SAMPLE SELECTION BIAS: CORRECTING FOR VARIABLES THAT AREN'T THERE

Cathleen D. Zick, University of Utah[1]

ABSTRACT

The potential for sample selection bias threatens both the internal and external validity of consumer research. This paper summarizes the statistical problems associated with indirectly censored samples, outlines the two-equation system that is used to solve these problems, and works through a consumer example of correcting for sample selection bias.

Statistical problems frequently cross disciplinary boundaries. Sociologists, economists, psychologists, and political scientists all face problems of assessing measurement reliability, deciding on appropriate levels of Type I and Type II errors, and determining the generalizability of their empirical results. Yet when a researcher finds a solution to a unique statistical problem, quite often it remains the "private property" of the researcher's field simply because of poor interdisciplinary communication. Sample censoring is an example of such disciplinary provincialism.

Technically, sample censoring occurs whenever some of the observations for the dependent variable that correspond to known values of the independent variables are unobservable. General censoring can be further categorized into two types: direct (or explicit) censoring and indirect (or incidental) censoring [2, p. 388]. Direct censoring occurs whenever observations are excluded because the value of the dependent variable is below some threshold. For example, suppose a researcher wants to examine what factors influence a household's expenditures on durable goods. If households that do not purchase any durables during the period in question are excluded from the sample then direct censoring has occurred. Corrections for direct censoring can be handled in a Tobit framework and those who would like to know more about Tobit are urged to read Jean Kinsey's paper [19]. Indirect censoring occurs whenever the selection into a sample is based on the value of an exogeneous variable that does not play a role in the relationship of concern to the researcher. The statistical problems associated with this second type of censoring and their solutions are the focus of this paper.

[1]Assistant Professor, Department of Family and Consumer Studies. The author wishes to thank Robert N. Mayer and Ken R. Smith for their helpful comments. Data used in the analysis were collected by the University of Utah Survey Center during November and December 1984. The author bears exclusive responsibility for any errors or ambiguities that remain in this work.

THE ISSUE

Recently, economists have devoted a great deal of energy to the issue of censored samples. Yet researchers in other fields by and large continue to ignore the statistical problems presented by these nonrandom sample selection criteria. In the economics literature, the most frequently cited example of incidental sample censoring (or sample selection bias as it is sometimes called) concerns the estimation of market wage rates for married women. When a woman is employed outside of the home the researcher can observe her market wage rate. However, what happens if the woman is a full-time homemaker? In that case, observations on the variable of interest (i.e., the wage rate that the homemaker could generate in the marketplace) are not available. Deletion of these missing observations from the analysis result in sample censoring because the factors that determine a woman's labor supply are for the most part the same factors that influence the market wage she commands.

Other examples of sample censoring are more subtle. Consider the researcher who is interested in identifying how drinking habits affect the number of traffic citations received over the preceding five years. If its the case that people who drink heavily are also more likely to be in fatal traffic accidents, then the researcher will have a problem with incidental censoring. The researcher will not be able to observe the number of traffic citations for those individuals who died in traffic accidents and thus the available sample will be censored.

Another case of censoring arises for the researcher interested in identifying the impact of various life cycle factors on desired housing expenditures. Most models view desired housing expenditures to be a function of a series of life cycle factors that are known for all households [4, 8, 24, 28]. However, suppose the researcher only has data on desired expenditures for those households that own their home. Renters have been systematically excluded because they have missing data on desired housing expenditures. And their exclusion corresponds to known values for life cycle characteristics that are associated with both housing tenure and desired level of housing expenditures.

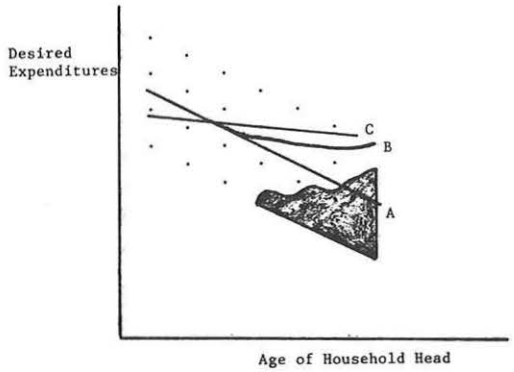
STATISTICAL PROBLEMS CAUSED BY SAMPLE CENSORING

Two statistical problems arise from sample censoring if the researcher is using a regression framework to analyze data. The first is that regression coefficients estimated with censored data systematically misstate the true slope of the

regression line for the larger, uncensored population (e.g., the owners and renters). Most researchers recognize this limitation. Indeed, it is quite common to see caveats about the external validity of the empirical work done with restricted samples. For instance, in the example described above, the researcher might preface the empirical findings by noting that the results show how various factors affect desired expenditures given that the respondent owns a home.

The second, less frequently recognized statistical problem associated with sample censoring is that the estimated regression coefficients are in all likelihood biased. This point can be best illustrated with the help of Figure 1. Suppose a researcher has data on desired durable expenditures for those respondents that recently purchased a durable. Furthermore, a respondent's desired expenditure level depends on his/her age. The relationship between desired expenditures and age is plotted in Figure 1. This diagram indicates that older consumers would like to spend less on durables than younger consumers. In addition, older consumers are less likely to have made a recent durable purchase. Thus observations on the dependent variable are simply missing for many older consumers. These consumers are then excluded from the sample. Their exclusion is represented by the shaded area.

Figure 1[a]



[a]Adapted from a graphical presentation done by Berk [2, p. 387].

If all of the data had been available, the fitted regression line between desired expenditures and age would have been line A. However, with the systematic exclusion of some data points from the sample, the new relationship between desired expenditures and age is nonlinear and represented by curve B. It becomes clear that any attempt to fit a straight line to this nonlinear relationship will produce some systematic specification error. Suppose that the researcher fits line C with least squares regression using the censored data. A visual inspection indicates that for low values of age the residuals are positive and for high values the residuals are negative. In this example, respondent's age is negatively correlated with the error term and as a result, causal effects are mistakenly attributed to the respondent's age when they are in fact the result of random error. In this situation the estimated age coefficient (as well as the estimated intercept term) will be biased [20, p.393] and therefore the researcher cannot treat the results as representative of even the restricted subsample of of consumers who made a recent durable purchase.

In sum, indirect sample censoring threatens the internal as well as the external validity of a researcher's empirical work. It is difficult to ascertain the extent of the bias (if any) that sample censoring introduces into an analysis especially when a researcher is dealing with a multivariate model. However, the first step is for the researcher to recognize when the possibility of biased estimates due to sample censoring arises.

Incidental censoring can occur for a variety of reasons that in most instances are beyond the researcher's immediate control. (See Richard Berk's 1983 article [2] for an good discussion of the various types of censoring.) For instance, wage data are simply not available for women who are not employed outside of the home. Similarly, desired expenditure data may only be collected from a sample of individuals who have made recent purchases.

Berk [2, p.392] points out that sample censoring is frequently an inadvertant result of the manner in which a questionnaire is constructed. For instance, the researcher interested in variations in complaining behavior may first screen respondents by asking them if they have complained about a product to a state agency recently. Once those who respond in the negative have been filtered out, then the researcher asks the respondent among other things, how much total time has he/she spent trying to obtain redress for the faulty product. This type of questionnaire construction is desirable in the sense that it may make it easier for some respondents to answer the questions (i.e., respondents no longer have to read through a series of questions that do not apply to their specific situations). Yet at the same time such filtering techniques can create the potential for incidental censoring.

STATISTICAL SOLUTIONS TO THE CENSORING ISSUE

During the last ten years statistical techniques have been developed by two econometricians, James Heckman [13, 15] and Tekeshi Amemiya [1], which allow a researcher to correct for possible sample censoring. The basic goal of both techniques is to develop an instrumental variable that when included in the regression equation, corrects for the biases introduced by sample censoring. The Heckman technique is discussed here because it is the more tractable one.

Many formal presentations of the correction for incidental censoring are available elsewhere [6, 15, 18], and what is presented here is no substitute for these more technical discussions.

Rather, what follows is meant to give the reader a basic understanding of how to correct for selectivity bias. This presentation draws heavily on Heckman's 1979 work [15].

CORRECTING FOR INDIRECT SAMPLE CENSORING

The procedure begins by thinking of the sample selection issue as a two equation system:

$$Y_{1i} = X_{1i}B_1 + u_{1i} \qquad i=1,\ldots N \qquad (1)$$

$$Y_{2i} = X_{2i}B_2 + u_{2i} \qquad i=1,\ldots N \qquad (2)$$

where,

Y_{1i} = the endogenous variable of interest (e.g., desired durable expenditures).

X_{1i} = a vector of exogenous variables that are posited to affect Y_{1i} (e.g., age, number of family members, etc.).

Y_{2i} = an endogenous variable that operates as the selection criterion for observation of Y_{1i} (e.g., the consumer has recently purchased a major durable).

X_{2i} = a vector of exogenous variables that are posited to affect the selection criterion. Some, if not all of these variables also appear in the X_{1i} vector (e.g., income affects both the probability of making a purchase and a consumer's desired expenditure).

Furthermore, the error structure of equations (1) and (2), is assumed to be as follows (where $E(\cdot)$ is used to denote the expected value):

$$E(u_{ji}) = 0 \qquad j=1,2 \quad i=1,\ldots N \qquad (3)$$

$$E(u_{ji}^2) = \sigma_j^2 \qquad j=1,2 \quad i=1,\ldots N \qquad (4)$$

$$E(u_{1i}u_{2i}) = \sigma_{12} \qquad i=1,\ldots N \qquad (5)$$

Thus, the expected value of the disturbance term in each equation is zero, but the disturbances are correlated across equations.[2] This covariance link across equations (σ_{12}) indicates that the dependent variable in equation (1) (e.g., desired expenditures) is affected by the dependent variable in equation (2) (e.g., whether or not the consumer makes a purchase). Thus, equation (5) embodies the concept of sample censoring in this mathematical formulation.

Given these assumptions, the population regression function for the equation of interest, equation (1), can be written,

$$E(Y_{1i} \mid X_{1i}) = X_{1i}B_1 \qquad i=1,\ldots N \qquad (6)$$

However, Heckman shows [15] that the equation for the censored subsample of available data is,

$$E(Y_{1i} \mid X_{1i}, \text{ selection rule}) = X_{1i}B_1 + E(u_{1i} \mid \text{selection rule}) \qquad i=1,\ldots C \qquad (7)$$

where,

C = N - the observations lost to sample censoring.

Note, if the censoring rule mandates that the dependent variable in equation (1) be observed only if the dependent variable in equation (2) is greater than zero (the situation one has when a desired expenditures question is asked only of those who have made recent purchases), equation (7) can be written as:

$$E(Y_{1i} \mid X_{1i}, Y_{2i}>0) = X_{1i}B_1 + E(u_{1i} \mid u_{2i}>-X_{2i}B_2) \qquad (8)$$

where,

$$E(u_{1i} \mid u_{2i}>-X_{2i}B_2) = \left[\frac{\sigma_{12}}{(\sigma_2^2)^{\frac{1}{2}}}\right]\lambda_i \qquad (9)$$

and,

$$\lambda_i = f\left[\frac{-X_{2i}B_2}{(\sigma_2^2)^{\frac{1}{2}}}\right] \Big/ \left(1 - F\left[\frac{-X_{2i}B_2}{(\sigma_2^2)^{\frac{1}{2}}}\right]\right) \qquad (10)$$

In equation (10), $f(\cdot)$ represents the probability density function of $-X_{2i}B_i/(\sigma_2^2)^{\frac{1}{2}}$, which is assumed to be normally distributed. Correspondingly, $F(\cdot)$ represents the cumulative density function of $-X_{2i}B_2/(\sigma_2^2)^{\frac{1}{2}}$, which is also assumed to be normally distributed. With this information, equation (8) can be rewritten as

$$E(Y_{1i} \mid X_{1i}; Y_{2i}>0) = X_{1i}B_1 + \left[\frac{\sigma_{12}}{(\sigma_2^2)^{\frac{1}{2}}}\right]\lambda_i \qquad (11)$$

Lambda (λ) is called the inverse of the Mill's Ratio or what is more commonly known as a hazard rate. It represents the probability that an observation is excluded from the sample of interest conditional on the larger, uncensored sample at risk [27, pp. 8-10].

From this formulation, it is apparent that Heckman has translated the problem of censoring observations on the dependent variable into one of correcting for an omitted independent variable. Lambda is the omitted variable that if included as a regressor in equation (1), satisfies the condition that the independent variables and the error term be uncorrelated. Thus, a researcher can avoid biased regression estimates when using

---

[2] Under this specification the residuals of both equations are assumed to be distributed bivariate normally. Other distributions could be specified however the reader should recognize that this would change the specifics of the correction term that Heckman derives.

censored data by including lambda as one of the regressors.[3]

AN EXAMPLE OF CORRECTING FOR SAMPLE CENSORING

Consider how the wording of the following survey question might result in sample censoring. During November/December 1984, the University of Utah Survey asked a random sample of Utahns the following question:

> How often would you say you wear seat belts when you are the driver of the car?

Out of 809 individuals aged 18 years or older who responded to this question, approximately 5% said that they did not drive and therefore could not adequately answer the question. A researcher interested in the determinants of seat belt usage would certainly be tempted to delete the small percentage of nondrivers from the sample and proceed to estimate a regression model on the censored subsample that remains. To explore the extent of the bias that such a strategy would introduce, a seat belt usage model is estimated here with and without correcting for the censoring of nondrivers.

From the preceding section, it is evident that correcting for selectivity bias involves several steps. First, one needs to obtain estimates of the hazard rate (λ) associated with driving. To do this, a probit equation,[4] in the spirit of equation (2) is estimated.[4] The dependent variable in this equation is the respondent's driving status (i.e., the respondent either drives or does not drive). It is posited that whether or not an individual drives is a function of the respondent's sex, marital status, employment status, residential location, age, and education.[5] The results of this probit equation appear in Table 1.

TABLE 1. Parameter Estimates of the Probit Equation for Respondent's Driving Status[a] (t-statistics in parentheses).

Variable	Coefficients
Constant	2.91 (3.43)*
Respondent's Sex (1=female; 0=male)	-.668 (-2.21)*
Respondent's Age (in years)	-.0311 (-4.67)*
Respondent's Marital Status (1=married; 0=otherwise)	.459 (1.95)
Urban/Rural Residence (1=urban; 0=rural	-.300 (-1.31)
Respondent's Employment Status (1=employed; 0=otherwise)	.489 (1.84)
Respondent's Education (in years)	.0410 (.871)

$\chi^2 = 86.24$**
N = 809

*Statistically significant at the .05 level.

**The critical value for $\chi^2_{(\alpha=.01,\ \upsilon=6)}=16.8$

[a]The dependent variable takes on a value of 1, if the respondent drives and 0 if the respondent does not drive.

Once the probit equation is estimated, the next step is to use the coefficients from this equation to calculate λ_i (i.e., the probability that a respondent will not drive, given the original sample), for each sample respondent using the specification given by equation (11).[6] Having completed this calculation, λ_i can now be entered as an independent variable in the substantive equation of interest (the seat belt usage equation). In this illustrative exercise, λ_i is entered as an independent variable along with age, education, religiousity, political orientation, gender, marital status, income, and health insurance coverage,[7] in the equation that estimates seat belt usage.

---

[3]Heckman uses the specification of equation (11) to note the two instances where there is no need to correct for selectivity bias [14, p.13]. First, if $\sigma_{12} = 0$. That is, if the disturbances that affect sample selection are independent of the disturbances that affect the substantive equation, then least squares parameter estimates of equation (1) are unbiased. Second, if $\lambda = 0$, that is, if the sample selection criterion allows all population observations to have an equal chance of being included in the sample then again, least squares estimates of equation (1) are unbaised. A researcher who chooses not to correct for sample selection bias is thus implicitly (if not explicitly) making one or both of these assumptions.

[4]Hazard rates can be calculated from a variety of models including linear probability models, logistic models, and probit models. The choice of a model and a corresponding estimation technique is decided by the assumptions one makes about the error structures of the equations. (See Hozencik [17] and Kinsey [19] for discussions of the error structure assumptions in some of the qualitative choice estimation techniques.) In this presentation it has been assumed that the error terms for equations (1) and (2) are distributed bivariate normal. This assumption is consistent with the probit estimation technique [15] and that is why probit is used here.

[5]Since this empirical work is done strictly for illustrative purposes, justification for the inclusion of these independent variables in the probit equation is omitted.

[6]Several computer software packages contain algorithms that with some adaptation can be used to calculate lambda. One of the more popular packages that has this capability is GLIM.

[7]Again, since the empirical work is being done for illustrative purposes, no justification for the

To assess the importance of correcting for selectivity bias in this illustration, the seat belt usage equation is also estimated without correcting for incidental censoring. The results for both equations are presented in Table 2.

TABLE 2. Ordinary Least Squares Parameter Estimates for the Seat Belt Use Equations [a] (t-statistics in parentheses).

Independent Variable	Uncorrected Coefficients	Corrected Coefficients
Constant	-110 (-15.5)*	-108 (-15.0)*
Annual household income (in $)	8.41×10^{-5} (.110)	-2.00×10^{-6} (-.0379)
Respondent's sex (1=female; 0=male)	-1.04 (-.524)	-1.50 (-.682)
Respondent's marital status (1=married; 0=otherwise)	-.971 (-.409)	-.250 (-.0951)
Respondent's education (in years)	2.38 (5.24)*	2.54 (5.50)*
Respondent's age (in years)	-.0142 (-.266)	-.130 (-1.41)
Respondent's religious involvement (1=active; 0=otherwise)	9.00 (4.13)*	9.20 (4.27)*
Respondent's health insurance (1=has insurance; 0=otherwise)	10.6 (3.19)*	11.8 (3.51)*
Respondent's political orientation (1=liberal; 0=otherwise) [b]	4.96 (1.88)	5.36 (2.06)*
Respondent's political orientation (1=middle-of-the-road; 0=otherwise) [b]	-3.44 (-1.46)	-2.84 (-1.22)
Lambda (λ)		15.3 (1.01)
R^2	.095	.11
F	10.1**	9.58**

*Statistically significant at the .05 level.

**The critical value for $F_{(\alpha=.01;\ 9,744)} = 2.41$.

[a] The dependent variable is defined as the percentage of time that a respondent wears a seat belt when driving.

[b] The omitted group in this analysis consists of those respondents who classified themselves as political conservatives.

At first glance the differences in the coefficients do not appear particularly dramatic (only one variable is statistically significant in the corrected equation that is not statistically significant in the uncorrected equation and none of the coefficients change much in magnitude).

However, the case for correction becomes more compelling if one recalls that a mere 5% of the sample is censored. Intuitively, one would think that the potential for biased coefficients would be positively correlated with the percentage of the sample that is censored. Thus, one would expect to see only negligible differences in the estimates associated with 5% censoring. The fact that the estimates change as much as they do must reflect a fairly strong correlation between the error terms in the selection equation and the substantive equation of interest.

Selection bias in the seat belt usage equation is further supported by the fact that the tests of statistical significance improve for six of the nine corrected coefficients. Kmenta [20, p.394] notes that tests of significance err on the conservative side when a relevant explanatory variable is omitted from the specification of a regression equation. The improvement in the levels of statistical significance in the corrected equation again argues for the need to correct for sample censoring in this example.

CONCLUSIONS

Economists have begun to use Heckman's procedure to correct for incidental sample censoring in a variety of situations. Indeed, corrections are now automatically made in the estimation of wage equations for groups like married women [5, 6, 12, 15]. Furthermore, the economic applications are spreading. Lee and Trost [21] correct for sample censoring in their investigation of housing demand. Duncan and Hoffman [7] correct for selectivity bias in estimating the economic gains from remarriage for divorced women. In both instances these economists found fairly dramatic differences in the estimated coefficients after correcting for incidental censoring.

Unfortunately, recognition of the statistical problem created by sample censoring and its remedies appears to be slow in spreading to other disciplines. There are a few exceptions. In psychology, Glass, McLanahan and Sorensen [11] find significant differences in the estimated psychological consequences of divorce before and after correcting for the selectivity bias introduced by divorcing. Within the realm of sociology, Berk and Shih [3] find substantial differences in citizens' evaluations of the criminal justice system after correcting for non-response to their questionnaire.

The analysis of many consumer issues might be hindered by censoring problems because consumer research often involves the use of nonrandom subsamples. Consumer researchers using nonrandom samples should begin their empirical work by investigating the question of whether sample censoring is a relevant issue. If a selection process can be modeled and if there is reason to believe that the correlation between the error terms in the resulting selection and substantive equations is nonzero, correction for selectivity bias should be made using the Heckman procedure.

---

list of the independent variables is presented here. Readers who are interested in knowing more about the determinants of seat belt usage are referred to [9, 10, 16, 22, 25].

It is important that researchers in the consumer field alert themselves to the possible problems associated with censoring and the steps that can be taken to correct those problems. By correcting for sample selection bias when it is appropriate, consumer researchers can make marked improvements in both the internal and the external validity of their empirical work.

REFERENCES

1. Amemiya, Takeshi. "Multivariate Regression and Simultaneous Equation Models When the Dependent Variable is Truncated Normal." Econometrica 42:999-1011.

2. Berk, Richard. "An Introduction to Sample Selection Bias in Sociological Data." American Sociological Review 48:386-398.

3. Berk, Richard and Anthony Shih. "Measuring Citizen Evaluations of the Criminal Justice System: A Final Report to the National Institute of Justice." Santa Barbara, CA:SPRI, 1982.

4. Bryant W. Keith. "A Portfolio Analysis of Rural Wage-Working Families' Assets and Debts." CEH Working Paper, September 1980.

5. Cogan, John. "Married Women's Labor Supply: A Comparison of Alternative Estimation Procedures," in Female Labor Supply: Theory and Estimation. Edited by James P. Smith. Princeton: Princeton University Press, 1980.

6. Duncan, Greg J. and Mary E. Corcoran. "Do Women 'Deserve' to Earn Less than Men?" in Years of Poverty, Years of Plenty. Edited by Greg J. Duncan. Ann Arbor: Survey Research Cetner, 1984.

7. Duncan, Greg J. and Saul Hoffman. "A Reconsideration of the Economic Consequences of Marital Dissolution." Institute for Social Research. Draft Manuscript, August 1984.

8. Dunkelberg, William and Frank P. Safford. "Debt in the Consumer Portfolio: Evidence From a Panel Study." American Economic Review 61:598-613.

9. Fhaner, Gurilla and Monica Hane. "Seat Belts: Factors Influencing Their Use: A Literature Survey." Accident Analysis and Prevention 5:27-43.

10. Fhaner, Gurilla and Monica Hane. "Seat Belts: Relations Between Beliefs, Attitude and Use." Journal of Applied Psychology 59:472-82.

11. Glass, Irene, Sara McLanahan, and Aage Sorensen. "Psychological Adjustment to Divorce." in Life Course Dynamics. Edited by Glen E. Elder Jr. Ithaca: Cornell University Press, forthcoming, May 1985.

12. Gerner, Jennifer L. and Cathleen D. Zick. "Time Allocation Decisions in Two-Parent Families." Home Economics Research Journal 12:145-158.

13. Heckman, James. "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models." Annals of Economic and Social Measurement 5:475-92.

14. Heckman, James. "Sample Selection Bias as a Specification Error." NBER Working Paper No. 172, March 1977 (revised).

15. Heckman, James. "Sample Selection Bias as a Specification Error." Econometrica 45:153-61.

16. Helsing, Knud J. and George W. Comstock. "What Kinds of People Do Not Use Seat Belts?" American Journal of Public Health 67:1043-50.

17. Hrozenick, Diane. "Binomial Logit Analysis." American Council on Consumer Interests 1984 Proceedings. Edited by Karen P. Goebel. pp. 147-150.

18. Judge, George, William E. Griffiths, Carter R. Hill and Tsoung-Chao Lee. The Theory and Practice of Econometrics. New York: Wiely, 1980.

19. Kinsey, Jean. "Probit and Tobit Analysis." American Council on Consumer Interests 1984 Proceedings. Edited by Karen P. Goebel. pp. 155-161.

20. Kmenta, Jan. Elements of Econometrics New York: Macmillan Publishing Co., Inc. 1971.

21. Lee, L. F. and R. Trost. "Estimation of Some Limited Dependent Variable Models with Applications to Housing Demand." Journal of Econometrics 8:357-382.

22. Mayer, Robert N. and Cathleen D. Zick. "Mandating Behavioral or Technological Change: The Case of Auto Safety." Division of Social Science Research Working Paper, University of Utah, March 1985.

23. Maynes, E. Scott. Decision-Making for Consumers New York: Macmillan Publishing Co. Inc., 1976.

24. Motley, Brian. "Household Demand for Assets: A Model of Short-Run Adjustments." Review of Economics and Statistics 52:236-41.

25. O'Neill, Brian, Allan F. Williams, and Ronald S. Karpf. "Passenger Car Size and Driver Seat Belt Use." American Journal of Public Health 73:588-90.

26. Tobin, James. "Estimation of Relationships for Limited Dependent Variables," Econometrica 26:24-36.

27. Tuma, Nancy B. "Nonparametric and Partiallly Parametric Approaches to Event-History Analysis," in Sociological Methodology, 1982. Edited by Samuel Leinhardt. San Francisco: Jossey-Bass, pp. 1-60.

28. Zick, Cathleen D. "Human Capital Investment in a Family Portfolio Context." Unpublished dissertation, Cornell University, January 1982.