SAMPLE MODELS AND WEIGHTS

C. J. Skinner, University of Southampton
Dept of Social Statistics, University of Southampton, Southampton S017 1BJ UK

Key Words: Regression, Selection, Weight

1. Introduction

The question of how to take account of the sampling design when modelling sample survey data has received considerable attention in the literature (see e.g. Skinner et al, 1989). Most of this discussion assumes that the model or parameter of interest is 'prespecified' and attention is restricted to the issue of how to take account of the sampling design in the process of statistical inference.

To assume a particular specification for the model, for example an additive one, before even looking at the data, closes off the possibility of discovering alternative models, for example multiplicative ones. In contrast the practice of modern statistical modelling usually operates through an iterative process of tentative model specification and model checking. The latter may employ both formal statistical tests and informal graphical methods. While procedures for formal tests, for example for the presence of a quadratic term or an interaction, under complex sampling designs are well established, graphical methods are much less developed. One reason for this is that sampling designs with unequal selection probabilities can arbitrarily distort graphical displays.

The aim of this paper is to develop an approach in which such model building methods can be applied to data arising from a complex sampling design. A practical motivation is that many survey data analysts start from a modelling approach using standard software and it seems desirable to develop means of taking account of sampling designs which adapt this approach naturally. The broad approach will be to fit 'sample models' to sample data and then to combine these to make inference about population models. To provide a specific focus, we concentrate on the case of regression modelling.

Before particularising further, we outline the general perspective upon which this paper is based. First, it is taken for granted that in statistical modelling, the parameter of interest is a characteristic of the (super) population model and not of the finite population. Second, although it is also taken for granted that 'all models are wrong', the model building process will by its nature proceed by assuming certain models are true until model checking implies otherwise. Third, we shall primarily be concerned with the possible bias effects of a complex sampling design arising from informative/outcome dependent sampling (eg Hoem, 1989; Pfeffermann, 1993). Our approach might be viewed as an alternative to two existing approaches: the use of conventional weighting, and the use of model-based adjustment based on population information (Holt et al, 1980). We suggest that both these approaches suffer the limitations that they depend on the prespecification of population models. In addition, we suggest that conventional weighting does not naturally conform with standard model-building practice and may lead to inefficient estimation. Furthermore, the adjustment methods of Holt et al (1980) require population information which may not be readily available on a survey data file.

Let us turn then to regression modelling. We conceive of a regression model as representing the conditional distribution of a response variable y given a vector of covariates x. More precisely, randomly index the units i=1,...,N in a finite population P (Rubin, 1987, p.27) and let (y_i,x_i) be the pair of values of (y,x) associated with unit i. Suppose the (y_i,x_i) are outcomes of random vectors and denote the marginal distribution of each of these by $f_P(y,x)$. The subscript P emphasises that f_P is the population model and the fact that the marginal distribution are common follows from the random indexing (Rubin, 1987, p.32). The conditional distribution $f_P(y|x)$ defines the regression model of interest. A conventional approach to the initial specification of $f_P(y|x)$ for the simplest case of scalar x would involve the inspection of a scatter plot of y_i versus x_i for sample units i. A problem with this approach for sample survey data is that sampling designs can arbitrarily distort the pattern in a scatterplot and thus distort the specification of a model. All that can be fitted to the sample data is a sample model $f_s(y|x)$ and there is no reason in general for this to be the same as the population model of interest.

In order to address this problem we first refine our objectives. Note first that we have taken the object of inference to be either the model $f_P(y|x)$ or some parameter which helps to characterise the model. This model defines the distribution of an individual Y_i unconditional on all other Y_j for $j \neq i$ and thus might be termed a 'marginal regression model' following, for example, Liang et al (1992). The outcomes Y_i and Y_j for $i \neq j$ may be dependent but it will be important to assume that any such dependence is not an object of interest and may be treated as a nuisance. This assumption seems to accord with most uses of regression models for sample survey data. It is also implicit in most previous approaches, such as those of Kish and Frankel (1974) and Fuller (1975) mentioned earlier and pseudolikelihood methods (Skinner et al, 1989).

2. A General Approach

Let s denote the sample selected from P according to a sampling design which assigns a probability $p(\tilde{s})$ to the members $\tilde{s}$ of a set $\tilde{S}$ of subsets of P. The sample design, characterised by the values $(p(\tilde{s}); \tilde{s} \in \tilde{S})$, will be denoted p(.). The inclusion probability π_i of unit $i \in P$ is defined as usual by summing $p(\tilde{s})$ across samples $\tilde{s}$ which contain i. We allow for the possibility that p(.) represents both unit nonresponse as well as deliberate sampling, so long as it can be assumed that the π_i, $i \in s$, are known.

Suppose the data available to the analyst consist of the values (y_i,x_i,w_i) for units i in the sample s together with whatever 'sampling information' - stratum and primary sampling unit identification for example - is needed for variance estimation. The weights w_i are taken to be reciprocals of the inclusion probabilities π_i : $w_i = \pi_i^{-1}$. For simplicity, we assume no item nonresponse on y_i and x_i.

Let $z_i = (y_i, x_i)$ and $Z_P = (z'_1, ..., z'_N)'$, where x_i is a row vector and $P = \{1,...,N\}$. Let the population model be the random process which generates Z_P as an outcome (with P held fixed). This model cannot be 'built' empirically since we do not observe values of z_i for all $i \in P$. Instead let Z_s be the observed submatrix of Z_P obtained by selecting those rows in Z_P corresponding to units in s. We should like to define the sample model as the random process, induced by the population model, which generates Z_s. However, to do this we also need to define how sample designs p(.) are jointly determined under the process that determines Z_P. In other words we need to define the joint distribution of p(.) and Z_P. One device, employed by some authors (eg Scott, 1977; Rubin, 1987, Chapter 2), is to suppose that p(.) is a deterministic function of an Nxq matrix D_P of 'design variables' (which may overlap with Z_P) and then write $p(.) = p(.|D_P)$. The joint distribution of p(.) and Z_P is then determined by the joint distribution of Z_P and D_P. The idea that the probability sampling design p(.) is itself the outcome of a random process may seem conceptually elaborate, but it does usefully imply that the π_i and hence the w_i are random and possibly correlated with the z_i. This corresponds to the practical observation from the data file that the weight w_i is 'just like' any other variable.

The matrix, D_P can often be naturally defined in terms of stratum identifiers and so forth. But for our purposes the introduction of D_P is unnecessary and it is sufficient to view Z_P and p(.) as joint outcomes of a random process. The sample s and sample data Z_s are then generated in two stages:

1. Z_P and p(.) are generated;

2. letting S be the random set taking values $\tilde{s} \in \tilde{S}$ with probabilities $p(\tilde{s})$, s is generated as an outcome of S, and Z_s is determined from Z_P and s.

The marginal population model (or distribution) $f_P(z_i)$ for any unit i is obtained simply by marginalising the distribution of Z_P. Conditioning on x_i gives the model $f_P(y_i|x_i)$ as before. One possible definition of the sample model might then be the conditional distribution of Z_s given s. In this case we may state, in analogous terminology to that of Rubin (1985), that p(.) provides an 'adequate summary' of the design in the sense that each z_i is conditionally independent of S given p(.). This follows since we may write

$$Pr(S=s|Z_P, p(.)) = p(s).$$

For some designs, p(s) is characterised by the weights w_i and these therefore provide an adequate summary (Rubin, 1985). Smith (1988) suggests, however, that such designs are unusual and that an adequate summary usually requires augmenting the w_i by further design information. We suggest that this is unnecessary if the sample model is defined differently.

For motivation, consider the sample cumulative distribution of y:

$$G(y) = \sum_{i=1}^{N} I_i\, \delta(y_i \leq y) \;/\; \sum_{i=1}^{N} I_i$$

where I_i is the sample indicator function such that $I_i=1$ if $i \in s$ and $I_i=0$ otherwise and $\delta(.)$ is the indicator function such that $\delta(A)=1$ if A is true and $\delta(A)=0$ otherwise.

Randomly indexing units as before, the $I_i\delta(y_i \leq y)$ have a common marginal distribution, equal to 1 with probability $\pi \Pr(y_i \leq y) | I_i=1)$, where $\pi=E(I_i)$, and 0 otherwise. It follows, under an appropriate law of large numbers, that $G(y) \rightarrow \Pr(y_i \leq y | I_i=1)$. Such a result may be generalised to other symmetric functions of the sample observations. For analyses based on estimators of this form, the sample observations may therefore be treated as exchangeable outcomes, each with marginal distribution

$$f_s(z_i) = f_p(z_i | I_i=1) \qquad (2.1)$$

and consistent estimates of this distribution may be obtained directly from the sample. We take $f_s(z_i)$ as our definition of the **marginal sample model.** This contrasts with the earlier definition considered, of the conditional distribution of Z_s given s. This would have implied a marginal distribution for z_i of $f_p(z_i | s) = f_p(z_i | I_1,\ldots,I_N)$.

Given our definition and focus on the marginal distribution of z_i we may overcome the problem raised by Smith (1988). Note first that since the weight w_i may be viewed just as any other variable we may write, by extension of (2.1):

$$f_s(z_i, w_i) = f_p(z_i, w_i | I_i=1).$$

The following result shows that the weights w_i provide an adequate summary of the design if one is interested only in the marginal model.

Proposition 1

The conditional sample and population distributions of z_i given w_i are identical:

$$f_s(z_i | w_i) = f_p(z_i | w_i)$$

Corollary 1

The same is true for the conditional distribution of y_i given x

$$f_s(y_i | x_i, w_i) = f_p(y_i | x_i, w_i)$$

The implication of Corollary 1 is that we may fit a regression model to the sample data with y_i as the response and both x_i and w_i as covariates and, subject to the limitations of model fitting, may treat the characteristics of the fitted marginal model as corresponding to the marginal population model.

We suggest that the process of model fitting follow conventional (model-based) procedures in regression. The possible dependence between different units in the sample model may reflect either dependence in the population model or dependence induced by selection. It does not matter which from the point of view of interpretation if only the marginal model is of interest. Methods of taking account of such dependence in statistical inference might, for example, include Taylor series or resampling methods as in pseudo likelihood methods (eg. Binder, 1989; Skinner et al, 1989).

Let us assume then that we can made adequate inference about $f_p(y|x,w)$ or about some parametric characteristics of this model. Now if it turns out that $f_p(y|x,w)$ is free of w, that is

$$f_p(y|x,w) = f_p(y|x) \qquad (2.2)$$

then we are finished, since $f_p(y|x)$ is precisely the model of interest. Given the uncertainty in model fitting we shall not be able to judge that (2.2) holds with certainty, however. Rather we may test the hypothesis that (2.2) holds within a fitted family of models. For example, if the model fitting indicates that the parametric family of models

$$Y|x,\ w \sim N(\alpha + x\beta + w\gamma,\ \sigma^2)$$

provides an adequate fit to $f_s(y|x,w)$ then we may test the hypothesis that $\gamma=0$ using a t (or F) test. Or if $\alpha + x\beta + w\gamma$ is augmented by an interaction term $wx\lambda$ then an F test of $H{:}\gamma=0,\ \lambda=0$ corresponds to the approach of DuMouchel and Duncan (1983) for deciding whether to use a weighted or unweighted estimator (see also Fuller, 1984). Under the null hypothesis that $\gamma=0,\ \lambda=0$ both F tests would have the same size but would have different powers. Results of Das Gupta and Perlman (1974) indicate that neither test will be uniformly most powerful.

If the null hypothesis is accepted then one might proceed to fit a regression model of y on x ignoring the weights. The possible impact of the pre-test procedure on subsequent inference (see eg Bancroft and Han 1977) seems worthy of exploration but will not be pursued here.

If the null hypothesis is rejected then it is recommended that the analyst consider whether there are not other covariates, which have been used in the sampling design or might affect nonresponse and which could be included in the model to remove any conditional dependence of y on w given x. In this way, the weights provide a diagnostic aid for choosing an appropriate model, as suggested by DuMouchel and Duncan (1983). This approach can be practically useful in encouraging the analyst to think about alternative models.

Suppose then that, even after consideration of alternative choices of covariates x, there is still evidence that y depends on w as well as x. How can we get from the fitted model $f_p(y|x,w) = f_s(y|x,w)$ to the model $f_p(y|x)$ of interest?

The two models are related by:

$$f_p(y|x) = \int f_p(y|x,w)\ f_p(w|x)dw \quad (2.3)$$

which suggests that we need to fit the model $f_p(w|x)$. In general $f_p(w|x) \neq f_s(w|x)$ and so we cannot simply fit a further regression model to the sample data for w on x to obtain $f_p(w|x)$. Instead we use the following results.

Proposition 2

The population distribution of w_i is obtained by weighting the sample distribution

$$f_p(w_i) = w_i\, f_s(w_i) \;/\; E_s(w)$$

where $E_s(w) = \int w\, f_s(w)\, dw$

Corollary 2

The same is true conditional on x:

$$f_p(w_i|x_i) = w_i\, f_s(w_i|x_i) \;/\; E_s(w|x_i)$$

where $E_s(w|x_i) = \int w\, f_s(w|x_i)\, dw$

It follows from (2.3) and Corollary 2 that a general approach to fitting the model $f_p(y|x)$ is obtained from the following four steps:

1. fit the regression model of y on x and w, $f_s(y|x,w)$ [$= f_p(y|x,w)$], to the sample data;

2. fit the regression model $f_s(w_i|x_i)$ to the sample data;

3. weight this model to obtain $f_p(w|x)$ as in Corollary 2;

4. combine the fitted models $f_p(y|x,w)$ and $f_p(w|x)$ via (2.3).

This approach is very general, applying to parametric, semiparametric or nonparametric regression models, to linear or nonlinear models and to continuous or discrete y. However, it clearly requires fleshing out in at least two respects. First, in regression modelling it is often desirable to have only a few parameters to represent the model and it is not clear how parameters of the fitted $f_s(y|x,w)$ and $f_s(w|x)$ models can be combined in a simple and interpretable way. Secondly, practical inference procedures for point estimation, variance estimation and so forth, need to be developed. We suggest how this general approach may be implemented under some specific conditions in the next section.

3. A Specific Approach based on a Log-linear Regression Model for the Weights

In this section we consider representing $f_s(w|x)$ as a log-linear regression model:

$$\log w_i = x_i\, \gamma + \varepsilon_i \qquad (3.1)$$

There are two particular reasons for the log transformation.

(a) We shall see that the effect of weighting in step 3 of Section 2 is particularly simple if the sample model is of form (3.1), provided the distribution of ε_i is free of x_i.

(b) Often the inclusion probability π_i can be expressed as a product of terms, each of which may depend on different x factors. For example, we may have $\pi_i = \pi_{1i}\, \pi_{2i}\, \pi_{3i}$, where π_{1i} is the probability of selecting the primary sampling unit (PSU) which contains i and may depend on area-level factors, π_{2i} is the probability of selecting individual i given that the PSU is selected and this may depend on the number of households at the dwelling within which the individual lives and π_{3i} may be the probability of individual i responding which may depend on other covariates. By taking $\log w_i$ or equivalently $-\log \pi_i$ we may reasonably expect the x factors to enter equation (3.1) in an additive way.

To elaborate upon reason (a), let $u = \exp(\varepsilon)$ and suppose that the probability density function of u in the sample is $f^u_s(u)$, which does not depend on x. The distribution of u in the population after weighting may be shown to be

$$f^u_p(u|x) = u\ f^u_s(u) \ /\ E_s(u). \qquad (3.2)$$

Hence it does not depend on x and the effect of weighting is simply to change the distribution of $\varepsilon_i = \log(u_i)$ in (3.1) according to (3.2).

In particular we may write

$$E_p(\log\ w|x) = E_s(\log\ w|x)+k. \qquad (3.3)$$

where

$$k = E_p(\varepsilon)-E_s(\varepsilon)$$
$$= E_s[\varepsilon \exp(\varepsilon)]/E_s[\exp(\varepsilon)] - E_s(\varepsilon).$$

If $\hat{\gamma}$ is the least squares estimate of γ and $e_i = \log w_i - x_i\ \gamma$, the ith residual, then k may be estimated by

$$\hat{k} = \sum_s e_i \exp(e_i) \ /\ \sum_s \exp(e_i). \qquad (3.4)$$

This estimator does not depend on any distributional assumption about ε_i. Note that, as usual in least squares estimation, $\sum_s e_i = 0$ so that $\hat{k}$ implicitly estimates $E_s(\varepsilon)$ to be zero.

To illustrate the impact of this adjustment suppose first that the fitted regression mean model at step 1 is

$$E_p(y|x,w) = (\alpha+\beta x)(1+\lambda \log w) \qquad (3.5)$$

for scalar x and that the model fitted at step 2 is

$$\log w = \gamma_0 + x\ \gamma_1 + \varepsilon$$

Then from (3.3) and (3.5) the combined model obtained at step 4 is

$$E_p(y|x) = (\alpha + \beta x)\ [1+\lambda(\gamma_o+x\gamma_1+k)]$$
$$= E_s(y|x) + k\ \lambda(\alpha + \beta x).$$

The example above assumes that w only appears as log w in the expression for $E_p(y|x,w)$. This will lead to the simplest possible expressions for $E_p(y|x)$, but clearly is a restrictive requirement. A broader class of models may be represented by a polynomial mean function

$$E_p(y|x,w) = \sum_{j=0}^{J} b_{s_j}(x)w^j.$$

In this case we may use the result that

$$E_p(w^j|x) = E_s(w^{j+1}|x) \ /\ E_s(w|x)$$

and in particular

$$E_p(w|x) = E_s(w^2|x) \ /\ E_s(w|x)$$
$$= [1+c.v^2(w|x)]\ E_s(w|x).$$

where c.v(w|x) is the coefficient of variation of w given x. Now under model (3.1).

$$c.v(w|x) = c.v(\exp(\varepsilon)|x)$$

which is constant if the distribution of ε is free of x. Hence the mean function of w on x is simply multiplied by a constant under weighting.

For general j we obtain

$$E_p(w^j|x) = \exp[(j+1)x\gamma]$$

$$E_s(u^{j+1}) \;/\; \exp(x\gamma)\; E_s(u)$$

$$= \exp[jx\gamma]\; E_s(u^{j+1})/E_s(u)$$

which may be estimated by

$$\exp[jx\hat{\gamma}] \sum_{i\in s} \exp[(j+1)e_i] / \sum_{i\in s} \exp(e_i). \tag{3.6}$$

Let us summarise then the nature of steps (1) - (4) of Section 2 for our simplified approach to modelling the mean function $E_p(y|x)$,

1) Fit a regression model

   $$E_s(y|x,w) = \mu_s(x,w)$$

   to the sample data (y_i, x_i, w_i), $i\in s$. Check whether dependence on w can be avoided by suitable choice of x. If not, represent dependence of $\mu_s(x,w)$ on w preferably in terms of logw

   $$\mu_s(x,w) = b_{s_0}(x) + b_{s_1}(x) \log w$$

   or otherwise in terms of a polynomial

   $$\mu_s(x,w) = \sum_{j=0}^{J} b_{sj}(x)\, w^j$$

2) Check whether the log-linear model assumption in (3.1) is adequate. If so, fit the model to the sample data (w_i, x_i), $i\in s$, using least squares estimation.

3) Either estimate $E_p(\log w|x)$ using the additive adjustment + $\hat{k}$ in (3.3) and (3.4) or estimate the required $E_p(w^j|x)$ using (3.6)

4) Substitute into the function fitted in (1).

This approach results in a single (point) estimate of the regression function $E_p(y|x)$. Variance estimators of parameters, or of the value of the regression function at a given x, may be obtained for example by replication methods (eg Wolter, 1985).

4. Example 1 - A Simple Simulation Study

To illustrate the properties of the proposed procedure in a simple setting, 'population' values (x_i,w_i), i=1,...,N=1000 were generated independently from the model

$$x_i \sim U(0,1)$$

$$\log w_i = 1+3x_i+\varepsilon_i,\quad \varepsilon_i \sim U(-1,1). \tag{4.1}$$

where U(a,b) denotes the uniform distribution on the interval [a.b]. Sample indicator values I_i were then generated independently such that $\Pr(I_i=1) = w_i^{-1}$. For sample units ($I_i=1$), values y_i were then generated independently from

$$y_i = x_i+\log w_i+\delta_i,\quad \delta_i \sim N(0,1). \tag{4.2}$$

It follows from (4.1) and (4.2) that the population model relating y_i and x_i is $y_i = 1 + 4x_i + (\varepsilon_i+\delta_i)$ and from (3.3) that the sample model is $y_i = k + 4x_i + u_i$, where k is some constant, u_i is independent of x_i and $E(u_i) = 0$.

We consider three estimators of the coefficients α and β of the linear population regression model

$E_p(y|x) = \alpha + \beta x$:

a) ordinary least squares (OLS);

b) weighted least squares (WLS) with weights w_i;

c) the procedure proposed in Section 3 involving (1) the fitting of $E_s(y|x,w) = \theta_0 + \theta_1 x + \theta_3 \log w$, (2) the fitting of $E_s(\log w|x) = \phi_1 + \phi_2 x$, (3) setting $\hat{\alpha} = \hat{\theta}_0 + \hat{\theta}_3(\hat{\phi}_1 + \hat{k})$, $\hat{\beta} = \hat{\theta}_1 + \hat{\theta}_3 \hat{\phi}_2$, where $\hat{k}$ is obtained from (3.4).

The procedure was repeated 1000 times and the empirical bias, variance and MSE of the three estimators of α and β are presented in Table 1.

It may be seen that the sample selection induces no bias in the ordinary least squares estimator of β (the simulation standard error of -0.005 is 0.013). This is to be expected since the sample model of y given x is linear with the same slope as in the population model. The ordinary least squarse estimator of α is biased, however, and this bias is corrected by the proposed procedure. The weighted least squares estimators of α and β are both approximately unbiased (only approximately since the simulation involves repeated generations from the model) but their variances are substantially greater than those of the proposed procedure in each case.

5. Conclusions

This paper has focussed on the fitting of 'marginal' regression models to survey data. It has been argued that as a first approach it is sensible to augment the set of covariates by the weights variable and to seek a model specification where the weight has no significant effect. If this proves impossible, an alternative approach to conventional weighting has been presented. This involves the fitting of two regression models: one, where the weight is a covariate and one where the weight is the dependent variable. A principal advantage of this approach is that only 'sample models' need to be specified to fit to sample data. This may make use of conventional iterative model-building methods. In contrast the conventional weighting approach requires the population model to be prespecified. A further advantage is a gain in efficiency.

Acknowledgements

Thanks to Danny Pfeffermann for comments. Research was supported by ESRC grant H519 25 5005.

References

Bancroft, T A and Han, C P (1977) Inference based on conditional specification: a note and a bibliography. *Intern. Statist. Rev. 45*, 117-27.

Binder, D A (1983) On the variances of asymptotically normal estimators from complex surveys. *Intern. Statist. Rev. 51*, 279-292.

Das Gupta, S and Perlman, M D (1974) Power of the noncentral F-test: effect of additional variates on Hotelling's T^2-test. *J. Amer. Statist. Assoc. 69*, 174-180.

Du Mouchel, W H and Duncan, G J (1983) Using sample survey weights in multiple regression analysis of stratified samples. *J Amer. Statist. Assoc. 78*, 535-543.

Fuller, W A (1975) Regression analysis for sample surveys. *Sankhya, C37*, 117-32.

Fuller, W A (1984) Least squares and related analysis for complex survey designs. *Survey Methodology 10*, 97-118.

Hoem, T M (1989) The issue of weights in panel surveys of individual behaviour. In Kasprzyk, D, Duncan, G, Kalton, G and Singh, M P (eds) *Panel Surveys*. New York: Wiley.

Holt, D, Smith, T M F and Winter, P D (1980) Regression analysis of data from complex surveys. *J Roy Stat Assoc, A143*, 474-487.

Kish, L and Frankel, M R (1974) Inference from complex samples (with discussion) *J Roy. Statist. Soc, Series B, 36*, 1-37.

Liang, K Y, Zeger, S L and Quaqish, B (1992) Multivariate regression analysis for categorical data (with discussion) *J Roy. Statist. Soc., B54*, 3-40.

Pfeffermann, D (1993) The role of sampling weights when modelling survey data. *Int. Statist. Rev.*

Rubin, D B (1985) The use of propensity scores in applied Bayesian inference. In *Bayesian Statistics 2*, (J M Bernardo, M H DeGroot, D V Lindley and A F M Smith, eds). Amsterdam: North Holland, 463-472.

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

Scott, A J (1977) Some comments on the problem of randomisation in survey sampling. *Sankhya C39*, 1-9.

Skinner, C J, Holt, D and Smith, T M F, eds (1989) *Analysis of Complex Surveys*. Chichester: Wiley.

Smith, T M F (1988) To weight or not to weight, that is the question. In *Bayesian Statistics 3* (J M Bernardo, M H DeGroot, D V Lindley and A F M Smith, eds). Oxford University Press, 437-451.

Wolter, K M (1985) *Introduction to Variance Estimation*. New York: Springer-Verlag.

Table 1

Estimation procedure	Parameter estimated					
	α			β		
	Bias	Variance	MSE	Bias	Variance	MSE
OLS	-0.312	0.023	0.120	-0.005	0.179	0.179
WLS	0.017	0.054	0.055	-0.073	0.383	0.388
Proposed	-0.005	0.027	0.027	-0.005	0.179	0.179