A RAND NOTE

Theory Testing in a World of Constrained Research Design

Ross M. Stolzenberg, Daniel A. Relles

June 1990

RAND

The research described in this report was supported by The RAND Corporation as part of its program of public service.

Published by The RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138

A RAND NOTE

N-3184-RC

Theory Testing in a World of Constrained Research Design

Ross M. Stolzenberg, Daniel A. Relles

June 1990

RAND

Because censored sampling is often unavoidable in much sociological data analysis, computationally simple corrections of censoring bias would be useful. Heckman's correction is simple to compute, widely used, and proven asymptotically correct under certain assumptions, but its limitations in practical situations are not well known in sociology. Here, we overview prior criticisms of Heckman's estimator, and we consider the case in which its normality assumptions are satisfied, censoring rates are high, and sample sizes are small. Results of 14,400 analyses of computer-generated simulation data suggest that Heckman's method performs well under certain circumstances, but that it very frequently worsens estimates, especially under conditions that are likely to be present in sociological data. Thus, the technique is probably not a general cure for censoring bias in sociology, except perhaps where strong theory permits certain strong assumptions. We reconsider censored sampling correction strategies in the context of statistical analysis as a theory-building tool, with emphasis on research strategy in the presence of irremediable censoring bias.

Theory Testing in a World of Constrained Research Design

The Significance of Heckman's Censored Sampling Bias Correction for Nonexperimental Research

ROSS M. STOLZENBERG
Graduate Management Admission Council
DANIEL A. RELLES
RAND Corporation

Censored sampling, truncation bias, and restriction of range are some names used by sociologists, economists, and psychologists to describe a complex of statistical problems obtained

AUTHORS' NOTE: This research was supported by the Trustees of the Graduate Management Admission Council (GMAC) under the terms of Research Agreement A-7BF-035. The GMAC encourages researchers to formulate and freely express their own opinions. Opinions expressed here are those of the authors and are not necessarily shared by the GMAC or any other organization or individual.

Reprinted from *Sociological Methods and Research,* Vol. 18, No. 4, May 1990, pp. 395–415,

from sample data that systematically exclude certain elements from the population they are supposed to represent (Berk, 1983; Cronbach, 1970; Heckman, 1976, 1979; Kendall and Buckland, 1971; Sorensen, 1976; Tuma and Hannan, 1984, 1978). Current sociological interest in censored sampling is very strong, and efforts to apply censored sampling correction methods are no longer unusual. Indeed, Berk (1983) suggests that censored sampling bias is so common that it should be suspected until proven absent, and that efforts to correct for censored sampling effects should be routine.

For nonexperimental branches of sociology, economics, psychology, and political science, simple and accurate corrections of censored sampling bias would be an important advance in research design, not just the neatening of a methodological loose end. Research in these fields often suffers because survey respondents tend to absent themselves from unpleasant circumstances that are nonetheless important to social science theories. For example, in studies of the wage benefits of schooling, sample censoring occurs because persons who would have low earnings if they worked for pay are less likely to hold paying jobs than otherwise similar persons who command high wage rates; lacking measurable earnings, nonworkers are thus censored from such studies. (See Little and Rubin, 1987, for a complete taxonomy of missing data problems, including those due to censoring and related causes.)

Because it is often impractical (and sometimes unethical and illegal) to apply data gathering strategies that would overcome respondents' tendencies to avoid unpleasantness and ignore perceived trivialities, sample censoring is inevitable in much nonexperimental social science research. Thus, it would be enormously valuable to have a routine statistical technique for correcting sample censoring bias. Although several methods can be used to rectify sample censoring bias in regression-type analyses (see Gronau, 1974; Lewis, 1974), Heckman's (1976, 1979) technique predominates in sociological studies that attempt such corrections. For example, Table 1 lists 11 *American Sociological Review* articles that use that technique to correct for suspected censoring bias. Indeed, the current widespread view that censoring problems are soluble without statistical virtuosity or ex-

TABLE 1: Recent Publications in *American Sociological Review* Using Heckman's Estimator for Sample Selection Bias Correction

Authors	*Subject*
Berk, Bridges, and Shih (1981)	Mental retardation
Marini (1984)	Childbearing
Hagan and Parker (1985)	Crime
Jasso (1985)	Coital frequency
Piliavin, Gartner, Thornton, and Matsueda (1986)	Crime
Allison and Long (1987)	Job mobility
Sanders and Nee (1987)	Minority employment
Tienda and Lii (1987)	Minority employment
Tienda, Smith, and Ortiz (1987)	Gender inequality
England, Kilbourne, Farkas, and Dou (1988)	Gender inequality
Snipp and Sandefur (1988)	Minority employment

traordinary data appears due entirely to acceptance of Heckman's method, which requires only straightforward application of probit and regression analyses.[1]

Because Heckman's estimator is so widely accepted in sociology, it is extremely important to question the extent to which it is effective enough to be considered a general solution to censoring problems. Although criticism of Heckman's estimator are not well known in sociology, we are not the first to question the accuracy of corrections made with Heckman's estimator, and there is a small literature that criticizes it on the basis of theory and the outcomes of experiments with Monte Carlo data. For example, Nelson (1984) finds Heckman's estimator less efficient than the maximum likelihood estimator for censoring bias correction. Paarsch (1984) finds that Heckman's estimator is sensitive to deviations from the assumption of normal errors. Little (1983) and Goldberger (1980) criticize Heckman's estimator, and Little and Rubin (1987) generally suggest that it is inferior to maximum likelihood methods. Duan et al. (1984: 283) argue that Heckman's estimator "has poor statistical and numerical properties and relies on untestable assumptions." And Lillard et al. (1986) find great instability of Heckman's correction in applications to CPS income data.

While we think it would be useful merely to bring these critiques to the attention of sociologists, we also note that past considerations of Heckman's estimator do not explicitly consider the situation in which normality can be assumed, censoring is severe, and samples are relatively small. In practice, this situation is often encountered in analyses of the relationship between grades received by students in school and their performance on preadmission measures of academic ability (this was the specific empirical situation that motivated us to perform analyses reported below). However, some of these conditions are common in other settings, such as studies of job mobility within organizations, where samples tend to be small and only a small proportion of people at each authority level advance to the next level. A priori, we expect that satisfaction of normality assumptions will work to the advantage of Heckman's estimator, as will the presence of severe censoring, which exacerbates censoring bias and permits the correction to be large relative to any additional error or bias introduced by a correction method. However, we lack a priori expectations about small sample size because Heckman's estimator is consistent but not unbiased, and may require large samples to work well, even when its assumptions are satisfied and censoring is severe.

Our strategy is computer simulation. Because asymptotically correct techniques sometimes fail to produce numerically accurate results, or produce only trivial improvements at considerable additional cost and difficulty, there is a tradition of using computer-generated simulation (Monte Carlo) data to test the accuracy of statistical methods (see Wonnacott and Wonnacott, 1970: 398; Kennedy and Gentle, 1980: 233-235; Hammersley and Handscomb, 1964). With Monte Carlo data, relevant parameters are used to construct the data, and so are known with certainty before analysis begins. Application of statistical techniques to such data illuminates the power of a method to discover those know parameters. We report results of some 14,400 separate applications of Heckman's method to Monte Carlo data, which we censor according to the selection model on which Heckman's estimator is based.

The next section describes the data used here, the nature of its censoring problem, and the results of using Heckman's estimator. In the third section, conclusions are drawn.

SIMULATION STUDIES

Our simulation model is defined by precisely the same mathematical structure that Heckman (1976: 476; 1979: 154) uses to derive his method: Let equation (1) be a regression equation of interest, where X_1 is the regression-independent variable, Y_1 is the regression-dependent variable, and ζ is the regression error term. X_1 and ζ are uncorrelated and marginally distributed as $N_{0,1}$ — that is, normal with zero mean and unit variance.

$$Y_1 = b_0 + b_1 X_1 + A\zeta \qquad [1]$$

For convenience, let parameters $b_0 = 0$ and $b_1 = 1$. Parameter A determines the squared correlation between X_1 and Y_1, which varies inversely with A.

Following Heckman, define the selection equation as

$$Y_2 = X_2 + B\varepsilon \qquad [2]$$

where X_2 and ε are uncorrelated, each with marginal distribution $N_{0,1}$. T is a number such that Y_1 is observed if $Y_2 > T$, and Y_1 is not observed otherwise. Y_3 is a dummy variable equal to 1 if $Y_2 > T$ and equal to 0 otherwise. T determines the average probability of selection (observation). B is a determinant of the effect of X_2 on the probability of selection. In Heckman's method, a probit analysis is estimated in which Y_3 is the dependent variable and X_2 is the independent variable. B determines the extent to which this probit analysis fits the data. More specifically, the expected squared correlation between X_2 and Y_2 is given by

$$\mathrm{Var}\,[B\varepsilon] \,/\, \mathrm{Var}\,[X_2 + B\varepsilon] = B^2 / (1 + B^2)$$

When B is smaller, the probit fits better; when larger, worse.

r_{12} is the correlation between X_1 and X_2. r_{12} is a parameter that we fix by construction in these simulations. $r_{\varepsilon\zeta}$ is the correlation between error terms ε and ζ, and it too is a parameter that we fix by construction. Values of these and other parameters used in these simulations are given in Table 2, along with brief definitions.

TABLE 2: Simulation Model Parameter Values

Parameter	*Brief Definition*	*Values*	*Notes*
b_0	Regression constant	0	
T	Selection rate parameter	$1.282(1 + B^2)^{0.5}$	10% selection rate
b_1	Regression coefficient	1	
A	Regression error variance	4	Regression $R^2 = .2$
		1	Regression $R^2 = .5$
		1/4	Regression $R^2 = .8$
B	Selection error variance	1/9	Selection $R^2 = .9$
		1	Selection $R^2 = .5$
		9	Selection $R^2 = .1$
$(r_{12})^2$	Squared correlation between X_1 and X_2	0	
		.25	
		.50	
		.75	
$(r_{\varepsilon\zeta})^2$	Squared correlation between ε and ζ	0	
		.25	
		.50	
		.75	

Given values of b_0, b_1, A, B, T, r_{12}, and $r_{\varepsilon\zeta}$, we perform the following steps: (1) Generate random data consistent with parameters in Table 2; (2) censor those data according to the selection equation; (3) regress Y_1 on X_1 in the censored data to determine how badly censoring biases estimates of b_1; (4) fit to the uncensored data a probit model in which Y_3 is the dependent variable and X_2 is the independent variable, and use the results of that probit analysis to estimate λ, Heckman's censored sampling correction factor; and (5) regress Y_1 on X_1 and λ, as prescribed by Heckman's correction method, to determine the extent to which that method improves the accuracy with which b_1 is estimated from the censored data.

We use a data set of n = 500 before sample censoring. We allow $(r_{12})^2$ and $(r_{\varepsilon\zeta})^2$ each to vary over four different values, and we let B and A each vary over three different values, for a total of 144 different combinations of parameter values.

We set $T = 1.282\,(1+B^2)^{0.5}$, to censor out approximately 90% of the cases (i.e., we set the average selection probability for each data set at 10%). Censoring of 90% is selected for both methodological and substantive reasons: The methodological purpose is to achieve a level of censoring bias sufficiently large to (1) be unambiguously distinct from ordinary estimation error, (2) permit observation of bias reduction as well as complete bias correction, and (3) make bias corrections large relative to any additional error introduced by the correction method. The substantive reason for choosing a 90% censoring level is that severe censoring is likely to be found in many situations in which sociologists might apply this method. For example, consider analyses of thefts based on criminals convicted of those crimes (surely more than 90% avoid conviction), studies requiring cooperation of very powerful or wealthy persons (who are notorious for their lack of cooperation with research), or studies requiring individuals to relive or experience unpleasant or stigmatizing events of behavior (few consent to do so).[2]

To ensure that the unique attributes of any given set of random numbers do not affect our conclusions, we apply each set of design parameters to 100 different random data sets, yielding 14,400 (= 100 × 144) separate applications of Heckman's method. Further, we take an additional step to guarantee that difference among parameters, rather than differences among data sets, are the unambiguous causes of differences in outcomes accross the 144 cells of our research design: The same 100 sets of random numbers are used to generate the data sets used in each cell of that design. These calculations are computer-intensive; although executed with efficient Fortran subroutines, they consumed more than six hours of central processor (CPU) time on an IBM 3033 mainframe computer running under MVS.

RESULTS

It would be cumbersome and confusing to display the performance of OLS and Heckman-corrected estimation in all 144 cells of our study design, so Tables 3 and 4 summarize the way the performance of these two estimators varies as the four varying parameters of the research design change (results for all 144 cells are presented in the appendix). Columns (1) and (2) of these tables indicate parameters of simulations that reported in each row. Column (3) indicates the percentage of

TABLE 3: **Average Results of 14,400 Simulations by Different Levels of Regression and Selection Equation R^2**

Values of Simulation Parameters		*Simulations Where Correction Increases Absolute Error of Estimate (%)*	*Mean Absolute Error of Estimated b_1*		*Estimated Bias (Mean Error of Estimated b_1)*	
Selection R^2	*Regression R^2*		*OLS Not Corrected*	*With Correction*	*OLS Not Corrected*	*With Correction*
(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
(Results of 1,600 simulation analyses reported in each line)						
.90	.20	55	.23*	.25	–.05	.05
.50	.20	46	.29	.26*	–.10	.05*
.10	.20	49	.29	.29	–.05*	.06
.90	.50	54	.11*	.13	–.03	.02*
.50	.50	46	.15	.13*	–.05	.02*
.10	.50	48	.14	.14	–.03	.03
.90	.80	55	.06	.06	–.01	.01
.50	.80	46	.07	.07	–.02	.01*
.10	.80	48	.07	.07	–.01*	.02
Total		50	.16	.16	–.04	.03*

*Less-biased estimate for each pair of corrected and uncorrected estimates.

simulations in which Heckman's correction makes estimates of b_1 less accurate than the uncorrected OLS estimate of that parameter (recall that b_1 is fixed at unity by design). The fourth and fifth columns of these tables indicate the average absolute difference between the estimate of b_1 and the true value of this parameter; comparing column (5) to column (4) indicates the extent to which Heckman's correction has increased or decreased the accuracy of estimates of the regression coefficient b_1. Finally, columns (6) and (7) estimate the bias of the uncorrected and corrected estimates of b_1, the differences between the parameters and the expected values of their estimators. Here, we estimate that bias with the mean signed deviation of estimates of b_1 from its true value, unity, which is fixed by construction in these simulations. Because these means are based on large n's (1,600 per line of Table 3, 900 per line of Table 4, and 14,400 for the total reported at the bottom of Table 3), they are reasonable approximations to the

TABLE 4: Average Results of 14,400 Simulations by Different Levels of Correlation Between Regression and Selection Error Terms

		Simulations Where Correction Increases Absolute Error of Corrected Estimate (%)	*Mean Absolute Error of Estimated b_1*		*Estimated Bias (Mean Error of Estimated b_1)*	
$(r_{12})^2$	$(r_{\varepsilon\zeta})^2$		*OLS Not Corrected*	*With Correction*	*OLS Not Correction*	*With Correction*
(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
		Each line reports results of 900 simulations				
0.00	0.00	50	0.13*	0.14	0.02	0.02
0.25	0.00	52	0.14	0.14	0.02	0.02
0.50	0.00	55	0.15*	0.16	0.03	0.03
0.75	0.00	64	0.16*	0.22	0.03*	0.04
0.00	0.25	46	0.13*	0.14	0.02	0.02
0.25	0.25	58	0.14*	0.15	0.01*	0.04
0.50	0.25	57	0.15*	0.17	−0.02*	0.04
0.75	0.25	62	0.17*	0.22	−0.07	0.05*
0.00	0.50	50	0.12	0.12	0.02	0.02
0.25	0.50	50	0.14	0.14	−0.02*	0.04
0.50	0.50	50	0.16	0.15*	−0.08	0.03*
0.75	0.50	47	0.21	0.21	−0.16	0.04*
0.00	0.75	44	0.11	0.10*	0.01	0.01
0.25	0.75	42	0.13	0.12*	−0.05	0.02*
0.50	0.75	35	0.18	0.13*	−0.14	0.02*
0.75	0.75	32	0.28	0.18*	−0.26	0.03*

*Less-biased estimate for each pair of corrected and uncorrected estimates.

true mean, or expectation, of the bias. In passing, a finding of bias is not surprising, as Heckman's estimator is consistent but not unbiased. However, the technique is motivated solely by a desire to reduce bias, so comparison of the biases of corrected and uncorrected estimates would seem to be of intrinsic interest.

The big news from these simulations is found in the total row at the bottom of Table 3, and can be stated simply: (1) On average, Heckman's correction improved approximately one-half of the estimates and worsened the rest—see column 3; (2) the average absolute error of the Heckman-corrected estimates is the same as the average absolute error of the OLS estimates—compare columns 4 and 5; and (3) on average, corrected estimates are slightly less biased than uncorrected estimates. Corrected and uncorrected estimates are biased in opposite directions, and the average bias of the corrected estimate is three-quarters as large as the average bias of the uncorrected OLS estimate (0.03 vs. -0.04; compare columns 6 and 7).

The body of Table 3 shows how these simulation results change as the goodness of fit (R^2 statistics) in the regression and selection equations are systematically varied. Each row of Table 3 reports the results of 1,600 different simulations. This table shows no discernible pattern of relationship between these goodness-of-fit measures and the performance of Heckman's estimator. This result suggests that the method is not ensured of success if the selection model fits the data especially well (as indicated by a good fit of the probit analysis), and that the method is not ensured of failure if the selection model fits the data only modestly. In practice, the selection model could fit poorly because the selection process includes much randomness, or because the model does not properly specify the selection process.

Table 4 shows how Heckman's adjustment is affected by variation in the correlation between error terms in the regression and selection equations (ε and ζ), and by changes in the regression equation R^2. Each row of Table 4 reports results of 1,200 different simulation analyses. Notice that when the squared correlation between ε and ζ is .25 or 0, the censoring correction always increases mean absolute error and leads to less accurate estimates of b_1 in more than half of the simulations. However, when the squared correlation between ε and ζ is .50 or .75, the mean absolute error obtained with Heckman's correction is often better (four of six lines in Table 4) and never worse than that obtained with OLS.

When the squared correlation between ε and ζ is .25 or 0, the Heckman correction also yields higher average bias than uncorrected OLS in four out of six lines of Table 4. But when this correlation is 0.50 or more, mean bias is reduced by the Heckman correction, sometimes dramatically (as in the bottom three rows of Table 4). However, even when the correction works best at reducing bias in our simulations, it still increases absolute error of coefficient estimates one-third of the time (see bottom row of Table 4).

On the average, Heckman's method performs no better than uncorrected ordinary least squares in our simulations, sometimes exacerbating bias, and worsening the accuracy of estimates almost as often as it improves them. In our simulations, we find that even when the method yields less bias than uncorrected OLS, it introduces a bias of its own, in the opposite direction of censoring bias. However, the method seems to reduce bias consistently in our simulations when two conditions are met simultaneously: (1) When the error term in the regression equation is, by sociological standards, very highly correlated with the error term in the selection equation; and (2) when the independent variable in the regression equation is very highly correlated with the independent variable in the selection equation.

Specifically, in our simulations we find that whenever the sum of these squared correlations is 1.0 or more, Heckman's correction lowers the absolute level of bias in the estimate of b_1. Whenever the sum of these squared correlations is 1.25 or higher, Heckman's correction dramatically reduces the absolute level of bias in our simulations. However, even in situations in which the correction works very well on average at reducing bias, it still worsens the absolute error of estimates in a third or more of the simulations.

Do these findings suggest ways to use Heckman's methods effectively? Perhaps. Although our simulations indicate that Heckman's correction often worsens estimation if $r_{\varepsilon\zeta}$ and r_{12} are not both very high, residuals in the selection equation ordinarily are not observable, making it very difficult indeed to estimate $r_{\varepsilon\zeta}$.[3] Thus, it is difficult or impossible to obtain empirical information that would confirm the applicability of Heckman's correction to a particular situation.

However, strong theory may provide the necessary information. For example, if theory strongly indicates that Y_1 and censoring are both caused by the same factors, and if theory also suggests that errors in the selection process and errors in the determination of Y_1 are very similar, then one might be justified in assuming that r_{12} and $r_{\varepsilon\zeta}$ are sufficiently strong to warrant application of Heckman's method. We think that such a situation exists in the analysis of women's earnings:

Theory strongly suggests that women's wages and women's probability of labor force participation are determined by their human capital. Theory also suggests that individual women have unique characteristics that cause some to earn more or less than would be expected on the basis of their human capital, and which make those women more or less likely to participate in the labor force than would be expected on the basis of their human capital. Thus, in terms of equations (1) and (2), human capital is both X_1 and X_2, causing r_{12} to be at least very large and maybe unity. The unique individual characteristics that affect earnings and labor force participation are the residuals ε and ζ, causing $r_{\varepsilon\zeta}$ to be very large. Seen in this light, it does not seem surprising that Heckman (1980) obtained reasonable results in applying his method to wage determination of married women. However, Heckman's subject matter in this example is the beneficiary of exceptionally well-developed theory; few subjects examined by sociologists have comparable theoretical foundations on which to build assumptions about the structure of unobservable errors.

CONCLUSIONS

In our Monte Carlo studies, we found that Heckman's method reduced the accuracy of coefficient estimates as often as it improved them. And while the correction worked very well under certain circumstances, those conditions are difficult or impossible to verify empirically in naturally occurring sociological data, and seem likely to be too rare to justify automatic application of the Heckman estimator when censoring bias is suspected. Further, even under conditions in which the method worked well on average, it worsened estimates approximately one-third or more of the time. This finding reinforces earlier findings of instability of Heckman's correction in applications to CPS income data (Lillard et al., 1986), and it is consistent with complaints about the method's inefficiency. In short, our results extend rather than contradict the earlier literature of criticism of Heckman's estimator for censored data.

We close by briefly considering four questions about censoring and Heckman's method:

First, why have empirically verified problems with Heckman's correction not surfaced sooner in sociology? Part of the answer, we think, is that the technique is usually applied to situations in which censoring bias is only suspected, and its likely effects are merely the subject of speculation. In such situations, unexplainable or peculiar empirical findings are seldom published, for they tend to be interpreted as the fruit of flawed models, contaminated data, incorrect hypotheses, or some combination of all three problems. Thus, it ironically may be true that journal publications are a severely censored sample of applications of Heckman's censored sampling correction method. If so, the cases censored out of this sample are those in which Heckman's method yields inexplicable or nonsensical results.

Second, is censoring a pervasive catastrophic problem? Censoring bias can be catastrophic, but it is less certain that it is frequently catastrophic. One might argue that Heckman's major unique contribution to the analysis of censoring effects is his observation that censoring bias is a case of omitted variable bias. Seen from the perspective of omitted variables, censored sampling bias may seem less exotic and difficult: Survey data seldom contain all variables one would like to include in analyses, and nonexperimental researchers regularly and successfully evaluate the probable consequences of omitting variables that they would like to include.

Further, censoring problems are often a consequence of substantive decisions about the population about which one wishes to make inferences. For example, consider censoring in a common type of educational policy study: One wishes to estimate effects of students' background characteristics on their college grades. Perhaps some censoring occurs because a few college students are ill at the time of final exams and do not have complete grade information when data are gathered at the end of the semester; perhaps those students tended to be ill during the preceding semester too, with consequent damage to their grades. Other students attend college but perform poorly and drop out before grades are given and are thus censored from the sample. Others graduate high school but know that they lack the ability, training, motivation or other resources necessary for academic success and so never enter college and thus lack college grade data.

Still others attend high school but lack the ability or other characteristics needed to earn their diploma; most of these would do poorly in college, but do not matriculate and so lack college grade data. If one wishes to draw inferences about the effects of background characteristics on college grades for all of these groups, one faces a severe degree of censoring, and probably a severe degree of censoring bias. Conversely, the fewer of these groups about which one wishes to draw inferences, the less severe the censoring and, probably, the less severe the censoring bias. Thus, the severity of censoring, and probably the severity of censoring bias, is affected by substantive decisions about the population to which one wishes to draw inferences.

Third, is censored sampling bias a fatal flaw? Sometimes, but certainly not always. For example, consider the situation in which censoring (or something else) causes downward bias of coefficients that are hypothesized to be positive. If empirical results show a positive effect in spite of negative bias, one has more reason, not less, to entertain a hypothesis of positive effects. This is the cumbersome, traditional strategy for dealing with irremediable bias, of course, and it is inferior to less ad hoc procedures. But one should not lose sight of the fact that irremediable bias is commonplace and often successfully managed in the social sciences. These problems are neither more nor less manageable if they are caused by censored sampling rather than by something else.

Fourth, and finally, what role should Heckman's estimator play in assessing and correcting censored sampling bias? We think that role should be a small and infrequent one. Certainly the estimator sometimes works well, but our results suggest that in small, severely censored samples, even if normality assumptions are true, unless one has a strong enough theory to justify the assumption that selection errors and regression errors are *very* highly correlated, it is exceedingly difficult to know whether Heckman's estimator worsens or cures censoring bias.[4] Other studies have found the estimator less efficient than the maximum likelihood estimator, numerically wanting, and a poor performer when normality assumptions are violated. In short, there is considerable evidence that the method can easily do more harm than good, and that its careless or mechanical application runs much danger of producing vivid examples of the problems that Lieberson (1985) so aptly discussed when he cautioned against statistical "corrections" that in fact produce substantial distortions far worse than the problems they are designed to cure.

Appendix

Results of Simulations for All 144 Cells of Study Design

This table presents outcomes of simulations separately for all 144 cells in the simulation design. Columns A through D indicate values of four design parameters: regression R^2, selection R^2, squared correlation between regression and selection independent variables, and squared correlation between regression and selection error terms. Column E is the proportion of simulations in which OLS had lower mean absolute error than the Heckman-corrected estimate. Column F gives the difference between the mean OLS estimate and the true population value of the parameter. Column G gives the analogous statistic for the Heckman-estimated parameter. Column H gives the mean absolute difference between the OLS estimate and the true parameter; column I gives the mean absolute difference for the Heckman estimate.

A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
0.20	0.10	0.00	0.00	44.00	0.07	0.08	0.22	0.22
0.20	0.10	0.00	0.25	48.00	0.04	0.05	0.22	0.23
0.20	0.10	0.00	0.50	47.00	0.02	0.04	0.21	0.21
0.20	0.10	0.00	0.75	44.00	−0.01	0.01	0.21	0.21
0.20	0.10	0.25	0.00	44.00	0.08	0.07	0.25	0.25
0.20	0.10	0.25	0.25	57.00	0.03	0.07	0.26	0.27
0.20	0.10	0.25	0.50	49.00	−0.02	0.06	0.26	0.25
0.20	0.10	0.25	0.75	42.00	−0.07	0.05	0.27	0.24
0.20	0.10	0.50	0.00	49.00	0.09	0.08	0.29	0.30
0.20	0.10	0.50	0.25	52.00	−0.01	0.07	0.30	0.32
0.20	0.10	0.50	0.50	51.00	−0.12	0.04	0.29	0.29
0.20	0.10	0.50	0.75	39.00	−0.20	0.06	0.31	0.26
0.20	0.10	0.75	0.00	59.00	0.08	0.06	0.34	0.41
0.20	0.10	0.75	0.25	64.00	−0.09	0.08	0.32	0.40
0.20	0.10	0.75	0.50	48.00	−0.26	0.10	0.39	0.39
0.20	0.10	0.75	0.75	39.00	−0.44	0.10	0.50	0.37
0.20	0.50	0.00	0.00	52.00	0.05	0.05	0.25	0.25
0.20	0.50	0.00	0.25	41.00	0.05	0.05	0.25	0.24

(Appendix continues)

Appendix Continued

A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
0.20	0.50	0.00	0.50	50.00	0.05	0.04	0.21	0.21
0.20	0.50	0.00	0.75	42.00	0.03	0.02	0.19	0.16
0.20	0.50	0.25	0.00	55.00	0.05	0.05	0.25	0.25
0.20	0.50	0.25	0.25	53.00	0.01	0.06	0.25	0.25
0.20	0.50	0.25	0.50	51.00	−0.04	0.07	0.23	0.24
0.20	0.50	0.25	0.75	41.00	−0.12	0.04	0.21	0.18
0.20	0.50	0.50	0.00	57.00	0.08	0.07	0.26	0.29
0.20	0.50	0.50	0.25	56.00	−0.08	0.03	0.27	0.29
0.20	0.50	0.50	0.50	43.00	−0.22	0.02	0.29	0.25
0.20	0.50	0.50	0.75	27.00	−0.34	0.01	0.37	0.21
0.20	0.50	0.75	0.00	64.00	0.07	0.08	0.27	0.37
0.20	0.50	0.75	0.25	51.00	−0.17	0.07	0.33	0.38
0.20	0.50	0.75	0.50	33.00	−0.40	0.04	0.43	0.34
0.20	0.50	0.75	0.75	18.00	−0.61	0.02	0.61	0.27
0.20	0.90	0.00	0.00	54.00	−0.01	0.00	0.23	0.23
0.20	0.90	0.00	0.25	51.00	0.01	0.01	0.23	0.23
0.20	0.90	0.00	0.50	55.00	0.04	0.05	0.21	0.22
0.20	0.90	0.00	0.75	46.00	0.03	0.03	0.17	0.17
0.20	0.90	0.25	0.00	56.00	−0.01	0.00	0.22	0.24
0.20	0.90	0.25	0.25	64.00	0.01	0.07	0.23	0.25
0.20	0.90	0.25	0.50	49.00	−0.01	0.07	0.24	0.24
0.20	0.90	0.25	0.75	44.00	−0.06	0.04	0.18	0.17
0.20	0.90	0.50	0.00	61.00	0.01	0.02	0.23	0.26
0.20	0.90	0.50	0.25	62.00	−0.01	0.09	0.23	0.28
0.20	0.90	0.50	0.50	56.00	−0.09	0.07	0.24	0.25
0.20	0.90	0.50	0.75	38.00	−0.17	0.04	0.23	0.20
0.20	0.90	0.75	0.00	70.00	0.03	0.06	0.23	0.36
0.20	0.90	0.75	0.25	70.00	−0.08	0.10	0.23	0.36
0.20	0.90	0.75	0.50	59.00	−0.19	0.08	0.26	0.33
0.20	0.90	0.75	0.75	38.00	−0.30	0.05	0.31	0.27
0.50	0.10	0.00	0.00	44.00	0.04	0.04	0.11	0.11
0.50	0.10	0.00	0.25	48.00	0.02	0.03	0.11	0.11
0.50	0.10	0.00	0.50	47.00	0.01	0.02	0.11	0.11
0.50	0.00	0.00	0.75	44.00	−0.01	0.01	0.11	0.11
0.50	0.10	0.25	0.00	44.00	0.04	0.04	0.13	0.13
0.50	0.10	0.25	0.25	57.00	0.01	0.03	0.13	0.13
0.50	0.10	0.25	0.50	49.00	−0.01	0.03	0.13	0.12
0.50	0.10	0.25	0.75	42.00	−0.04	0.02	0.13	0.12
0.50	0.10	0.50	0.00	48.00	0.04	0.04	0.15	0.15
0.50	0.10	0.50	0.25	52.00	0.00	0.03	0.15	0.16
0.50	0.10	0.50	0.50	51.00	−0.06	0.02	0.14	0.15
0.50	0.10	0.50	0.75	39.00	−0.10	0.03	0.16	0.13
0.50	0.10	0.75	0.00	59.00	0.04	0.03	0.17	0.20

Appendix Continued

A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
0.50	0.10	0.75	0.25	64.00	−0.04	0.04	0.16	0.20
0.50	0.10	0.75	0.50	48.00	−0.13	0.05	0.19	0.20
0.50	0.10	0.75	0.75	39.00	−0.22	0.05	0.25	0.19
0.50	0.50	0.00	0.00	53.00	0.02	0.02	0.12	0.12
0.50	0.50	0.00	0.25	41.00	0.03	0.03	0.12	0.12
0.50	0.50	0.00	0.50	50.00	0.03	0.02	0.11	0.11
0.50	0.50	0.00	0.75	42.00	0.02	0.01	0.09	0.08
0.50	0.50	0.25	0.00	55.00	0.02	0.02	0.12	0.12
0.50	0.50	0.25	0.25	53.00	0.00	0.03	0.12	0.12
0.50	0.50	0.25	0.50	51.00	−0.02	0.04	0.11	0.12
0.50	0.50	0.25	0.75	41.00	−0.06	0.02	0.11	0.09
0.50	0.50	0.50	0.00	57.00	0.04	0.04	0.13	0.15
0.50	0.50	0.50	0.25	56.00	−0.04	0.02	0.13	0.14
0.50	0.50	0.50	0.50	43.00	−0.11	0.01	0.14	0.12
0.50	0.50	0.50	0.75	27.00	−0.17	0.01	0.18	0.11
0.50	0.50	0.75	0.00	64.00	0.03	0.04	0.14	0.18
0.50	0.50	0.75	0.25	51.00	−0.08	0.03	0.17	0.19
0.50	0.50	0.75	0.50	33.00	−0.20	0.02	0.21	0.17
0.50	0.50	0.75	0.75	17.00	−0.30	0.01	0.31	0.14
0.50	0.90	0.00	0.00	54.00	0.00	0.00	0.11	0.11
0.50	0.90	0.00	0.25	50.00	0.00	0.01	0.11	0.11
0.50	0.90	0.00	0.50	54.00	0.02	0.02	0.11	0.11
0.50	0.90	0.00	0.75	46.00	0.02	0.01	0.09	0.08
0.50	0.90	0.25	0.00	56.00	0.00	0.00	0.11	0.12
0.50	0.90	0.25	0.25	64.00	0.00	0.03	0.12	0.13
0.50	0.90	0.25	0.50	49.00	−0.01	0.04	0.12	0.12
0.50	0.90	0.25	0.75	44.00	−0.03	0.02	0.09	0.09
0.50	0.90	0.50	0.00	61.00	0.00	0.01	0.12	0.13
0.50	0.90	0.50	0.25	62.00	−0.01	0.05	0.11	0.14
0.50	0.90	0.50	0.50	56.00	−0.05	0.04	0.12	0.13
0.50	0.90	0.50	0.75	38.00	−0.09	0.02	0.12	0.10
0.50	0.90	0.75	0.00	70.00	0.01	0.03	0.11	0.18
0.50	0.90	0.75	0.25	70.00	−0.04	0.05	0.12	0.18
0.50	0.90	0.75	0.50	59.00	−0.09	0.04	0.13	0.17
0.50	0.90	0.75	0.75	38.00	−0.15	0.03	0.16	0.13
0.80	0.10	0.00	0.00	43.00	0.02	0.02	0.05	0.06
0.80	0.10	0.00	0.25	46.00	0.01	0.01	0.05	0.06
0.80	0.10	0.00	0.50	47.00	0.00	0.01	0.05	0.05
0.80	0.10	0.00	0.75	44.00	0.00	0.00	0.05	0.05
0.80	0.10	0.25	0.00	44.00	0.02	0.02	0.06	0.06
0.80	0.10	0.25	0.25	57.00	0.01	0.02	0.06	0.07
0.80	0.10	0.25	0.50	49.00	−0.01	0.01	0.06	0.06

(Appendix continues)

Appendix Continued

A	B	C	D	E	F	G	H	I
0.80	0.10	0.25	0.75	42.00	−0.02	0.01	0.07	0.06
0.80	0.10	0.50	0.00	48.00	0.02	0.02	0.07	0.07
0.80	0.10	0.50	0.25	52.00	0.00	0.02	0.07	0.08
0.80	0.10	0.50	0.50	51.00	−0.03	0.01	0.07	0.07
0.80	0.10	0.50	0.75	39.00	−0.05	0.01	0.08	0.06
0.80	0.10	0.75	0.00	59.00	0.02	0.02	0.09	0.10
0.80	0.10	0.75	0.25	64.00	−0.02	0.02	0.08	0.10
0.80	0.10	0.75	0.50	48.00	−0.06	0.02	0.10	0.10
0.80	0.10	0.75	0.75	39.00	−0.11	0.02	0.13	0.09
0.80	0.50	0.00	0.00	52.00	0.01	0.01	0.06	0.06
0.80	0.50	0.00	0.25	41.00	0.01	0.01	0.06	0.06
0.80	0.50	0.00	0.50	49.00	0.01	0.01	0.05	0.05
0.80	0.50	0.00	0.75	42.00	0.01	0.00	0.05	0.04
0.80	0.50	0.25	0.00	55.00	0.01	0.01	0.06	0.06
0.80	0.50	0.25	0.25	53.00	0.00	0.02	0.06	0.06
0.80	0.50	0.25	0.50	51.00	−0.01	0.02	0.06	0.06
0.80	0.50	0.25	0.75	41.00	−0.03	0.01	0.05	0.05
0.80	0.50	0.50	0.00	56.00	0.02	0.02	0.07	0.07
0.80	0.50	0.50	0.25	56.00	−0.02	0.01	0.07	0.07
0.80	0.50	0.50	0.50	43.00	−0.06	0.01	0.07	0.06
0.80	0.50	0.50	0.75	27.00	−0.09	0.00	0.09	0.05
0.80	0.50	0.75	0.00	64.00	0.02	0.02	0.07	0.09
0.80	0.50	0.75	0.25	51.00	−0.04	0.02	0.08	0.09
0.80	0.50	0.75	0.50	33.00	−0.10	0.01	0.11	0.08
0.80	0.50	0.75	0.75	18.00	−0.15	0.01	0.15	0.07
0.80	0.90	0.00	0.00	54.00	0.00	0.00	0.06	0.06
0.80	0.90	0.00	0.25	50.00	0.00	0.00	0.06	0.06
0.80	0.90	0.00	0.50	55.00	0.01	0.01	0.05	0.05
0.80	0.90	0.00	0.75	46.00	0.01	0.01	0.04	0.04
0.80	0.90	0.25	0.00	56.00	0.00	0.00	0.06	0.06
0.80	0.90	0.25	0.25	64.00	0.00	0.02	0.06	0.06
0.80	0.90	0.25	0.50	49.00	0.00	0.02	0.06	0.06
0.80	0.90	0.25	0.75	44.00	−0.02	0.01	0.05	0.04
0.80	0.90	0.50	0.00	61.00	0.00	0.01	0.06	0.07
0.80	0.90	0.50	0.25	62.00	0.00	0.02	0.06	0.07
0.80	0.90	0.50	0.50	56.00	−0.02	0.02	0.06	0.06
0.80	0.90	0.50	0.75	38.00	−0.04	0.01	0.06	0.05
0.80	0.90	0.75	0.00	70.00	0.01	0.01	0.06	0.09
0.80	0.90	0.75	0.25	70.00	−0.02	0.02	0.06	0.09
0.80	0.90	0.75	0.50	59.00	−0.05	0.02	0.07	0.08
0.80	0.90	0.75	0.75	38.00	−0.07	0.01	0.08	0.07

NOTES

1. When only the correlation coefficient is biased by censoring, correction methods are often available. These straightforward methods have been widely used by educational psychological studies (see Cronbach, 1970; Miner and Pearson, 1983). However, sociologists and economists tend to be concerned with regression coefficient bias rather than correlation coefficient bias.

2. These situations can also cause sample truncation, which is more serious than sample censoring, and has no known analytical cure. Truncation is similar to censoring, except that no data whatsoever are available on truncated cases.

3. Several points are relevant in passing: First, $r_{ed} = 0$ implies no censoring bias because it indicates no correlation between errors in selection and regression equations. Second, a squared correlation of 0.25 is substantial by sociological standards, and suggests considerable censoring bias. And, third, the correlation between the true X_1 and the true X_2 may be difficult to calculate because of uncertainty that the independent variables used in one's regression and probit models are the true causes of Y_1 and Y_2. Misspecification or omission of key variables could easily lead to improper calculation of r_{12}. This problem would tend to underestimate r_{12}.

4. Unfortunately, we can recommend no censoring correction as uniformly effective. Heckman's correction is one of several described in Chapter 11 of Little and Rubin (1987), including maximum likelihood and predictive Bayesian methods. This topic remains an active subject in statistical journals as well.

REFERENCES

ALLISON, P. D. and J. S. LONG (1987) "Interuniversity mobility of academic scientists." Amer. Soc. Rev. 52: 643-652.

BERK, R. (1983) "An introduction to sample selection bias in sociological data." Amer. Soc. Rev. 48: 386-398.

BERK R. A., W. P. BRIDGES, and A. SHIH (1981) "Does IQ really matter — a study of the use of IQ scores for the tracking of the mentally retarded." Amer. Soc. Rev. 46: 58-71.

CRONBACH, L. (1970) Essentials of Psychological Testing (3rd ed.). New York: Harper & Row.

DUAN, N., W. MANNING, C. MORRIS, and J. NEWHOUSE (1984) "Choosing between the sample selection model and the multi-part model." J. of Business and Econ. Statistics 2: 283-289.

ENGLAND, P., B. S. KILBOURNE, G. FARKAS, and T. DOU (1988) "Explaining occupational sex segregation and wages — findings from a model with fixed effects." Amer. Soc. Rev. 53: 544-558.

GOLDBERGER, A. (1980) "Abnormal selection bias." Madison: University of Wisconsin Department of Economics. (unpublished)

GRONAU, R. (1974) "Wage comparisons — a selectivity bias." J. of Pol. Economy 82: 1119-1144.

HAGAN, J. and P. PARKER (1985) "White-collar crime and punishment — the class structure and legal sanctioning of securities violations." Amer. Soc. Rev. 50: 302-316.

HAMMERSLEY, J. and D. HANDSCOMB (1964) Monte Carlo Methods. London: Methuen.

HECKMAN, J. (1976) "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models." Annals of Econ. and Social Measurement 5: 475-492.

HECKMAN, J. (1979) "Sample selection bias as a specification error." Econometrica 45: 153-161.

HECKMAN, J. (1980) "Sample selection bias as a specification error," pp. 206-248 in J. P. Smith (ed.), Female Labor Supply. Princeton, NJ: Princeton Univ. Press.

JASSO, G. (1985) "Marital coital frequency and the passage of time – estimating the separate effects of spouses' ages and marital duration, birth and marriage cohorts, and period influences." Amer. Soc. Rev. 50: 224-241.

KENDALL, D. and W. BUCKLAND (1971) A Dictionary of Statistical Terms. New York: Hafner.

KENNEDY, W. and J. GENTLE (1980) Statistical Computing. New York: Decker.

LEWIS, H. G. (1974) "Comments on selectivity biases in wage comparisons." J. of Pol. Economy 82: 1145-1155.

LIEBERSON, S. (1985) Making it Count. Berkeley: Univ. of California Press.

LILLARD, L., J. P. SMITH, and F. WELCH (1986) "What do we really know about wages? the importance of nonreporting and census imputation." J. of Pol. Economy 94: 489-506.

LITTLE, R. (1983) "A note about models for selectivity bias." Los Angeles, CA: UCLA School of Medicine Department of Biomathematics. (unpublished)

LITTLE, R. and D. RUBIN (1987) Statistical Analysis with Missing Data. New York: Wiley.

MARINI, M. M. (1984) "Women's educational attainment and the timing of entry into parenthood." Amer. Soc. Rev. 49: 491-511.

MINER, J. and J. PEARSON (1983) "An evaluation of business and economics doctoral student selection." Collegiate News and Views (Fall): 15-19.

NELSON, F. (1984) "Efficiency of the two-step estimator for models with endogenous sample selection." J. of Econometrics 24: 181-196.

PAARSCH, H. (1984) "A Monte Carlo comparison of estimators for censored regression models." J. of Econometrics 24: 197-213.

PILIAVIN, I., R. GARTNER, C. THORNTON, and R. L. MATSUEDA (1986) "Crime, deterrence, and rational choice." Amer. Soc. Rev. 51: 101-119.

SANDERS, J. M. and V. NEE (1987) "Limits of ethnic solidarity in the enclave economy." Amer. Soc. Rev. 52: 745-767.

SNIPP, C. M. and G. D. SANDEFUR (1988) "Earnings of American Indians and Alaskan natives – the effects of residence and migration." Social Forces 66: 994-1008.

SORENSEN, A. (1976) "Estimating rates from retrospective questions," pp. 209-223 in D. Heise (ed.), Sociological Methodology 1977. San Francisco: Jossey-Bass.

TIENDA, M. and D. T. LII (1987) "Minority concentration and earnings inequality – blacks, Hispanics, and Asians compared." Amer. Soc. Rev. 93: 141-165.

TIENDA, M., S. A. SMITH, and V. ORTIZ (1987) "Industrial restructuring, gender segregation, and sex differences in earnings." Amer. Soc. Rev. 52: 195-210.

TUMA, N. and M. HANNAN (1978) "Approaches to the censoring problem in analysis of event histories," pp. 209-240 in K. Schuessler (ed.), Sociological Methodology 1979. San Francisco: Jossey-Bass.

TUMA, N. and M. HANNAN (1984) Social Dynamics: Models and Methods. New York: Academic Press.

WONNACOTT, R. J. and T. H. WONNACOTT (1970) Econometrics. New York: John Wiley.

Ross M. Stolzenberg is a sociologist, applied statistician, and vice president of research of the Graduate Management Admission Council. He has served on the faculties of Harvard University, Johns Hopkins University, and the University of Illinois. In addition to statistical methods, his research focuses on the connections between employment, organizations, and higher education.

Daniel A. Relles is a mathematical statistician and associate head of the Department of Economics and Statistics at the RAND Corporation. He was associate editor of the Journal of the American Statistical Association, *and his current research at RAND focuses on statistical computing, health economics policy, and military manpower planning.*

RAND/N-3184-RC