Daniela Cocchi* and Michel Mouchart**

Quasi-linear Bayes estimation in stratified finite populations

*Dip. Sc. Statistiche "P. Fortunati" Università di Bologna
**CORE, Université Catholique de Louvain

Serie Ricerche n.4

Dipartimento di Scienze Statistiche "Paolo Fortunati"
Università degli studi di Bologna
1992

Finito di stampare nel mese di Agosto 1992
presso le Officine Grafiche Tecnoprint
Via del Legatore 3, Bologna

1 Introduction

In finite populations inference, the random effects analysis of variance model has known particular relevance for a long time, especially within the Bayesian context. In particular the contribution of Scott and Smith (1969) on this subject is fundamental, together with the review of Ericson (1988). A common characteristic of the developments contained in those papers is the essentially normal context, with or without noninformative priors. The model can conveniently correspond to a two-stage sampling, with a simpler modelization, or to stratified sampling.

While further analysing this setting out (Cocchi, 1985; Cocchi and Mouchart, 1990), we realised that some features deserve particular care. First of all, since the model is very strongly linked to the normal context, a solution able to keep departures from normality into account may have some interest, but, instead of a normal solution, a least squares approximation with the same two moments may be an answer to the preceding point. To this aim, the choice of the coordinates for the solution of the problem is fundamental. In particular, the proposal of a projection made directly on the observations maintains rigourously the solution within the normal context. Also the extent to which a hierarchical formulation of the model is specified is relevant, together with the way of assigning prior evaluations.

So, with reference to the mentioned points, we are developing an extension of the idea of Bayesian least squares approximations to the finite population context. Particularly in such context, in fact, when an exact distribution is considered, the computation of the posterior moments suffers of the complexity of the likelihood function and the difficulty of integrating the nuisance parameters, in particular the variances, and an approximate solution reveals to be interesting. Since a preminent characteristic of a Bayesian solution is the derivation of the joint distribution of observations and parameters, or at least of its first two moments when least squares approximations are evaluated, and such result is obtained with successive integrations, it has been possible to see clearly when the various prior evaluations enter in the solution.

The solution is more or less complex according to the distributional assumptions made, to the choice of the statistic on which to condition and to the fact that the solution be exact or approximated. We shall see that the specification of only first and second order moments is sufficient to build the joint distribution of parameters and data. The differences between alternative solutions appear in the number of prior evaluations to be elicitated and in the fact that, even if only expectations and variances are to be computed, moments of order up to the fourth can be involved.

The solution here proposed, which keeps departures from normality into account, by means of the prior elicitation of 3rd and 4th order moments, permits to consider the issue of robustness in a different way from that usually considered for superpopulation models in finite populations (see, for instance, Royall and Pfefferman, 1982 and Bolfarine *et al.*, 1987)

In Section 2 the hypotheses necessary to deal with the problem are listed in a stepwise manner. In particular, it is stressed how they aim at obtaining admissible reductions of the model and at looking for operational simplifications. In Section 3, the idea of finding a solution by least squares approximations is developed. In Section 4 a Bayesian least squares approximation which considers the conditioning to an appropriate statistic is developed. The solution, able to keep departures from normality into account, is also compared with the purely normal one. In Section 5, a comparison between different solutions to the model studied in this paper, some of which have already been developed in the literature, is performed. First of all, heuristic and analytical differences are explored, then the comparison is continued with an application which uses simulated data.

2 Fundamental hypotheses on the model

As in Cocchi and Mouchart (1990) the basic model may be presented in the framework of a hierarchical linear form:

$$\mu = \mathbf{Z}_1 \mu_0 + \delta \tag{1}$$

$$\eta = \mathbf{Z}_2 \mu + \varepsilon \tag{2}$$

where μ and δ are p-vectors, η and ε are N-vectors, μ_0 is a q-vector, $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$ is constituted by matrices of appropriate dimensions.

As far as notation is concerned, greek letters denote unobservables, latin letters denote observables, matrices and functions are denoted by bold capital letters, small bold letters are used for vectors and scalars are denoted by italics. The framework being Bayesian, all letters represent random variables (assumed defined on a probability space). Thus the Bayesian modelling starts from the joint distribution of all the

involved variables and introduces, in a stepwise manner, a sequence of hypotheses aiming at obtaining admissible reductions of the underlying model and at looking for operational simplifications.

The motivation for the detailed conditions that will be discussed in what follows stands in the fact that structural hypotheses are typically formulated in terms of conditional independence and are typically not testable from an empirical point of view. Therefore it is important to write them out explicitly and to understand their logical meaning, in particular in properly understanding their implications.

The interest is focused on a finite population model, where N stands for the size of the population. Since in this paper we concentrate on an Anova II type model, $\mathbf{Z}_2$ is a $N \times p$ matrix with the following block diagonal structure: $\mathbf{Z}_2 = [\mathbf{i}_{N_k} \quad \mathbf{e}_k'] = \text{diag}\,[\mathbf{i}_{N_k}]$, where $\mathbf{i}_{N_k} = (1, 1, \quad \dots \quad 1)' \in \Re^{N_k}$ and $\mathbf{e}_k$ is the k-th column of $\mathbf{I}_p$. The N_k' are such that $\Sigma_k N_k = N$, $1 \le k \le p$, while $\mathbf{Z}_1 = \mathbf{i}_p$ is a p-vector of units and μ_0 is assumed to be a scalar, *i.e.* $q = 1$.

The model is therefore suitable both for describing a situation of two stage sampling, with a simple modelization, since hierarchy stands in the way of selection, and for describing stratification, with a hierarchical modelization, p being the number of groups to which each element of the population can be allocated by means of matrix $\mathbf{Z}_2$. The consideration of superpopulation models begins to be an appreciated tool also in small domain estimation, starting from the contribution of Holt, Smith and Tomberlin (1979). In particular, the solution we are developing is a natural extension of the contribution of Lui and Cumberland (1989).

In this section we wish to stress the role and the relevance of the hypotheses necessary for the solution: details more oriented to the description of the superpopulation model can be found in Cocchi and Mouchart (1990, 1989).

2.1 *Structural modeling*

The variables of the model are $(\eta, \mu, \mathbf{Z})$. In particular, η and μ are unobservable and indeed the purpose of sampling is to observe a part of vector η. The modeling we present in this section has a double aim: we want to build an admissible reduction of the joint distribution of $(\eta, \mu, \mathbf{Z})$ to the conditional model $(\eta, \mu \mid \mathbf{Z})$, in this way operating the first of two

Bayesian cuts. With a second Bayesian cut we essentially look for obtaining easier analytical computations. The conditions concerning the parameters of the model are stated hierarchically in what follows.

2.1.1 A first Bayesian cut

First, let the joint distribution of (η, μ, Z) be characterized by a parameter $\tau = (\theta, \psi)$, where ψ is a parameter sufficient for the marginal process generating Z and θ is a parameter sufficient for the conditional process generating $(\eta, \mu \mid Z)$. More specifically, ψ and θ are defined by the properties:

$$Z \perp\!\!\!\perp \tau \mid \psi \tag{3}$$

$$(\eta, \mu) \perp\!\!\!\perp \tau \mid Z, \theta \quad . \tag{4}$$

The hypothesis of cut between (η, μ) and Z reads as:

$$\text{H1.} \quad \psi \perp\!\!\!\perp \theta. \tag{5}$$

In what follows we consider that θ contains the parameters of interest, while ψ constitutes a nuisance parameter. In this way, Z is exogenous for the inference on θ, and it is admissible to concentrate the attention on $p(\eta, \mu \mid Z, \theta)$. (On this subject see also Florens and Mouchart, 1985 and Florens, Mouchart and Rolin, 1990).

2.1.2 A second Bayesian cut

A second Bayesian cut between μ and η conditional on Z is based on a decomposition of θ relying on the definitions:

$$\mu \perp\!\!\!\perp \theta \mid Z, \theta_1 \tag{6}$$

$$\eta \perp\!\!\!\perp \theta \mid Z, \theta_2, \mu \tag{7}$$

which respectively mean that θ_1 is a parameter sufficient for describing the process generating $\mu \mid Z$ and θ_2 is a parameter sufficient for describing the process generating $(\eta \mid Z, \mu)$. The second hypothesis of cut assumes:

$$\text{H2.} \quad \theta_1 \perp\!\!\!\perp \theta_2 \mid Z. \tag{8}$$

The statistical meaning of (1) is now made explicit by the assumption:

$$\text{H3.} \quad \mu \perp\!\!\!\perp Z_2 \mid Z_1, \theta, \tag{9}$$

meaning that $\mu \mid Z, \theta$ does not depend on Z_2, and is therefore characterized by (1) under the usual implicit assumption that δ has zero expectation and is distributed independently of (Z_1, μ_0), *i.e.*:

$$E(\mu \mid Z, \theta) = E(\mu \mid Z_1, \theta_1) = Z_1\mu_0, \tag{10}$$

Consequently, $\theta_1 = (\Phi_1, \mu_0)$, where Φ_1 is the (multivariate) distribution function of the vector δ. Along with (6), H3 implies

$$\mu \perp\!\!\!\perp (Z, \theta) \mid Z_1, \theta_1. \tag{11}$$

In the same way the statistical meaning of (2) is explicated by the hypothesis

$$\text{H4.} \quad \eta \perp\!\!\!\perp Z_1 \mid Z_2, \theta, \mu \tag{12}$$

i.e. $\eta \mid Z, \mu, \theta$ does not depend on Z_1 and is therefore characterized by (2) under the corresponding implicit assumptions that ε has zero expectation and is distributed independently of (Z_2, μ), and therefore

$$E(\eta \mid Z, \theta, \mu) = E(\eta \mid Z_2, \theta_2, \mu) = Z_2\mu. \tag{13}$$

Consequently, $\theta_2 = (\Phi_2)$, where Φ_2 is the multivariate distribution function of the vector ε. Again, along with (7), H4 implies

$$\eta \perp\!\!\!\perp (Z, \theta) \mid Z_2, \theta_2, \mu. \tag{14}$$

Note that Z_1 needs to be not disjointed from Z_2. When it happens, an identification problem may possibly arise.

Remark 1. Property (3), with H1 and H2 imply (see Florens, Mouchart and Rolin, 1990, Corollary 7.6.4):

$$\theta_1 \perp\!\!\!\perp \theta_2 \perp\!\!\!\perp Z, \tag{15}$$

in particular:

$\theta_1 \perp\!\!\!\perp \theta_2$ and $\theta \perp\!\!\!\perp Z$. (16)

Remark 2. Note also that (6) and H2 are equivalent to $\theta_2 \perp\!\!\!\perp (\theta_1, \mu) \mid Z$ and therefore imply:

$$\mu \perp\!\!\!\perp \theta_2 \mid Z. \quad (17)$$

2.1.3 Conditions of mutual independence

Whereas conditions of cut allow for decomposing inferences and eventually concentrating the inference about the parameters of interest on the relevant part of the sampling process, conditions of mutual independence allow for easier accumulation of statistical information. These hypotheses concern the processes generating both $(\eta \mid \mu, Z)$ and $(\mu \mid Z)$.

For what concerns $(\eta \mid \mu, Z)$, the basic hypothesis is that Z_2 "sifts" η conditionally on (θ, μ), meaning that (see Florens, Mouchart and Rolin, 1990, Definition 7.6.11)

H5. $\underset{(0 \le \lambda \le N)}{\perp\!\!\!\perp} \eta_\lambda \mid Z_2, \theta_2, \mu$ (18)

and, partitioning $Z_2 = (z_{2,1} \dots z_{2,\lambda} \dots z_{2,N})'$:

H6. $\eta_\lambda \perp\!\!\!\perp Z_2 \mid z_{2,\lambda}, \theta_2, \mu$ (19)

i.e. conditionally on (Z_2, θ_2, μ), the η_λ's are mutually independent (H5) and the individual $z_{2,\lambda}$'s are "allocated" to the individual η_λ (H6).

Similarly, the model assumes that Z_1 sifts μ, *i.e.*:

H7. $\underset{(0 \le k \le p)}{\perp\!\!\!\perp} \mu_k \mid \theta_1, Z_1$ (20)

and, partitioning $Z_1 = (z_{1,1} \dots z_{1,k} \dots z_{1,p})'$,

H8. $\mu_k \perp\!\!\!\perp Z_1 \mid z_{1,k}, \theta_1$ (21)

Note that H5 and H7 imply that both $V(\eta \mid Z_2, \mu, \theta_2)$ and $V(\mu \mid Z_1, \theta_1)$ are diagonal matrices. Finally, we complete the structural model by assuming homoscedasticity both for η and μ, *i.e.*, from H5:

H9. $V(\eta \mid Z, \theta, \mu) = V(\eta \mid Z_2, \theta_2, \mu) = \sigma_2^2 I_N$ (22)

and, from H7:

H10. $V(\mu \mid Z_1, \theta_1) = \sigma_1^2 I_p$. (23)

2.2 Sampling model

A basic feature of finite population models is that the individual characteristics η_λ's are only partially observable. More precisely, a sampling mechanism will select n labels (out of N elements of the population): $s = \{s_1 \ \dots \ s_n\} \subset \{1 \ \dots \ N\}$. With such notation, we retain only distinct labels, dropping repetitions out. It will be convenient to represent the sampling results by a $(n \times N)$ selection matrix $S = (e_{s_1} \ \dots \ e_{s_n})'$, where e_{s_i} is the s_i-th column of I_N. If we also define a $(N-n) \times N$ selection matrix $\bar{S}$ to represent the unsampled labels, we see that the sampling result s determines a $(N \times N)$ permutation matrix S^* as follows : $S^* = (S', \bar{S}')'$.

When sampling is introduced in the presence of an auxiliary variable, also the so called hypothesis of partial design noninformativity, first introduced by Scott (1977) and also discussed in Sugden and Smith (1984), is very natural:

H11. $S \perp\!\!\!\perp (\eta, \mu, \theta) \mid Z$. (24)

The former hypothesis, together with H5, implies:

$$S\eta \perp\!\!\!\perp \bar{S}\eta \mid Z, S, \theta, \mu \quad (25)$$

i.e. gives the possibility of reordering the distribution of η conditional on (Z, S, θ, μ) according to the fundamental partitioning of the population in sampled $(y = S\eta)$ and unsampled $(\bar{S}\eta)$ units.

Remark 3. Assumption H11 may be obtained through a somewhat more involved argument. Let us denote by χ the distribution $p(s \mid Z, \mu, \theta, \eta)$. Thus, χ characterizes the sampling design. So, (partial) noninformativity of labels means:

$$S \perp\!\!\!\perp (\eta, \mu, \theta) \mid Z, \chi$$

and (partial) non informativity of the design means:

$$\chi \perp\!\!\!\perp (\eta, \mu, \theta) \mid Z.$$

The joint consideration of the two partial noninformativities are equivalent to:

$(S,\chi) \perp\!\!\!\perp (\eta,\mu,\theta) \mid Z$

which clearly implies H11.

Remark 4. Together, H11 and H2 of a cut conditional on **Z** also imply a cut conditional on (**Z**, S):

$$\mu \perp\!\!\!\perp \theta \mid Z, S, \theta_1 \tag{6bis}$$

$$\eta \perp\!\!\!\perp \theta \mid Z, S, \mu, \theta_2 \tag{7bis}$$

$$\theta_1 \perp\!\!\!\perp \theta_2 \mid Z, S. \tag{8bis}$$

Remark 5. Similarly , (17) also holds conditionally on S:

$$\theta_2 \perp\!\!\!\perp \mu \mid Z, S. \tag{17bis}$$

3 Least squares approximations

The hierarchical model just described involves two kinds of parameters: the vector (μ', η') of incidental parameters (or latent variables) and the vector θ of structural parameters. In this paper the inference effort is concentrated on γ, a vector built from the $(N - n + p)$-vector of the unobserved individual characteristics $\bar{S}\eta$ and the (partially) incidental parameter μ: *i.e.* $\gamma = [(\bar{S}\eta)', \mu']'$. The data are (Z, S, y).

More specifically we want to approximate the posterior expectation $E(\gamma \mid Z, S, y)$, our motivations being the care for the well known possible pathologies of the likelihood function (eventually leading to negative unconstrained maximum likelihood estimates of variance), the aversion for too model-dependent procedures and the possibilities of taking non-normalities into account (such as those detected by third and fourth moments) within a reasonable computational cost.

3.1 *The choice of a set of approximating functions*

Looking for approximations requires the specification of both an approximation criterion and a set of approximating functions. Here, the criterion is taken to be the least squares principle. As far as approximating functions are concerned, a naive procedure would evaluate:

$$\hat{E}(\gamma \mid Z, S, y) = \arg\inf_{l} E[\| E(\gamma \mid Z, S, y) - l(Z, S, y) \|^2] \tag{26}$$

$$= E(\gamma) + C[\gamma, (Z, S, y)][V(Z, S, y)]^{-1}[(Z, S, y) - E(Z, S, y)]$$

where l runs over the set of all linear functions of (Z, S, y) - and an implicit vectorization of (Z, S, y) is assumed. Note that $\hat{E}(\gamma \mid Z, S, y)$ may be interpreted as the conditional expectation of γ given (Z, S, y) computed through a normal approximation with the same first two moments as the actual distribution of (γ, Z, S, y).

Such a naive procedure suffers from two basic defects. First, it does not seem natural to consider only linear functions of binary variables such as (Z, S), neither it is natural to approximate the actual distribution of (Z, S) (marginally or conditionally on γ or on y) by a normal distribution. Secondly, considering only linear functions of y has been shown in this context (see Cocchi and Mouchart, 1990) to imply that the prior information on the moments of order higher than the first one will not be revised in the process of inference on γ. These two difficulties will be overcome by replacing y by a suitable statistic t, the choice of which will be discussed later, and by considering the set of all functions that are linear in t with coefficients being arbitrary (but measurable) functions of (Z, S). More specifically we shall approximate $E(\gamma \mid Z, S, y)$ by:

$$\hat{E}^{Z,S}(\gamma \mid t) = \arg\inf_{l_{Z,S}} E[\| E(\gamma \mid Z, S, t) - l_{Z,S}(t) \|^2 \mid Z, S] \tag{27}$$

$$= E(\gamma \mid Z, S) + C(\gamma, t' \mid Z, S)[V(t \mid Z, S)]^{-1}[t - E(t \mid Z, S)]$$

where $l_{Z,S}$ runs over all the functions which are linear in t with coefficients being measurable functions of (S, Z).

Note that $\hat{E}^{Z,S}(\gamma \mid t)$ may now be interpreted as the conditional expectation of γ given t computed through a normal approximation of the actual distribution of $(\gamma, t \mid Z, S)$. For further comments on this kind of approximations see *e.g.* Mouchart and Simar (1984a); in the particular case where t is an unbiased estimator, see also Mouchart and Simar (1980, 1983, 1984b) and, in the framework of credibility theory, De Vylder (1982) and Norberg (1986).

We shall also be interested in evaluating an upper bound for the expected posterior variance of γ, obtained through a "semi-linear variance of γ", defined as:

$$\hat{V}^{Z,S}(\gamma \mid t) = V[\gamma - \hat{E}^{Z,S}(\gamma \mid t) \mid Z, S] \tag{28}$$

$$= V(\gamma \mid Z,S) - C(\gamma, t' \mid Z,S)\,[V(t \mid Z,S)]^{-1} C(t, \gamma' \mid Z,S).$$

Indeed, it is immediate to show that in general:

$$E[V(\gamma \mid Z,S,y) \mid Z,S] \le \hat{V}^{Z,S}(\gamma \mid y) \quad \text{a.s.} \tag{29}$$

with equality if and only if the approximation is exact in the sense that:

$$E[\gamma \mid Z,S,y] = \hat{E}^{Z,S}(\gamma \mid y) \qquad \text{a.s.} \tag{30}$$

We first remark that the semi-linear inference on γ (*i.e.* the evaluation of $\hat{E}^{Z,S}(\gamma \mid t)$ and $\hat{V}^{Z,S}(\gamma \mid t)$) is entirely characterized by the semi-linear inference on μ, in view of the following Lemma.

Lemma 3.1. Under the assumptions of Section 2:

$$\hat{E}^{Z,S}(\bar{S}\eta \mid t) = \bar{S}Z\,\hat{E}^{Z,S}(\mu \mid t) \tag{31}$$

$$\hat{V}^{Z,S}(\bar{S}\eta \mid t) = \bar{S}Z\,[\hat{V}^{Z,S}(\mu \mid t)]\,Z'\bar{S}' + E(\sigma_2^2 \mid Z,S)\,I_{(N-n)}. \tag{32}$$

Furthermore,

$$\hat{C}^{Z,S}(\bar{S}\eta, \mu \mid t) = \bar{S}Z\,\hat{V}^{Z,S}(\mu \mid t). \tag{33}$$

where the linear covariance $\hat{C}^{Z,S}(\bar{S}\eta, \mu \mid t)$ is defined as:

$$\hat{C}^{Z,S}(\bar{S}\eta, \mu \mid t) = C[\bar{S}\eta - \hat{E}^{Z,S}(\bar{S}\eta \mid t), \mu - \hat{E}^{Z,S}(\mu \mid t) \mid Z,S]. \tag{34}$$

The proof follows the same route as Lemma 3.1 of Cocchi and Mouchart (1990) and is heavily based upon H5.

3.2 *The choice of the statistic on which conditioning*

Let us now discuss the specification of the statistic t. Remember that the relevance of this choice is to concentrate the sample information

(Z, S, y) into (Z, S, t) and to consider a normal approximation of the distribution of $(\mu, t \mid Z, S)$. The relevance and the impact of such a summary may be appreciated by noticing that alternative proposals for t induce different sets of prior information to be used in evaluating $\hat{E}^{Z,S}(\mu \mid t)$. More specifically, if t is a polynomial of order d, prior information on sampling moments of order $2d$ will be used.

As a first proposal, the choice of $t = y$ is actually equivalent to:

$$t_0 = Z'S'y, \tag{35}$$

the vector of group totals, in the sense that (see Cocchi and Mouchart, 1990):

$$\hat{E}^{Z,S}(\mu \mid y) = \hat{E}^{Z,S}(\mu \mid t_0) \tag{36}$$

and, which is more important, that makes use of prior information on the first two moments only, making no use of any information reflecting a suspicion of non-normality (such as prior information on 3rd and 4th order moments, for instance). Furthermore, as mentioned earlier, the use of the mere t_0 leads to revise the prior information on μ_0 but not on higher moments, not even the second order ones. This experience lead to enlarge the statistic t_0. A natural suggestion is to augment it with reasonable estimators of the structural parameters which would characterize Φ_1 and Φ_2 if they were assumed to be normal, *i.e.* σ_1^2 and σ_2^2. This means to enrich t_0 with the usual "between" and "within" sums of squares:

$$t_1 = \sum_{1 \le k \le p} n_k(\bar{y}_k - \bar{y})^2 = y'Q_1y \tag{37}$$

$$t_2 = \sum_{1 \le k \le p} \; \sum_{1 \le i \le n_k} (y_{k(i)} - \bar{y}_{k(i)})^2 = y'Q_2y \tag{38}$$

where $k(i)$ is the group which the i-th observation belongs to and:

$$\bar{y} = n^{-1}i'_n y = n^{-1}i'_p t_0 = n^{-1}n'\bar{y} \tag{39}$$

$$Q_1 = SZ\Delta_n^{-1}Z'S' - n^{-1}i_n i'_n = SZ(\Delta_n^{-1} - n^{-1}i_p i'_p)Z'S' \tag{40}$$

$$Q_2 = I_n - SZ\Delta_n^{-1}Z'S' \tag{41}$$

$$Q_1 + Q_2 = I_n - n^{-1}i_n i'_n.$$

with:

$$\Delta_n = \mathbf{Z}'\mathbf{S}'\mathbf{S}\mathbf{Z} = \text{diag}\,[n_k] \quad :p \times p \tag{42}$$

$$\mathbf{n} = \Delta_n \mathbf{i}_p = [n_k] \quad :p \times 1 \tag{43}$$

$$\bar{\mathbf{y}} = [\bar{y}_k] = \Delta_n^{-1}\mathbf{t}_0 \qquad :p \times 1 \tag{44}$$

Consequently we shall base our L.S. approximation on the $(p+2)$-dimensional statistic $\mathbf{t}' = (\mathbf{t}_0', t_1, t_2)$ which incorporates (up to multiplicative constants) unbiased estimators of μ_0 and σ_2^2 along with a reasonable estimator of σ_1^2, for which no positive unbiased estimator is known to exist (see Rao and Kleffe, 1980, pp. 11-12).

Since the statistic t, used in the L.S. approximation $\hat{E}^{Z,S}(\gamma, \mathbf{t})$, involves sample moments up to order d (here $d=2$), hypotheses and prior information on population moments up to order $2d$ (here $2d=4$) are required. Thus θ_1 not only involves μ_0 and σ_1^2, but also:

$$\alpha_j = E[(\mu_k - \mu_0)^j \mid \mathbf{Z}, \theta] = E[(\mu_k - \mu_0)^j \mid \mathbf{Z}_1, \theta_1] = E(\delta_k^j \mid \mathbf{Z}_1, \theta_1) \quad j = 3,4 \tag{45}$$

which are characteristics of Φ_1. Similarly, θ_2 not only involves σ_2^2, but also:

$$\beta_j = E[(\eta_\lambda - \mathbf{e}_\lambda'\mathbf{Z}\mu)^j \mid \mathbf{Z}, \mu, \theta] = E[(\eta_\lambda - \mathbf{e}_\lambda'\mathbf{Z}\mu)^j \mid \mathbf{Z}_2, \mu, \theta_2] = E(\varepsilon_\lambda^j \mid \mathbf{Z}_2, \mu, \theta_2) \quad j = 3,4 \tag{46}$$

which are characteristics of Φ_2, where $\mathbf{e}_\lambda'$ is the λ-th row of $\mathbf{I}_N$, and therefore $\mathbf{e}_\lambda'\mathbf{Z}\mu$ is the expectation of the λ-th individual value of the η's in the population conditionally on $(\mu, \mathbf{Z}_2, \theta_2)$. Thus $(\mu_0, \sigma_1^2, \alpha_3, \alpha_4) \subset (\mu_0, \Phi_1) = \theta_1$ and $(\sigma_2^2, \beta_3, \beta_4) \subset \Phi_2 = \theta_2$.

4 The main result

4.1 Statement of the main result

From Lemma 3.1, the L.S. approximation of $E(\mathbf{A}'\gamma \mid \mathbf{Z}, \mathbf{S}, \mathbf{y})$ for any matrix A on the basis of the $(p+2)$-dimensional statistic $\mathbf{t}$ boils down to evaluate:

$$\hat{E}^{Z,S}(\mu \mid \mathbf{t}) = E(\mu \mid \mathbf{Z}, \mathbf{S}) + C(\mu, \mathbf{t}' \mid \mathbf{Z}, \mathbf{S})\,[V(\mathbf{t} \mid \mathbf{Z}, \mathbf{S})]^{-1}\,[\mathbf{t} - E(\mathbf{t} \mid \mathbf{Z}, \mathbf{S})] \tag{47}$$

$$\hat{V}^{Z,S}(\mu \mid \mathbf{t}) = V(\mu \mid \mathbf{Z}, \mathbf{S}) - C(\mu, \mathbf{t}' \mid \mathbf{Z}, \mathbf{S})\,[V(\mathbf{t} \mid \mathbf{Z}, \mathbf{S})]^{-1} C(\mathbf{t}, \mu' \mid \mathbf{Z}, \mathbf{S}). \tag{48}$$

In this section, we present and comment the components of these formulae, the derivation of which appears with some details in Cocchi and Mouchart (1989, Appendix D, along with some useful identities in Appendix C).

This solution requires, as an input, the prior evaluation of the following quantities:

$$m_0 = E(\mu_0 \mid \mathbf{Z}, \mathbf{S}) \qquad M_0 = V(\mu_0 \mid \mathbf{Z}, \mathbf{S})$$
$$v_1 = E(\sigma_1^2 \mid \mathbf{Z}, \mathbf{S}) \qquad v_2 = E(\sigma_2^2 \mid \mathbf{Z}, \mathbf{S})$$
$$V_1 = V(\sigma_1^2 \mid \mathbf{Z}, \mathbf{S}) \qquad V_2 = V(\sigma_2^2 \mid \mathbf{Z}, \mathbf{S})$$
$$a_3 = E(\alpha_3 \mid \mathbf{Z}, \mathbf{S}) \qquad a_4 = E(\alpha_4 \mid \mathbf{Z}, \mathbf{S})$$
$$b_3 = E(\beta_3 \mid \mathbf{Z}, \mathbf{S}) \qquad b_4 = E(\beta_4 \mid \mathbf{Z}, \mathbf{S})$$
$$c_{0,1} = C(\mu_0, \sigma_1^2 \mid \mathbf{Z}, \mathbf{S})$$
$$g_\eta = \frac{b_4}{V_2 + v_2^2} - 3 \qquad g_\mu = \frac{a_4}{V_1 + v_1^2} - 3$$

In the last column of Table 1 (see the appendix) we present the elements of the vectors and matrices appearing in (47) and (48) and in the third column we give some intermediary evaluations in order to offer some hint on the underlying manipulations. In the appendix, we list some intermediary results which show, in particular, the role of the hypotheses stated in Section 2.

In Table 1, the symbol * denotes the Hadamard componentwise product of matrices. It is useful to notice the following two identities:

$$\mathbf{n} * \mathbf{n} = \Delta_n \mathbf{n} = (n_k\, n_k) \quad :p \times 1,$$

$$n^{\cdot} = \mathbf{i}_p' \Delta_n^{-1} \mathbf{i}_p = \Sigma\, n_k^{-1}.$$

An elegant analytical version of the result, analogous to that obtained in Cocchi and Mouchart (1990) is not easy to derive, essentially because of the presence of $V[(t_1\, t_2)' \mid \mathbf{Z}, \mathbf{S}]$. For this reason a solution employing numerical methods is required. In order to sketch anyway some peculiarity of the solution, let us partition the inverse of $V(\mathbf{t} \mid \mathbf{Z}, \mathbf{S})$ and $C(\mu, \mathbf{t}' \mid \mathbf{Z}, \mathbf{S})$ as:

$$[V(\mathbf{t} \mid \mathbf{Z}, \mathbf{S})]^{-1} = \begin{bmatrix} \mathbf{V}^{00} & \mathbf{v}^{01} & \mathbf{v}^{02} \\ \mathbf{v}^{10} & v^{11} & v^{12} \\ \mathbf{v}^{20} & v^{21} & v^{22} \end{bmatrix} \tag{49}$$

$$C(\mu, \mathbf{t}' \mid \mathbf{Z}, \mathbf{S}) = [\mathbf{C}_0 \quad \mathbf{c}_1 \quad \mathbf{0}]. \tag{50}$$

Note that v^{22} does not actually enter (47) nor (48), and therefore (47) can be rewritten as follows:

$$\hat{E}^{Z,S}(\mu \mid \mathbf{t}) = E(\mu \mid \mathbf{Z}, \mathbf{S}) + [\mathbf{C}_0\mathbf{V}^{00} + \mathbf{c}_1(\mathbf{v}^{01})'] [\mathbf{t}_0 - E(\mathbf{t}_0 \mid \mathbf{Z}, \mathbf{S})] \tag{51}$$
$$+ [\mathbf{C}_0\mathbf{v}^{01} + \mathbf{c}_1 v^{11}] [t_1 - E(t_1 \mid \mathbf{Z}, \mathbf{S})] + [\mathbf{C}_0\mathbf{v}^{02} + \mathbf{c}_1 v^{12}] [t_2 - E(t_2 \mid \mathbf{Z}, \mathbf{S})].$$

After some rearrangements, taking the elements of Table 1 into account, and defining:

$$\mathbf{C}_{\bullet} = [\mathbf{C}_0\mathbf{V}^{00} + \mathbf{c}_1(\mathbf{v}^{01})']\Delta_{\mathbf{n}}.$$

$$\mathbf{C}_{\bullet\bullet} = [\mathbf{C}_0\mathbf{v}^{01} + \mathbf{c}_1 v^{11} \quad \mathbf{C}_0\mathbf{v}^{02} + \mathbf{c}_1 v^{12}] \quad : p \times 2 \tag{52}$$

we obtain:

$$\hat{E}^{Z,S}(\mu \mid \mathbf{t}) = (\mathbf{I}_p - \mathbf{C}_{\bullet})\mathbf{i}_p m_0 + \mathbf{C}_{\bullet}\bar{\mathbf{y}} + \mathbf{C}_{\bullet\bullet} \begin{bmatrix} t_1 - [v_2(p-1) + v_1(n - n^{-1}\mathbf{n}'\mathbf{n})] \\ t_2 - v_2(n-p) \end{bmatrix} \tag{53}$$

4.2 On the interpretation of the main result

Let us now interpret the main result by means of some comments.

(a) *On the double averaging*

The least squares approximation $\hat{E}^{Z,S}(\mu \mid \mathbf{t})$ appears as a (matrix)-weighted average between the prior expectation $E(\mu \mid \mathbf{Z}, \mathbf{S}) = \mathbf{i}_p m_0$ and $\bar{\mathbf{y}}$, the unbiased estimator of μ, plus a correction term proportional to the difference between the sums of squares (t_1, t_2) and their prior expectations $E(t_i \mid \mathbf{Z}, \mathbf{S})$, $i = 1, 2$. Note that the prior expectation can be a function of $(\mathbf{Z}, \mathbf{S})$ without complicating the result.

(b) *On the validity of an approximated result rather than an exact one*

The present result provides an example of the situation where a statistician has to face the choice between an exact solution - *viz.* $E(\mu \mid \mathbf{Z}, \mathbf{S}, \mathbf{y})$ - with respect to a completely specified model that will typically be rather crude (*i.e.* often a model with a normality assumption and a specific prior distribution) and an approximate solution $\hat{E}^{Z,S}(\mu \mid \mathbf{t})$ for a class of models embodying, according to the different situations, more flexibility with respect to the specification of non-normal features such as expected asymmetry and/or kurtosis.

Another aspect of the same argument consists in remarking that the validity of approximating the actual distribution of $(\mu, \mathbf{t} \mid \mathbf{Z}, \mathbf{S})$ by a normal one may crucially depend on the choice of $\mathbf{t}$, and a central limit argument may suggest that taking the space of individual observations (as done with $\mathbf{y}$) as coordinates of the sample might be less appropriate than taking sample moments.

(c) *On the presence of 3rd and 4th order moments*

In fact, a main characteristics of the result emerges, *i.e.* both the weighting of the elements of $\bar{\mathbf{y}}$ through $\mathbf{C}_{\bullet}$ and the weighting of the differences $(t_i - E(t_i \mid \mathbf{Z}, \mathbf{S}))$, $i = 1, 2$, via $\mathbf{C}_{\bullet\bullet}$ depend on the expected non-normality of both the $(\eta \mid \mathbf{Z}, \mu, \theta)$-process and the $(\mu \mid \mathbf{Z}, \theta)$-process by the presence of g_η and g_μ. This kind of result is analogous to the multiple regression case in infinite populations as analysed in Mouchart and Simar (1984a). It may be stressed that basing the L.S. approximation on $\mathbf{t}$ rather than on $\mathbf{y}$ provides the possibility of taking into account expected non-normalities through (a_3, b_3) (*i.e.* expected lack of symmetry) or through (g_η, g_μ) (*i.e.* expected plati- or lepto-kurtosis) with a small effort. Indeed this only requires the specification of the functions of expected moments listed above, while the difference of computational cost between (47)-(48) and some similar results conditional on normality is almost negligible: Section 5 will discuss some of such results.

(d) *On the consequences of sampling under exchangeability conditions*

Note that (47) and (48) do not involve the full matrix $\mathbf{Z}$ but only the submatrix selected by $\mathbf{S}$, *viz.* $\mathbf{SZ}$, and this is due to the fact that $\mathbf{t}$ is a function of $\mathbf{y} = \mathbf{S}\eta = \mathbf{SZ}\mu + \mathbf{S}\varepsilon$. In other words $\hat{E}^{Z,S}(\mu \mid \mathbf{t}) = \hat{E}^{SZ,S}(\mu \mid \mathbf{t})$. Furthermore, the information contained in $\mathbf{SZ}$ and in $\mathbf{S}$ is used only for defining the sample size within each subpopulation, *i.e.* n_k (see (42) and (43)), and for allocating each sampled element to one of the subpopulations (see (40)-(43) and Table 1).

Remember indeed that the identity $SZ = \mathrm{diag}[\mathbf{i}_{\cdot k}]$ allows one to allocate each sampled element to a subpopulation and therefore to reorder the sampled labels in SZ, and permits to compute the partial averages in $\bar{\mathbf{y}}$ or the partial totals in $\mathbf{t}_0$. This symmetric treatment of the individual values within the same subpopulation is due to the partial exchangeability of the distribution of $(S^{*}\eta, \mu \mid Z, S, \theta)$ (see Cocchi and Mouchart (1990), formulae (2.25) and (2.26)).

Note however that the descriptive inference on the unobserved ζ_{λ}' - *viz.* $\hat{E}^{Z,S}(\bar{S}\eta \mid \mathbf{t})$ and $\hat{V}^{Z,S}(\bar{S}\eta \mid \mathbf{t})$ obtained through Lemma 3.1 - actually uses the complete knowledge of **Z** because the knowledge of $SZ, \bar{S}Z, \bar{S}$ and S is evidently equivalent to know S and Z.

(e) *On the inference on the population mean*

A trivial by-product of the former results is the descriptive inference on linear combinations of η such as the population total $\sum_{1 \le \lambda \le N} \eta_{\lambda}$ or the population average $N^{-1} \sum_{1 \le \lambda \le N} \eta_{\lambda}$. Further details are in Cocchi and Mouchart (1990).

4.3 The normal case and the departure from normality

When $a_3 = b_3 = c_{0,1} = 0$, *i.e.* a priori expected symmetry in the distributions of μ and of $(\eta \mid \mu)$ and prior uncorrelation between μ_0 and σ_1^2 occur, the $(p+2)\times(p+2)$ variance matrix $V(\mathbf{t} \mid Z, S)$ becomes block diagonal and $\mathbf{c}_1 = C(\mu, t_1 \mid Z, S)$ also vanishes. This implies that $\mathbf{C}_{\cdot\cdot} = \mathbf{0}$ and eventually $\mathbf{t}_0$ become L.S. sufficient, in the sense that $\hat{E}^{Z,S}(\mu \mid \mathbf{t}) = \hat{E}^{Z,S}(\mu \mid \mathbf{t}_0)$; in other words, t_1 and t_2 improve the L.S. approximation $\hat{E}^{Z,S}(\mu \mid \mathbf{t})$ in comparison to $\hat{E}^{Z,S}(\mu \mid \mathbf{t}_0)$ only if the third moments of the distribution of η and μ do not behave as if they were symmetric with probability 1 and/or $c_{0,1} \neq 0$. This feature will be commented in Section 5, when comparing this result with similar ones obtained earlier in the literature.

The prior information on Φ_1 and Φ_2, involved by their normality with probability one, implies that $a_3 = g_{\mu} = 0$ and $b_3 = g_{\eta} = 0$, but this feature alone does not simplify substantially (47) and (48).

Moreover, in the normal case, the structural parameter reduces to $\theta_1 = (\mu_0, \sigma_1^2)$, $\theta_2 = \sigma_2^2$ and the parameter under normality is $\theta^{N} = (\mu_0, \sigma_1^2, \sigma_2^2)$.

In general, the roles respectively played by the prior information concerning symmetry and kurtosis are different: in fact, b_3 appears only in $C(\mathbf{t}_0, t_j \mid Z, S)$, and a_3 in $C(\mu, t_1 \mid Z, S)$ and in $C(\mathbf{t}_0, t_1 \mid Z, S)$, while 4th order moments concern only $C(t_i, t_j \mid Z, S)$.

As a general comment, properties on symmetry are more important than properties of kurtosis, and this agrees with known results on the forms on distributions, as shown also in Ferreri (1968).

5 A comparison between different solutions of the model

5.1 An analytical comparison

Several papers, in particular Scott and Smith (1969), Cocchi (1985) and Cocchi and Mouchart (1990), have already considered models involving assumptions essentially similar to those made explicit in Section 2. These models differ in their distributional assumptions (required normality or not, proper or not prior distribution, complete specification of such distribution or not) and in their solutions, which can be exact or approximate. Eventually these solutions also differ in the amount of prior and sampling information they use.

In Table 2 (see the appendix) the main characteristics of the different contributions are presented with a homogeneous notation. The proposed solutions are different because of various features. First of all, the distributional assumptions vary from the normality of the joint distribution of μ and η to the mere knowledge of the first two moments of such distribution. As a second point, the obtained result varies from the exact evaluation of the first two moments of the posterior distribution of μ to its Bayesian least squares approximation. Also the choice of the statistic on which conditioning the calculations is relevant, and a consequence of the differences in the former aspects is the need of different elicitation efforts for prior evaluations.

The progressive relaxation of the strength of the distributional assumptions that can be noticed along the second column of the table is compensated, in the third column, by the increase of the prior evaluations necessary to solve the problem. Furthermore, an exact solution is allowed in the first two cases thanks to the normal hypothesis and the conditioning on the population values of variances between and within groups. When the two conditions are abandoned, the search for an exact solution is difficult; nevertheless, an appropriate choice of the coordinates of the problem permits to derive Bayesian least squares approximations, as happens for the last two solutions. It so happens that analytical exact solutions can be computed only conditionally on the variances σ_1^2, σ_2^2, while the computation of an analytical approximate solution, on the basis of the knowledge of the first two moments of the joint distribution of $(\eta, \mu \mid Z, S, \theta)$, permits, on the contrary, the integration of the structural parameter. A further possibility, which is not practiced here, would consist in an exact numerical solution.

The remarkable difference between the last two solutions of Table 2 comes from the choice of the statistic on which conditioning. All solutions, in fact, are conditional on the vector of group totals t_0. The further conditioning on t_1 and t_2 increases the set of the prior elicitations necessary for the solution from those of group (a) to those of group (b), and also, by means of the elements of the group (c), takes departures from normality into account.

Cocchi (1985) gives an exact solution under very strong assumptions (*viz.* normality of $(\eta, \mu \mid Z, S, \theta)$ and known θ^N) which is linear in t_0, *i.e.* $E(\mu \mid Z, S, t_0, \theta^N) = \hat{E}^{Z,S,\theta^N}(\mu \mid t_0)$. Moreover, when comparing the different solutions, the interpretation is important. For instance, Cocchi (1985) and Scott and Smith (1969) are not Bayesian in a strict sense, since they are conditional on θ. But they can be interpreted in the framework of the present result as the particular case where $M_0 = V_1 = V_2 = c_{0,1} = 0$. Furthermore, the solution based on (47) and (48) uses more intensively the prior expected values of the within and between variances. In fact, in the former solutions they intervene only as weights to be attributed respectively to the sample group averages and to their prior evaluations.

The results are summarized in the last three columns of Table 2, where with the symbol Δ_x we denote the diagonal matrix canonically associated to the vector x. All solutions can be seen as least squares ones, but the first three are also exact solutions in the normal case. So, when normality is assumed, it is possible to compare all the solutions. On the contrary, when the normal assumption is abandoned, the only solution to practice is that introduced by (47) and (48), since also the approximate solution of Cocchi and Mouchart (1990) does not vary according to Φ_1 and Φ_2. For such solution, it is possible to evaluate the differences in the result according to the distributional assumption chosen for the characteristics of interest and for the vector of group averages.

In the four cases, the result consists in an averaging between the vector $\bar{y}$, the unbiased estimator of μ, and a vector, the components of which are proportional to some estimate of μ_0. In Cocchi (1985) this last quantity is exactly μ_0, and the weights are, as usual in Bayesian methods, proportional to precisions. In Scott and Smith (1969) the estimator of μ_0 is a weighted average of the components of $\bar{y}$. In Cocchi and Mouchart (1990) the least squares approximation of the posterior expectation of μ is obtained through a double averaging, *i.e.* a weighted average between the vector of sample means $\bar{y}$ and an overall mean, which is itself a weighted average between m_0, the prior expectation of μ_0 and m, a weighted average of the sample data whereas in this last solution we turn back to a single averaging between $\bar{y}$ and a vector proportional to m_0, the prior expectation of μ_0, while t_1 and t_2 introduce a new correction term.

5.1.1 The perfect information condition

For what concerns the expected values of the quantities to be introduced in the simulations, there is to choose whether we suppose to be in a condition of perfect information, and in this last case the expected values coincide with the real ones, or the condition is not fulfilled.

The perfect information condition concerns the conjectures on some characteristics of the parameter θ. It can be splitted into the conjectures on the expected values of θ^N, which are essential also in the first three solutions, and into those on the remaining parts of Φ_1 and Φ_2, which are specific to the fourth solution. For instance, under perfect information on the distributions, for the normal distribution and for any symmetric one, the expectations of the third moment are null, and those of the fourth moment are, for the normal,

$$a_4 = 3E(\sigma_1^4 \mid Z, S)$$

$$b_4 = 3E(\sigma_2^4 \mid Z, S)$$

and, taking the t distribution as an example of nonnormality

$$a_4 = \left(\frac{6}{v-4} + 3\right) E(\sigma_1^4 \mid Z, S)$$

$$b_4 = \left(\frac{6}{v'-4} + 3\right) E(\sigma_2^4 \mid Z, S)$$

where

$$E(\sigma_j^4 \mid Z, S) = V(\sigma_j^2 \mid Z, S) + [E(\sigma_j^2 \mid Z, S)]^2 \quad j = 1, 2.$$

For such distributions, as we have already said, the error of specification consists in the erroneous evaluation of the third and fourth order moments, while for other distributions the problem of the specification error occurs also for the expectation of the first two moments.

The object of the comparison is to evaluate the numerical sensitivity of the different procedures with a same set of data and to evaluate statistical robustness with respect to alternative models. The main difficulty stands in making comparable procedures which involve different amount of prior information.

Only when both μ and η are normal, it is possible to compare all the solutions, the exact and the approximate ones. On the contrary, even when only one of the two distributions is not normal, the computation of the exact solutions, which are possible only in the normal case, is meaningless. In such case the only element to check is the specification error for the present solution, since also the approximate solution of Cocchi and Mouchart (1990) does not vary as the scenarios vary.

5.2 A numerical comparison

In this paragraph we check, by means of simulated data, the behaviour of the four different solutions for different data generation processes and elicitation of priors.

The attention is focused on the normal data generation process. In such a context, the part θ^N of the structural parameter can be chosen arbitrarily. In Cocchi and Mouchart (1990) the influence of variations in the ratio σ_2^2/σ_1^2 has been tested. The developments of the present solution increase the possibilities of comparison between the different choices.

We consider a population of 400 elements which can be attributed to 5 different groups, with size $N_1 = 30, N_2 = 100, N_3 = 40, N_4 = 80, N_5 = 150$. From each group a sample is drawn, the total sample size being 80, under the following alternative hypotheses: constant sampling fraction equal to 1/5 and equal sample size of 16 within groups.

Different values of θ^N have been tried, *i.e.* $\mu_0 = 100$ and the following values of σ_1^2 and σ_2^2:

$\sigma_1^2 = 400$	$\sigma_2^2 = \{25, 100, 400, 1600, 6400\}$
$\sigma_1^2 = 1600$	$\sigma_2^2 = \{100, 400, 1600, 6400, 25600\}$
$\sigma_1^2 = 25600$	$\sigma_2^2 = \{100, 400, 1600, 6400, 25600\}$

The first line corresponds to the idea of low variability between groups, and the third line to that of high variability between groups. The values experimented for σ_2^2 correspond to different conjectures on the ratio σ_1^2/σ_2^2, which varies from 4 to 1/16. On the basis of the possible associations of variances within and between groups 15 different scenarios, normal in μ and η, have been generated.

(a) *Simulations in the normal case with complete perfect information*

The first set of simulations concern a situation of perfect information on the whole parameter θ: indeed, the solutions found under such condition constitute a landmark for any comparison. Under that condition, the expected values coincide with the population ones, *i.e.* the elements of θ^N are fixed as $m_0 = \mu_0$, $v_1 = \sigma_1^2$ and $v_2 = \sigma_2^2$. The values of a_j and b_j, $j = 3,4$, are those following from the normal assumption which appear in the former subsection.

On the contrary, the conjectures on the values of the variances of m_0, v_1 and v_2, necessary in the solution here proposed, have to be arbitrarily fixed. After some trials, which did not alterate the general comments on the result, we decided to explore further the case where the coefficients of variation $V_1^{1/2}/v_1$ and $V_2^{1/2}/v_2$ are equal to 2 and $M_0^{1/2}/m_0$ is 1/2.

Also the values of $c_{0,1}$ are not kept constant. In particular the covariances between σ_1^2 and μ_0 giving values of the linear correlation coefficient equal to $\{0, \pm .2, \pm .5, \pm .9\}$ have been checked.

(b) *A general comment on the result*

We do not report the details of the simulations performed on the 15 scenarios but we want to stress that the first one, where the sources of variability between and within groups are at the minimum level checked, shows the coincidence between all the solutions. The low structural variability within and between groups and the exactness of the conjectures on θ^N induces the numerical posterior independence of the μ_k's for all solutions, while these averages are analytically independent only in Cocchi (1985). When the structural variability of the η_λ's increases, the variance-covariance matrix of the Scott and Smith's (1969) solution begins to be different from the first matrix, while the same matrices in Cocchi and Mouchart (1990) and in the present solution begin to assume intermediate values between the two. In general, the same considerations hold for all the scenarios checked.

In the whole, the analysis of the results shows that the solution where the influence of the population values on the general mean is more influent on the final result is Cocchi (1985), while in the other solutions the weight of the sampling result is more relevant. Since, in the normal case, it is possible to compare the four solutions, we can remark that, for the same θ^N, the posterior variance of the general mean has a minimum value, that of the solution of Cocchi (1985), which is exact and computed conditionally on θ^N. Such variance increases in the solution of Cocchi and Mouchart (1990) and in the present solution, reaching its maximum with the solution of Scott and Smith (1969), which is also exact, conditional on θ^N, but computed through a diffuse prior on μ. The

Scott and Smith's solution constitutes an upper limit, as we shall see, only in the case of perfect information on all the conjectures to elicitate. When the posterior group averages and their variance-covariance matrices are compared, the different modes of interaction between sample values and prior values can be appreciated. Such differences are smoothed in the computation of the posterior of the general mean. Indeed, when comparing solutions in the case of perfect information, a feature to remark is the greater influence of the different assumptions on the vector of the posterior group means rather that on the general mean.

Also the effect of the sample sizes is strong: if the n_k are equal - and the sampling fractions are different - the posterior variances of the μ_k's will be equal; therefore the differences between the posterior variances is to be attributed only to the variation in the sample sizes in the different groups.

We also notice that, in the case of proportional sampling, the solution of Scott and Smith (1969), which employs the whole information available when computing the posterior expectations of the group averages and obtains a different value for each of them, gives a posterior general mean which is equal to the corresponding sample value.

(c) *Consequences of the specification error on the superpopulation mean*

In order to evaluate the consequences of the relaxation of the perfect information on the general mean, the values $m_0 = \{70, 100, 130\}$ have been alternatively checked, choosing, in the different trials, the values preserving the coefficients of variation fixed before. The analysis concerning the specification error for m_0 does not deserve particular comments. In fact, when the other components of the prior evaluations are those of perfect information, the error modifies only the balancing, usual in the Bayesian analysis, between prior values and values generated by sampling.

(d) *A detailed example*

We do not report results on the whole set of trials, but we comment in detail the case coming from the scenario where $\sigma_1^2 = 1600$ and $\sigma_2^2 = 25600$, with constant sampling fraction. For such data set, the prior specifications with perfect information for the different parameters are:

$(m_0 = 100\ M_0 = 2500\ v_1 = 1600\ V_1 = 6400\ v_2 = 25600\ V_2 = 102400)$
$(a_3 = 0\ a_4 = 7699200\ b_3 = 0\ b_4 = 1966387200)$

The following table, together with the population group averages and the sample group averages (note that the described case presents a high structural variability within groups), shows the posterior expectations (exact or approximate according to the solution, and therefore denoted by a generic symbol) of the group averages and the general mean, together with the posterior variance (which also can be exact or approximate) of the general mean. We do not reproduce the posterior variance-covariance matrix of the group averages.

The comments made on the whole sets of simulations on the 15 scenarios according to the two sampling schemes can be verified in the example, where it can be noticed how the introduction of the covariance between μ_0 and σ_1^2 (here considered through the corresponding linear correlation coefficient) tends to decrease the posterior values with respect to those determined when such covariance is supposed null.

μ_t	124.02	95.03	70.15	55.20	76.38			
$\bar{y}_t$	158.14	109.52	78.05	104.34	104.87	$\bar{y} = 107.24$		
	$E(\mu_1)$	$E(\mu_2)$	$E(\mu_3)$	$E(\mu_4)$	$E(\mu_5)$	$E(\bar{\eta})$	$V(\bar{\eta})$	
CC85	115.86	105.29	92.68	102.17	103.18	104.17	161.23	
SS69	121.88	108.97	98.21	106.32	106.06	107.24	256.0	
CM90	120.5779	108.1721	97.011	105.4174	105.4369	106.576	235.4655	
CM91	120.5779	108.1721	97.011	105.4174	105.4369	106.576	235.4655	$\rho_{a1} = 0$
CM91	120.5143	108.1333	96.9528	105.3738	105.4065	106.5436	235.4648	$\rho_{a1} = .2$
CM91	120.4189	108.0749	96.8653	105.3081	105.3608	106.495	235.4611	$\rho_{a1} = .5$
CM91	120.2912	107.9969	96.7482	105.2203	105.2997	106.43	235.4512	$\rho_{a1} = .9$
CM91	120.6413	108.2108	97.0691	105.4610	105.4672	106.6082	235.4648	$\rho_{a1} = -.2$
CM91	120.7363	108.2989	97.1562	105.5263	105.5126	106.6566	235.4611	$\rho_{a1} = -.5$
CM91	120.8627	108.3462	92.2721	105.6133	105.5731	106.7210	235.4512	$\rho_{a1} = -.9$

(e) *The error in the evaluation of the variances expectation*

If the trial of different values of m_0 gives results which do not deserve particular comments, the interpretation of the elicitation error for the variances expectations is not so simple. In the first part of the following table, where $\sigma_1^2 = 25600$ and $\sigma_2^2 = 6400$, with constant sampling fraction, the two exact solutions obtained conditionally on σ_1^2 and σ_2^2 are shown

the two exact solutions obtained conditionally on σ_1^2 and σ_2^2 are shown
(we are checking a case where the variability between groups is very high). In the second part of the table, where the solution with perfect information is put in evidence, the posterior evaluations obtained with erroneous conjectures for v_1 and v_2 are reported.

The most striking feature of all the essays is given by the greater relevance of σ_2^2 with respect to σ_1^2. When σ_2^2 is exactly conjectured, the results are closer to those of perfect information, for which the solution of Scott and Smith (1969) constitutes an upper limit. On the contrary, aberrant values of v_1 tend to alterate the different posterior evaluations of the μ_t's, even if compensating in the posterior expectation of the general mean: if very high values, with respect to the correspondent population ones, of v_2 bring to overcome the upper limit for the variance constituted by the Scott and Smith solution, exceptionally small values for the variance expectation within groups dominate the sample information, producing a posterior variance which is strongly influenced by the prior evaluation and is lower than the bound of the exact solution. As a general comment, we can say that the prior evaluation of σ_1^2 ought to be given very cautiously, since the errors concerning it will mostly influence, as we have seen, posterior results.

μ_t		196.10	80.10	-19.41	-79.18	5.52		
$\bar{y}_t$		213.16	87.35	-15.46	-54.61	19.77	$\bar{y} = 32.77$	
		$E(\mu_1)$	$E(\mu_2)$	$E(\mu_3)$	$E(\mu_4)$	$E(\mu_5)$	$E(\bar{\eta})$	$V(\bar{\eta})$
CC85		208.63	87.51	-11.96	-52.23	20.43	33.39	63.22
SS69		206.61	86.88	-13.49	-53.01	20.01	32.77	64.00
v_1	v_2							
25600 0	6400	207.9758	87.3030	-12.4566	-52.4857	20.2935	33.1865	63.4702
25600	400	212.8201	87.3459	-15.2664	-54.4778	19.7988	32.7938	3.9979
25600	25600	194.6739	87.1788	-4.4331	-46.4720	21.8305	34.4287	247.8653
6400	25600	154.3906	83.8310	11.7742	-30.4418	25.2334	36.2244	240.3667
57600	25600	204.7439	87.4320	-9.8473	-50.6725	20.8080	33.6716	251.4686
57600	6400	210.9376	87.3696	-13.9996	-53.6077	20.0289	32.9952	63.7102
57600	400	213.0155	87.3501	-15.3671	-54.5494	19.7821	32.7818	3.9989
6400	400	211.6489	87.2852	-14.8210	-54.1398	19.8636	32.8205	3.9958
6400	6400	192.1908	86.3511	-6.3645	-47.4945	21.2698	33.6278	62.9468

(f) *The possibility of keeping non normality into account*

In order to begin dealing with non-normality, let us briefly consider the example of a data generation process given, with constant sampling fraction, $\sigma_1^2 = 1600$ and $\sigma_2^2 = 25600$, by a t distribution with 8 d.f. both for the η_λ's and the μ_k's. Normal and non normal solutions can be compared since the values of t_8 and the corresponding normal values are the same percentiles of the respective distributions.

The condition of perfect information with respect to Φ_1 and Φ_2 corresponds to the adequate choice of the values of a_j, b_j, $j = 3,4$ reported in paragraph 5.1.

$\mu_k(t)$	125.09	94.86	68.63	51.82	75.35			
$\bar{y}_k(t)$	165.94	48.01	147.34	42.05	144.44	$\bar{y} = 109.55$		
	$E(\mu_1)$	$E(\mu_2)$	$E(\mu_3)$	$E(\mu_4)$	$E(\mu_5)$	$E(\bar{\eta})$	$V(\bar{\eta})$	
CM90	136.7756	77.8076	127.472	74.8266	126.0223	104.8757	282.1287	
CM91	136.7756	77.8076	127.472	74.8266	126.0223	104.8757	282.1287	$\rho_{\lambda 1} = 0$
CM91	136.7738	77.8058	127.4702	74.8247	126.0505	104.8742	282.1278	$\rho_{\lambda 1} = .2$
CM91	136.7709	77.8029	127.4673	74.8219	126.0176	104.8719	282.1236	$\rho_{\lambda 1} = .5$
CM91	136.7668	77.7988	127.4633	74.8178	126.0135	104.8687	282.1121	$\rho_{\lambda 1} = .9$

Also for the t distribution, which is symmetric, there are no differences between the solution of Cocchi and Mouchart (1990) and the solution here discussed.

(g) *The specification error in the form of the distribution*

The difference between the two solutions just mentioned can on the contrary be pointed out also in the normal case, when the conjectures on a and b are erroneous. In fact, in the example where, with constant sampling fractions, $\sigma_1^2 = 400$ and $\sigma_2^2 = 1600$, with perfect information on θ^N, but where $a_3 = b_3 = 300$ and $a_4 = b_4 = 0$, we obtain the following results for the two solutions:

μ_k	112.01	97.51	85.07	77.60	88.19		
$\bar{y}_k$	120.54	101.14	87.05	89.89	95.31	$\bar{y} = 96.75$	
	$E(\mu_1)$	$E(\mu_2)$	$E(\mu_3)$	$E(\mu_4)$	$E(\mu_5)$	$E(\bar{\eta})$	$V(\bar{\eta})$
CM90	111.5547	100.6254	90.7244	91.5245	95.6383	96.7616	15.9018
CM91	111.6634	100.6816	90.8162	91.5863	95.6900	96.8121	15.9022

(h) *Concluding remarks*

The simulations performed are a contribution to the understanding the role of the differences existing between different solutions, not all necessarily Bayesian, either of exact or approximated type, to the same superpopulation model for finite populations. We think that the role of the available information within different contexts has been remarked, by means of the different role of the parameters and prior evaluations and their impact on the posteriors.

Appendix

Most of the results of the third column of Table 1 derive from the following theorems, proved in Appendix D of Cocchi and Mouchart (1989) and reported here without proof.

In particular, noting that t_1 and t_2 may be rewritten as:

$$t_j = \mathbf{y}'\mathbf{Q}_j\mathbf{y} = \eta'\mathbf{S}'\mathbf{Q}_j\mathbf{S}\eta = \eta'\mathbf{F}_j\eta, \qquad j = 1,2 \tag{54}$$

where

$$\mathbf{F}_1 = \mathbf{S}'\mathbf{S}\mathbf{Z}(\Delta_n^{-1} - n^{-1}\mathbf{i}_p\mathbf{i}'_p)\mathbf{Z}'\mathbf{S}'\mathbf{S} \qquad :N \times N \tag{55}$$

$$\mathbf{F}_2 = \mathbf{S}'[\mathbf{I}_n - \mathbf{S}\mathbf{Z}\Delta_n^{-1}\mathbf{Z}'\mathbf{S}']\mathbf{S}. \qquad :N \times N \tag{56}$$

the evaluation of $C(t_0, t_i \mid \mathbf{Z}, \mathbf{S}, \theta, \mu) = \mathbf{Z}'\mathbf{S}'\mathbf{S}\, C(\eta, \eta'\mathbf{F}_i\eta \mid \mathbf{Z}, \mathbf{S}, \theta, \mu)$, $i = 1,2$, relies on the following simple results:

Lemma 1. Let $\mathbf{r} = (r_1 \dots r_n)'$. If $\perp\!\!\!\perp\, r_i$, $E(r_i) = 0$ and $E(r_i^3) = k_3\ \forall i$, then, for any matrix $\mathbf{W}$, $C(\mathbf{r}, \mathbf{r}'\mathbf{W}\mathbf{r}) = k_3\mathbf{d}_\mathbf{W}$, where $\mathbf{d}_\mathbf{W} = (w_{11}\ w_{22} \dots w_{nn})'$.

Corollary 2. Under H5 and H9, for any matrix $\mathbf{W}$:

$C(\eta, \eta'\mathbf{W}\eta \mid \mathbf{Z}, \mathbf{S}, \theta, \mu) = 2\sigma_2^2\mathbf{WZ}\mu + \beta_3\mathbf{d}_\mathbf{W}$.

Similarly, the evaluation of $C(t_i, t_j \mid \mathbf{Z}, \mathbf{S}, \theta, \mu) = C(\eta'\mathbf{F}_i\eta, \eta'\mathbf{F}_j\eta \mid \mathbf{Z}, \mathbf{S}, \theta, \mu)$, $i, j = 1, 2$, relies on the following theorems.

Lemma 3. Let $\mathbf{r} = (r_1 \dots r_n)'$, $\mathbf{U} = \mathbf{U}'$ and $\mathbf{V} = \mathbf{V}'$. If $\perp\!\!\!\perp\ r_i$, $E(r_i) = 0$, $E(r_i^2) = k_2$, and $E(r_i^4) = k_4\ \forall i$, then $C(\mathbf{r}'\mathbf{Ur}, \mathbf{r}'\mathbf{Vr}) = (k_4 - 3k_2^2)\mathbf{d}'_\mathbf{U}\mathbf{d}_\mathbf{V} + 2k_2^2 \operatorname{tr} \mathbf{UV}$, where $\mathbf{d}_\mathbf{U} = (u_{11}\ u_{22} \dots u_{nn})'$ and $\mathbf{d}_\mathbf{V} = (v_{11}\ v_{22} \dots v_{nn})'$.

Corollary 4. Under H5, (22), H9, (47) and (48):

$$C(\eta'\mathbf{U}\eta, \eta'\mathbf{V}\eta \mid \mathbf{Z}, \mathbf{S}, \theta, \mu) = (\beta_4 - 3\sigma_2^4)\mathbf{d}'_\mathbf{U}\mathbf{d}_\mathbf{V} + 2\sigma_2^4 \operatorname{tr}\mathbf{UV} +$$

$$+2\beta_3(\mathbf{d}'_\mathbf{U}\mathbf{V} + \mathbf{d}'_\mathbf{V}\mathbf{U})\mathbf{Z}\mu + 4\sigma_2^2\mu'\mathbf{Z}'\mathbf{UVZ}\mu$$

The following properties of the distribution of $(\mu \mid \mathbf{Z}, \mathbf{S}, \theta)$ permit the passage to the last column of Table 1.

Lemma 5. Under H3, H7 and H9 we have both $\perp\!\!\!\perp\ \delta_k \mid \mathbf{Z}, \mathbf{S}, \theta$ and $\perp\!\!\!\perp\ \mu_k \mid \mathbf{Z}, \mathbf{S}, \theta$.

Lemma 6. Under (45), (46) and (23)

$$V(\mu'\mathbf{Z}'\mathbf{F}_1\mathbf{Z}\mu \mid \mathbf{Z}, \mathbf{S}, \theta) = (\alpha_4 - 3\sigma_1^4)(n - n^{-1}\mathbf{n}'\mathbf{n}) + 2\sigma_1^4 \operatorname{tr}(\mathbf{Z}'\mathbf{F}_1\mathbf{Z})^2$$

Tables

Table 1. Moments requested for (47) and (48)

		the conditioning is respectively on	
components of (47) and (48)	constituted by	$(\mathbf{Z}, \mathbf{S}, \theta, \mu)$ *i.e.* intermediate results	$(\mathbf{Z}, \mathbf{S})$ *i.e.* elements of (47) and (48)
$E(\mu \mid .) : p \times 1$		μ	$m_0\mathbf{l}_p$
$E(\mathbf{T} \mid .) : (p+2) \times 1$	$E(\mathbf{T}_0 \mid .) : p \times 1$	$\Delta_\mathbf{n}\mu$	$m_0\mathbf{n}$
	$E(T_1 \mid .)$	$\sigma_2^2(p-1) + \mu'(\Delta_\mathbf{n} - n^{-1}\mathbf{nn}')\mu$	$v_2(p-1) + v_1(n - n^{-1}\mathbf{n}'\mathbf{n})$
	$E(T_2 \mid .)$	$\sigma_2^2(n-p)$	$v_2(n-p)$
$C(\mu, \mathbf{T}' \mid .)$ $: p \times (p+2)$	$C(\mu, \mathbf{T}'_0 \mid .) : p \times p$	$\mathbf{0}$	$v_1\Delta_\mathbf{n} + M_0\mathbf{l}_p\mathbf{n}'$
	$C(\mu, T_1 \mid .) : p \times 1$	$\mathbf{0}$	$a_3(\mathbf{n} - n^{-1}\mathbf{n}{*}\mathbf{n}) + c_{0,1}(n - n^{-1}\mathbf{n}'\mathbf{n})\mathbf{l}_p$
	$C(\mu, T_2 \mid .) : p \times 1$	$\mathbf{0}$	$\mathbf{0}$
$V(\mu \mid .) : p \times p$		$\mathbf{0}$	$v_1\mathbf{I}_p + M_0\mathbf{l}_p\mathbf{l}_p'$
$V(\mathbf{T} \mid .)$ $: (p+2) \times (p+2)$	$V(\mathbf{T}_0 \mid .) : p \times p$	$\sigma_2^2\Delta_\mathbf{n}$	$v_1\Delta_\mathbf{n}^2 + v_2\Delta_\mathbf{n} + M_0\mathbf{nn}'$
	$V(T_1 \mid .)$	$(\beta_4 - 3\sigma_2^4)(n^{:} - n^{-1}(2p-1))$ $+2\sigma_2^4(p-1) +$ $4\beta_3(\mathbf{l}_p - n^{-1}p\mathbf{n})'\mu$ $+4\sigma_2^2\mu'(\Delta_\mathbf{n} - n^{-1}\mathbf{nn}')\mu$	$(V_2 + v_2^2)[g_\mathbf{n}(n^{:} - n^{-1}(2p-1))$ $+2(p-1)] + (V_1 + v_1^2)[g_\mu(n - n^{-1}\mathbf{n}'\mathbf{n})$ $+2\operatorname{tr}(\mathbf{Z}'\mathbf{F}_1\mathbf{Z})^2]$ $+(p-1)^2V_2 + V_1(n - n^{-1}\mathbf{n}'\mathbf{n})^2$ $+4v_1v_2(n - n^{-1}\mathbf{n}'\mathbf{n})$
	$V(T_2 \mid .)$	$(\beta_4 - 3\sigma_2^4)(n + n^{:} - 2p)$ $+2\sigma_2^4(n-p)$	$(V_2 + v_2^2)[g_\mathbf{n}(n + n^{:} - 2p)$ $+2(n-p)] + (n-p)^2V_2$
	$C(T_1, T_2 \mid .)$	$(\beta_4 - 3\sigma_2^4)[p(n^{-1}+1) - n^{:} - 1]$ $+2\beta_3(n^{-1}p\mathbf{n} - \mathbf{l}_p)'\mu$	$(V_2 + v_2^2)g_\mathbf{n}(p(n^{-1}+1) - n^{:} - 1)$ $+V_2(p-1)(n-p)$
	$C(\mathbf{T}_0, T_1 \mid .) : p \times 1$	$2\sigma_2^2(\Delta_\mathbf{n} - n^{-1}\mathbf{nn}')\mu + \beta_3(\mathbf{l}_p - n^{-1}\mathbf{n})$	$[(b_3\mathbf{I}_p + a_3\Delta_\mathbf{n})(\mathbf{I}_p - n^{-1}\Delta_\mathbf{n})$ $+c_{0,1}(n - n^{-1}\mathbf{n}'\mathbf{n})\Delta_\mathbf{n}]\mathbf{l}_p$
	$C(\mathbf{T}_0, T_2 \mid .) : p \times 1$	$\beta_3(\mathbf{n} - \mathbf{l}_p)$	$b_3(\mathbf{n} - \mathbf{l}_p)$

Table 2. A comparison between different solutions to model (1)-(2)

Solution	Sampling assumptions	Prior specifications	Result	Posterior expectation	Posterior variance	where
CC85	$(\eta,\mu \mid Z,S,\theta) \sim N$	θ^N known	$E(\mu \mid Z,S,t_0,\theta)$ $V(\mu \mid Z,S,t_0,\theta)$	$\sigma_2^2\omega\mu_0+(I_p-\sigma_2^2\Delta_\omega)\bar{y}$	$\sigma_1^2\sigma_2^2\Delta_\omega$	$\omega=[\omega_k]=\left[(n_k\sigma_1^2+\sigma_2^2)^{-1}\right]$
SS69	$(\eta,\mu \mid Z,S,\theta) \sim N$	$p(\mu_0 \mid Z,S,\sigma_1^2,\sigma_2^2)=\text{const}$	$E(\mu \mid Z,S,t_0,\sigma_1^2,\sigma_2^2)$ $V(\mu \mid Z,S,t_0,\sigma_1^2,\sigma_2^2)$	$\sigma_2^2\omega\gamma+(I_p-\sigma_2^2\Delta_\omega)\bar{y}$	$\sigma_1^2\sigma_2^2\Delta_\omega+\sigma_2^4(\omega'n)^{-1}\omega\omega'$	$\gamma=(\omega'\Delta_n\bar{y})/(\omega'\Delta_n i_p)=(\omega't_0)/(\omega'n)$
CM90	first two moments of $(\eta,\mu \mid Z,S,\theta)$	$E(\mu_0 \mid Z,S)=m_0$ $V(\mu_0 \mid Z,S)=M_0$ $E(\sigma_1^2 \mid Z,S)=v_1$ (a) $E(\sigma_2^2 \mid Z,S)=v_2$	$\hat{E}^{Z,S}(\mu \mid t_0)$ $\hat{V}^{Z,S}(\mu \mid t_0)$	$v_2 w[hm_0+(1-h)g]+(I_p-v_2\Delta_w)\bar{y}$	$v_1v_2\Delta_w+v_2^2M_0 h w w'$	$w=[w_k]=[(n_kv_1+v_2)^{-1}]$ $h=(1+M_0w'n)^{-1}$ $g=(w'\Delta_n\bar{y})/(w'\Delta_n i_p)=(w't_0)/(w'n)$
CM91	first two moments of $(\eta,\mu \mid Z,S,\theta)$	(a) + $V(\sigma_1^2 \mid Z,S)=V_1$ $V(\sigma_2^2 \mid Z,S)=V_2$ (b) $C(\sigma_1^2,\mu_0 \mid Z,S)=c_{0,1}$ + $a_3=E(\alpha_3 \mid Z,S)$ $a_4=E(\alpha_4 \mid Z,S)$ (c) $b_3=E(\beta_3 \mid Z,S)$ $b_4=E(\beta_4 \mid Z,S)$	$\hat{E}^{Z,S}(\mu \mid t)$ $\hat{V}^{Z,S}(\mu \mid t)$	$(I_p-C_\circ)i_pm_0+C_\circ\bar{y}$ $+C_{\circ\circ}\begin{bmatrix} t_1-E(t_1 \mid Z,S) \\ t_2-E(t_2 \mid Z,S)\end{bmatrix}$	for which it is difficult to find an analytical formulation which can be compared to the former ones	$C_\circ=[C_0V^{00}+c_1(v^{01})']\Delta_w$ $C_{\circ\circ}=[C_0v^{01}+c_1v^{11}$ $C_0v^{02}+c_1v^{12}]\quad :p\times 2$

References

H. Bolfarine, C.A. De Bragança Pereira and J.Rodrigues (1987), Robust prediction in finite populations: a Bayesian perspective, *Sankhyā*, 49, Series B, (1), 23-35.

D. Cocchi (1985), Bayesian inference in a finite population with a categorical auxiliary variable, *Statistica*, 45, (3).

D. Cocchi and M.Mouchart (1989), *Approximations of Bayesian solutions in finite population models*, CORE DP 8905.

D. Cocchi and M. Mouchart (1990), Linear Bayes estimation in finite populations with a categorical auxiliary variable, *Statistics*, 3, 437-454.

F. De Vylder (1982), Constrained credibility estimators in the regression model, *Scand. Actuarial J.*, 23-37.

W.A. Ericson (1988), Bayesian inference in finite populations, Ch. 9 of P.R. Krishnaiah and C.R. Rao (eds.), *Handbook of Statistics*, vol. 6, Sampling, North Holland, pp. 213-246.

C. Ferreri (1968), *Di alcune ricerche su un nuovo schema metodologico per l'analisi delle distribuzioni statistiche di frequenza*, Ingrana, Palermo.

J.P. Florens and M. Mouchart (1985), Conditioning in dynamic models, *Journal of Time Series Analysis*, 6 (1), 15-34.

J.P. Florens, M. Mouchart and J.-M. Rolin (1990), *Elements of Bayesian Statistics*, Marcel Dekker Inc., New York.

D. Holt, T.M.F. Smith and T.J. Tomberlin (1979), A model based approach to estimation for small subgroups of a population, *Journal of the American Statistical Association*, 355-358.

K.J. Lui and W.G. Cumberland (1989), A Bayesian approach to small domain estimation, *Journal of Official Statistics*, 143-156.

M. Mouchart and L. Simar (1980), Least squares approximation in Bayesian analysis, In *Bayesian Statistics*, Proceedings of the First International Meeting held in Valencia, Spain, May 28-June 2, 1979, Valencia University Press.

M. Mouchart and L. Simar (1983), Theory and applications of least squares approximation in Bayesian analysis. In *Specifying Statistical Models*, Proceedings of the Second Franco-Belgian Meeting of Statisticians. Edited by J.P. Florens, M. Mouchart, J.-P. Raoult, L. Simar and A.F.M. Smith, Lecture Notes in Statistics, 16, Springer-Verlag, New York.

M. Mouchart and L. Simar (1984a), A note on Least-Squares approximation in the Bayesian analysis of regression models, *J. R. Statist. Soc.* B, 46, (1), 124-133.

M. Mouchart and L. Simar (1984b), Bayesian predictions: non parametric methods and least squares approximations. In *Alternative Approaches to Time Series Analysis*, Proceedings of the Third Franco-Belgian Meeting of Statisticians. Edited by J.P. Florens, M. Mouchart, J.P. Raoult and L. Simar, Presses Universitaires de Saint Louis, Bruxelles.

R. Norberg (1986), Hierarchical credibility: analysis of a random effect linear model with nested classification, *Scand. Actuarial Journal*, 204-222.

C.R. Rao and J. Kleffe (1980), Estimation of variance components, in P.R. Krishnaiah, ed., *Handbook of Statistics*, 1, North Holland, 1-40.

R.M. Royall and D.Pfeffermann (1982), Balanced samples and robust Bayesian inference in finite population sampling, *Biometrika*, 69 (2), 410-409.

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

A.J. Scott and T.M.F. Smith (1969), Estimation in multi-stage surveys, *Journal of the American Statistical Association*, 64, 830-840.

R.A. Sugden and T.M.F. Smith (1984), Ignorable and informative designs in survey sampling inference, *Biometrika*, 71 (3), 495-506).