Survey Methodology, December 1992
Vol. 18, No. 2, pp. 187-208
Statistics Canada

Robust Model-Based Methods for Analytic Surveys

T.M.F. SMITH and E. NJENGA[1]

ABSTRACT

This paper reviews the idea of robustness for randomisation and model-based inference for descriptive and analytic surveys. The lack of robustness for model-based procedures can be partially overcome by careful design. In this paper a robust model-based approach to analysis is proposed based on smoothing methods.

KEY WORDS: Analytic surveys; Robustness; Smoothing methods.

1. INTRODUCTION

The concept of robustness in finite population inference from both the randomisation and model-based viewpoints is examined. In his seminal paper on a unified theory of sampling from finite populations Godambe (1955) not only proved his famous non-existence theorem but also made suggestions for robust finite population inference. He proposed a superpopulation model for the unit variables y_i and suggested that strategies, that is the choice of both design and estimator, should be based on the model expectation of the sampling variance. He then imposed p-unbiasedness to obtain optimum strategies. These ideas were amplified in several papers including Godambe (1982) and Godambe and Thompson (1977). The results obtained include the apparent optimality of πps sampling and the Horvitz-Thompson (1952) estimator. But the inefficiency of this strategy in multipurpose surveys is well known so we find these results on optimality and robustness less convincing than the apparently negative results on the foundations of inference.

The lack of robustness of many model-based procedures is well known, see Hansen *et al.* (1983), and much of the work of Royall and his colleagues, for example Royall and Herson (1973a,b) has been devoted to constructing robust model-based strategies. After reviewing this work we propose a robust model-based method for estimating many complex statistics employed in the multivariate analysis of survey data which adjusts for the effects of selection. Our proposal is not a strategy but is a procedure which can be employed for the analysis of survey data after the sample is drawn.

2. FORMAL STRUCTURE

In order to examine robustness we must first structure finite population inference in the formal manner pioneered by Godambe (1955). We consider a population of N units with label set $U = \{1, 2, \ldots, N\}$. Attached to unit i is a vector of values, $\underset{\sim}{y}_i$, which will be measured on the sample units, and $\underset{\sim}{y}_U = (\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_N)$ denotes the finite population matrix of values. A sample, s, is a subset of U drawn according to some rule. We are concerned here with rules based only on prior information, $\underset{\sim}{z}_i$, available on all the units in the population. Let $\underset{\sim}{z}_U$ denote the prior information for the whole population, and let $p(s \mid \underset{\sim}{z}_U)$ denote the sampling rule.

[1] T.M.F. Smith, University of Southampton, United Kingdom; E. Njenga, Kenyatta University, Kenya.

Since the rule does not depend on $\underset{\sim}{y}_U$ it is uninformative. If $p(s \mid \underset{\sim}{z}_U)$ is a random sampling rule then it determines a probability distribution over ξ, the set of all samples, which is the basis for randomisation inference. The sample data comprises $d_s = \{(i,\underset{\sim}{y}_i)\colon i \epsilon s\}$. Let y_s denote the matrix of sample values, then an estimator is a function of the data, d_s, and of the prior information, $\underset{\sim}{z}_U$, which includes auxiliary information. We denote by E_p, V_p, expectations and variances with respect to the distribution $p(s \mid \underset{\sim}{z}_U)$.

In a model-based approach it is further assumed that the population values $\underset{\sim}{y}_U$ are random variables. A major problem with this approach is to specify a parametric probability model for the joint distribution of all these random variables, which must be based on all the prior information including that on the structures of, and relationships between, the units in the population. So models must reflect hierarchical groupings (clusters) and block groupings (strata), as well as correlations between the variables. This structure is potentially so complex that attention is usually restricted to means and covariance matrices. In general let $f(\underset{\sim}{y}_U \mid \underset{\sim}{z}_U;\underset{\sim}{\lambda})$ denote the conditional finite population distribution, where $\underset{\sim}{\lambda}$ is a vector of unknown parameters. For predictive inference about finite population values, such as totals, this is a sufficient specification. For analytic inference about parameters in the marginal distribution of $\underset{\sim}{y}$ we must additionally specify the marginal distribution of the prior values $\underset{\sim}{z}_U$. Let $f(\underset{\sim}{z}_U;\underset{\sim}{\phi})$ denote this distribution, then the marginal distribution of $\underset{\sim}{y}_U$ is

$$f(\underset{\sim}{y}_U;\underset{\sim}{\theta}) = \int f(\underset{\sim}{y}_U \mid \underset{\sim}{z}_U;\underset{\sim}{\lambda})f(\underset{\sim}{z}_U;\underset{\sim}{\phi})d\underset{\sim}{z}_U, \tag{2.1}$$

where $\underset{\sim}{\theta} = g(\underset{\sim}{\lambda},\underset{\sim}{\phi})$ is the parameter of analytic interest.

Applying the sampling rule to the population generates the data, d_s. The joint distribution of the data, d_s, and prior values, $\underset{\sim}{z}_U$, is

$$\begin{aligned} f(d_s,\underset{\sim}{z}_U;\underset{\sim}{\lambda},\underset{\sim}{\phi}) &= p(s \mid \underset{\sim}{z}_U) \int f(\underset{\sim}{y}_U \mid \underset{\sim}{z}_U;\underset{\sim}{\lambda})f(\underset{\sim}{z}_U;\underset{\sim}{\phi})d\underset{\sim}{y}_{\bar{s}} \\ &= p(s \mid \underset{\sim}{z}_U)\, f(\underset{\sim}{y}_s \mid \underset{\sim}{z}_U;\underset{\sim}{\lambda})\, f(\underset{\sim}{z}_U;\underset{\sim}{\phi}), \end{aligned} \tag{2.2}$$

where $\bar{s}$ denotes units not in s. This distribution is the basis of a model-based approach to inference. We let E_m, V_m, denote expectations and variances with respect to the model.

An implication of (2.2) is that the sampling rule, $p(s \mid \underset{\sim}{z}_U)$, must be completely known to the person making the inference, as must the values of $\underset{\sim}{z}_U$. Absence of knowledge may render $p(s \mid \underset{\sim}{z}_U)$ informative about the unobserved values $\underset{\sim}{y}_{\bar{s}}$, see Scott (1977), Sugden and Smith (1984), in which case it cannot be taken outside the integral in (2.2).

In this general set-up, embracing both random selection and modelling of values, randomisation inference corresponds to the case where the values $\underset{\sim}{y}_U$ are unknown constants and the model distribution becomes degenerate at the point $\underset{\sim}{y}_U$. The only probability remaining is that in $p(s \mid \underset{\sim}{z}_U)$, and this distribution over the set ξ of all possible samples is the basis of randomisation inference. Note that the randomisation distribution is completely specified by knowledge of the sampling rule and of the prior values, $\underset{\sim}{z}_U$. It does not depend on any unknown parameters or on the survey values, $\underset{\sim}{y}_U$. This renders $p(s \mid \underset{\sim}{z}_U)$ uninformative because there is less information in $p(s \mid \underset{\sim}{z}_U)$ than in $\underset{\sim}{z}_U$ itself. This accounts for the negative nature of Godambe's results about randomisation inference.

In contrast model-based inference depends solely on the model component of (2.2), since $p(s \mid \underset{\sim}{z}_U)$ contains no information about $\underset{\sim}{y}_{\bar{s}}$. Predictive inferences about $\underset{\sim}{y}_{\bar{s}}$ are made using the conditional distribution, $f(\underset{\sim}{y}_u \mid \underset{\sim}{y}_s,\underset{\sim}{z}_U;\underset{\sim}{\lambda})$, independent of the randomisation distribution, $p(s \mid \underset{\sim}{z}_U)$. The sampling rule is still important at the design stage, for it affects efficiency and robustness, but it has no rôle to play at the inference stage. Random sampling also provides a guarantee that the sampling rule is in fact uninformative, providing a scientifically acceptable sampling procedure. Model-based inferences may not be robust, however, because they may depend strongly on the choice of model, as demonstrated by many authors including Hansen *et al.* (1983).

A compromise solution is to employ both components of (2.2), the model and the randomisation distribution, in the choice of estimator. This was proposed by Godambe (1955) as a positive response to his negative results. He proposed using as a criterion the model expectation of the randomisation variance, namely $E_m \; V_p(t_s)$, where t_s is an estimator of a finite population total T. To find an optimum solution in a particular class of models Godambe restricted the choice of t_s to the class of p-unbiased estimators. This restriction has been much criticized and subsequently several authors, including Brewer (1979), Särndal (1980), Isaki and Fuller (1982), Little (1983), have proposed replacing exact unbiasednesses by some form of approximate unbiasedness. This is usually expressed in the form of asymptotic design unbiasedness which requires the construction of a hypothetical sequence of finite populations with sizes tending to infinity. Although one may feel unhappy with this mathematical construction the suggestion that strategies, chosen before drawing the sample, should be based on considerations of the average under a model of a repeated sampling procedure is perfectly acceptable. The controversial issue is the choice of distribution for making inferences after the sample has been drawn.

3. ROBUSTNESS

Robustness is not a well defined concept in statistics. The Encyclopedia of Statistical Sciences, (Kotz and Johnson 1988), states that:

> *"a robust procedure performs well not only under ideal conditions but also under departures from the ideal."*

It goes on to say that both the nature of departures from the ideal and the meaning of "*performs well*" must be specified. With this broad definition in mind we now examine robustness for randomisation and model-based inference for finite population totals. The general perception is that randomisation inference is robust and that model-based inference is not.

Godambe's negative results can be interpreted to mean that randomisation inference is impossible in general. This is certainly true for heterogeneous populations, such as Royall's axe, ass and box of horseshoes, or for populations with a few very extreme values, but for homogeneous populations the evidence overwhelmingly shows that randomisation inference is not only possible but also works in a well defined sense.

Employing randomisation inference implies abandoning certain statistical principles, such as the likelihood principle, and replacing them by an appeal to the central limit theorem. The assertion is that under repeated random sampling using the specified rule $p(s \mid z_U)$

$$\frac{t_s - T}{\hat{V}_p(t_s)} \sim N(0,1), \tag{3.1}$$

for any t_s which is approximately p-unbiased for T, where both N and n are large, but n/N is small. Although proved formally only under SRS and related schemes, empirical evidence shows that the randomisation coverage properties of 95% confidence intervals of the form

$$t_s \pm 1.96\sqrt{\hat{V}_p(t_s)}, \tag{3.2}$$

where $\hat{V}_p(t_s)$ is a consistent estimator of $V_p(t_s)$, are approximately correct except for extreme designs or heterogeneous populations.

Godambe and Thompson (1977) express their views about this approach in the following terms.

> *"The use of such a confidence interval may be interpreted as follows:*
>
> *I: We are fairly sure a priori that y belongs to that subset of R^N for which the interval covers $T(y)$ for 95% of all possible samples.*
>
> *II: There is no way that the sampled y-values, in conjunction with whatever other information we may have about the population, have altered the conviction in I. Thus even after sampling we believe that if the design were implemented again and again on this population the interval would cover $T(y)$ approximately 95% of the time.*
>
> *The robustness of the interval arises of course from the fact that only very weak and essentially informal conditions are required for the validity of its interpretation in the sense of I and II."*

Very similar views are expressed by Hansen *et al.* (1983).

> *"For probability-sampling designs the computed confidence intervals, for samples large enough, are valid in the sense that the randomization probability that the confidence intervals contain the value being estimated is equal to or greater than the nominal confidence coefficient, independent of the distribution of the characteristics among the elements of the population from which the sample is drawn."*
>
> *"Robustness is usually understood to mean that inferences made from a sample are insensitive to violations of the assumptions that have been made. In principle, and ordinarily in fact, robustness is achieved in probability-sampling surveys by the use of sampling with known probabilities (i.e., randomization) and consistent estimators, and using a large enough sample that the central limit theorem applies, so that the estimates can be regarded as approximately normally distributed."*

Note that this concept of robustness does not appear to require any specification of ideal conditions or of departures from the ideal. Random sampling and consistent estimation are all that is required. Brewer and Särndal (1983) are quite explicit:

> *"Probability sampling methods are robust by definition; since they do not appeal to a model, there is no need to discuss what happens under model breakdown."*

How can a statistical procedure be so robust?

The reason is that the entire procedure is under the control of the statistician, no attempt is made to introduce "nature" into the structure. The randomisation distribution has a known form and does not depend on unknown parameters. There is no need to make an inference about $p(s \mid z_U)$. Similarly the framework for inference is chosen by the statistician, it is repeated sampling using $p(s \mid z_U)$. Different statisticians may use different sampling rules and estimators but the procedure represented by (3.1) gives approximately correct coverage properties in every case, and so is robust. This is an example of criterion robustness. However, any given procedure may not be efficient for the totals of some variables. We have already highlighted the well known inefficiency of the Horvitz-Thompson estimator which occurs when the survey variable is negatively correlated with the size variable. The search for efficiency robustness over a wide range of variables leads frequently to the recommendation that the design should be a stratified SRS design, see, for example, Godambe (1982), Hansen *et al.* (1983).

In model-based inference the statistician is playing the game of modelling "nature". Probability distributions such as $f(\underset{\sim}{y}_U \mid \underset{\sim}{z}_U;\underset{\sim}{\lambda})$ are chosen by the statistician but their true form is unknown, as also are the values of the parameters. If an estimator, t_s of T, is chosen then its expected value and variance will depend on the choice of model. Deviations from the model may lead to changes in the mean and variance and hence to changes in confidence intervals based on applying the central limit theorem to the model residuals. In model-based inference the robustness due to the central limit theorem is more limited than that in randomisation inference since it applies only to the residuals. Some model deviations can be controlled by choosing an appropriate design, as in Royall and Herson (1973a,b), but there can never be complete robustness. The framework for inference is also completely different. Instead of employing the unconditional distribution based on repeated sampling model-based inference employs the conditional distribution given the selected sample s.

Can these two positions ever be reconciled? Before sampling, when choosing strategies, they can. Both schools of thought have the same prior information, $\underset{\sim}{z}_U$, and both use models to suggest designs and estimators and choose strategies based on the overall mean squared error

$$E_m E_p (t_s - T)^2. \qquad (3.3)$$

Randomisers usually impose a constraint such as approximate p-unbiasedness while modellers may impose approximate model unbiasedness and the two positions can be reconciled by choosing a sample design such that the model-unbiased estimator is also p-unbiased. This strategy utilizes the full structure of (2.2) and gets the best of both worlds.

After sampling there appears to be little hope of reconciliation. The two frameworks for inference are quite different, one being based on an unconditional distribution the other on a conditional distribution. Royall and Cumberland (1981) have demonstrated convincingly how much difference this can make. Incidentally they have also demonstrated the lack of robustness of some of the conventional model-based variance estimators.

One case where reconciliation is possible occurs in stratified sampling. Both randomisers and modellers have converged on stratified sampling as a robust design, and for SRS within strata model-based and p-based inferences coincide. This provides evidence for one of the few positive results in sample surveys:

Theorem: Stratification is a good thing.

Proof: See Cochran (1977, Ch.5).

Stratification allows us to look at the problem of robustness more closely. If both a randomiser and a modeller adopt the same stratification, and both also adopt the same SRS design within strata, then for a given sample they will both make identical inferences. Now suppose on the basis of further analysis or evidence it is agreed that an extra level of stratification should have been used. How does this affect the respective inferences? The modeller now has to say that the original model was misspecified and hence that inferences from that model would be biased. Both the estimator and the variance of the original model would be wrong. The randomiser, however, can say that the extra information is interesting, and could be used to post-stratify the original results, but that it can also be ignored if necessary because the original inferences are still valid in the sense defined in (3.2). All that has happened is a possible loss of efficiency. In one case the original inference is condemned as not being robust, in the other case the same inference is apparently robust. The modellers bias, when averaged over repeated samples, is transformed for the randomiser into a component of sampling variance, or a loss of efficiency. So if initially randomisers and modellers start from the same position then deviations from that position are interpreted differently. In one case it is a bias in the other case a variance. Can this really be called robust in one case and not robust in the other?

4. ANALYTIC INFERENCE

In analytic inference the target for inference is no longer a known function of the finite population values, $\underset{\sim}{y}_U$, so that even if $n = N$ there is still residual uncertainty in the inference. Examples are tests of hypotheses, where the null hypothesis of no difference is meaningless in a fixed finite population. Possible targets for inference are the parameters $\underset{\sim}{\lambda}$, $\underset{\sim}{\phi}$, of the model (2.2), or functions of them such as $\underset{\sim}{\theta}$ in (2.1). Other targets are the parameters in finite populations related to the given finite population in some known way, perhaps through a spatial or time series structure. Methods for analytic inference have recently been reviewed by Skinner *et al.* (1989).

The starting point for analytic inference is the specification of the superpopulation model which aims to show how the finite population is related to the superpopulation. A common assumption is that the finite population is generated as IID random variables from a superpopulation. Whether this can be justified for populations with structure, such as clustering or stratification, is debatable. In this paper we assume that it is true, at least within broadly defined strata. With this assumption a SRS from the finite population is itself an IID sample from the superpopulation and inferences can be made directly from the sample to the superpopulation. If the sample is not a SRS, but is drawn using a design $p(s \mid \underset{\sim}{z}_U)$ which uses the information in $\underset{\sim}{z}_U$, then the achieved sample is no longer an IID sample from the superpopulation. This is the problem of selection and the effect of selection must be taken into account in the final inference.

The superpopulation model establishes a hierarchy,

$$\text{superpopulation} \supset \text{finite population} \supset \text{sample.}$$

If the finite population is IID from the superpopulation then finite population parameters, such as means, are related to the corresponding superpopulation parameters by

$$\bar{y}_U = E_m(\bar{y}_U) + O_p(N^{-1/2}). \tag{4.1}$$

Since N is usually very large an inference about $\bar{y}_U$ is a good approximation to an inference about $E_m(\bar{y}_U)$. Inferences about $\bar{y}_U$ using the p-weights associated with the sampling rule $p(s \mid \underset{\sim}{z}_U)$ are the basis of the randomisation approach to analytic inference. Note that this approach depends strongly on the IID assumption for the finite population.

For more complex analyses, such as logistic regression analysis, the pseudo-MLE approach in Skinner *et al.* (1989, sec. 3.4.4.) and Binder (1983) can be used to define both the finite population parameter of interest and the randomisation estimator. The finite population parameter is usually defined through an estimating equation, see Godambe (1960) and Godambe and Thompson (1986). As in Section 3 confidence intervals are based on the unconditional distribution generated by repeated random sampling.

Model-based analytic inference is based on the complete model of the survey population $\underset{\sim}{y}_U$, the design variables $\underset{\sim}{z}_U$, and the sample selection rule $p(s \mid \underset{\sim}{z}_U)$, that is

$$f(\underset{\sim}{y}_U,\underset{\sim}{z}_U,s;\underset{\sim}{\lambda},\underset{\sim}{\phi}) = f(\underset{\sim}{y}_U \mid \underset{\sim}{z}_U;\underset{\sim}{\lambda})\, f(\underset{\sim}{z}_U;\underset{\sim}{\phi})p(s \mid \underset{\sim}{z}_U)\,. \tag{4.2}$$

For random sampling rules the selection scheme leaves the conditional distribution $f(\underset{\sim}{y}_U \mid \underset{\sim}{z}_U;\underset{\sim}{\lambda})$ unchanged, but changes the marginal distribution of $\underset{\sim}{z}_U$ from $f(\underset{\sim}{z}_U;\underset{\sim}{\phi})$ before selection to

$$g_s(\underset{\sim}{z}_U;\underset{\sim}{\phi}) = f(\underset{\sim}{z}_U;\underset{\sim}{\phi})p(s \mid \underset{\sim}{z}_U) \tag{4.3}$$

after selection. Thus inferences about $\underset{\sim}{\lambda}$ are unaffected by selection but inferences about ϕ, and hence about $\theta = g(\underset{\sim}{\lambda},\underset{\sim}{\phi})$, the parameters of the marginal distribution $f(\underset{\sim}{y}_U;\underset{\sim}{\theta})$, are affected by selection. For these latter inferences the sample data cannot be treated as though it were a SRS from the superpopulation model.

If we assume that the superpopulation distributions are multivariate normal then

(i) $E(\underset{\sim}{y} \mid \underset{\sim}{z})$ is linear in $\underset{\sim}{z}$, and

(ii) $V(\underset{\sim}{y} \mid \underset{\sim}{z}) = \underset{\sim}{K}$, independent of $\underset{\sim}{z}$.

Under these assumptions of linearity and homoscedasticity a model-based estimator of the covariance matrix, $\underset{\sim}{\Sigma}_{yy}$, of $\underset{\sim}{y}$ is given by

$$\hat{\underset{\sim}{\Sigma}}_{yy} = \underset{\sim}{V}_{yys} + \underset{\sim}{b}_{yz} \quad (\underset{\sim}{V}_{zzu} - \underset{\sim}{V}_{zzs})\underset{\sim}{b}_{yz}^T, \tag{4.4}$$

as shown in Skinner *et al.* (1989 Section 6.4), where $\underset{\sim}{V}_{yys}$, $\underset{\sim}{V}_{zzs}$, $\underset{\sim}{b}_{yz}$ are sample covariance matrices and a matrix of regression coefficients based on treating the sample data as IID from the conditional distribution $f(\underset{\sim}{y}_U \mid \underset{\sim}{z}_U;\underset{\sim}{\lambda})$. We call (4.4) the Pearson adjusted estimator after Pearson (1903).

Theoretical and empirical studies by Pfeffermann and Holmes (1985), Holmes (1987) and Njenga (1990), have shown that model-based inferences from (4.4) are not robust to departures from the assumptions of linearity and homoscedasticity. Nathan and Holt (1980) proposed a p-weighted version of (4.4) as a more robust alternative. This estimator is formed by replacing all the equally weighted sums in (4.4) by the corresponding p-weighted sums. The resulting estimator is called the probability weighted maximum likelihood estimator (*pwml*). The properties of this estimator have been studied empirically and theoretically in Holmes (1987), Njenga (1990) and in Skinner, Holt and Smith (1989, Ch.8). It was found to have similar unconditional properties to alternative p-weighted estimators, such as the Horvitz-Thompson estimator of $\underset{\sim}{\Sigma}_{yy}$, and superior conditional properties. In the simulation study in Section 6 the *pwml* estimator is taken to represent the entire class of p-weighted estimators. Since the p-weighted version of $\underset{\sim}{V}_{zzs}$ in (4.4) is a design consistent estimator of $\underset{\sim}{V}_{zzu}$ the resulting estimator is a design consistent estimator of $\underset{\sim}{\Sigma}_{yy}$. We now investigate a new robust model-based procedure.

5. A NONPARAMETRIC MOMENT-BASED ESTIMATOR

In this section we attempt to overcome the lack of robustness of model-based estimators such as (4.4) which depend strongly on assumptions of linearity and homoscedasticity. If the finite population is realized as IID observations from the superpopulation and if interest centres on the superpopulation parameters $\underset{\sim}{\mu}_y, \underset{\sim}{\Sigma}_{yy}$ in the marginal distribution of $\underset{\sim}{y}$, then the approach we adopt uses the fact that the sample data are IID from the conditional distribution $f(\underset{\sim}{y} \mid \underset{\sim}{z})$

while the design variables $\underset{\sim}{z}_U$ are an IID sample of size N from the marginal distribution of $\underset{\sim}{z}$. For simplicity we assume that only one design variable has been used, such as a measure of size, so that z is a scalar random variable.

We assume that the conditional mean and covariance matrix of $\underset{\sim}{y}$ given z are smooth functions of z of unknown form. Let

$$E(\underset{\sim}{y} \mid z) = \underset{\sim}{\mu}(z), \tag{5.1}$$

$$V(\underset{\sim}{y} \mid z) = \underset{\sim}{\Sigma}_{yy}(z). \tag{5.2}$$

These parametric functions can be estimated using some form of nonparametric estimation such as linear smoothing. Examples of linear smoothing methods are kernel estimation, see, for example, Gasser and Muller (1979), local regression, see, for example, Cleveland (1979), and smoothing splines, see, for example, Silverman (1985). We propose estimating the functions in (5.1) term by term using the kernel estimator

$$\hat{\underset{\sim}{\mu}}(z) = \sum_{j\epsilon s} W_k(z,z_j)\underset{\sim}{y}_j. \tag{5.3}$$

We constrain the sum of the weights to be unity so that the estimator is a weighted average and employ the Gaussian kernel with k being the bandwidth. These estimators have been extensively studied and a recent review is Gasser and Engel (1990).

The structure in (5.1) and (5.2) implicitly assumes that we can write

$$\underset{\sim}{y}_j = \underset{\sim}{\mu}(z_j) + \underset{\sim}{\epsilon}_j, \quad j\epsilon s, \tag{5.4}$$

so that

$$\hat{\underset{\sim}{\epsilon}}_j = \underset{\sim}{y}_j - \hat{\underset{\sim}{\mu}}(z_j), \quad j\epsilon s. \tag{5.5}$$

Thus

$$\hat{\underset{\sim}{\epsilon}}_j\hat{\underset{\sim}{\epsilon}}_j^T = (\underset{\sim}{y}_j - \hat{\underset{\sim}{\mu}}(z_j))(\underset{\sim}{y}_j - \hat{\mu}(\underset{\sim}{z}_j))^T \tag{5.6}$$

is an estimator of $\underset{\sim}{\Sigma}_{yy}(z_j)$. Applying a linear smoother to each term $\sigma_{ab}(z_j)$ of $\underset{\sim}{\Sigma}_{yy}(z_j)$ gives

$$\hat{\sigma}_{ab}(z) = \sum_{j\epsilon s} W_h(z,z_j)\hat{\epsilon}_{ja}\hat{\epsilon}_{jb}, \tag{5.7}$$

where $W_h(z,z_j)$ is a kernel with band width h which will usually be wider than the band width k chosen for the estimation of the conditional mean, (5.3).

The estimates of the marginal moments then employ the standard results that

$$\underset{\sim}{\mu}_y = \underset{\sim}{E}_z(\mu(z)), \tag{5.8}$$

$$\underset{\sim}{\Sigma}_{yy} = E_z(\underset{\sim}{\Sigma}_{yy}(z)) + V_z(\underset{\sim}{\mu}(z)). \tag{5.9}$$

Now

$$\underset{\sim}{\mu}_y = \int \underset{\sim}{\mu}(z)f(z)dz,$$

and our proposed estimator is

$$\hat{\underset{\sim}{\mu}}_y = \int \hat{\mu}(z)\hat{f}(z)dz. \tag{5.10}$$

Since N is large we propose using the empirical p.d.f. (Parzen 1962), given by

$$\begin{aligned} d\hat{F}(z) &= \hat{f}(z) = 1/N, \quad \text{if} \quad z = z_j, \quad j = 1, \ldots, N, \\ &= 0 \quad , \quad \text{otherwise} . \end{aligned} \tag{5.11}$$

Substituting in (5.10) gives the estimator

$$\hat{\underset{\sim}{\mu}}_y = N^{-1} \sum_{j=1}^{N} \hat{\mu}(z_j). \tag{5.12}$$

To estimate $\underset{\sim}{\Sigma}_{yy}$ we adopt a similar procedure for the first term of (5.9). The second term can be written

$$V_z(\underset{\sim}{\mu}(z)) = \int (\underset{\sim}{\mu}(z) - \underset{\sim}{\mu}_y)(\underset{\sim}{\mu}(z) - \underset{\sim}{\mu}_y)^T f(z)dz. \tag{5.13}$$

For our estimator we propose

$$\hat{V}_z(\underset{\sim}{\mu}(z)) = N^{-1} \sum_{j=1}^{N} (\hat{\underset{\sim}{\mu}}(z_j) - \hat{\underset{\sim}{\mu}}_y)(\hat{\underset{\sim}{\mu}}(z_j) - \hat{\underset{\sim}{\mu}}(y))^T. \tag{5.14}$$

Thus the proposed estimator of is $\underset{\sim}{\Sigma}_{yy}$ is

$$\hat{\underset{\sim}{\Sigma}}_{yy} = N^{-1}\left[\sum_{j=1}^{N} \{\hat{\underset{\sim}{\Sigma}}_{yy}(z_j) + (\hat{\underset{\sim}{\mu}}(z_j) - \hat{\underset{\sim}{\mu}}_y)(\hat{\underset{\sim}{\mu}}(z_j) - \hat{\underset{\sim}{\mu}}_y)^T\} \right]. \tag{5.15}$$

Njenga (1990) examines the asymptotic statistical properties of these estimators.

One of the main reasons for estimating $\underset{\sim}{\Sigma}_{yy}$ is to carry out some form of multivariate analysis, such as a regression analysis between two or more of the components of $\underset{\sim}{y}$. In the next section we report the results of a simulation study in which the simple regression coefficient between two y-variables is estimated from stratified random samples with different sampling fractions.

6. ESTIMATING A REGRESSION COEFFICIENT A SIMULATION STUDY

Let $\underset{\sim}{y} = (y_1,y_2)^T$ with mean $\underset{\sim}{\mu}_y = (\mu_1,\mu_2)^T$ and covariance matrix

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}.$$

We are interested in estimating a function of $\underset{\sim}{\Sigma}_{yy}$, the simple linear regression coefficient,

$$B_{12} = \sigma_{12}/\sigma_2^2. \tag{6.1}$$

The elements of $\underset{\sim}{\Sigma}_{yy}$ will be estimated using:

(i) the Pearson adjusted estimator of $\underset{\sim}{\Sigma}_{yy}$ based on (4.4),

(ii) the probability weighted version of (4.4),

(iii) a kernel estimator based on (5.14).

The corresponding estimators of B_{12}, or of its finite population equivalent B_{12U}, are denoted $\hat{B}_{12,ml}$, $\hat{B}_{12\,,pwml}$ and $\hat{B}_{12,nw}$ respectively. The estimator $\hat{B}_{12,ml}$ is indexed by *"ml"* because it is also the MLE under a multivariate normal model. The estimator $B_{12,nw}$ is indexed "*nw*" after Nadaraya (1964) and Watson (1964). The first two estimators were chosen because of their good performance in previous simulation studies, see Skinner *et al.* (1989, Ch.8).

We carried out three types of simulation study. In the first simulation study we generated a multivariate normal population to compare the performance of the new estimator with the maximum likelihood estimator which is optimal for this population. In the second simulation study we generated a quadratic homoscedastic population to compare the estimators when only the linearity assumption is violated. In the last simulation study we compared the estimators when the structure of the population is unknown, *i.e.* we used a 'real' population. In these simulation studies we carried out both conditional and unconditional analyses. The former allow us to assess whether a particular estimator is good in some samples and poor for others whereas the latter averages over all possible samples for a particular design.

The new estimator uses the Gaussian Kernel

$$W_k(z_i,z_j) = c_i \exp\{-(z_i - z_j)^2/2k^2\}, \quad i\epsilon U, \quad j\epsilon s,$$

where $c_i = 1/\sum_{j\epsilon s} \exp\{-(z_i - z_j)^2/2k^2\}$. A simulation with different values of the band width k showed that the mean squared error was relatively constant for a wide range of values of k and that this was achieved by trading off bias against variance. We selected values for k that gave relatively small values for the bias for each stratified sample design.

Since the 'real' population available to us was 6,962 observations from the 1975 UK Family Expenditure Survey we constructed all three populations to be of this size with mean vector and covariance matrix

$$\underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_z \end{bmatrix}, \quad \underset{\sim}{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1z} \\ & \sigma_2^2, & \sigma_{2z} \\ & & \sigma_z^2 \end{bmatrix}.$$

The actual values of $\underset{\sim}{\Sigma}$ are shown in Table 6.1.

The design variable is based on the expenditure on food, the independent variable is the total income and the dependent variable is the total expenditure. This finite population was stratified into five strata according to increasing values of the design variable, such that the first stratum contains 1,393 units with lowest values of z, second, third, fourth contain 1,392 units each and the fifth contains the last 1,393 units with the highest z values.

Table 6.1
Parameter Values from the Real Population

Variable		S.D.	Correlation matrix		
y_1	Expenditure on all items	0.668	1		
y_2	Total income	0.849	0.75	1	
z	Expenditure on food	0.658	0.41	0.28	1

Table 6.2
Stratified Sample Designs

Sample design		n_1	n_2	n_3	n_4	n_5	Symbol
D1	Proportional allocation	20	20	20	20	20	Δ
D2	Increasing allocation	5	9	16	30	40	∇
D3	U-shaped allocation	40	8	4	8	40	+

The sample designs used were based on those used by Holt, Smith and Winter (1980). Denote a stratified random sampling design by $(n_1 \ldots n_5)$ with n_h units selected from the h^{th} stratum, $h = 1, \ldots, 5$, then the designs are shown in Table 6.2, together with the symbols used in the plots.

For the various stratified sample designs we selected 1,000 independent samples of size 100 from the finite population. The sampling distribution of the various statistics under investigation were estimated from these 1,000 repeated samples. We obtain the unconditional results by averaging the statistics under investigation over all the 1,000 samples.

To assess the conditional properties of the estimators the 1,000 samples were divided into 20 groups of 50 samples each according to increasing values of $\Delta_{zz}^F = (S_{zzs} - S_{zz})/S_{zz}$ for the *nw* and *ml* estimators where

$$S_{zz} = N^{-1}\Sigma_U(z_i - \bar{z}_U)^2, \quad S_{zzs} = n^{-1}\ \Sigma_s(z_i - \bar{z}_s)^2,$$

$$\bar{z}_U = N^{-1}\Sigma_U z_i, \quad \bar{z}_s = n^{-1}\Sigma_s z_i,$$

and of $\Delta_{zz}^{*F} = (S_{zzs}^* - S_{zz})/S_{zz}$ for the *pwml* estimators where

$$S_{zzs}^* = \Sigma_s w_i(z_i - \bar{z}_s^*)^2, \quad \bar{z}_s^* = \Sigma_s w_i z_i, \quad w_i = (N\pi_i)^{-1} \text{ and } \pi_i$$

denotes the probability of including the i^{th} unit in the sample such that the first group contained the 50 samples with the smallest values of Δ_{zz}^F (or Δ_{zz}^{*F}) and so on up to the 20th group which contains the 50 samples with the largest values of Δ_{zz}^F (or Δ_{zz}^{*F}). We assume that the variation in Δ_{zz}^F (or Δ_{zz}^{*F}) within each group is small. The conditional distribution of the various estimators given Δ_{zz}^F (or Δ_{zz}^{*F}) can then be plotted.

The biases, standard deviations and mean square errors reported in simulation studies 1 and 2 are computed around the value of B_{12U} in the finite population generated from the model. This enables them to be compared with the values generated from the real finite population in simulation study 3.

Table 6.3

Unconditional Absolute Biases of the Three Estimators of B_{12}
$N = 6{,}962$, $n = 100$ True Value $B_{12} = 0.595$

Sample design	Absolute biases of		
	$\hat{B}_{12,ml}$	$\hat{B}_{12,pwml}$	$\hat{B}_{12,nw}$
D1	0.0003	0.0003	0.0185
D2	0.0007	0.0019	0.0269
D3	0.0026	0.0018	0.0159

Table 6.4

Unconditional Standard Deviation of the Three Estimators of B_{12}

Sample design	Standard deviations		
	$\hat{B}_{12,ml}$	$\hat{B}_{12,pwml}$	$\hat{B}_{12,nw}$
D1	0.0500	0.0500	0.0507
D2	0.0522	0.0693	0.0531
D3	0.0486	0.0710	0.0503

Table 6.5

Unconditional Mean Square Errors of the Three Estimators of B_{12}

Sample design	Mean square errors		
	$\hat{B}_{12,ml}$	$\hat{B}_{12,pwml}$	$\hat{B}_{12,nw}$
D1	0.0025	0.0025	0.0029
D2	0.0027	0.0048	0.0035
D3	0.0024	0.0050	0.0028

Simulation Study 1

In the first simulation study the 6,962 finite population values were generated from a multivariate normal distribution with correlation matrix given in Table 6.1. These data should be favourable to the estimator $\hat{B}_{12,ml}$.

The unconditional biases, standard deviations and mean squared errors are shown in Tables 6.3, 6.4 and 6.5.

As expected the estimator $\hat{B}_{12,ml}$ is best in terms of mean squared error. The new estimator $\hat{B}_{12,nw}$ does surprisingly well, it has a large bias but a similar standard deviation. The size of the bias for a very smooth (linear) population is consistent with the results in other studies, see Gasser and Engel (1990). A very wide bandwidth is needed to capture a very smooth function.

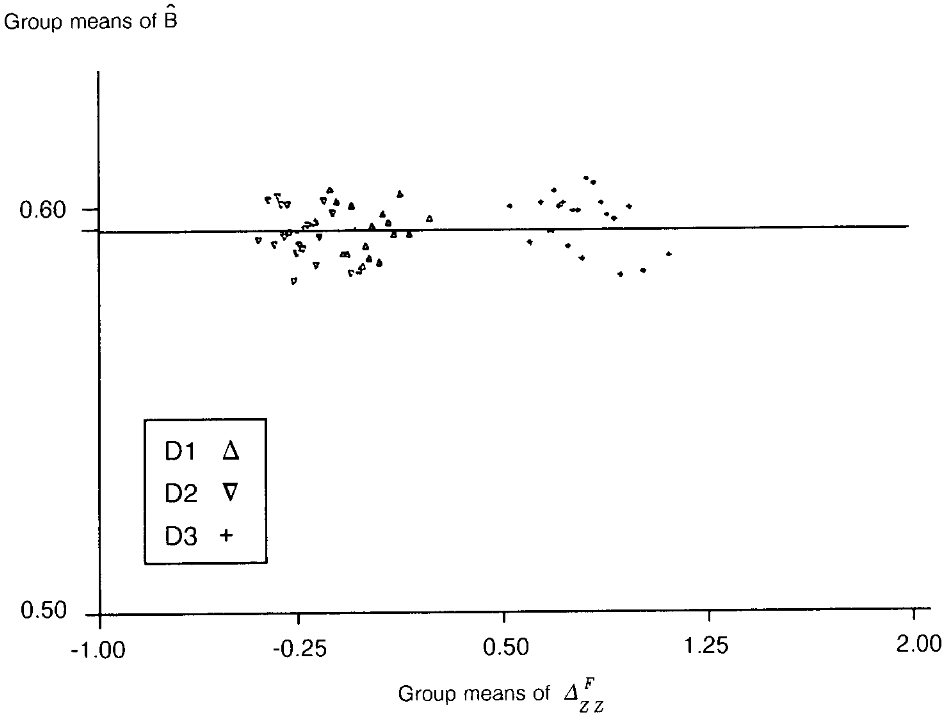

Figure 6.1 Scattergram of group means of $\hat{B}_{12,ml}$

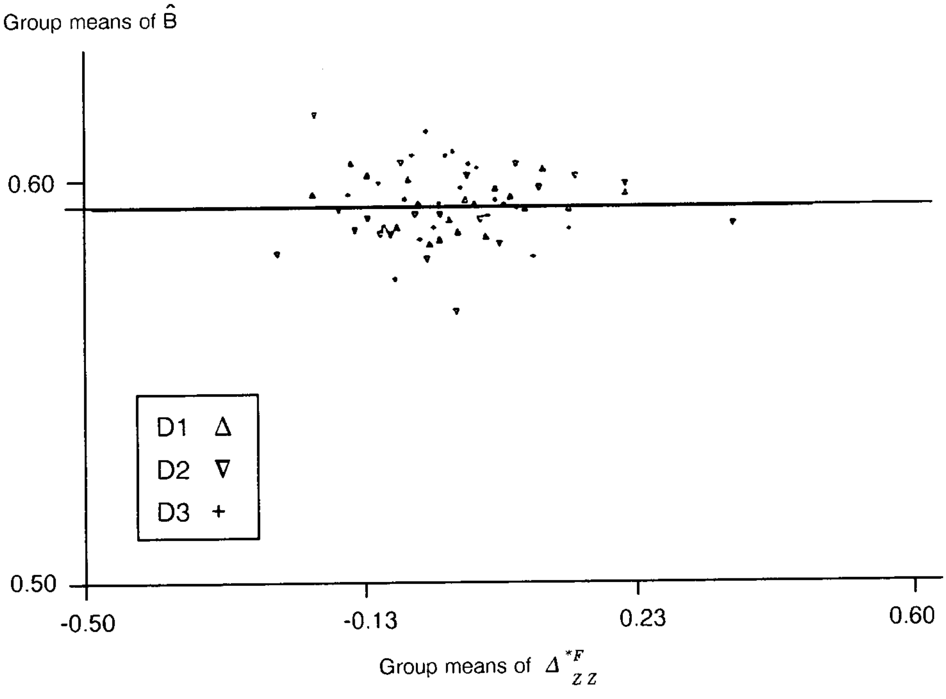

Figure 6.2 Scattergram of group means of $\hat{B}_{12,pwml}$

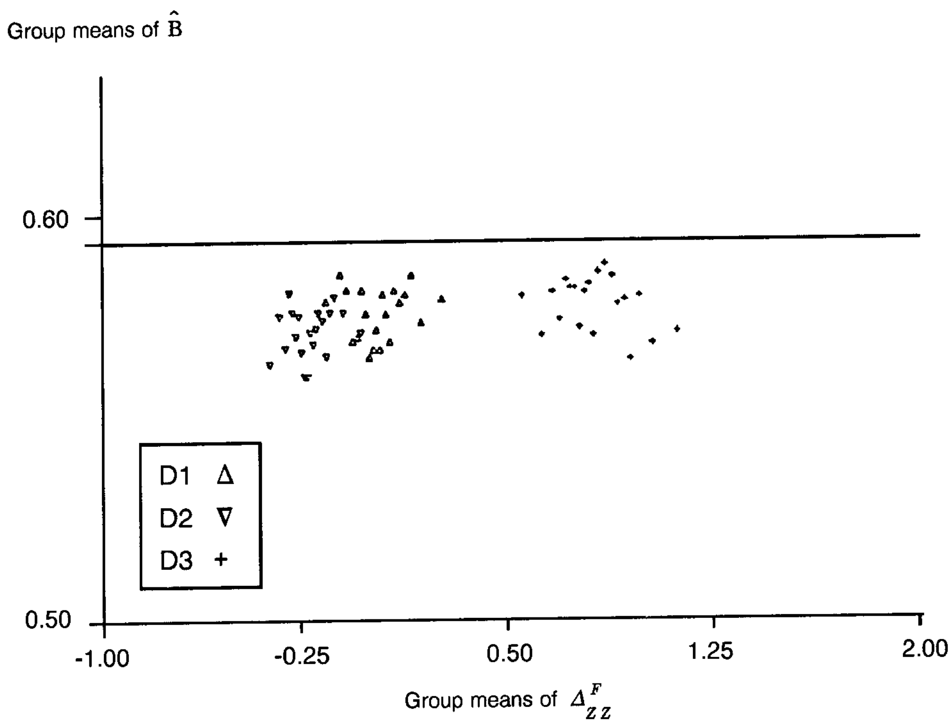

Figure 6.3 Scattergram of group means of $\hat{B}_{12,nw}$

The conditional plots are shown in Figures 6.1, 6.2 and 6.3. These plots show that there is no additional pattern to the bias beyond the absolute level of bias shown in Table 6.3. Previous studies have shown consistent patterns of bias for SRS estimators and simple p-weighted estimators, see Skinner *et al.* (1989, Chs. 7 and 8).

Simulation Study 2

Repeated sampling from a quadratic homoscedastic population

This simulation study is similar to one carried out by Holmes (1987). We generated 6,962 finite population values of (y_{1i}, y_{2i}, z_i) $i = 1 \ldots 6{,}962$ by first generating a value of z_i from the uniform distribution $U(0,10)$. Using this generated value of z_i the corresponding values of y_{1i} and y_{2i} are obtained from the relationships;

$$y_{2i} = m_2 + H_2 z_i + R_2 z_i^2 + \epsilon_{2i}$$

and

$$y_{1i} = m_1 + H_1 z_i + R_1 z_i^2 + \epsilon_{1i},$$

where ϵ_{2i} and ϵ_{1i} are random variables from normal distributions with mean zero and constant variance, and $R_1 \neq 0$, $R_2 \neq 0$. Following Holmes (1987) we chose the parameters in these expressions so that the regressions of y_1 and y_2 on z are monotonically increasing functions of z and the regression of y_1 on y_2 is approximately linear so that the regression coefficient B_{12} will be a meaningful parameter to estimate.

Table 6.6

Unconditional Standard Deviation of the Three Estimators of B_{12}
$N = 6{,}962$, $n = 100$ True Value $B_{12} = 0.857$

Sample design	Absolute biases of		
	$\hat{B}_{12,ml}$	$\hat{B}_{12,pwml}$	$\hat{B}_{12,nw}$
D1	0.0119	0.0119	0.0171
D2	0.0923	0.0132	0.5556
D3	0.0124	0.0098	0.0104

Table 6.7

Unconditional Standard Deviation of the Three Estimators of B_{12}

Design	Standard deviations		
	$\hat{B}_{12,ml}$	$\hat{B}_{12,pwml}$	$\hat{B}_{12,nw}$
D1	0.0877	0.0877	0.0877
D2	0.0972	0.1230	0.1150
D3	0.0785	0.1110	0.0797

Table 6.8

Unconditional Mean Square Errors of the Three Estimators of B_{12}

Sample design	Mean square errors		
	$\hat{B}_{12,ml}$	$\hat{B}_{12,pwml}$	$\hat{B}_{12,nw}$
D1	0.0078	0.0078	0.0080
D2	0.0180	0.0153	0.0164
D3	0.0063	0.0124	0.0065

The unconditional results of the three estimators of the regression coefficient are given in Tables 6.6, 6.7 and 6.8.

We see from the tables that the ml estimator is severely biased and very inefficient for the increasing allocation design D2, but is approximately unconditionally unbiased and efficient for the designs D1 and D3. The *pwml* estimator as expected is approximately unconditionally unbiased across all the sample designs considered. Though more biased than the *pwml* estimator, the *nw* estimator is less biased than the *ml* estimator for the unequal probability designs. We also see that the *nw* estimator is more efficient than *ml* for the design D2 and approximately equally efficient for design D3. It is also more efficient than the *pwml* estimator for the U-shaped design D3.

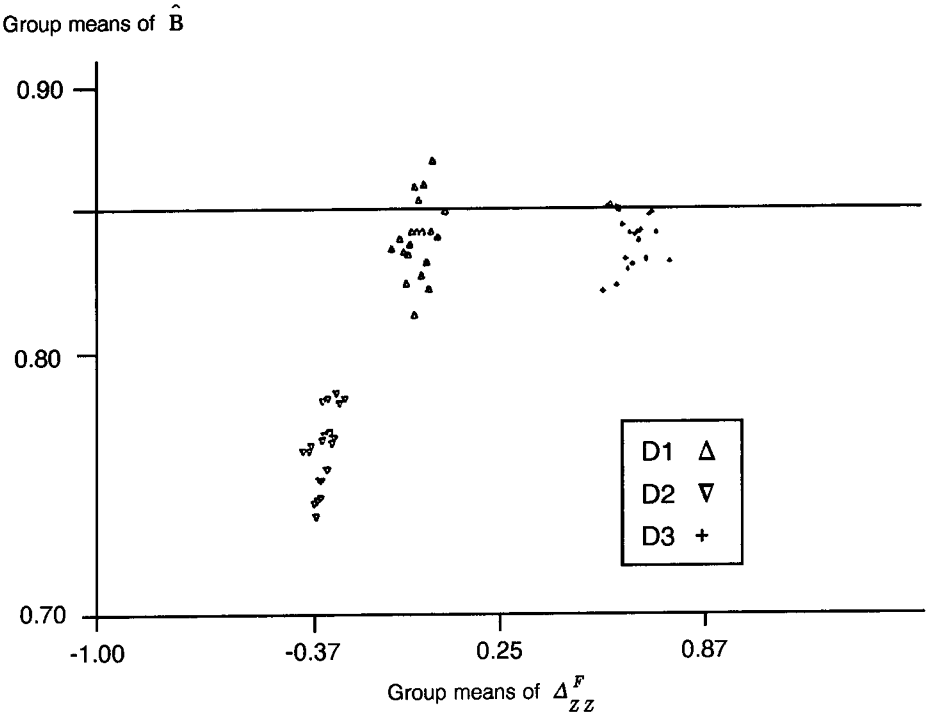

Figure 6.4 Scattergram of group means of $\hat{B}_{12,ml}$

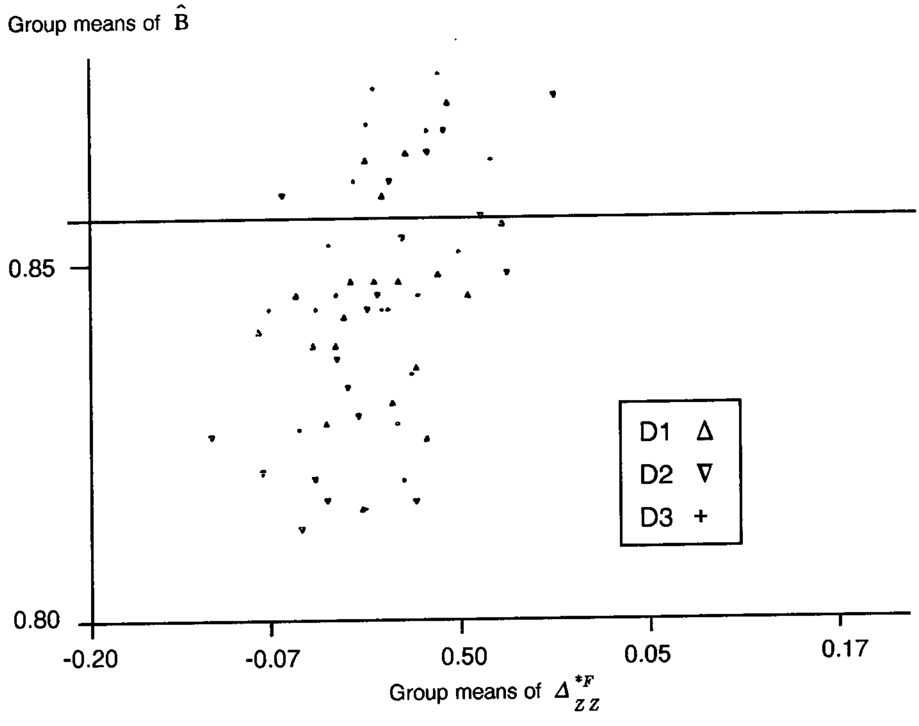

Figure 6.5 Scattergram of group means of $\hat{B}_{12,pwml}$

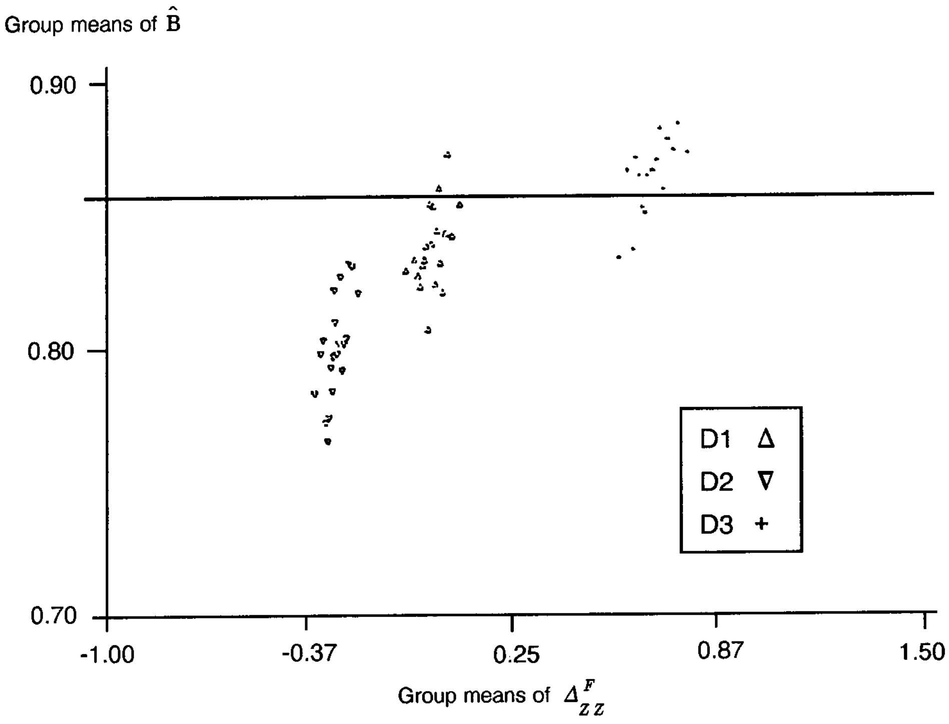

Figure 6.6 Scattergram of group means of $\hat{B}_{12,nw}$

The plots of the conditional analysis are shown in Figures 6.4, 6.5 and 6.6.

We see from Figure 6.4 that the *ml* estimator is approximately conditionally unbiased for the design D1 and D3, and has no additional conditional bias for the design D2. From Figure 6.5 we see that the *pwml* estimator has no additional conditional bias for any of the designs. We see from Figure 6.6 that the *nw* kernel estimator has only a small additional conditional bias within each of the three probability designs.

Simulation Study 3

Repeated sampling from a multivariate 'Real' population

In this simulation study we employ the 6,962 actual data points from the Family Expenditure Survey for the finite population. We consider the same variables as in section 3.1 and sample repeatedly from this population to investigate the robustness properties of the three regression estimators. We expect the real population to violate all the normality assumptions.

The unconditional results are shown in Tables 6.9, 6.10 and 6.11, and we see that the *nw* kernel estimator is the most efficient and is approximately unconditionally unbiased across all the probability designs. The *ml* estimator is less biased and more efficient than the *pwml* estimator for the unequal probability designs.

The plots of the conditional analyses are shown in Figures 6.7, 6.8 and 6.9.

We see from Figure 6.7 that the *ml* estimator is approximately conditionally unbiased for the designs D1 and D2 but has a slight conditional bias for design D3. From Figure 6.8 we see that the *pwml* estimator has no additional conditional bias for any of the designs. From Figure 6.9 we see that the *nw* kernel estimator is approximately conditionally unbiased for the three probability designs.

Table 6.9

Unconditional Absolute Biases of the Three Estimators of B_{12}
$N = 6{,}962$, $n = 100$ True Value $B_{12} = 0.595$

Sample design	Absolute biases of		
	$\hat{B}_{12,ml}$	$\hat{B}_{12,pwml}$	$\hat{B}_{12,nw}$
D1	0.0245	0.0245	0.0056
D2	0.0260	0.0408	0.0060
D3	0.0128	0.0355	0.0072

Table 6.10

Unconditional Standard Deviation of the Three Estimators of B_{12}

Sample design	Standard deviation		
	$\hat{B}_{12,ml}$	$\hat{B}_{12,pwml}$	$\hat{B}_{12,nw}$
D1	0.111	0.111	0.111
D2	0.106	0.132	0.108
D3	0.111	0.122	0.111

Table 6.11

Unconditional Mean Square Errors of the Three Estimators of B_{12}

Sample design	Mean square errors		
	$\hat{B}_{12,ml}$	$\hat{B}_{12,pwml}$	$\hat{B}_{12,nw}$
D1	0.0130	0.0130	0.0121
D2	0.0120	0.0192	0.0117
D3	0.0125	0.0161	0.0123

We conclude from these simulation studies that the new estimator $\hat{\beta}_{12,nw}$ has performed well. When the assumptions of linearity and homoscedasticity are violated it appears to be robust across a variety of designs, to have good efficiency and to have reasonable conditional as well as unconditional properties. We know from previous studies that $\hat{\beta}_{12,pwml}$ performs as well as more conventional p-weighted estimators unconditionally and has far better conditional properties. The fact that in this study the new estimator $\hat{B}_{12,nw}$ apparently has better properties than the *pwml* estimator, which was chosen to represent the class of p-weighted estimators because of its performance in other simulation studies, suggests that it is an approach that could be considered in analytic studies of a small number of key parameters.

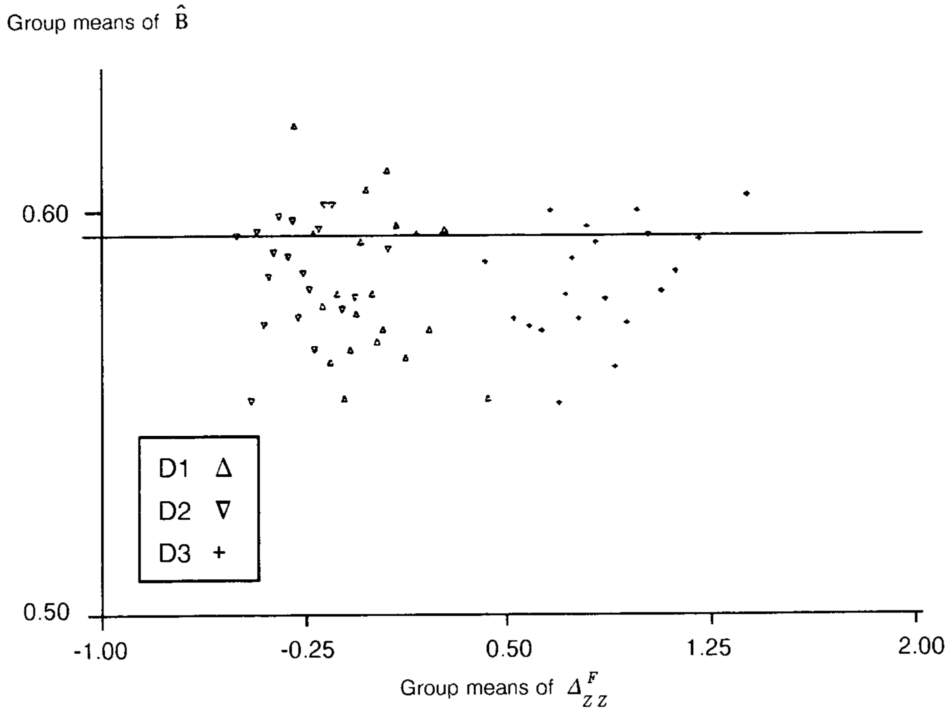

Figure 6.7 Scattergram of group means of $\hat{B}_{12,ml}$

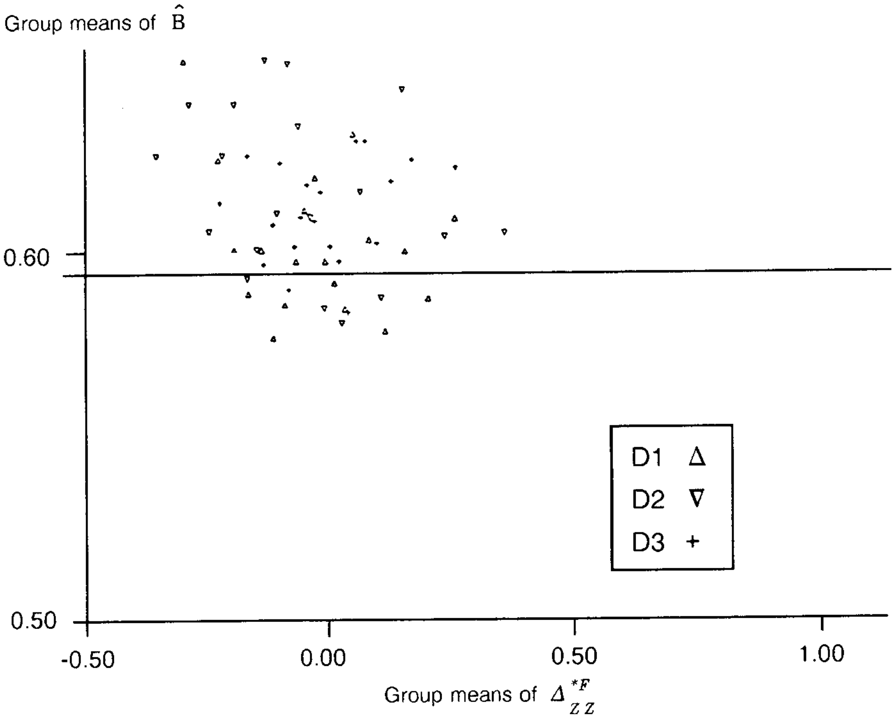

Figure 6.8 Scattergram of group means of $\hat{B}_{12,pwml}$

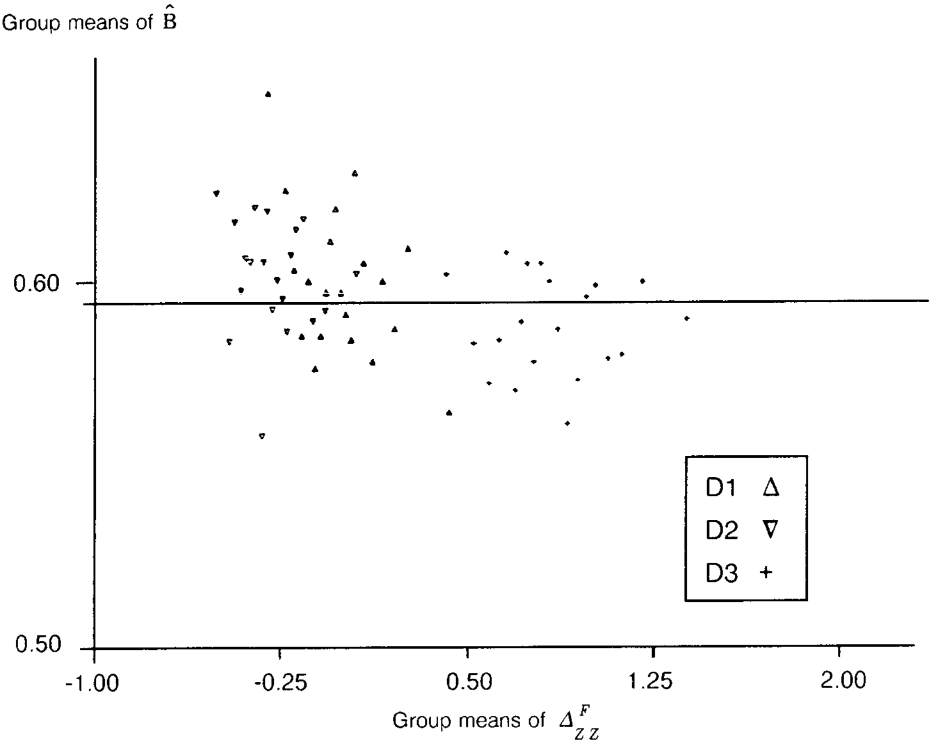

Figure 6.9 Scattergram of group means of $\hat{B}_{12,nw}$

ACKNOWLEDGEMENTS

The authors wish to thank an anonymous referee for many helpful comments which improved the presentation of the paper. E. Njenga was supported by a grant from the British Council.

REFERENCES

BINDER, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51, 279-292.

BREWER, K.R.W. (1979). A class of robust designs for large scale surveys. *Journal of the American Statistical Association*, 74, 911-915.

BREWER, K.R.W., and SÄRNDAL, C.-E. (1983). Six approaches to enumerative survey sampling. *Incomplete Data in Sample Surveys*, (Vol. 3). New York: Academic Press, 363-368.

CLEVELAND, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74, 829-836.

GASSER, T., and MULLER, H.G. (1979). Kernel estimation of regression functions. *Smoothing Techniques for Curve Estimation*, (Eds. T. Gasser and M. Rosenblatt). New York: Springer-Verlag, 23-68.

GASSER, T., and ENGEL, J. (1990). The choice of weights in kernel regression estimation. *Biometrika*, 77, 377-381.

GODAMBE, V.P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society*, B, 17, 269-278.

GODAMBE, V.P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 31, 1208-1211.

GODAMBE, V.P. (1982). Estimation in survey sampling: Robustness and optimality. *Journal of the American Statistical Association*, 77, 393-403.

GODAMBE, V.P., and THOMPSON, M.E. (1977). Robust near optimal estimation in survey practice. *Bulletin of the International Statistical Institute*, 47, 129-146.

GODAMBE, V.P., and THOMPSON, M.E. (1986). Parameters of superpopulation and survey population: their relationships and estimation. *International Statistical Review*, 54, 127-138.

HANSEN, M.H., MADOW, W.G., and TEPPING, B.J. (1983). An evaluation of model dependent and probability sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78, 776-793.

HOLT, D., SMITH, T.M.F., and WINTERS, P.D. (1980). Regression analysis of data from complex surveys. *Journal of the Royal Statistical Society*, Ser. A, 143, 474-487.

HOLMES, D. (1987). The effect of selection on the robustness of multivariate methods. Unpublished Doctoral thesis, University of Southampton, U.K.

HORVITZ, D.G., and THOMPSON, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

ISAKI, C.T., and FULLER, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

KOTZ, S., and JOHNSON, N.L. (1988). *Encyclopedia of Statistical Sciences*, (Vol. 8). New York: John Wiley, 157.

LITTLE, R.J.A. (1983). Estimating a finite population mean from unequal probability samples. *Journal of the American Statistical Association*, 78, 596-604.

NADARAYA, E.A. (1964). On estimating regression. *Theory of Probability Application*, 9, 141-142.

NATHAN, G., and HOLT, D. (1980). The effect of survey design on regression analysis. *Journal of the Royal Statistical Society*, B, 42, 377-386.

NJENGA, E.G. (1990). Robust estimation of the regression coefficients in complex surveys. Unpublished Ph.D. thesis, University of Southampton.

PARZEN, E. (1962). On the estimation of the probability density function and mode. *Annals of Mathematical Statistics*, 33, 1065-1076.

PEARSON, K. (1903). On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions Royal Society of London*, A, 200, 1-66.

PFEFFERMANN, D.J., and HOLMES, D. (1985). Robustness consideration in the choice of method of inference for regression analysis of survey data. *Journal of the Royal Statistical Society*, A, 148, 268-278.

ROYALL, R.M., and CUMBERLAND, W.G. (1981). An empirical study of the ratio estimator and estimators of its variance. *Journal of the American Statistical Association*, 76, 66-88.

ROYALL, R.M., and HERSON, J. (1973a). Robust estimation in finite populations I. *Journal of the American Statistical Association*, 68, 880-889.

ROYALL, R.M. and HERSON, J. (1973b). Robust estimation in finite populations II. *Journal of the American Statistical Association*, 68, 890-893.

SÄRNDAL, C.-E. (1980). On π-inverse weighting versus best linear weighting in probability sampling. *Biometrika*, 67, 639-650.

SCOTT, A.J. (1977). On the problem of randomization in survey sampling. *Sankhyā*, C, 39, 1-9.

SKINNER, C.J., HOLT, D., and SMITH, T.M.F. (1989). *Analysis of Complex Surveys*. New York: Wiley.

SILVERMAN, B.W. (1985). Some aspects of the spline smoothing approach to nonparametric curve fitting. *Journal of the Royal Statistical Society*, B, 47, 1-52.

SUGDEN, R.A., and SMITH, T.M.F. (1984). Ignorable and informative designs in survey sampling inference. *Biometrika*, 71, 495-506.

WATSON, G.S. (1964). Smooth regression analysis. *Sankhyā*, A, 359-372.