IMPUTATION OF MULTIVARIATE DATA ON HOUSEHOLD NET WORTH

Steven G. Heeringa, Roderick J.A. Little, Trivellore E. Raghunathan
Steven Heeringa, Inst. for Social Res., POB 1248, Univ. of Michigan, Ann Arbor, MI 48106-1248

Keywords: Household net worth; limited variables; missing data; bracketed responses; interval censoring; mixed Normal distributions; Bayes' posterior simulation; multivariate imputation; imputation variance.

1. Introduction

Survey questions that ask respondents to report financial amounts -- particularly dollar values for income, assets, and liabilities -- are subject to high rates of item missing data. Researchers are using special questionnaire formats to address the missing data problem for these financial variables. Loosely termed "bracketed response questions," these question formats collect an interval-scale observation whenever a respondent is unable or unwilling to provide an exact response to a financial amount question (Heeringa, 1993; Kennickell, 1996). Bracketing question sequences reduce the rate of completely missing data for net worth variables; however, the resulting measures are a coarsened mixture of single valued responses, "bracketed" or interval-censored responses, and completely missing data.

The purpose of the paper is to investigate methods of imputing amounts for bracketed net worth variables, to empirically compare the distributional properties of data imputed by the alternative methods and to estimate the imputation variance that each method adds to the completed data set.

2. HRS Net Worth Variables: Bracketing and Missing Data Rates

Table 1 presents rates of bracketing and missing data for 12 of 23 variables needed to compute total household net worth for households interviewed in Wave 1 of the Health and Retirement Survey (HRS) (Juster and Smith, 1994). The left-hand panel of Table 1 identifies the individual net worth component variables. The central panel, labeled "Does item apply?", provides estimates of the percentage of HRS Wave 1 sample households (unweighted) that reported having each asset or liability (i.e., a nonzero amount value is assumed). For households that report owning a particular asset or having a particular type of debt, the right-hand panel of Table 1 describes the distribution of response types: actual value, bracketed value or missing data value. Clearly the use of bracketed response questions reduces the loss of information due to item nonresponse.

3. Model and Method for Multivariate Imputation of Net Worth Components

We assume in this and the following sections that conditional on the observed data, responses to specific HRS net worth variables are coarsened at random (CAR; Heitjan and Rubin, 1991). Therefore, we are concerned here with the data model and not the probability mechanism that generates the missing values or interval censoring of the multivariate data. We begin by defining the survey data:

$$\underset{\sim}{Y}_i = (Y_{i1},\ldots,Y_{ip}) = \textit{survey values of}$$

$$j = 1,\ldots p \quad \textit{net worth variables for the ith case};$$

$$\underset{\sim}{T}_i : \quad T_{ij} = \begin{cases} 1, & \textit{if } Y_{ij} > 0 \\ 0, & \textit{if } Y_{ij} = 0 \end{cases}$$

$$(\textit{assumed known here});$$

$$\underset{\sim}{R}_i : \quad R_{ij} = \begin{cases} 1, & \textit{if } Y_{ij} \textit{ observed} \\ 0, & \textit{if } Y_{ij} \textit{ missing}. \end{cases}$$

If $R_{ij} = 0$ (missing data) the bracketed response question sequence also provides upper and lower limits for the value of the missing amount.

One feature of the multivariate vector of net worth component measures that proves to be problematic in defining the data model is that the univariate distribution of each variable contains a significant point mass at the zero value. Tobin (1958) coined the label "limited variables" to describe observations of this type. Table 2 illustrates the multivariate nature of the problem by presenting the bivariate distribution of zero/non-zero values for two net worth components. We propose an extension of the general location model (Olkin & Tate, 1961) to such multivariate data with finite probability of zero values ("limited" data). To illustrate, we consider the *mixed normal model* for the bivariate case:

$$(\underset{\sim}{Z}_i | \underset{\sim}{T}_i \in s) \sim N_2(\mu_{\{s\}}, \Sigma_{\{s\}})$$

$$\underset{\sim}{Y}_i = \begin{cases} (Z_{i1}, Z_{i2}) & \textit{if } \underset{\sim}{T}_i = (1,1) \\ (Z_{i1}, 0) & \textit{if } \underset{\sim}{T}_i = (1,0) \\ (0, Z_{i2}) & \textit{if } \underset{\sim}{T}_i = (0,1) \\ (0,0) & \textit{if } \underset{\sim}{T}_i = (0,0) \end{cases}$$

We use {s} to denote sets of patterns defined by the vectors T. (There are four such patterns in the bivariate example of Table 2.)

Table 1
HRS Wave 1 Net Worth Components
Distribution of Responses by Response Type (n = 7607 respondent households)

HRS Wave 1 Household Asset or Liability	Does Item Apply?				If Item Applies To Household				
	Total	Yes	No	DK	Total n	Total %	Actual Value	Bracketed Value	Missing Value
Real Estate(not home)	100%	23.1%	76.3%	0.6%	1759	100%	74.7%	21.0%	3.4%
Vehicles, Pers Prop	100%	-	-	-	7607	100%	87.4%	10.6%	1.9%
Business	100%	16.1%	83.4%	0.5%	1226	100%	68.2%	26.0%	5.8%
IRA, Keogh	100%	37.1%	62.2%	0.7%	2825	100%	73.7%	21.5%	4.9%
Stock, Mutual Funds	100%	26.5%	72.6%	0.9%	2015	100%	67.4%	26.6%	6.0%
Checking, Savings	100%	77.5%	21.5%	1.0%	5895	100%	73.3%	21.5%	5.2%
CDS, Savings Bonds	100%	24.6%	74.3%	1.1%	1870	100%	70.6%	22.6%	6.8%
Bonds	100%	5.9%	93.2%	0.9%	445	100%	69.7%	19.8%	10.6%
Other Valuables	100%	15.0%	83.9%	1.1%	1143	100%	72.1%	21.6%	6.4%
Other Debt	100%	38.2%	60.8%	1.0%	2908	100%	90.4%	5.7%	3.9%
Home value	100%	68.0%	31.9%	0.1%	5176	100%	95.4%	1.2%	3.4%
1st Mortgage	100%	43.6%	56.0%	0.4%	3314	100%	92.4%	0.8%	6.8%

The complete data log likelihood for the mixed normal model is:

$$\ln L(\mu_{\{s\}}, \Sigma_{\{s\}} / \underset{\sim}{Z}, \underset{\sim}{T}) \propto$$
$$\sum_{i=1}^{n} -\frac{1}{2}\left\{\left[2\ln(2\pi)+\ln|\Sigma_{\{s\}}|\right] + (\underset{\sim}{Z}_i-\mu_{\{s\}})\Sigma_{\{s\}}^{-1}(\underset{\sim}{Z}_i-\mu_{\{s\}})^T\right\}$$

The general form of the mixed normal model permits the means and variances/covariances for all non-zero valued variables within each pattern set, {s}, to vary independently of those in other sets. Reduced forms of this model arise through various assumptions concerning common means and variances/covariances for net worth variables across different patterns of zero/non-zero value observations. The general location model is one such sub-model that assumes common variances/covariances across all patterns, $\Sigma^{(s)} = \Sigma$ (Olkin and Tate, 1961; Little and Schluchter, 1985; Little and Su, 1987; Heeringa, 1995; Schafer, 1996). The general location model cannot be applied directly to the Y's, since the assumption of a constant covariance matrix is untenable -- for example, the variance of a component Y_{ij} is zero when $T_{ij}=0$. This motivates the introduction of the Z's, with components Z_{ij} that are treated as missing in cells where $T_{ij}=0$ (see also Little and Su, 1987). The Z's can be legitimately modeled via the general location model, with means of the unobserved components of Z in cells with $T_{ij}=0$ constrained to avoid superfluous unidentified parameters. Maximum likelihood and Bayesian estimation and imputation under the mixed normal model for limited data treat the Z_{ij} as unobserved (missing) whenever Y_{ij} is zero, missing or known to lie inside a bracket.

A multivariate extension of Tobin's (1958) left-censored normal distributional model, which we term the truncated normal model, is an alternative for multivariate data with zero values. This model makes the same distributional assumptions for the Z's but different assumptions relating the Z's to the Y's, namely:

$$\underset{\sim}{Y}_i = \begin{cases} (Z_{i1}, Z_{i2}) \text{ if } Z_{i1}>0, Z_{i2}>0 \\ (Z_{i1}, 0) \text{ if } Z_{i1}>0, Z_{i2}\le 0 \\ (0, Z_{i2}) \text{ if } Z_{i1}\le 0, Z_{i2}>0 \\ (0,0) \text{ if } Z_{i1}\le 0, Z_{i2}\le 0 \end{cases}$$

This set of assumptions leads to truncated normal distributions for the non-zero assets, rather than normal distributions as in the mixed normal model. We do not discuss the truncated normal model further here since the mixed normal model for the log asset amounts provides a better fit to the HRS net worth data.

Our ongoing research is aimed at the development of ML and Bayes algorithms for the mixed normal models for the net worth data. Meanwhile, we have applied an existing multivariate imputation algorithm to the HRS Wave 1 net worth data. We describe here a general purpose algorithm that has been developed at the University of Michigan Survey Research Center for multivariate imputation of mixed categorical and continuous variables. Details of the algorithm and the SAS-based software program are described elsewhere (Raghunathan, 1997). The basic strategy is to create imputations through a sequence of univariate regressions. The type of regression model used depends on the variable (continuous, binary, multinomial) being imputed and the covariates include all other variables observed or imputed on that individual. The imputations are defined as draws from the predictive distribution specified by the

Table 2
HRS Wave 1 Net Worth Components
Patterns of Zero/Non-Zero Values
Business and Stock Values

Pattern	Business	Stock	n	%
1	>0	>0	496	6.5
2	>0	0	717	9.4
3	0	>0	1532	20.2
4	0	0	4862	63.9
Total	-	-	7607	100.0

above-mentioned regression model with a noninformative prior for the regression parameters. The iterative algorithm will produce correct draws from the Bayes' predictive posterior distribution of the missing data under the general location model when the mixed categorical and truly continuous data conform to a pattern of missing data where the categorical variables are more observed than the continuous. Unmodified application of the algorithm to the net worth component variables with their "limited" zero/non-zero continuous values can produce only approximate draws from the true Bayes' posterior under the mixed normal model described in this section.

The program provides for variables (like assets and liabilities) that have a finite probability of being zero and a continuous distribution for cases that have non-zero values. These variables are declared to be of mixed type and a two-phase logistic/normal regression procedure is used in imputation. Here, the zero amounts are presumed to be known constants and not reimputed. Only the normal regression estimation cycle of the iterative algorithm is used to estimate the conditional distribution of the non-zero amounts. Imputation of missing and bracketed amounts is performed by taking random draws from the estimated conditional distributions. The algorithm is adapted to take into account bracketing information and conditions on the information (complete or incomplete) in other asset items and hence is truly multivariate.

4. An Empirical Comparison of Alternative Imputation Methods

The HRS Wave 1 public use data set includes imputations of holding status and missing amounts for 23 net worth variables. The imputation of holding was not repeated for this empirical exercise but fixed at the actual and imputed values in the public use data set. Likewise, the values of 11 of the 13 non-bracketed housing equity and debt items were not reimputed in this exercise but were fixed at the actual values or hot deck imputations that are contained in the HRS Wave 1 public use data set. The remaining 12 major net worth components listed in Table 1 were reimputed using two deterministic imputation methods and two stochastic imputation approaches. Each imputation method made full use of the bracketing information that was available. Results presented in this paper are based on imputations performed on the natural log transformation of the positive values for each of the net worth component variables. To assess the uncertainty associated with the imputation of the bracketed and missing data, the two stochastic imputation processes were independently replicated a total of 16 times. At the completion of each imputation pass for the 12 selected variables, the total net worth for each household was calculated by computing the appropriate sums (assets) and differences (liabilities) of its 23 net worth components.

The following are brief descriptions of the alternative imputation methods that are examined in this empirical comparison.

1) Complete Case Analysis: A total of 4566 (60.0%) of the 7607 cooperating HRS Wave 1 households provided complete information on holding and amounts for each of the 23 asset and liability components. (Zero values count as non-missing values.) Using only these 4566 complete cases, estimates of the distribution of total household net worth can be computed and compared to the distributions that result from the imputation methods.

2) Mean/Median Substitution: Substitution of the observed mean or median value of a characteristic has also been used as a simple method of imputing missing data in continuous variables. A modification of this method that can be applied to univariate imputation of the HRS net worth variables is to impute the mean or median of observed values that fall within the known bounds for bracketed observations. If an observation was completely missing (no bracket information), the overall mean or median of the observed cases was imputed to the case.

3) Hot Deck Method: Each asset and liability component is imputed separately by classifying observed and missing cases into adjustment cells based on the available bracketing information and covariate information including the age, race, sex and marital status of the household head. Each missing value is then imputed using the value of a randomly selected observed case within the same hot deck cell.

4) Approximate Bayes' method: This method creates 16 multiple imputes of the missing values using the approximate Bayes' algorithm (Section 3) applied to the multivariate distribution of the 12 net worth components that were reimputed in this exercise. Three variations of the multivariate imputation method are applied: (i) an essentially unrestricted version that limits the imputed log-amount for cases to be no more than 18.42 (the log-transformed equivalent of $99,999,999); (ii) a restricted version that limits the imputed log-amount for cases in the highest bracket to be no greater than the highest observed value in the data set; and (iii) a version applied to the data ignoring bracketing information.

5. Results from Application of Imputation Methods to the HRS Wave 1 Net Worth Data

Table 3 presents results of the empirical comparison of descriptive analysis of household net worth based on HRS Wave 1 complete cases and data sets imputed using the four imputation methods described in Section 4. Each analysis computes the population weighted estimate of the mean, standard deviation, median and other selected quantiles of the distribution of net worth in U.S. households represented by the HRS Wave 1 sample. The estimates of these descriptive statistics were computed using SAS PROC UNIVARIATE (SAS, 1990). Population weighted estimates of the quantiles of the net worth distribution were obtained by specifying the integer-valued HRS analysis weight variable as the parameter of the FREQ keyword of the UNIVARIATE procedure. Distributional statistics for the hot deck and approximate Bayes' method are the multiple imputation estimates (averages) based on 16 independent replications. The standard errors reported for the distributional statistics are correct standard errors estimated using the Jackknife Repeated Replications (JRR) method that reflect the influences of the analysis weighting as well as the stratification and clustering of the complex multistage sample design for the HRS. Table 3 also includes estimates of the imputation variance for estimates based on the stochastic hot deck and approximate Bayes' imputation methods. The proportional increase in variance from imputation uncertainty is estimated and is included in the columns of Table 3 labeled the "imputation effect."

Complete case analysis: Complete case analysis appears to markedly underestimate the distribution of household net worth for HRS households, yielding lower estimates of the overall mean, standard deviation and percentile values. The apparent underestimation that occurs in complete case analysis can be explained by the fact that HRS respondents' propensity to use brackets increases with the value of the net worth component.

Mean/Median substitution methods: Compared to stochastic imputation alternatives, mean and median value substitution imputation methods also appear to lead to underestimation of the mean and percentiles of the full net worth distribution. The standard deviation of the imputed household net worth distribution produced by these deterministic imputation methods is attenuated compared to standard deviation in net worth amounts imputed by the stochastic hot deck and approximate Bayes' alternatives.

Approximate Bayes' Algorithm with Restrictions vs. the Hot Deck Method: Since the hot deck method can never impute an asset value greater than the largest observed value, the closest comparison is between the hot deck method and the "restricted" application of the approximate Bayes' algorithm. Comparing results for these two methods in column sets (4) and (6) of Table 3, there is very little difference in the imputed distributions of household net worth. There is slight evidence that the hot deck method results in some shrinkage from the tails of the distribution to the overall mean. The explanation for this minor difference in results may lie in how bracketed and completely missing data are treated under the two imputation alternatives. In cases of completely missing data -- from 2% to 12% of all cases depending on the net worth component, the hot deck method performs a random draw from the full empirical distribution of observed cases. The expected value of these draws is the overall mean of the observed amounts for the variables. The corresponding treatment of completely missing amounts under the multivariate approximate Bayes' procedure is to make draws from the estimated conditional distribution of the variable in question. The expectation for these draws is not the overall mean but is conditional on the observed or imputed values of other net worth variables.

Approximate Bayes' Without Restrictions: Relaxing the restrictions on the maximum values that may be imputed by the multivariate algorithm has the expected effect. From column (5) of Table 3, the distribution based on unrestricted net worth component imputations has a higher estimated mean, standard deviation and maximum value. Relaxing the restrictions on the possible ranges for imputed values produces far greater instability in the estimates of these statistics. The estimated quantiles Q25, Q50, Q75, Q90, Q95 and even Q99 are very similar for the restricted and unrestricted imputations.

Approximate Bayes' Ignoring the Bracketing Information: The importance of the bracketing information to the multivariate imputation of net worth components is clearly seen in a Column (7) of Table 3. The mean, median and other quantiles of the HRS net worth distribution appear to be seriously underestimated in comparison to the imputation approaches that use the bracketing information. In fact the imputed distribution bears a close resemblance to that estimated from complete cases alone.

We explain this result as follows. Lacking bracketing information, the approximate Bayes' imputations are unconstrained draws from a conditional distribution defined by regression of the variable to be imputed on the remaining components and covariates. The explained variance (R^2) of the regression that defines this conditional distribution for each individual component can be very low in which case the expected value of the conditional posterior will be very close to the overall mean of the variable for the observed cases. The algorithm therefore converges to a joint distribution that

Table 3

HRS Wave 1 Net Worth Imputations

Estimated Distribution for Total Household Net Worth Under Imputation Alternatives

	Complete Data		Mean Substitution		Median Substitution		Hot Deck Method			A-Bayes' Method--Unrestricted			A-Bayes' Method--Restricted			A-Bayes' Method--No Brackets		
	1		2		3		4			5			6			7		
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	MI Estimate	MI Std. Error	Imp. Effect	MI Estimate	MI Std. Error	Imp. Effect	MI Estimate	MI Std. Error	Imp. Effect	MI Estimate	MI Std. Error	Imp. Effect
N	4,566		7,607		7,607		7,607		7.607	-			7,607			7,607		
Mean	186,825	9,146	213,538	7,661	195,850	7,280	232,480	9,421	1.202	249,536	13,315	1.41	240,279	9,450	1.15	191,314	7,711	1.031
Std Dev	417,288	27,468	443,665	29,399	412,971	29,308	491,073	42,836	1.746	695,735	227,289	2.56	526,446	41,325	1.36	421,807	30,841	1.179
Q25	15,300	2,787	28,398	2,024	27,800	2,847	29,515	2,440	1.023	28,381	2,140	1.012	28,442	2,117	1.008	23,928	2,262	1.020
Q50	78,000	3,254	97,314	4,393	90,220	2,339	100,847	5,023	1.013	99,393	4,634	1.013	99,276	4,603	1.017	86,318	3,631	1.021
Q75	195,500	10,018	218,000	7,021	203,330	6,706	240,443	8,606	1.039	241,097	9,954	1.029	241,066	9,816	1.029	202,933	8,306	1.020
Q90	408,500	15,312	471,571	23,578	424,000	29,715	514,968	30,895	1.021	538,814	23,815	1.039	537,546	24,447	1.038	416,706	21,510	1.038
Q95	663,000	27,982	779,561	33,155	697,275	27,923	839,781	56,599	1.137	897,342	59,209	1.073	895,904	59,662	1.086	657,020	27,643	1.047
Q99	1,995,000	107,074	2,142,131	62,035	1,837,000	133,056	2,317,818	215,996	1.405	2,576,533	246,483	1.360	2,505,904	229,494	1.382	1,857,447	261,327	1.065
Max	6,202,000	322,000	9,096,500	469,445	8,632,275	402,275	9,644,756	3,070,000	1.915	24,458,054	15,700,000	8.803	10,907,357	2,951,000	1.485	8,844,044	2,005,492	1.025

resembles the complete data and despite its iterative nature is incapable of correcting the bias in the imputation since it is not supplied the most informative "predictors" -- the bracketing information.

Imputation variance: Imputation variance effects for most distributional statistics are relatively modest. For example, the estimated increase due to imputations (that reflect bracketing information) in variance of estimates of the 25th percentile ranges from .8% - 2.3%. Imputation variances for estimates of percentiles increase gradually moving from left to right in the distribution. Total variance for imputed estimates of the 95th percentile amounts are from 8.6% to 13.7% greater than the estimated variances that ignore the imputation uncertainty. Imputation variance effects are clearly greatest for the estimated means, standard deviation, Q99 and maximum value -- those statistics that are influenced by the uncertainty in imputations for values in the upper tails of the component asset and liability distributions.

6. Summary and Conclusions

The results of this empirical comparison show that complete case analysis and simple deterministic imputation by mean or median substitution appear to result in a serious underestimation of the mean and percentiles of the distribution of household net worth. If a simple methodology is needed, a more sensible approach would be to use the univariate hot deck method.

The estimated distributions of net worth that result from hot deck univariate imputation and approximate Bayes' multivariate imputation of bracketed and missing values for the net worth component variables are very similar. The similarity of the distributions that result from these two very different imputation methods can be explained by the fact that each makes efficient use of the highly informative bracketing information when this information is available. The results of the empirical comparison suggest that the hot deck method may produce a small amount of shrinkage of the distribution -- a likely result of the weak predictive power of the univariate hot deck method when no bracketing information is available. In contrast, the multivariate method draws predictive strength from the other variables to impute completely missing observations.

The importance of the bracketing information to imputation and estimation of composite household net worth is clearly seen throughout this empirical comparison. When bracketing information is available for some but not all missing data for the net worth component variables, the choice of a stochastic imputation method may depend more on how completely missing data are imputed than in the handling of imputations for bracketed missing values. Here the multivariate imputation method has the advantage since it makes full use of the information (including bracketing) that has been provided for the conditioning variables that define the predictive distribution for the imputation draws.

Bibliography

Curtin, R.T., Juster, F.T., & Morgan, J.N. (1989). "Survey estimates of wealth: An assessment of quality," *in* R.E. Lipsig & H.S. Tiu (eds.), *The Measurement of Saving, Investment and Wealth.* Chicago: The University of Chicago Press.

Heeringa, S.G. (1993). "Imputation of item missing data in the Health and Retirement Survey." Invited paper. *Proceedings of the Survey Methods Section, American Statistical Association*, pp. 107-116.

Heeringa, S.G. (1995). "Application of generalized iterative Bayesian simulation methods to estimation and inference for coarsened household income and asset data." Invited paper. *Proceedings of the Section on Survey Methods,* American Statistical Association, pp. 42-51.

Heitjan, D.F., & Rubin, D.B. (1991). "Ignorability and coarse data," *The Annals of Statistics*, Vol. 19, No. 4, pp. 2244-2253.

Juster, F.T., & Smith, J.P. (1994). "Improving the quality of economic data: Lessons from the HRS." HRS Working Paper Series #94-027, presented at NBER Summer Institute on Health and Aging. Cambridge (MA), July, 1994.

Kennickell, A.B. (1996), "Using range techniques with CAPI in the 1995 Survey of Consumer Finances." To appear in *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Little, R.J.A. (1976). "Inference about means from incomplete multivariate data," *Biometrika*, Vol. 63, pp. 593-604.

Little, R.J.A. (1993). "Pattern-mixture models for multivariate incomplete data," *Journal of the American Statistical Society*, Vol. 88, No. 421, pp. 125-134.

Little, R.J.A., & Schluchter, M.D. (1985). "Maximum likelihood estimation for mixed continuous and categorical data with missing values," *Biometrika*, Vol. 72, No. 3, pp. 492-512.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* Wiley: New York.

Little, R.J.A., & Su, H.L. (1987), "Missing data adjustments for partially scaled variables," *Proceedings of the Section on Survey Research Methods,* American Statistical Association, pp. 644-649.

Olkin, i., and Tate, R.F. (1961). "Multivariate correlation models with mixed discrete and continuous variables," *Annals of Mathematical Statistics*, Vol. 32, pp. 448-465.

Raghunathan, T.E., Lepkowski, J., Van Hoewyk, J., and Solenberger, P. (1997), "A multivariate technique for imputing missing values using a sequence of regression models." Technical Report, Survey Methodology Program, Survey Research Center, ISR, University of Michigan.

Rubin, D.B. (1976). "Inference and missing data (with discussion)," *Biometrika*, Vol. 63, pp. 581-592.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

SAS Institute, Inc. (1990), *SAS® Procedures Guide,* pp. 617-634.

Schafer, J.L. (1996). *Analysis of Incomplete Multivariate Data.* London: Chapman and Hall.

Tobin, J. (1958). "Estimation of relationships for limited dependent variables," *Econometrica*, Vol. 26, pp. 24-36.