holdings on the basis of its "income" (see for :ample Orr, 1970).

The identification problem is the source of the empirical objecon. At the aggregate level, it is difficult to isolate supply-induced ıanges in the money stock outstanding. The most recent compreensive study of money demand by Goldfeld (1973) concentrates on ıgle equation estimates of the demand for money.

These comments underscore the need to research the issue at ıe individual level. The use of such data avoids the aggregation and lentification problems discussed above. Three studies at the houseold level have been undertaken by Lee (1966), Peterson (1974), and rescher (1977).

The interest of this paper is to draw special attention to the ata truncation problems of these household studies. All of these :udies, including the present, rely on demand deposit balances as manner of money holdings. It has been impossible to collect data n cash and currency holdings at the household level. For this eason, the existence of households in the surveys that report no deıand deposits poses a methodological problem. Does one include all he observations in the survey of households including those which eported zero demand deposits? Such a procedure could result in a runcation bias because households that report zero demand depoits might be doing so because they are constrained from holding ıegative demand deposits. On the other hand, exclusion of these ıouseholds could result in a different bias if household behavior is lifferent when demand deposits are positive and when they are ıegative.

These points can be illustrated in the following diagram. Conider only the money-income relationship (in logarithms) where the ısual increasing relationship is assumed. The x's represent the observed data points. Below Y_0, households would like to hold negative money balances represented by the 0's but only the corresponding zero balances are observed. These points have been drawn with the hypothesis that after income level Y_1, the usual positive income elasticity of money demands holds. Between Y_0 and Y_1, zero demand deposits are observed because the start-up costs of opening a checking account (such as minimum balance requirements) prevent the opening of such an account. Below Y_0, how-

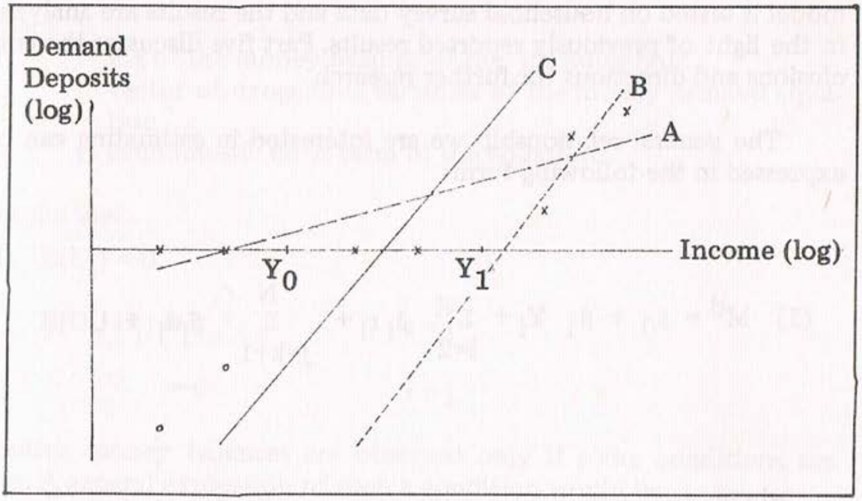

Diagram 1

ever, these fixed costs are not operative and the usual positive elasticity could be estimated only if negative money balances could be observed.

A regression procedure that used all the observations could be drawn as the line labeled A. It exhibits a smaller income elasticity because the zero observations have been included in the estimation. It also suggests a smaller fixed costs to money holdings, with positive holdings starting at a low level of income. A regression procedure that used only the positive observations (line B) would be able to estimate the proper income responsiveness as long as actual behavior is consistent with equal income elasticities for positive and negative demand deposits. However, it would tend to suggest that the fixed costs of starting up a checking account are rather high.

Line C would be an estimate that would neither underestimate nor overestimate both the fixed costs and the income elasticity of money balances. A procedure whereby the information from the observed x's about the possible location of the 0's in the diagram will be required to estimate such a relationship.

By making assumptions about the process by which positive money balances are observed, a procedure for estimating the possible location of the 0's is proposed in the next section. The third part is a brief description of the data set used. In the fourth section, this nodel is tested on household survey data and the results are analyzed n the light of previously reported results. Part five discusses the conclusions and directions for further research.

The general relationship we are interested in estimating can be xpressed in the following form:

$$(1) \quad M^d = \beta_0 + \beta_1 Y_i + \sum_{i=2}^{K} \beta_i r_i + \sum_{j=k+1}^{N} \beta_j d_j + U$$

where :

M^d : money demanded

Y : disposable income

r_i : rate of return to money and alternative assets

d_j : dummy variable for household characteristics

U_j : an econometric error term

The economic variables are in logarithms so that the coefficients of these variables measure the corresponding response elasticities.

2. Estimation On a Truncated Sample

The problem of sample selection bias arose from the fact that after data editing, only 73 per cent of the households in the survey reported positive demand deposits. The question was: What are the consequences of estimating equation (1) on the basis of the positive observations only instead of the whole sample?[1] The answer turned out to be dependent on the process by which the subsample is chosen — hence the problem has come to be called one of sample selection.

Following Heckman (1977), the problem of estimation bias can be illustrated in the following model. Write out a general formulation of the money demand equation for N observations as:

$$(2) \quad M_i = X_i^1 \beta + U_i \qquad i = 1, \ldots, N$$

[1] The same situation applies in many other fields. For example, what is the bias in estimating a migration decision model on the basis of data on migrants alone?

where :
M_i : log of the money balance of the *ith* household
X_i : vector of exogenous variables of the money demand equation
U_i : econometric error term of the equation

Assume that:

(3) $E(U_i) = 0$

$$E(U_i, U_j) = \sigma_u^2 \qquad i = j$$

$$= 0 \qquad i \neq j$$

Positive money balances are observed only if some conditions are met. A general expression of such a condition would be:

(4) $M_i > 0$ if and only if $L_i > 0$

(5) $L_i = Z_i' \alpha + V_i$

where L_i is the variable, possibly unobserved, that determines whether a positive value will be observed for money balances. Z_i are the variables that explain L_i and V_i is the stochastic error term. Assume that:

(6) $E(V_i) = 0$

$$E(V_i, V_j) = \sigma_v^2 \qquad i = j$$

$$= 0 \qquad i \neq j$$

The regression function for the whole sample when there are no problems in observing all possible values of the dependent variable is simply

(7) $E(M_i \mid X_i) = X_i\beta \qquad i = 1, \ldots, N.$

However, when only a subsample of the data is available, the regression function becomes

(8) $E(M_i \mid X_i,$ sample selection rule) $= X_i \beta + E(U_i \mid$ sample selection rule) and this applies to observations $i = 1, \ldots, K$ that permit positive observations only. If the conditional expectation of U_i is zero, application of least squares permits the unbiased estimation of the coefficients accompanied by a loss in efficiency.

If we apply conditions (4) and (5), this conditional expectation ill be:

$$9)\quad E(U_i \mid \text{sample selection rule}) = E(U_i \mid L_i > 0)$$
$$= E(U_i \mid V_i > -Z_i'\alpha)$$

nd unless U_i and V_i are independent, this expectation will not be ero. Because of equation (8), the regression equation (7) becomes:

$$10)\ E(M_i \mid X_i, L_i > 0) = X_i\beta + E(U_i' \ V_i > Z_i'\alpha)$$

o that in effect, the estimation of equation (7) is the estimation of n equation with a missing explanatory variable.

When the zero observations are caused *only* by the infeasibility of holding negative money balances, we have the "Tobit" model (Tobin, 1958; Amemiya, 1973.) A Tobit is a special case of the sample election model with $M_i \equiv L_i$, $\beta \equiv \alpha$, $X_i \equiv Z_i$, and $U_i = V_i$ so hat data are missing when $M_i < 0$. The bias is guaranteed when he sample is limited to positive values because X_i is identical to Z_i and therefore the conditional mean of U_i is not orthogonal to X_i because of equation (8). If there were no fixed costs in the relevant range, a "Tobit" would remove the bias in the line A of Diagram 1.

The principal problem in the estimation of the demand for money by households based on equation (9) is the estimation of the conditional expectation on the left-hand side. This involves the estimation of parameters of the joint distribution of the random variable U_i and V_i.

Under the assumption that U and V are jointly normal with covariance σ_{uv}, it has been shown that (Gronau, 1974):

$$(11)\ E(U_i \mid V_i > -Z_i\alpha) = \frac{\sigma_{uv}}{\sigma_v}\lambda_i$$

$$(12)\ \lambda_i = \frac{\phi(W_i)}{1 - \Phi(W_i)}$$

where ϕ is the density and Φ the distribution of the standard normal variable W_i defined by

$$(13) \quad W_i = \frac{Z_i\alpha}{\sigma_v}$$

The variable $\lambda_i \equiv \lambda(W_i)$ is the inverse of the Mills ratio. Its numerator is a measure of the probability of occurrence of a particular observation and its denominator is the probability that an observation with characteristics Z_i is selected into the observed sample.

Some comments on the nature of the inverse of the Mills ratio, $\lambda(W_i)$, are appropriate. From its formula (12), $\lambda(W_i)$ is the ratio of the value at W_i of a standard normal and the area of the tail of the same distribution from W_i. It takes on nonnegative values. Because its denominator is the probability of sample selection, λ is a decreasing function of this probability. Observations with larger chances of selection (for example, those with higher incomes for our case of money demand) are assigned a smaller value of λ. When the probability of selection is high for all observations, that is, when little sample selection exists, λ tends toward zero which means that the problem of misspecification is minor. Otherwise, the use of λ as an additional variable is called for.

Substituting equation (10) into equation (9) yields

$$(14) \quad E(M_i \mid X_i, L_i > 0) = X_i\beta + \frac{\sigma_{uv}}{\sigma_v} \lambda(W_i).$$

A consistent estimate of $\lambda(W_i)$ can be used as an additional regressor in equation (14) which can then be estimated by applying ordinary least squares on the positive observations only.[2] Such a procedure allows the unbiased estimation of the vector β and $\frac{\sigma_{uv}}{\sigma_v}$.

A consistent estimate of $\lambda(W_i)$ can be obtained by first estimating W_i of equation (13) and then calculating $\lambda(W_i)$ by applying the equations of a standard normal function. If the pro-

[2] Heckman (1977) has shown that the ordinary regression estimates of the standard errors of the coefficients are less than the actual error of these estimates.

ability that the variable is positive originates from a standard normal distribution, this probability is equal to:

$$(15)\quad \int_{-\alpha}^{\frac{Z'_i\alpha}{\sigma_v}} \frac{1}{\sqrt{2\pi}} e^{-t^2} dt$$

so that a probit log likelihood function of the form

$$L = \prod_{i=1}^{N} [\phi(W_i)]^{d_i} [\phi(-W_i)]^{\lambda d_i}$$

can be applied to all the observations to consistently estimate $\frac{Z'_i\alpha}{\sigma_v} W_i$. The event d_i is the event "observation of M_i" because based on equation (4), M_i is observed if and only if L_i is positive.

We are now in a position to apply these ideas to the estimation problem at hand. There are many alternative hypotheses regarding the process by which the sample of positive demand deposit holdings is generated. The principal interest of this paper is not to discriminate between these alternative hypotheses (operationally defined by alternative specifications of equation (5)), but to measure the effect of a particular specification of the sample selection process on the money demand equation.

Limitations of computer resources in estimating the probit model[3] of equation (14) constituted the binding constraint[4] on the testing of alternative specifications of equation (5). With more than a thousand observations, a maximum of three explanatory variables could be handled in the probit estimation. Because the main equation has more than three explanatory variables, a straight "Tobit" model (where the variables in the probit function are identical to those in the main equation) could not be estimated.

---

[3]The probit estimation program, "The General Multivariate Analysis Program," was written by Fred Nold and John Lewis at Stanford University. The assistance of Howarth Bouis in adapting the program is gratefully acknowledged.

[4]Probit program can usually handle a huge number of observations because these are usually applied to experimental data where the level of treatment is fixed. This permits the input of grouped data. With data from a survey, the "treatments" are variables such as income which vary across a continuum in the sample.

The data on money balances consisted of balances on demand deposits of survey respondents. Respondents were asked whether they had any demand deposits. If they answered yes, respondents were asked how much they had in these deposits at the time of the interview.

The question of how a person decides to start a checking account (for United States residents) has much to do with the substitutability between cash and demand deposits. In a market economy such as that of the U.S., everyone has access to cash in exchange for the sale of goods and services. Cash is acceptable for practically all transactions. In contrast, the household must make a conscious decision to open a checking account. In the period when the data used in this paper was gathered, checking accounts did not pay any interest and were actually subject to service charges. In addition, in many areas in the United States, a check is not really as good as cash; the check is accepted as a means of payment only if accompanied by proper identification on the part of the issuer.

A checking account does provide a few advantages to the user. There is less risk of loss from theft. The ownership of a checking account in the U.S. permits the safe payment of debt through the postal system. A checking account would be a tremendous convenience for households that can avail of consumer credit (in the use of credit cards, for example).

Consider the following hypothesis about the nature of the auxiliary relationship that helps determine whether positive checking account balances are observed:

(16) $$L_i = \alpha_0 + \alpha_1 Y_i + V_i$$

where L_i is an index of the financial sophistication[5] of household i which determines whether the household uses a checking account or not and Y_i the disposable income of household i. Equation (15) expressed the restriction that the unobserved variable "financial sophistication" is completely determined by the disposable income of a household except for a scale value (the constant term) and a random, unpredictable error, V_i.

[5]It is typical in qualitative response models not to specify the exact variable that actually determines the qualitative response or to think of this variable in most general terms (Amemiya, 1982). A very general approach, for example, is to think of the explanatory variables to the auxiliary equation as parameters of the individual's utility functions (Domencich and McFadden, 1975). When these parameters are of such sufficient size that the utility value crosses a threshold, the individual is observed to make the qualitative response.

It would be more appropriate for our purposes to include variables that describe household characteristics such as educational attainment as other explanatory variables in equation (15). However, the inclusion of these variables frequently either yielded ill-conditioned data which prevented the probit estimation procedure from converging or resulted in non-significant contributions of certain household variables. The first problem can be interpreted as evidence of extreme collinearity between these household characteristics and income so that income itself may well be adequately proxying the effect of these variables. The second observation could be interpreted to mean that these particular insignificant variables may not be important to the decision of the household to open a checking account. But because the variables could be included only three at a time, not too much importance can be attached to these results. It is more appropriate to evaluate the results of this paper on the basis of the strict hypothesis proposed in equation (16).

Only the parameters of the auxiliary equation required in the unbiased estimation of the main equation (1) are of interest. With the addition of the inverse of the Mills variable, equation (1) now becomes:

$$(1') \qquad M^d = \beta_0 + \beta_1 Y + \sum_{i=2}^{K} \beta_i r_i + \sum_{j=K+1}^{N} \beta_j d_j + \alpha\lambda + U$$

where λ is the inverse of the Mills variable and

$$(17) \qquad \alpha = \frac{\sigma_{uv}}{\sigma_V}.$$

The first stage of the procedure is to estimate the probit relationship

$$(18) \quad \text{Prob } (L > 0) = \text{Prob } (M^d > 0)$$

$$= N[A + BY + \in]$$

where N stands for the normal distribution, A and B are coefficients and $\in$ is the error term over both positive and non-positive observations of demand deposits. Given the estimates of A and B, the probability of occurrence of each positive observation in equation

(1′) can be calculated using equation (17). Using equation (12), this probability of occurrence can be used to calculate the inverse of the Mills ratio. Equation (1′) can then be estimated by ordinary least squares with the estimated coefficient of λ as the estimate of α of equation (17).

3. Description of Data Set

The data used in this study were collected by the Michigan Survey Research Center in the United States in a project called "The Survey of Consumer Finances" by Gary Hendricks and Kenwood C. Youmans. The data were collected at the household level in a national sample in the years 1967, 1968, 1969, and 1970. Data on household characteristics, income, financial assets including demand deposit balances were collected from the same 1434 households in these four years.

The data editing process resulted in 165 households dropped because of missing values (that is, "don't know" or "no response") for either income or demand deposits. Some households were also not included because of apparent unreasonableness in their reported net financial assets. Ninety-five cases were dropped because these reported missing values for any of the three liquid assets for which the survey gathered data: demand deposits, time deposits, or bonds. An additional 26 cases were dropped because the reported increase in their liquid asset balances between certain years was greater than the sum of disposable income net of housing payments (mortgage or rent) and inheritance and the decrease in the value of common stock owned and the decrease in the value of the residence (if there was a decrease). This criterion was a minimum guarantee that none of the included respondents implicitly reported negative liquid asset balances in any year.

The total number of observations (note that there will be four observations per household) was 4592 or 1148 households. The total number of observations with positive demand deposit balances was 3356.

The data from the survey were augmented by interest rates that were computed for each state in the United States. These variables are charges on demand deposits (DDCHR), interest on savings accounts and time deposits (TDINT), and the return on savings and loan shares.

The method used to construct these indicators is the same one which Feige (1964) used in deriving his "actual interest rates" on a state-by-state basis. Likewise, I computed 51 interest rates for each of the three variables (50 states plus Washington, D.C.).

"Actual interest rates" are computed by taking interest payments (or charges) as these appear in the state-by-state combined financial statements of financial institutions and dividing this number by the "average" balances of the accounts these interest rates were applied to. Raw data sources were: Controller of the Currency, *Annual Reports*, 1968—1969; and the Federal Home Loan Bank Board, *Combined Financial Statements*, 1968—1969.

These interest rates are *not* marginal rates but average rates since "average" balances are derived by taking an arithmetic average between the beginning and ending balances of these accounts as these rates would actually be applicable through the whole year on which they are computed. Data from the Survey of Consumer Finances are gathered only within the first quarter of the same year. An additional assumption is therefore made that the relative positions of these interest rates across states were constant throughout that year.

The natural logs of these interest rates are referred to respectively as LDDCHR, LTDINT, and LSLINT.

After these fifty interest rates were computed, the states where the survey respondents came from were identified and the interest rates applicable to those states were assigned. The 4592 cases left in the main sample came from 35 different states so that in the cross section, there were 35 values of each interest variable.

In addition to these cross-section rates, various time-series rates were tried on the data with little success. For these rates, there were only four observations possible — one for each year in the cross section. Some of the rates tried were bond and stock yields, the rate on three month treasury bills. In order to exhibit some typical results, the regression reported in this study used the rate on three-month treasury bills (TR3MO) and its log is denoted by LTR3MO.

The distribution of these rates for the 4592 observations in the main sample can be seen in Table 1.

Table I — Distribution Of Constructed Interest Rates

	Mean	Median	Variance	Minimum	Maximum
DDCHR	0.510	0.479	0.048	0.177	1.044
LDDCHR	−0.764	−0.736	0.187	−1.730	0.043
TDINT	4.358	4.285	0.387	3.161	7.623
LTDINT	1.463	1.455	0.016	1.151	2.031
SLINT	4.582	4.588	0.119	0.443	6.228
LSLINT	1.517	1.523	0.016	−0.814	1.829
TR3MO	5.305	5.110	0.758	4.321	6.677
LTR3MO	1.656	1.630	0.025	1.464	1.899

4. Results

The estimate of the probability of observing positive demand deposits given the level of disposable income is the first result of interest:

$$\text{Prob }(L > 0) = N[\underset{(0.34)}{-8.22} + \underset{(0.04)}{1.02}\ \text{LDDISPY}]$$

Chi-square: 4508.9 Number of Observations = 4592

where LDDISPY stands for the logarithm of disposable income and the figures in the parentheses are the (asymptotic) standard errors of the estimates of the coefficients. These numbers show the probability that positive demand deposits as observed can be explained by the increase in the logarithm of disposable income at the 1 per cent significance level. An increase in the logarithm of disposable income by one standard deviation increases the probability of observing positive demand deposits by 13.5 per cent.

The results of the next stage of the estimation procedure are exhibited in Tables 2 and 3. For purposes of discussion, an estimate based on the positive observations only is given in the first column and the estimate based on the positive observations but now augmented by the inverse of the Mills variable is given in the second column. The large number of explanatory variables has necessitated reporting the OLS results in two tables. Table 2 contains the coeffi-

Table 2 — Cross-Section Demand For Money With Mills Variable—I

Explanatory Variable	Estimated Coefficient (Standard Error)	
	Without Mills Variable	With Mills Variable
Age of Head of Household	0.034* (0.003)	0.034* (0.002)
Family Size	–0.127* (0.017)	–0.129* (0.016)
Dummy for Bond Ownership (1 if true)	–0.110* (0.055)	–0.109* (0.054)
Dummy for Stock Ownership (1 if true)	0.229* (0.057)	0.212* (0.057)
Dummy for Home Ownership (1 if true)	0.007 (0.065)	0.009 (0.062)
Dummy for Managerial Occupation (1 if true)	0.192* (0.062)	0.198* (0.061)
Dummy for Wife in Labor Force (1 if true)	–0.240* (0.052)	–0.214* (0.050)
Dummy for Urban Location (1 if live-in-city)	–0.492* (0.054)	–0.487* (0.052)
Dummy for Expected Financial Condition Next Year (1 if worse or uncertain)	0.206* (0.077)	0.201* (0.074)
Dummy for Husband's Education (1 if college graduate)	0.090 (0.068)	0.094 (0.068)
Dummy for Wife's Education (1 if college graduate)	0.218* (0.087)	0.205* (0.088)

*Significant at the 5% level.

cients for most of the personal characteristics of the households while Table 3 concentrates on the economic variables. For each column in Tables 2 and 3 therefore, there is only one regression procedure being reported.

Our first comment concerns the results exhibited in Table 2. The demographic variables used in this cross-section regression were chosen because these had been previously used in earlier studies. All of them, except for two, (home ownership and husband's education), are significant at the 5 per cent level.

We can now interpret these results one by one. Age of the household head seems to increase demand deposit balances in a regression that includes the income level while family size has a significant negative sign. Both these results confirm the findings of other cross-section studies on the demand for money (for example Peterson, 1977). The negative coefficient on family size is evidence in favor of some economies of scale in money holdings.

The positive coefficient on age is paradoxical in the light of other findings of this study. Age cannot be a proxy for increasing financial sophistication which we have asserted to be indexed by the inverse of the Mills variable and income itself. Its coefficient does not change very much with the addition of the Mills variable. One possible explanation is that of human capital in the sense of Karni (1974). In a perfectly competitive market, wage will reflect payments to human capital. Since wage is a measure of the opportunity cost of time and since money balances are thought to reduce transactions in time, higher human capital (associated with greater age particularly if most training is acquired by learning by doing) would be accompanied by higher money balances. Attempts to directly derive wage rates of income earners were unsuccessful and prevented a more direct test of this explanation.

The next two variables have to do with ownership of other financial assets — in particular bond ownership and stock ownership. The significant negative coefficient on the bond ownership dummy variable suggests that households reduce their average money holdings when their money balances increase. This is not a direct test of the substitution relationship between bonds and money which can be carried out only with interest rates and a multi-equation model that reflects the Slutzky restrictions. The significant positive coefficient on the stock ownership dummy, on the other hand, suggests a complementary relationship between demand deposits and stock ownership.

Table 3 — Cross-Section Demand For Money With Mills Variable—II

Explanatory Variable	Estimated Coefficient (Standard Error)	
	Without Mills Variable	With Mills Variable
The Constant Term	—1.393* (0.583)	—1.684* (0.570)
Log of Charges on Demand Deposits	—0.135* (0.061)	—0.106* (0.059)
Log of Interest on Time Deposits	0.131 (0.229)	0.009 (0.221)
Log of Dividend Rate on Savings and Loan Shares	—0.274 (0.210)	—0.165 (0.192)
Log or Rate on 3-Month Treasury Bills	0.174 (0.181)	0.014 (0.174)
Log of Real Disposable Income	0.636* (0.057)	0.699* (0.056)
Inverse of Mills Ratio		—0.040* (0.014)

*Significant at the 5% level.

Total number of observations:		4592
Number of Positive Observations:		3356
Standard Error of the Estimate:	1.345	1.139
Adjusted Squared Multiple Correlation:	0.186	0.180
F-Statistic:	43.13	44.33

The next significant variable is the dummy for managerial occupation. It suggests that managerial personnel tend to hold higher money balances. The picture that emerges from this result and that of the previous one of complementarity between stock ownership and larger money balances seems to be that households engaging in more complicated financial dealings require higher money balances.

Because one would expect that these households would be more careful in their money management, these latter results are more relevant to the cash versus demand deposits than to the money versus other liquid asset substitution issue. These households probably use checking accounts for many transactions other households would use cash for.

The use of the dummy of the wife in the labor force was an attempt to measure the opportunity cost of the household's time including that of the housewife. This variable is significant but unexpectedly negative. It seems to point to the issue in studies on labor supply of women (for example, Heckman, 1977) that, apart from opportunity cost factors, there are probably permanent factors (such as taste and background) that are important in determining whether a woman seeks a job or not. It seems that these factors swamp the opportunity cost factors. The negative sign will be discussed in conjunction with the Mills variables below.

Urban location tends to result in significantly lower money balances. This is consistent with the result that money balances tend to be higher when the expected financial condition of the household is expected to be worse or uncertain. Because agricultural income tends to be subject to greater variance than salary income, the increase in money balances from rural folk is evident of the precautionary motive for holding money balances.

The dummy variable for a college degree for the husband is not a significant explanatory variable for money balances. This contradicts earlier findings such as that of Peterson (1974). The variable that is significant is the dummy on whether the wife has completed college. The positive sign is consistent with the value of time and human capital theories. This finding points to the importance of the housewife's participation in household financial management even for United States data.

Turning now to Table 3, the initial general comment is that the interest variables do not fare well in the estimation of the model. The only significant variable is the own interest rate on demand deposits which has a significant sign. The size of the estimated elasticity compares very well with previously reported estimates which range from 0.10 to 0.16.

Taken at face value, it seems that the substitution behavior between demand deposits and time deposits or savings and loan shares is not very strong. This finding however could also be due to the weakness inherent in most calculations of cross-section interest rates.

Disposable income has a very significant coefficient between 0.64 and 0.70. This value is significantly less than one and lends support to the idea of economies of scale in money holdings. In fact, the coefficient appears to be quite close to the theoretical elasticity of one-half from the Baumol-Tobin transactions theory.

The Mills variable has a significant coefficient of -0.40. The negative sign does not contradict our hypothesis that the probability of observing positive demand deposit balances is a function of the household's financial sophistication. The greater the probability of a household's having a checking account, the smaller the balances kept in this account tend to be. That is, σ_{uv} from equation (17) is negative. When average checking balances are high, the financial sophistication of that household tends to be low, and vice versa.

What is the effect of the Mills variable on the coefficients of the other variables? The bias from estimating the equation from positive observations alone seems to be most evident in the estimates of the constant term and the income elasticity. The constant term for the regression with the Mills variable is smaller while the income elasticity is higher.

My interpretation of this result is that households try to keep as low a balance in their checking accounts as they possibly can, so that empirically, one observes a lot of observations at low money balance levels of different levels of income. A regression only based on positive observations will interpret these observations as evidence of a relatively low income elasticity of money demand and possibly low fixed costs in financial management. The addition of the Mills variable correctly permits the interpretation of the data as evidence of more aggressive money management by households.

Diagram 2 tries to clarify this point. The difference between Diagrams 1 and 2 is the additional set of observed values marked by *'s. These *'s mark those households that truly keep only the minimum necessary for transactions purposes in their checking account. A regression based on positive observations only would exhibit a higher intercept and a lower income elasticity. The inclusion of the Mills variable would reflect the correct higher income elasticity for the whole population. It is a reasonable conjecture that the income elasticity for money demand is at least relatively high for households on the lower end of the income scale.

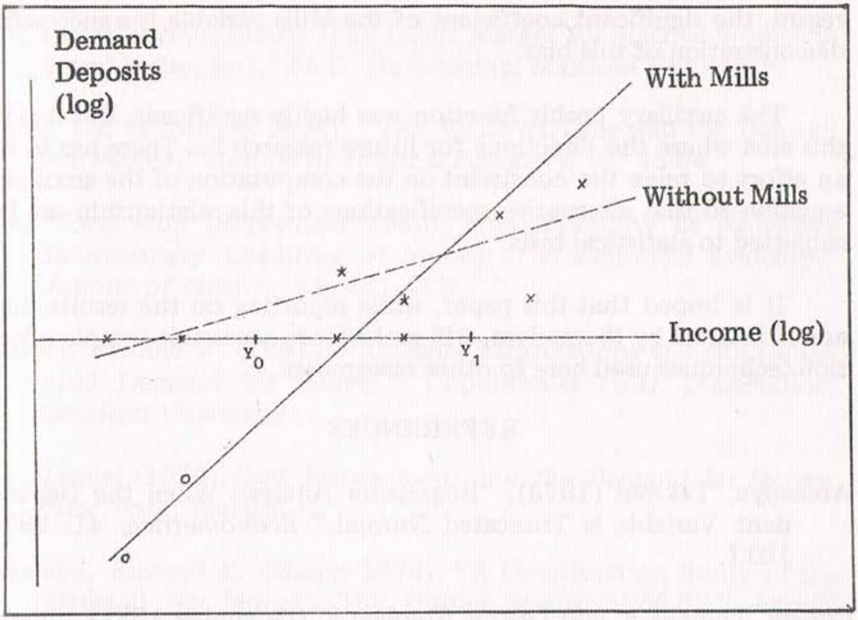

Diagram 2

The overall regression has an adjusted correlation coefficient of only 18 per cent. This means that there exist other explanatory variables whose effect could be very random that have not been taken into account in the estimated equation. Increasing the explained variance beyond 20 per cent in a household cross-section study has always been difficult. For our purpose of testing various hypotheses about the demand for money, the highly significant F-statistic in the order of more than 40 suffices to lend credence to our conclusions.

5. Conclusions

Previous findings on the effect of personal characteristics on the money demand function survive in this study relatively intact. There is evidence that the elasticity of money demand from this cross-section sample is less than one, supporting the hypothesis of economies of scale in money holdings. Only limited success can be reported in measuring interest elasticity of money demand at the household level.

The principal focus of this paper was the attempt to quantify the effects of truncation bias when household data are used. In this regard, the significant coefficient of the Mills variable is a successful demonstration of this bias.

The auxiliary probit function was highly significant. But it is in this area where the directions for future research lie. There has to be an effort to relax the constraint on the computation of the auxiliary equation so that alternative specifications of this relationship can be subjected to statistical tests.

It is hoped that this paper, while reporting on the results that are of interest by themselves, will make more accessible sample selection techniques used here to other researchers.

REFERENCES

Amemiya, Takeshi (1973), "Regression Analysis When the Dependent Variable is Truncated Normal," *Econometrica*, 41: 997-1017.

Cooley, Thomas F. and LeRoy, Stephen F. (December 1981), "Identification and Estimation of Money Demand," *American Economic Review*, 71: 825-844.

Demoncich, T.A. and McFadden, Daniel (1975), *Urban Travel Demand*, Amsterdam: North-Holland.

Drescher, Sol (1977), "The Demand for Money: A Household Production Function Approach," Ph.D. Dissertation, City University of New York.

Feige, Edgar L. (1964), *The Demand for Liquid Assets: A Temporal Cross-Section Analysis*, Englewood Cliff, N.J.: Prentice Hall.

Goldfeld Steven M. (1973), "The Demand for Money Revisited," *Brookings Papers on Economic Activity*, 3: 577-638.

Gronau, Reuben (1974), "Wage Comparisons — A Selectivity Bias," *Journal of Political Economy*, (November/December): 1119-1144.

Hamburger, Michael J. (1977), "The Behavior of the Money Stock — Is There a Puzzle?' *Journal of Monetary Economics.*

Heckman, James J. (1977), "Sample Selection Bias as a Specification Error with an Application to the Estimation of Labor Supply Functions," Mimeographed.

Ho, Teresa Jayme (1980), "The Labor Market for Married Women in Rural Philippines," Ph.D. Dissertation, Stanford University.

Karni, Edi (February 1974), "The Value of Time and the Demand for Money," *Journal of Political Economy*, VI: 45-64.

Lee, Tong Hun (September 1966), Substitutability of Non-Bank Intermediary Liabilities of Money: The Empirical Evidence," *Journal of Finance*, 21: 441-457.

Montes, Manuel F. (1981), "Rational Expectations and the Household Demand for Money," Unpublished Ph.D. Dissertation, Stanford University.

Orr, Daniel (1970), *Cash Management and the Demand for Money*, New York: Praeger.

Peterson, Richard E. (March 1974), "A Cross-Section Study of the Demand for Money: The United States, 1960-62," *Journal of Finance*, 29: 73-88.

Tobin, James (January 1958), "The Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26: 24-36.