It is well-known that, under the logit model for binary response, the random sampling and response-based sampling maximum likelihood estimators coincide for all parameters except the intercept. Citing this coincidence, many researchers have assumed the logit model and analyzed data from response-based samples as if those data were obtained by random sampling. We argue that this practice should be avoided unless the researcher really believes the logit specification. One preferable alternative is the weighted maximum likelihood estimator of Manski and Lerman (1977). Random sampling maximum likelihood analysis does not have a natural interpretation when the true response function is not logit. Weighted maximum likelihood analysis estimates a constrained best predictor of the binary response and so remains interpretable.

The Logit Model and Response-Based Samples

YU XIE
CHARLES F. MANSKI
University of Wisconsin—Madison

In binary response analysis, a sampling process is termed "response-based" if it stratifies on the response.[1] Response-based samples are potentially very useful in sociological research. Many social phenomena are rare events; hence a random sample will not find enough cases with positive response to the event under study for effective statistical analysis. For example, a researcher who wants to study the determinants of crime commission would not find many respondents with criminal histories if he or she were to conduct a national survey based on random sampling. A more effective approach would be to obtain a random sample of those with criminal records (from the court archives, for example) and

AUTHORS' NOTE: *Computations were performed at the Department of Sociology and the Center for Demography and Ecology, University of Wisconsin—Madison.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 17 No. 3, February 1989 283-302

another sample of the general population without criminal records. The combined sample constitutes a response-based sample.

Manski and Lerman (1977) demonstrated that the maximum likelihood estimator for random samples and exogenously stratified samples is generally inconsistent when applied to response-based samples. As a correction, they proposed a simple procedure—a weighted maximum likelihood estimator—which they proved to be consistent. Since then, various estimators have appeared. Manski and McFadden (1981) introduced a conditional maximum likelihood method. See Bye et al. (1987) for a sociological application. Cosslett (1981a, 1981b) studied the full information maximum likelihood estimator under response-based sampling. Hsieh et al. (1985) extended the domain of these methods in various respects. A text review is given by Amemiya (1985: 319-338).

The purpose of the foregoing literature is to find response-based sampling estimators for *general* binary response models. The word *general* is emphasized because for the logit model the full-information maximum likelihood estimator under response-based sampling happens to coincide with that under random sampling for all parameters except the intercept. Moreover, the appropriate maximum likelihood estimate for the intercept can be obtained post hoc. Versions of this result are reported in the articles listed above and also in Bishop et al. (1975: 63) and Prentice and Pyke (1979).[2]

Citing this coincidence, many researchers have assumed the logit model and analyzed data from response-based samples as if those data were obtained by random sampling. Gortmaker (1979) is a sociological example. Researchers using the logit model have generally not argued that the true response function is logit. Rather, they have applied the conventional wisdom that maximum likelihood estimates of logit models tend to be very similar to corresponding estimates of other binary response models such as the probit. See, for example, Cox (1970: 28) and Maddala (1983: 23). Given this, the argument goes, one might as well choose the convenient logit form.

Is this practice justified? Should researchers analyzing data from response-based samples simply specify the logit model and apply the random sampling maximum likelihood estimator? We have found that, unless one really believes the logit specification, the answer is negative.

The conventional wisdom on maximum likelihood logit analysis is well-grounded in random samples but not in response-based samples. In random samples, the maximum likelihood estimate of a logit model has an appealing interpretation whether or not the logit specification is correct. That is, the fitted logit model estimates a constrained best predictor of binary response. In a certain sense, the fitted model approximates the true probabilistic response function optimally. Discussions of this fact appear in Efron (1978), Hastie (1987), and Manski and Thompson (1989).

In response-based samples, maximum likelihood logit analysis loses its best predictor interpretation. If the logit specification is not correct, the fitted model does not necessarily approximate the true response function well. Hence the practice of specifying the logit model and estimating by random sampling maximum likelihood is dangerous unless one really believes the logit specification. We will elaborate on this in the next section.

The foregoing discussion leaves the applied researcher in an uncomfortable position. One rarely, if ever, knows whether the response function has the logit or some other form. How then should one proceed given response-based data? Happily, answers are available.

One approach is to apply some semiparametric method whose validity does not require knowledge of the exact form of the response function. Various such methods have been proposed in recent years for use in random and exogenously stratified samples. See Manski (1988) for a survey. Of these, the maximum score estimator has been proved applicable to response-based samples (see Manski, 1986). It may be that response-based sampling versions of other semiparametric estimators can be developed.

Second, one may select some response function specification (say logit) and apply a method that is interpretable whether or not that specification is correct. The weighted maximum likelihood estimator of Manski and Lerman (1977) is such a method. By weighting the observations appropriately, this method makes the response-based sample behave asymptotically as if it were a random sample. Hence the weighted maximum likelihood estimate of a logit model possesses the same best predictor interpretation as does maximum likelihood logit analysis in random samples.

This article elaborates on the second answer. First we formally state the problem of estimation from response-based samples and cite some key results from the literature. Then we qualitatively compare the limiting behavior of weighted and random sampling maximum likelihood estimates of a logit model when the true response function may not be logit. After that, we report quantitative findings on the asymptotic bias of the weighted and unweighted estimators. Finally we present a Monte Carlo experiment that provides evidence on small sample bias and precision.

ESTIMATION FROM RESPONSE-BASED SAMPLES

Let $x_i \epsilon R^M$ be a vector of variables measuring the ith individual's characteristics, where $i = 1, \ldots N$, and N is the sample size. Let the marginal density of x in the population be $p(x)$.[3] Let the binary response for the ith individual be $j \epsilon J$, where $J = 0, 1$ is the response set. The objective is to learn how the response probability $Pr(j \mid x)$ varies as a function of x. Assume tentatively that $Pr(j \mid x) = P(j \mid x, \theta^*)$, where P is a function known up to a vector of parameters $\theta^* \epsilon \Theta$, and Θ is a parameter space of finite dimension K. Given this, the problem of learning the response function reduces to one of estimating θ^*.

In the population the probability density of a (j, x) pair is

$$f(j, x) = P(j \mid x, \theta^*)p(x). \qquad [1]$$

We now want to distinguish three sampling schemes and derive the likelihood of an observation under each regime.

Random sampling. The likelihood of drawing an observation (j, x) is

$$\lambda_r = f(j, x) = P(j \mid x, \theta^*)p(x). \qquad [2]$$

Exogenous sampling. The population is stratified on the explanatory variable x. Let the sampling distribution of x be denoted g. An example is where the researcher oversamples small racial or ethnic groups. The likelihood of an observation is

$$\lambda_e = P(j \mid x, \theta^*)g(x). \qquad [3]$$

Response-based sampling. The population is stratified on the binary response variable j and the sampling proportions for $j = 0$ and $j = 1$ are different from those in the population. That is, we oversample one response group and undersample the other. Let N_j/N be the sampling proportion for response j. Let $Q(j)$ be the population probability of response j. The likelihood of an observation is then

$$\lambda_{rb} = Pr(x \mid j)\frac{N_j}{N} = \frac{P(j \mid x, \theta^*)\,p(x)\dfrac{N_j}{N}}{Q(j)}. \qquad [4]$$

The first equality of the equation follows from the definition of conditional probability. The second equality follows from Bayes Theorem.

Maximum likelihood estimation under response-based sampling is qualitatively different from such estimation under random or exogenous sampling. The usual presumption is that the research knows neither θ^* nor the marginal density p; thus p is a function-valued nuisance parameter. Inspection of equations 2 and 3 shows that under random and exogenous sampling the likelihood decomposes into the product of the response probability $P(j|x, \theta^*)$ and the density function p or g. In both cases, the maximum likelihood estimate for θ^* solves the problem.

$$\max_{\theta\epsilon\Theta} \sum_{i=1}^{N} \log P(j_i|x_i, \theta). \qquad [5]$$

On the other hand, inspection of equation 4 shows that, under response-based sampling, the likelihood does not decompose in this manner. The reason is that the quantity $Q(j)$ is implicitly dependent on both θ^* and on p, via the equation

$$Q(j) = \int P(j|x, \theta^*) p(x) dx, \qquad [6]$$

where the integration is over the entire space of x. Hence maximum likelihood estimation requires solution of the problem.

$$\max_{\theta\epsilon\Theta,\, \psi\epsilon\Psi} \sum_{i=1}^{N} \log \frac{P(j_i|x_i, \theta)\,\psi(x_i)}{\int P(j_i|x, \theta)\,\psi(x) dx}, \qquad [7]$$

where Ψ is the space of all possible density functions for x, $\psi(x)$.

It was shown by Manski and Lerman (1977) that, when a sample is response-based, ignoring the sampling process and applying the random sampling maximum likelihood method 5 generally yields inconsistent estimates. To correct for the inconsistency, Manski and Lerman (1977) proposed the weighted maximum likelihood estimator, which is easy to implement.[4]

The weighted maximum likelihood estimator solves the problem

$$\max_{\theta\epsilon\Theta} \sum_{i=1}^{N} w(j_i) \log P(j_i|x_i, \theta), \qquad [8]$$

where $w(j) = Q(j)/(N_j/N)$. Application of this estimator presumes knowledge of $Q(j)$ and N_j/N. The data reveal N_j/N. What is crucial is knowledge of the distribution of J in the population, $Q(j)$. Information about $Q(j)$ might be obtained in several ways. It could come from the census tabulations, from other social surveys, or from national estimates conducted by certain national organizations. Once $Q(j)$ is obtained, the weighted maximum likelihood estimator can be implemented using any computer software with a maximum likelihood estimation routine that allows for weighting.[5] A more detailed discussion of computer programs is given in the Appendix.

An exception to the general rule that random sampling maximum likelihood is inconsistent in response-based samples is the logit model with an intercept term. It can be shown that for this model, solution of problems 5 and 7 yields the same estimates for all components of θ^* except the intercept. Even for the intercept, the inconsistency problem is innocuous since the estimate can be corrected after estimation.[6] This result derives from the special mathematical form of the logistic distribution, which makes the logit model a member of the class of multiplicative intercept models (Hsieh et al., 1985).

ESTIMATION WHEN THE RESPONSE FUNCTION IS MISSPECIFIED

If the actual response function has the logit form, then one may estimate θ^* either by the weighted maximum likelihood method or by random sampling maximum likelihood, correcting the intercept ex post. Both approaches yield consistent, asymptotically normal estimates. The latter is asymptotically more efficient.

We are interested in the behavior of these approaches when the actual response function is not logit. Consider the weighted maximum likelihood estimate. Its probability limit is obtained by solving the following limiting version of equation 8:

$$\max_{\theta\epsilon\Theta} \int \left[P(0\,|\,x,\theta^*) \frac{N_0/N}{Q(0)} w(0) \log P^0(0\,|\,x,\theta) + P(1\,|\,x,\theta^*) \frac{N_1/N}{Q(1)} w(1) \log P^0(1\,|\,x,\theta) \right] p(x)dx, \qquad [9]$$

where P is the true probability function and P^0 is the logit probability function used in estimation. The weight $w(j)$ equals $Q(j)/(N_j/N)$. Hence, the problem reduces to

$$\max_{\theta\epsilon\Theta} \int \left[P(0\,|\,x,\theta^*) \log P^0(0\,|\,x,\theta) + P(1\,|\,x,\theta^*) \log P^0(1\,|\,x,\theta) \right] p(x)dx, \qquad [10]$$

the same limiting problem as solved by maximum likelihood under random sampling.

As is well-known, the solution to problem 10 is θ^* if the true response function P and the assumed one P^0 coincide. If P is not the same as P^0, problem 10 nevertheless retains a useful interpretation; it determines a constrained best predictor of the binary response. Efron (1978), Hastie (1987), Manski and Thompson (1989) observe that equation 10 can be rewritten as

$$\min_{\theta\epsilon\Theta} E[-\log\{1-|j-P^0(1\,|\,x,\theta)|\}], \qquad [11]$$

where the expectation is with respect to the true population density $P(j\,|\,x, \theta^*)p(x)$. Problem 11 defines a constrained best predictor of the binary response j, conditional on observation of x. Suppose that one wishes to obtain the best predictor of j within the class of predictor functions $P^0(1\,|\,x, \theta)$, $\theta \in \Theta$, when the loss incurred for prediction errors is $-\log\{1 - |j - P^0(1\,|\,x, \theta)|\}$. Then one would solve problem 11.

Now consider the random sampling maximum likelihood estimate of the logit model. Assume that as $N \rightarrow \infty$, the ratios N_0 / N and N_0 / N approach limits H_0 and H_1. Then under response-based sampling, the probability limit is obtained by solving the following limiting version of problem 5:

$$\max_{\theta\epsilon\Theta} \int \left[P(0 \mid x, \theta^*) \frac{H_0}{Q(0)} \log P^0(0 \mid x, \theta) \right. \qquad [12]$$

$$\left. + P(1 \mid x, \theta^*) \frac{H_1}{Q(1)} \log P^0(1 \mid x, \theta) \right] p(x)dx.$$

The intercept aside, the solution to problem 12 is θ^* if P is logit, as P^0. But problem 12 does not define a best predictor of j when P and P^0 do not coincide. The random sampling maximum likelihood logit estimate lacks any natural interpretation when the response function is not logit.[7]

ANALYSIS OF ASYMPTOTIC BIAS

This section provides a quantitative analysis of the asymptotic bias associated with the estimators 5 and 8 when the true response function is not logit. Assume that a researcher has a response-based sample and that the marginal distribution of the response $Q(j)$ (j = 0, 1) is known. Therefore, the weighted maximum likelihood estimator can be applied. The problem that the researcher faces is the following. If he or she knows that the true model is logit, he or she should run the random sampling maximum likelihood logit estimator in order to retain efficiency; if he or she does not know this, the discussion of the preceding section suggests that weighted maximum likelihood estimation is preferable.

In order to challenge the conventional wisdom most directly, we assume that the true response function is probit, a specification usually thought very close to the logit. In particular, we specify the following probit model: $y_i^* = a + bx_i + \epsilon_i$, where ϵ_i is distributed standard normal. We observe $y_i = 1$ if $y_i^* \geq 0$, and $y_i = 0$ otherwise. The parameters are set to $a = b = 1$. The independent variable x_i is distributed normally with a mean of α and a variance of one. The mean of x is allowed to vary in order to capture various ratios of $y = 1$ to $y = 0$. We choose three values for α: −2, −3, and −4. They give respectively 23.98%, 7.87%, and 1.69% of $y = 1$ cases in the population. Samples are assumed to be drawn as response-based, with equal numbers of $y = 1$ cases and of $y = 0$ cases. That is, $N_0/N = N_1/N = 0.5$.

For samples thus drawn, we obtain both the weighted and the random sampling maximum likelihood logit estimates. The limits of the estimators as the sample size approaches infinity are obtained by numerically solving equations 10 and 12, where $\theta = (a, b)$.

Table 1 gives the asymptotic biases of the two estimates for b.[8] Observe that generally the weighted logit estimate has smaller asymptotic bias. Such bias as the weighted logit method has is not due to the fact that the sample is response-based. It is rather due to the imperfect approximation of the probit model by the logit model. This is especially true when data are heavily concentrated at one of the tails (as in the case where $\alpha = -4$) since the normal distribution differs from the logistic distribution more at the tails than in the middle range (Johnson and Kotz, 1970: 1-21). The asymptotic bias of the weighted logit estimate under response-based sampling is the same as that of the usual unweighted logit estimate under random sampling. This is because, in the limit, the weighted logit estimator under response-base sampling solves the same problem as does the unweighted logit estimator under random sampling.

We also observe that the asymptotic bias of the random sampling logit estimator increases as α changes from −2 to −4, and the population probability Q_j further departs from the sampling probability H_j (0.5). This result is consistent with Cosslett's (1981a) conclusion. Treating response-based samples as if they

TABLE 1
Asymptotic Bias of Alternative Estimators when True Model is Probit

		Estimation Methods	
α	% of y=1	Unweighted Logit	Weighted Logit
-2	23.98	-0.02	-0.04
-3	7.87	0.12	0.05
-4	1.69	0.35	0.18

NOTE: Asymptotic biases for coefficient b are displayed. The true model is

$$y_i^* = a + bx_i + e_i,$$

where $a = b = 1$, $x_i \sim N(\alpha, 1)$. e_i is distributed as standard normal. We observe $y_i = 1$ if $y_i^* > 0$, and $y_i = 0$ otherwise. Asymptotic bias equals

$$plim_{N \to \infty}(b_N - b^*),$$

where b_N is the estimate based on N observations, b^* is the true parameter, and N is the sample size. Estimates were rescaled to be comparable to probit estimates.

were random samples becomes increasingly inappropriate as H_j moves away from Q_j.

There is no perfect way of rescaling parameters from logit to probit; our rule of $3^{1/2}/\pi$ is only intuitively appealing. It is possible that the comparisons reported in Table 1 might be confounded by the rescaling method that we chose. To compare further the performance of the two estimators, we carry out another exercise: comparing estimated probabilities of $y = 1$ as functions of the independent variable x. Specifically, we compute estimated probabilities $P(y = 1 \mid x, \hat{\theta})$, where $\hat{\theta} = (\hat{a}, \hat{b})$ is the limiting value of an intercept and slope estimate (a_n, b_n). The intercept term from the random sampling logit estimation is corrected using the formula (derived from Hsieh et al., 1985: 659):

$$a^* = a_{ul} + \log\left(\frac{Q(1)/H(1)}{Q(0)/H(0)}\right),$$

TABLE 2
Estimated Probability of $y = 1$ from Various Asymptotic Estimates (X 100)

	Estimation Methods			
	Unweighted Logit		Weighted Logit	
α	Case 1	Case 2	Case 1	Case 2
-2	51.0	14.9	50.4	15.0
	$(x = 1)$	$(x = 0)$	$(x = 1)$	$(x = 0)$
-3	57.5	15.0	53.2	14.5
	$(x = 2)$	$(x = 1)$	$(x = 2)$	$(x = 1)$
-4	73.2	19.0	61.1	15.1
	$(x = 3)$	$(x = 2)$	$(x = 3)$	$(x = 2)$

NOTE: Probabilities were calculated from asymptotic estimates. Case 1 and Case 2 are defined by varying the value of x. The x values were chosen as to make the true probabilities of $y = 1$ in the population 0.500 and 0.159 for the true probit model. For model specifications, see Table 1 and the text.

where α^* is the corrected intercept and α_{ul} is the estimated intercept. Otherwise, the estimates are not rescaled.

The results are given in Table 2. In Table 2, two values of x were chosen, giving probabilities of $y = 1$ of 0.500 and 0.159, respectively. These probabilities are computed under the correct model and the true parameters. We observe that the random sampling logit estimator leads to substantial overprediction in the case $\alpha = -4$, estimating 0.7832 and 0.190 instead of 0.500 and 0.159 in the population. Estimated probabilities from the weighted logit estimator are better—different from those in the population but not very far from them. This is so because the weighted logit estimator provides a best predictor of the probit function, within the constrained class of logit models.

THE MONTE CARLO EXPERIMENT

Results of asymptotic analysis do not suffice to guide data analysis. In practice, the sample size is finite, and the issue of efficiency can be important. In order to explore how the two estimators perform in small samples, we conducted a Monte Carlo experiment.

The Monte Carlo simulation experiment can be seen as an extension of the previous asymptotic analysis. The design of the experiment is simple: We repeatedly draw independent response-based samples of a fixed size according to the model specifications described in the last section and obtain the weighted maximum likelihood and random sampling maximum likelihood logit estimates for each sample. The sample size varies from 100, 200, to 1,000. For each model and sample size specification, 1,000 repetitions are performed. The actual distribution of the 1,000 estimates provides the basis for inference regarding the behavior of the estimators.

Table 3 presents the results of our analysis of the estimation of the slope coefficient *b*. For each sample size, the mean, the standard deviation, and the root mean square error (RMSE) of each estimator are displayed. The root mean square error is given because it is the measure that combines the bias and the variance of an estimator. The asymptotic limits of the estimates are also given for the purpose of comparisons.

From Table 3, we observe first of all that for samples of size $N = 1{,}000$ both of the estimators have more or less converged to their asymptotic values. In other words, the conclusions drawn from the asymptotic analysis are directly relevant to data analysis when the sample size is as large as 1,000. In particular, the random sampling logit estimator overall does not perform very well. For the smaller sample sizes of $N = 100$ and $N = 200$, however, this estimator is better than the weighted logit estimator by the RMSE criterion. This illustrates the trade-off for small samples between using an estimator that has larger asymptotic bias but smaller variance, on the one hand, and one that has smaller asymptotic bias but larger variance, on the other.

Because of the problem of rescaling, the results from Table 3 may be thought ambiguous. Following the approach of the last section, we would like to compare the estimates in a different way—comparing estimated probabilities of $y = 1$. For each sample size, we calculate the estimated probabilities by using

TABLE 3
Comparison of Estimates Under True Probit Models When Samples Are Response-Based: A Monte Carlo Experiment[a]

	Estimation Methods					
	Unweighted Logit			Weighted Logit		
N	Mean	St.Dev.	RMSE	Mean	St.Dev.	RMSE
$\alpha = -2$[b]						
100	1.03	0.22	0.23	1.02	0.23	0.23
200	1.01	0.14	0.15	0.99	0.15	0.16
1000	0.99	0.06	0.06	0.97	0.07	0.07
∞	0.98	0.00	0.02	0.96	0.00	0.04
$\alpha = -3$[c]						
100	1.18	0.25	0.30	1.15	0.33	0.36
200	1.15	0.17	0.22	1.09	0.21	0.23
1000	1.13	0.07	0.15	1.05	0.09	0.10
∞	1.12	0.00	0.12	1.05	0.00	0.05
$\alpha = -4$[d]						
100	1.47	0.33	0.57	1.57	0.64	0.86
200	1.41	0.21	0.46	1.33	0.36	0.48
1000	1.36	0.08	0.37	1.21	0.14	0.25
∞	1.35	0.00	0.35	1.18	0.00	0.18

a. Estimates of coefficient b are displayed. The true model is

$$y_i^* = a + bx_i + \epsilon_i,$$

where $a = b = 1$, $x_i \sim N(\alpha, 1)$, $\epsilon_i \sim N(0, 1)$, and x_i is independent of ϵ_i. We observe $y_i = 1$ if $y_i^* > 0$, and $y_i = 0$ otherwise. N is the sample size. In the population the ratio of $y = 1$ cases to $y = 0$ cases depends on the value of α. In all of our response-based samples, we sampled equal number of cases of $y = 1$ and $y = 0$. For each specification, 1,000 repetitions were performed. Estimates were rescaled to be comparable to probit estimates.

b. This gives in population roughly 23.98% cases $y = 1$ and 76.02% cases $y = 0$.

c. This gives in population roughly 7.87% cases $y = 1$ and 92.13% cases $y = 0$.

d. This gives in population roughly 1.69% cases $y = 1$ and 98.31% cases $y = 0$.

estimates from 1,000 repetitions. In this way, we obtain 1,000 estimated probabilities for each specification.

Table 4 reports the means, standard deviations, and RMSEs of these probabilities. As before, estimated probabilities from the

TABLE 4
Comparison of Estimated Probability of $y = 1$ from Various Estimates when the True Model is Probit (X 100)

	Estimation Methods											
	Unweighted Logit						Weighted Logit					
N	Mean	S.D.	RMSE	Mean	S.D.	RMSE	Mean	S.D.	RMSE	Mean	S.D.	RMSE
$\alpha = -2$												
		$x = 1$			$x = 0$			$x = 1$			$x = 0$	
100	52.0	7.6	7.9	14.5	2.6	2.9	51.6	7.8	7.9	14.6	2.6	2.9
200	51.7	5.1	5.4	14.8	1.7	2.0	51.2	5.3	5.4	14.9	1.8	2.0
1000	51.2	2.3	2.6	14.9	0.8	1.3	50.6	2.3	2.4	15.0	0.8	1.2
∞	51.0	0.0	1.0	14.9	0.0	1.0	50.4	0.0	0.4	15.0	0.0	0.9
$\alpha = -3$												
		$x = 2$			$x = 1$			$x = 2$			$x = 1$	
100	59.9	13.2	16.4	15.8	3.5	3.5	57.1	15.8	17.3	15.4	3.7	3.8
200	58.4	9.9	13.0	15.3	2.3	2.4	54.8	11.4	12.3	14.8	2.3	2.5
1000	57.5	4.6	8.8	15.0	0.9	1.3	53.4	5.3	6.2	14.5	0.9	1.6
∞	57.5	0.0	7.5	15.0	0.0	0.9	53.2	0.0	3.2	14.5	0.0	1.4
$\alpha = -4$												
		$x = 3$			$x = 2$			$x = 3$			$x = 2$	
100	75.1	16.5	30.1	24.2	13.1	15.5	72.3	23.1	32.1	26.8	19.9	22.7
200	73.9	13.1	27.3	21.3	7.8	9.5	65.7	20.0	25.4	19.4	10.3	10.9
1000	73.1	6.2	23.9	19.2	2.8	4.3	61.5	10.6	15.7	15.7	3.2	3.2
∞	73.2	0.0	23.2	19.0	0.0	3.1	61.1	0.0	11.1	15.1	0.0	0.8

NOTE: Probabilities were calculated from Monte Carlo and asymptotic estimates. The x values were chosen as to make the true probabilities of $y = 1$ in the population 0.500 and 0.159. For model specifications, see Table 3 and the text.

two logit estimates are presented. There are two cases with true probabilities of 0.500 and 0.159. Inspection of the table shows once again the trade-off between bias and precision in small samples. In large samples, weighted logit estimation generally yields superior results. In small samples, random sampling maximum likelihood performs somewhat better.

CONCLUSION

The logit specification is often chosen only for its convenience. When this is the case, choosing the logit specification does not permit the researcher to analyze response-based data as if they were generated by a random sampling process. Application of random sampling methods is proper only if the researcher believes that the true response function takes the logit form.

When the true response function is not logit, the random sampling maximum likelihood logit estimator yields results that do not describe the data in an interpretable way.

One solution to this problem is to apply the weighted maximum likelihood estimator of Manski and Lerman (1977). This article demonstrates that the weighted maximum likelihood estimator provides good results for reasonably large samples even when the response function is misspecified. The weighted maximum likelihood analysis is preferable because it estimates a constrained best predictor of the binary response. Alternatively, the researcher may apply a semiparametric method that avoids the need to specify the exact form of the response function.

APPENDIX

This appendix provides some information on computer programs. The weighted maximum likelihood estimator of Manski and Lerman (1977) can be implemented using any computer software with a maximum likelihood estimation routine that allows for weighting. Parameter estimates can be obtained as follows: (1) calculate a weight variable that equals

$$Q(1)\Big/\frac{N_1}{N} \text{ for } y = 1 \quad \text{and} \quad Q(0)\Big/\frac{N_0}{N}$$

otherwise, (2) specify the model to be estimated, and (3) estimate the model, multiplying each observation's contribution to the log-likelihood by the appropriate weight.

In general, the covariance matrix of the estimated parameters reported by the computer program performing this simple procedure is incorrect. See Manski and Lerman (1977), Cosslett (1981a), and Hsieh et al. (1985) for discussions of the correct covariance matrix. The formulas appearing in the above references vary, depending on the nature of the researcher's knowledge of $Q(j)$ and whether N_j/N is fixed by design. In the following, we will only discuss the simplest case, that is, when $Q(j)$ is exactly known, and N_j/N is allowed to vary around its known limit $H(j)$ (Manski and Lerman, 1977: Amemiya, 1985: 322-328).

Let $w(j) = Q(j)/H(j)$. The asymptotic covariance matrix of the estimator (θ) solving equation 8 can be estimated by:

$$\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} \qquad [13]$$

where

$$\mathbf{A} = \frac{1}{N}\sum_{i=1}^{N} w(j_i)\,\frac{\partial \log P(j_i|x,\hat{\theta})}{\partial\theta}\,\frac{\partial \log P(j_i|x,\hat{\theta})}{\partial\theta'} \qquad [14]$$

$$\mathbf{B} = \frac{1}{N}\sum_{i=1}^{N} w(j_i)^2\,\frac{\partial \log P(j_i|x,\hat{\theta})}{\partial\theta}\,\frac{\partial \log P(j_i|x,\hat{\theta})}{\partial\theta'} \qquad [15]$$

Notice that the $\mathbf{B}^{-1}$ matrix is the covariance matrix usually reported by the computer program performing a weighted maximum likelihood estimation.

For the probit and logit models, the formula is simple. In the case of the logit model, for example, $\partial \log P/\partial\theta = \theta(1 - P)$. Computer codes for obtaining the correct covariance matrix are available in Greene's LIMDEP manual (1986: 19.3-7). Below is an example implementing the weighted maximum likelihood logit estimator in LIMDEP. Adaptation to other standard computer software should be straightforward.

```
type  ;  This is to estimate a logit model via weighted ML $
read  ;nrec=200 ;nvar=2 ;file=udiskl:for010.dat
      ;names(x1=y,x2=x2) ;format=(F2.0,F8.5) $
create ;if (y=1) cbswt=0.0787/0.5 ;(else) cbswt=0.9213/0.5 $
namelist ;x=one,x2 $
logit ;lhs=y ;rhs=x ;wts=cbswt ;keep=yfit ;matrix(b=b,H=H) $
create ;derivs=xyfit $
create ;derivs=derivs^2*cbswt $
matrix ;G=xdot(x,derivs) ;V=qfrm(G,H) ;V=sinv(V) ;stat(b,V) $
stop
```

NOTES

[1] For general discussions of response-based samples, see Manski (1981) and Bye et al. (1987). Other terms for "response-based sample" are "choice-based sample" (Manski and Lerman, 1977) and "case-control sample" (Prentice and Pyke, 1979).

[2] In log-linear analysis, it is well-known that interaction terms are invariant to changes in marginal distributions and are consequently not affected by response-based sampling.

[3] For simplicity, we consider the case in which x has an "ordinary" density; that is, a density with respect to Lebesgue measure. In fact, x can have discrete components. If so, integrals appearing in the later equations should be changed to sums. No other changes are needed.

[4] Another easy-to-implement method is the Manski-McFadden (1981) conditional maximum likelihood estimator. In the case of a logit model, this procedure and the full-information maximum likelihood method 7 coincide. Thus it need not be considered here.

[5] We shall, for simplicity, assume that $Q(j)$ is known exactly. In fact, it is enough that one be able to estimate $Q(j)$, say from an auxiliary random sample. See Hsieh et al. (1985) for details.

[6] The appropriate correction is to add

$$\log\left(\frac{Q(1)/N_1}{Q(0)/N_0}\right)$$

to the estimated intercept. Note that implementation of the correction requires knowledge of the marginal population response probabilities $Q(j)$.

[7] Problem 12 can be written in the form of equation 11 but the expectation is not with respect to the true population density. It is, rather, with respect to the misspecified density in which the response probability is

$$\frac{P(j\,|\,x, \theta^*)H_j/Q_j}{P(0\,|\,x, \theta^*)H_0/Q_0 + P(1\,|\,x, \theta^*)H_1/Q_1}$$

rather than $P(j\,|\,x, \theta^*)$, and the marginal density of x is

$$p(x)\left[P(0\,|\,x, \theta^*)\frac{H_0}{Q_0} + P(1\,|\,x, \theta^*)\frac{H_1}{Q_1}\right]$$

rather than $p(x)$.

[8] To calculate these biases, the logit results have been multiplied by $3^{1/2}/\pi$. The logit and probit models employ different scale normalizations. Whereas the standard normal distribution has variance one, the standard logistic distribution has variance $\pi^2/3$.

REFERENCES

AMEMIYA, T. (1985) Advanced Econometrics. Cambridge, MA: Harvard Univ. Press.

BISHOP, Y.M.M., S. E. FIENBERG, and P. W. HOLLAND (1975) Discrete Multivariate Analysis: Theory and Practice. Cambridge: MIT Press.

BYE, B. V., S. J. GALLICCHIO, and J. M. LEVY (1987) "Estimation of discrete choice models in retrospective samples: application of the Manski and McFadden conditional maximum likelihood estimator." Soc. Methods & Research 15: 467-492.

COSSLETT, S. R. (1981a) "Efficient estimation of discrete-choice models," pp. 51-111 in C. F. Manski and D. McFadden (eds.) Structural Analysis of Discrete Data with Econometric Applications. Cambridge: MIT Press.

COSSLETT, S. R. (1981b) "Maximum likelihood estimator for choice-based samples." Econometrica 49: 1289-1316.

COX, D. R. (1970) The Analysis of Binary Data. London: Methuen.

EFRON, B. (1978) "Regression and ANOVA with zero-one data: measures of residual variation." J. of the Amer. Stat. Assn. 73: 113-121.

GORTMAKER, S. L. (1979) "Poverty and infant mortality in the United States." Amer. Soc. Rev. 44: 280-297.

GREENE, W. H. (1986) LIMDEP: User's Manual.

HASTIE, T. (1987) "A closer look at the deviance." Amer. Statistician 41: 16-20.

HSIEH, D. A., C. F. MANSKI, and D. McFADDEN (1985) "Estimation of response probabilities from augmented retrospective observations." J. of the Amer. Stat. Assn. 80: 651-662.

JOHNSON, N. L. and S. KOTZ (1970) Continuous Univariate Distributions—2. New York: John Wiley.

MADDALA, G. S. (1983) Limited-dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge Univ. Press.

MANSKI, C. F. (1981) "Structural models for discrete data: the analysis of discrete choice," pp. 58-109 in S. Leinhardt (ed.) Sociological Methodology 1981. San Francisco, CA: Jossey-Bass.

MANSKI, C. F. (1986) "Semiparametric analysis of binary response from response-based samples." J. of Econometrics 31: 31-40.

MANSKI, C. F. (1988) "Identification of binary response models." J. of the Amer. Stat. Assn. 83: 729-738.

MANSKI, C. F. and S. R. LERMAN (1977) "Estimation of choice probabilities from choice-based samples." Econometrica 45: 1977-1989.

MANSKI, C. F. and D. McFADDEN (1981) "Alternative estimators and sample designs for discrete choice," pp. 2-50 in C. F. Manski and D. McFadden (eds.) Structural Analysis of Discrete Data with Econometric Applications. Cambridge: MIT Press.

MANSKI, C. F. and T. S. THOMPSON (1989) "Estimation of best predictors of binary response." J. of Econometrics 40: 97-123.

PRENTICE, R. L. and R. PYKE (1979) "Logic disease incidence models and case-control studies." Biometrika 66: 403-411.

Yu Xie is a doctoral candidate in sociology at the University of Wisconsin—Madison. He is studying entry into scientific careers by combining the 1962 and 1973 Occupational Changes in a Generation Surveys with 1962 and 1972 Postcensal Surveys of scientific and technical personnel.

Charles F. Manski is a Professor of Economics at the University of Wisconsin—Madison. His recent book **Analog Estimation Methods in Econometrics** *was published by Chapman and Hall, 1988. His current research includes development of dynamic choice models and study of schooling behavior.*