ELSEVIER

Journal of Econometrics 79 (1997) 291–303

JOURNAL OF
Econometrics

Bayes WESML
Posterior inference from choice-based samples

Tony Lancaster

Department of Economics, Box B, Brown University, Providence, RI 02912, USA

Abstract

In this paper I show that the Weighted Exogenous Sampling Likelihood (WESML) estimator for choice-based samples due to Manski and Lerman (1977) can be given a Bayesian interpretation under a multinomial model for the data with improper Dirichlet priors. Bayesian posterior distributions of choice model parameters are computed to study the gains from choice-based sampling and the effect of knowledge of population marginal choice probabilities.

Keywords: Choice-based sampling; Case-control sampling; Bayesian bootstrap; Discrete choice; Probit
JEL classification: C21, C25; C42

0. Introduction and summary

This paper deals with inference from choice-based samples. The problem arises in the following way. There exists a finite collection of mutually exclusive choices. The econometrician is prepared to entertain a stochastic model showing the probability with which an agent makes each choice as a function of an observed covariate vector, x, which may vary over both choices and agents. This model, which describes the conditional probability of each choice given the covariate vector, is specified up to a finite parameter, θ. The familiar *logit* or *probit* models for binary choice are examples of such models. In contrast, the distribution of the covariates, x, is unknown to the investigator.

I acknowledge support from National Science Foundation award SES 9122477.

PII S0304-4076(96)00024-9

Data – observations on the choice, y, and the covariate x – are obtained by stratifying the population on the basis of the choice each member made, and then taking random samples of predetermined sizes from each choice stratum. This is what econometricians call choice-based sampling. A particular case is that in which the strata correspond to the choices and we call this pure choice-based sampling. With just two choices pure choice-based sampling amounts to taking a random sample of size N_1 from those members of the population who made choice 1, and then taking a random sample of size N_0 from those who made choice 0.

The sampling scheme, usually implicit, in textbook discussions of regression estimation is one of random sampling from the whole population, or possibly random sampling from within strata defined by the covariates or other exogenous variables. Under such schemes the distribution of the covariate enters the likelihood function multiplicatively and can be ignored so far as inference about θ is concerned. But under choice-based sampling – and more generally under any sampling scheme involving random sampling from within strata defined by the dependent variable of the model of interest – the covariate distribution does *not* factor out of the likelihood.[1] This implies that the inference problem is semiparametric. The likelihood function involves both an unknown finite parameter θ and an unknown function – the distribution of the covariate.

There are many solutions to this difficult inference problem, early work, by Cosslett among others, being lucidly summarised in Manski and McFadden (1981), and the most recent contribution provided by Imbens (1992). The most widely used estimator of θ was given as long ago as 1977 by Manski and Lerman.[2] It involves maximizing a reweighted version of the random or exogenous sampling likelihood function where the weights depend upon both the stratum sizes – the N_y – and the marginal choice probabilities Q_y. These latter are the fractions of the whole population making each choice. They are assumed known a priori or consistently estimated from an independent sample. Given knowledge of the weights the Manski–Lerman, or Weighted Exogenous Sampling Maximum Likelihood (WESML) estimator can be calculated using standard software. The WESML estimator is consistent for θ.

All existing econometric work on the choice-based sampling problem is written in the classical or frequentist tradition. The present paper offers a Bayesian approach to the problem. I do this not only out of idle curiosity and ideological conviction but also because, in inference from choice-based samples, prior information plays a crucial role. This is because 'knowledge' of the marginal choice probabilities is necessary to identification. Such knowledge cannot come from the choice-based sample – 100 observations of those who

[1] See, for example, Amemiya (1985) on this point.

[2] See also Hsieh et al. (1985).

travel by bus and 100 from those who do not reveals absolutely nothing about the fraction of the population who travel by bus.[3] Knowledge of the Q_y must come from outside the sample and it is in handling such extra-sample information that a Bayesian approach is particularly straightforward.

The essential points of this paper can be briefly summarised as follows. We 'solve' the semiparametric inference problem by postulating multinomial forms for the distributions of the covariate vector given the choice. The distributions are supported on L *known* points with probabilities $p_{x|y}$ that are *unknown*. Since the choice is also a discrete random variable the full data distribution is then completely discrete. We then *define* the parameter θ of the choice model as a function of the probabilities in this discrete data distribution – Eq. (2). We form the likelihood for the probabilities $p_{x|y}$ on the basis of the choice-based sample and the likelihood for the marginal choice probabilities[4] q_y using independent auxiliary information. Multiplying by a prior for these probabilities which can, in particular, incorporate prior information about the Q_y, yields the posterior distribution of all the probabilities appearing in the joint distribution of y and x. Since θ is defined by these probabilities this yields a posterior distribution for θ which is what is required. When the prior distribution for the conditional probabilities of the covariate given the choice is the improper Dirichlet distribution their posterior distribution assigns zero posterior probability density to points of support not realised in the sample. The resulting posterior distribution for θ is the Bayesian bootstrap distribution (Rubin, 1981). Its approximate posterior expectation is the WESML estimator, when the Q_y are 'known'. We thus achieve a Bayesian interpretation for this estimator and an exact and readily calculated posterior distribution that can be used for making Bayesian inferences in the spirit of the WESML procedure. Moreover this distribution can be calculated whatever the extent of our prior knowledge of the Q_y.

We provide a small numerical experiment to study the effect on the posterior distribution of θ of varying amounts of prior information about Q_y and various sample designs as measured by the relative magnitude of N_1 and N_0. The formal development of the posterior is given in Sections 1–3. Section 4 gives the numerical experiment, and Section 5 concludes with a criticism and some generalisations.

1. The inference problem

We shall consider binary choice with covariate vector x and choice indicator y. The vector $z = (y, x)$ is distributed over the relevant population with

[3] Except that it is neither zero nor one!

[4] Lower case q refers to the random variable whose true value is Q.

distribution function P, an element of a set of probability measures Π. The sample space for X is denoted $\mathscr{X}$. The conditional probability of choice 1 in the population given $X = x \in \mathscr{X}$ is specified up to a finite parameter θ as $Pr(y = 1|x, \theta) = P_{1x}(\theta)$ where $\theta \in \Theta$, a parameter space.

We assume that θ is uniformly identifiable relative to (Π, Θ) so we may write $\theta = t(P)$ where $t(*)$ maps Π onto Θ (Manski, 1988). We define θ as a functional of P by

$$\theta = \operatorname{argmax}_{c \in \Theta} \int y \log P_{1x}(c) + (1 - y) \log P_{0x}(c) \, \mathrm{d}P, \tag{1}$$

where $P_0 = 1 - P_1$. That is, for every $P \in \Pi$, θ maximises the expected log conditional likelihood. This is a *best fit* definition of θ in the sense that this parameter is defined to be the vector that minimises the Kullback–Liebler or information theoretic distance between the parameterized model and the true, but unknown, data distribution.

The right-hand side of (1) is $t(P)$. A probability distribution over Π, if one could be defined, induces a probability distribution over Θ. To get such a distribution we choose to restrict Π to be the set of multinomial distributions on $2(L + 1) < \infty$ points of support. Where L is a suitably large number. If L is sufficiently large this is no practical restriction (Efron, 1982). Under this restriction (1) becomes

$$\theta = \operatorname{argmax}_{c \in \Theta} \sum_{y} \sum_{x} p_{yx} [y \log P_{1x}(c) + (1 - y) \log P_{0x}(c)], \tag{2}$$

where $p_{yx} = Pr(Y = y, X = x)$ in the multinomial distribution $P \in \Pi$. The summations are over $y \in (0, 1)$, $x \in (x^0, x^1, \ldots, x^L) = \mathscr{X}$. We assume that both L and $\mathscr{X}$ are known a priori. A probability distribution over Π then requires only a distribution over $\{p_{yx}\}$, $y \in (0, 1)$, $x \in \mathscr{X}$.

2. The posterior distribution of $\{p_{yx}\}$

Assume data are gathered by pure choice-based sampling in which independent random samples of sizes N_y are taken from the two subsets of the population having $Y = y$, $y \in (0, 1)$. Factor p_{yx} as

$$p_{yx} = p_{x|y} q_y,$$

where $q_y = Pr(Y = y)$, $y \in (0, 1)$ are marginal choice probabilities in the population. The likelihoods from these random samples are

$$\mathscr{L}_y = \prod_{x \in \mathscr{X}} p_{x|y}^{n_{yx}}, \quad y = 0, 1, \tag{3}$$

where n_{yx} is the sample number of people making choice y and having $X = x$.

Next we assume that an independent, auxiliary, random sample of size N_q is taken from the whole population and only y is observed. This provides the basis for a posterior distribution for q_1. The likelihood function

$$\mathscr{L}_q = q_1^{n_1} q_0^{n_0}, \tag{4}$$

where n_y is the number of sampled people having $Y = y$ and $n_1 + n_0 = N_q$.

The total likelihood is the product of (3) and (4). Finally, assume that $\{p_{x|1}\}$, $\{p_{x|0}\}$ and q_1 are independent a priori, the former with the (improper) Dirichlet distributions

$$\prod_{x \in \mathscr{X}} p_{x|y}^{-1}, \quad y = 0, 1, \tag{5}$$

and the latter with the Dirichlet (Beta) distribution

$$q_1^{l_1} q_0^{l_0}, \quad l_0, l_1 \geqslant -1. \tag{6}$$

Then $\{p_{x|1}\}$, $\{p_{x|0}\}$, q_1 are independently distributed with Dirichlet distributions

$$p(\{p_{x|1}\}, \{p_{x|0}\}, q_1 | data) \propto \prod_{x \in \mathscr{X}} p_{x|1}^{n_{1x}-1} \prod_{x \in \mathscr{X}} p_{x|0}^{n_{0x}-1} q_1^{n_1+l_1} q_0^{n_0+l_0}. \tag{7}$$

This posterior distribution assigns zero probability density to points in $\mathscr{X}$ that were not observed realized in the sample.[5] It follows that neither L nor all elements of $\mathscr{X}$ need be specified, a priori, by the investigator. Note that

$$\mathscr{E}(p_{x|y} | data) = n_{yx}/N_y, \quad \mathscr{E}(q_1 | data) = (n_1 + l_1 + 1)/(N_q + l_0 + l_1 + 2). \tag{8}$$

3. The Bayesian bootstrap

The distribution (7) implies a distribution for θ defined by (2). While it is possible in principle to deduce the exact posterior distribution of θ and one might also, by taking multivariate normal approximations to the joint distributions of the $\{p_{y|x}\}$ and of q_1, calculate an approximation to the posterior distribution of θ, it is much simpler to sample, repeatedly, from this distribution. To sample from this distribution take a realization of $\{p_{x|1}\}$, $\{p_{x|0}\}$, q_1 from (7) and solve for θ. Many repetitions of this procedure provide the posterior

[5] To elaborate on this point, consider a proper Dirichlet prior for $p_{x|1}$, $x \in \mathscr{X}$, say, with the same parameter ε attached all L probabilities. Then by taking ε arbitrarily small the marginal joint posterior probability density of those $p_{x|1}$ that correspond to values of x contained in $\mathscr{X}$ but not observed in the choice 1 sample can be made arbitrarily close to zero for values of those probabilities greater than zero.

distribution of θ. A posterior distribution calculated in this way and using the prior distributions described in Section 2 is the Bayesian bootstrap distribution, Rubin (1981).

Since $\{p_{x|1}\}$, $\{p_{x|0}\}$, $\{q_y\}$ are independent it suffices to sample separately from the three Dirichlet distributions that are the components of (7). Rubin describes a method for sampling from such a distribution of the general form

$$p(\pi) \propto \prod_{k=1}^{K} \pi_k^{n_k - 1}. \tag{9}$$

Let $N = \sum_{k=1}^{K} n_k$. Draw $N - 1$ Uniform $(0, 1)$ variates and order them as $u_1, u_2, \ldots, u_{N-1}$. Let $\{g_n\}$ be the gaps defined as $g_1 = u_1$; $g_n = u_n - u_{n-1}$, $n = 2, \ldots, N - 1$; $g_N = 1 - u_{N-1}$. Partition the $\{g_k\}$ into K collections, the kth having n_k elements and let P_k be the sum of the gaps in the kth collection. Then $(P_1, \ldots, P_K)$ follows the Dirichlet distribution (9). This is the method used in the calculations reported in Section 5. For example, to generate a sample from the posterior distribution of $p_{x|1}$ we perform this calculation with the $\{n_k\}$ equal to the frequencies of each distinct value of the covariate realized in the sample of people who made choice 1, there being K such distinct values.

What is the connection between this calculation and the WESML estimator? When $N_q \to \infty$ the posterior distribution of q_1 concentrates on the true marginal probability, Q_1. When this is so *and when the* $\{p_{x|y}\}$ *are equal to their posterior expectations*, given by (8), the function $t(P)$ is

$$\frac{Q_1}{N_1} \sum_x n_{1x} \log P_{1x}(c) + \frac{Q_0}{N_0} \sum_x n_{0x} \log P_{0x}(c). \tag{9'}$$

This is the criterion function that defines the WESML estimator when the Q_y are known. It is a reweighted version of the θ log likelihood that would arise if the data were sampled randomly, or with exogenous stratification, the weights being Q_1/N_1 and Q_0/N_0.[6] The WESML estimator, with Q_y known, thus emerges as a first-order approximation to the posterior expectation of θ under the Dirichlet prior assumptions.

Realizations of the $\{p_{x|y}\}$ correspond approximately, (because the $\{p_{x|y}\}$ are not generally ratios of integers), to bootstrap samples from the x's realised in the two choice-based sub-samples. Thus the Bayesian bootstrap is approximately the same as a frequentist boot-strap for the WESML estimator when the marginal probability is known. The Bayesian procedure readily allows the incorporation of prior beliefs about the population marginal choice probability.[7]

[6] Under random sampling $N_1 \sim NQ_1$ and $N_0 \sim NQ_0$ and the weights are equal.

[7] The WESML procedure has been extended to allow to for uncertain prior information about the marginal probabilities in Hsieh et al. (1985).

4. An application

In this Section I use the Bayesian bootstrap to explore some aspects of the design of a choice-based sample. I assume a Probit model $P_{1x}(\theta) = \Phi(\theta_0 + \theta_1 x)$ for x scalar. I chose $\theta_0 = -2$, $\theta_1 = 1$ and let the support of X be 9 points equally spaced from -2 to 2. The distribution of X is symmetric and unimodal on this set. The implied marginal probability of choice 1 is $Q_1 = 0.08$ so the event is rare, a circumstance in which use of choice-based sampling might be reasonable. The experiments proceed by drawing a sample from the population described by the above distribution of X and conditional choice probability model according to the specified sampling scheme and then calculating the posterior distribution of the θ's given the data thus drawn.

Experiment 1. In this experiment I study the effect of varying precision of knowledge of Q_1 in combination with a small, balanced, choice-based sample in which $N_1 = N_0 = 20$.

I compare the case $N_q = 1000$ which corresponds to rather precise[8] knowledge of Q_1 with the case $N_q = 0$. In both instances $l_0 = l_1 = 0$ so that $N_q = 0$ implies a uniform prior distribution of q_1 on the unit interval. In frequentist terms, when Q_1 is unknown, which presumably corresponds to $N_q = 0$ and l_0, $l_1 = -1$, then the intercept in the choice model is identified only by functional form. In particular, had the choice model been logit rather than probit, the intercept would have been unidentifiable from this particular choice-based sample design. Any positive posterior precision for θ_0 when $N_q = 0$ will be a consequence of (a) the flat prior and (b) the non-linearity of the choice model log odds as a function of x, i.e. the departure from logit form.

The results are displayed in Figs. 1a and b.[9] The first row in each figure refers to the case $N_q = 1000$ and the second to $N_q = 0$. For both sets of calculations the n_{yx} were identical so the only difference between the rows lies in the amount of information about Q_1.

Fig. 1a tells us that the effect of (almost) knowing the population marginal probability is to improve dramatically the precision with which we can know the probit intercept, though it also skews the posterior distribution. Fig. 1(b) tells us that knowledge of the marginal probability has little effect on the precision of the slope, though it again introduces some non-Normality.

Experiment 2. In this experiment we compare the precision obtainable with a balanced choice-based sample to that to be found from a sample whose

[8] The posterior standard deviation of q_1 is 0.009.

[9] All density plots have been smoothed. The graphics were prepared using SPlus.

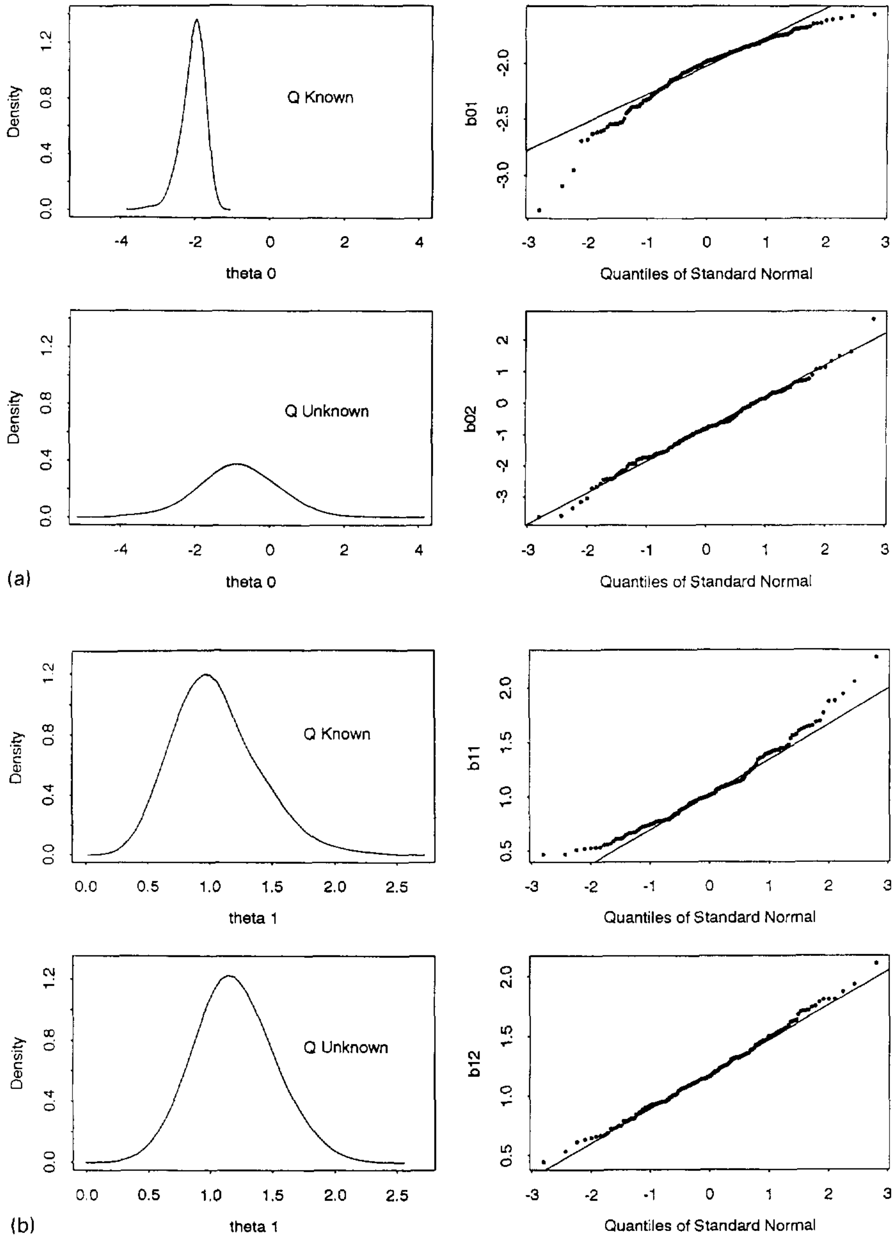
Density
0.0
0.4
0.8
1.2
Q Known
-4
-2
0
2
4
theta 0
b01
-3.0
-2.5
-2.0
-3
-2
-1
0
1
2
3
Quantiles of Standard Normal
Density
0.0
0.4
0.8
1.2
Q Unknown
-4
-2
0
2
4
theta 0
b02
-3
-2
-1
0
1
2
-3
-2
-1
0
1
2
3
Quantiles of Standard Normal
(a)
Density
0.0
0.4
0.8
1.2
Q Known
0.0
0.5
1.0
1.5
2.0
2.5
theta 1
b11
0.5
1.0
1.5
2.0
-3
-2
-1
0
1
2
3
Quantiles of Standard Normal
Density
0.0
0.4
0.8
1.2
Q Unknown
0.0
0.5
1.0
1.5
2.0
2.5
theta 1
b12
0.5
1.0
1.5
2.0
-3
-2
-1
0
1
2
3
Quantiles of Standard Normal
(b)

Fig. 1.

frequencies mimic random sampling in that the frequency of choice 1 observattions is the same as it is in the population. The purpose of the calculation is to assess the gain from balanced choice-based sampling.

The total sample size is 1000. In the balanced choice-based sample there are 500 observations from each choice. In the 'random' sample the fraction of choice 1 observations is equal to the marginal choice probability in the population, namely 0.08. So the sample sizes are 80 and 920. In both sets of calculations $N_Q = 1000$ so that the population marginal probability is known rather precisely. The results are shown in Fig. 2a, which refers to the intercept, and in Fig. 2b, which refers to the slope. The first rows give the balanced case and the second rows the 'random' sampling case.

Fig. 2a shows the gain in precision and in nearness to Normality of the posterior distribution of the intercept due to the balanced design. The reduction in inter-quartile range is about 25%. Fig. 2b shows very similar effects for the slope.

Fig. 3 shows the joint distributions of slope and intercept in the various experiments. It appears that knowledge of Q induces a marked negative correlation between slope and intercept. The effect seems strongest when the sample is small or large but unbalanced.

The results described above have the usual drawback of monte carlo work that they refer only to one particular experimental set-up. In addition they also refer to only a small number of sample realisations. We have replicated the calculations for alternative samples from the same model and found no reason to think the figures given are untypical. Alternative models would have different population x distributions, different values of θ and could of course have more than one covariate. They could also involve more flexible parametric forms than the probit. Bayesian bootstrapping such alternative models presents no difficulty other than that presented by repeated calculation of the WESML estimator. Such calculations might well be desirable for investigators proposing to design a survey in that they should provide insight into the likely posterior precision afforded by alternative designs.

5. Concluding remarks

5.1. Design

Our results suggest that knowledge of the marginal choice probabilities largely affects the precision with which the intercept in the choice model can be estimated and has little value for estimating the covariate effects. Precise knowledge of the marginal probability seems to introduce severe non-Normality into the posterior distributions except for large and balanced samples suggesting that large sample normal approximations need to be applied with caution.

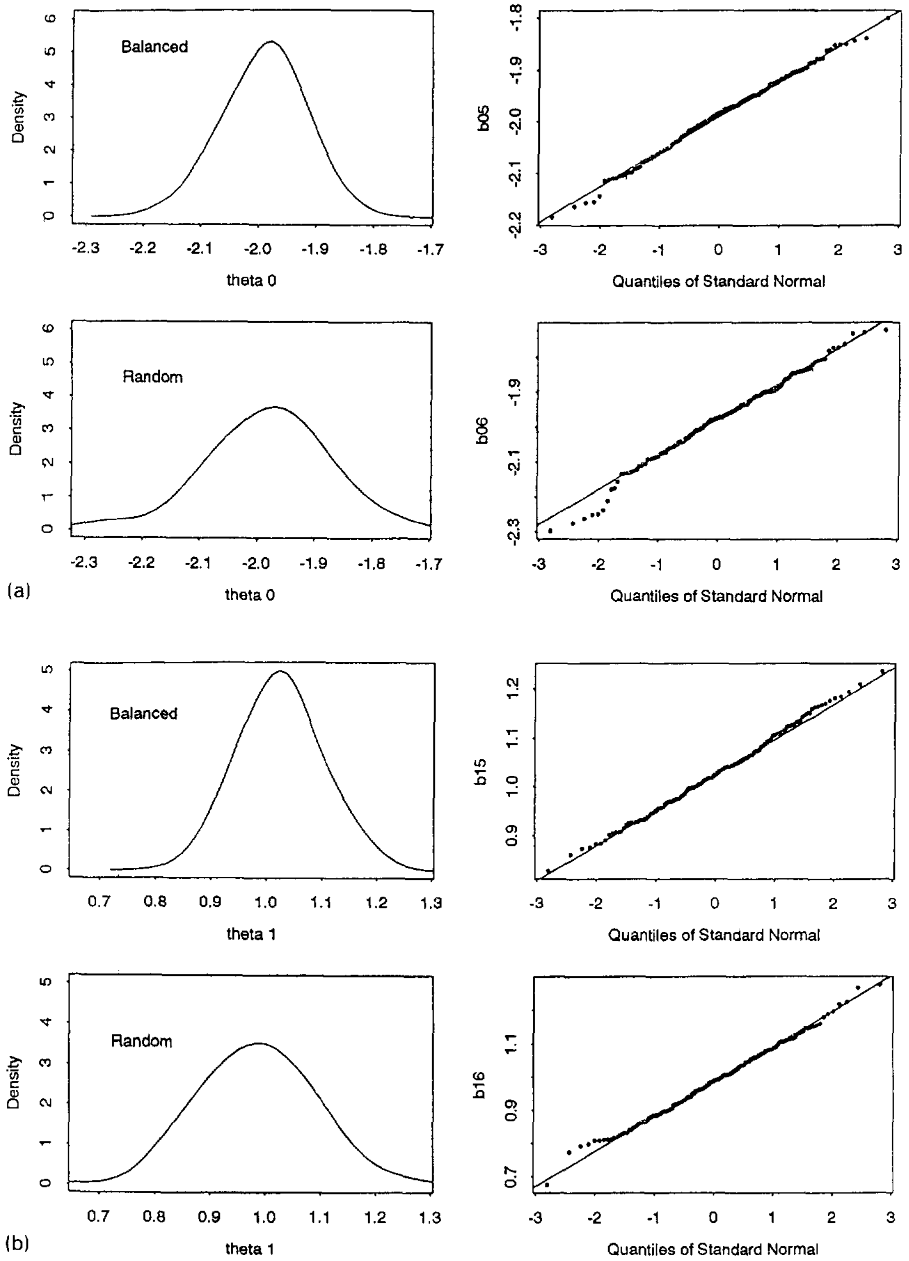
Balanced
Density
0 1 2 3 4 5 6
-2.3 -2.2 -2.1 -2.0 -1.9 -1.8 -1.7
theta 0
b05
-2.2 -2.1 -2.0 -1.9 -1.8
-3 -2 -1 0 1 2 3
Quantiles of Standard Normal
Random
Density
0 1 2 3 4 5 6
-2.3 -2.2 -2.1 -2.0 -1.9 -1.8 -1.7
theta 0
b06
-2.3 -2.1 -1.9
-3 -2 -1 0 1 2 3
Quantiles of Standard Normal
(a)
Balanced
Density
0 1 2 3 4 5
0.7 0.8 0.9 1.0 1.1 1.2 1.3
theta 1
b15
0.9 1.0 1.1 1.2
-3 -2 -1 0 1 2 3
Quantiles of Standard Normal
Random
Density
0 1 2 3 4 5
0.7 0.8 0.9 1.0 1.1 1.2 1.3
theta 1
b16
0.7 0.9 1.1
-3 -2 -1 0 1 2 3
Quantiles of Standard Normal
(b)

Fig. 2.

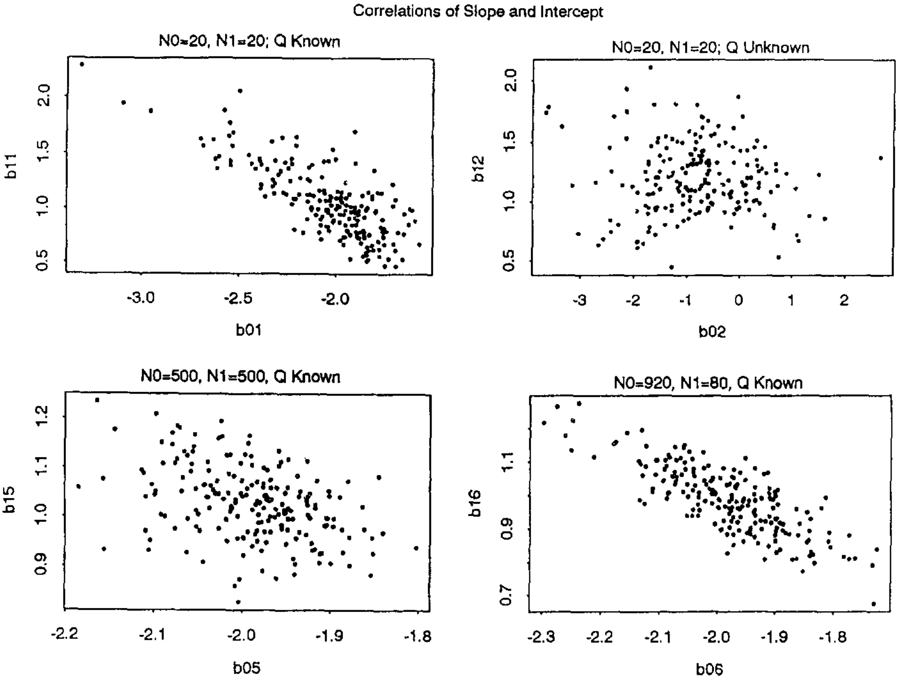

Fig. 3.

When the marginal choice probability is small the gain from a balanced sample – which is not necessarily the optimal design – as compared to a random sample can be of the order of a 30% reduction in the posterior standard deviation. For marginal probabilities smaller than 0.08 the gain is presumably greater.

It should be emphasised that these conclusions refer only to binary choice and to the probit model. Multinomial choice and/or models other than probit might lead to other conclusions.

5.2. Criticism and extensions

The Bayesian bootstrap does depend on the improper prior (5) without which it would be necessary to specify the support of the covariate distribution. This is the Bayesian counterpart of the standard criticism of the frequentist bootstrap that it essentially supposes the support of the data to be that which is observed in sample.[10] The Bayesian boot-strap, because of its dependence on the

[10] Rubin provides a fuller criticism of both the Bayesian and the standard bootstrap.

improper Dirichlet prior, cannot be regarded as a fully satisfactory inference procedure. Nevertheless, because of the ease with which it allows the investigator to incorporate the widely available, and essential, prior information about marginal choice probabilities, it is an appealing, if flawed, solution to the problem of Bayesian inference from choice-based samples.

The method described above extends to models with more than two choices and to any sampling design stratified on the choice set. It also extends to situations that do not fall strictly within the choice-based sampling framework. An example of this is the so-called contaminated sampling set-up in which the investigator has two random samples, one from the people who made choice 1 and the other from the whole population. *In the second sample only the covariate and not the choice is observed.* For example, one might have a random sample of female employees and a random sample of women of working age with no information on who worked. Since the fraction of women who participate in the labor force, Q_1, is a fairly accurately known macroeconomic statistic this information can be combined with sample data to construct posterior distributions of the parameters of a model for female labor force participation.[11]

An apparent drawback of the Bayesian bootstrap is that it does not let the investigator incorporate into the inference proper prior beliefs about θ. Ways of doing this are being addressed in current research.

The method also applies to the analysis of data gathered by random sampling and gives a way of incorporating prior information about the *marginal* distribution of the dependent variable into an analysis. Consider a random sample of size N from a multinomial population in which p_{yx} is the joint probability of $Y = y$ and $X = x$. The likelihood is

$$\mathscr{L}_1 = \prod_{y,x} p_{yx}^{n_{yx}},$$

where n_{yx} is the number of observations having $Y = y$, $X = x$. Suppose there exists an independent auxiliary random sample of size N_q from the marginal distribution of Y giving likelihood

$$\mathscr{L}_2 = q_1^{n_1} q_0^{n_0}.$$

Now factor p_{yx} as $p_{x|y} q_y$ and multiply the total likelihood, $\mathscr{L}_1 \mathscr{L}_2$, by the prior given by the product of (5) and (6). This gives a posterior distribution of essentially the same form as (7) and thus leads in the same way to a readily calculated Bayesian bootstrap posterior distribution for θ. This is a Bayesian analog of the procedures described in frequentist terms in Imbens and Lancaster (1991) in which the marginal information about y is provided by census data and

[11] Lancaster and Imbens (1992) give a frequentist version of this method.

the data for likelihood $\mathscr{L}_1$ represents a conventional microeconomic data set. The method thus allows the incorporation of macroeconomic data into microeconomic models.

In many econometric applications of discrete choice modelling, for example, studies of labor force participation, the investigator will usually have a good deal of prior information on the marginal frequency of the event in the sampled population. This should lead to rather precise prior distributions for q which should in turn lead to tighter posterior distributions for θ, though in practice the gain may be largely confined to the intercept as it was in the example of Section 4.

References

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

Efron, B., 1982. The Jacknife, the Bootstrap and other Resampling Schemes. SIAM, Philadelphia

Hsieh, D.A., Manski, C.F., McFadden, D., 1985. Estimation of response probabilities from augmented retrospective observations. Journal of the American Statistical Association 80, 651–662.

Imbens, G.W., 1992. An efficient method of moments estimator for discrete choice models with choice-based sampling. Econometrica 60, 1187–1214.

Imbens, G.W., Lancaster, T., 1994. Combining micro and macro data in microeconometric models. Review of Economic Studies 61, 655–680.

Lancaster, T., Imbens, G.W., 1996. Case-control studies with contaminated controls. Journal of Econometrics 71, 145–160.

Manski, C.F., 1988. Analog Estimation Methods in Econometrics. Chapman and Hall.

Manski, C.F., Lerman, S., 1977. Estimation of choice probabilities from choice-based samples. Econometrica 45, 1977–1988.

Manski, C.F., McFadden, D., 1981. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA.

Manski, C.F., McFadden, D., 1981. (Eds.), Alternative estimators and sample designs for discrete choice analysis. Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA.

Rubin, D., 1981. The Bayesian Bootstrap. The Annals of Statistics 9(1), 130–134.