Statistics Norway
Department for Coordination and Development

Li-Chun Zhang

Documents

SMAREST
A survey of SMall ARea ESTimation

Statistics Norway
Statistisk sentralbyrå

Preface The present survey of small area estimation techinques contains three parts which, while being dependent of each other, can be read separately. The first part provides an overview of a number of techniques in the literature. The second part concentrates on models for continuous survey variables, whereas the third part deals with categorical variables.

SMAREST (I): An overview

1 The growing demand on small area statistics for policy making, fund distribution, and local planning has generated considerable research interest in the last twenty or so years at many national statistical agencies, including Statistics Norway.

1.1 *Review articles, bibliogrphical notes:* Ghosh and Rao (1994) presented a review which was central in the recent years. Earlier ones included, among others, Chaudhuri (1992), Rao (1986), Purcell and Kish (1979), and Morrison (1971), which more or less covered the subject up to their respective time of appearance. Platek, Rao, Särndal, and Singh (1987) and Platek and Singh (1986), and Kalton, Kordos, and Platek (1993) brought together contributions from two international symposia on small area estimation in 1986 and 1992. Whereas Small Area Estimation Research Team (1983) contained a large reference list.

1.2 *Experiences from international statistical agencies*: On general practice there are, for instance, Singh, Gambino, and Mantel (1993) for U.S., Brackstone (1987), Statistics Canada (1987) for Canada, and Ansen, Hallen, and Ylander (1988) for Sweden. For empirical study, Drew, Singh, and Choudhry (1982) and Falorsi, Falorsi, and Russo (1994) compared different methods for the Labour Force Survey (LFS) in, respectively, Canada and Italy. Whereas Lundstrøm (1987) and Decaudin and Labat (1997) dealt with demographical statistics in Sweden and France. The debate on U.S. Census 1980 found Ericksen and Kadane (1985) and Ericksen, Kadane, and Tukey (1989) on the one side, and Freedman and Navidi (1986, 1992) as the sharping oppsite. A number of connected issues were discussed in Ericksen and Kadane (1987), Cressie (1989, 1992) and Isaki, Schultz, Smith, and Diffendal (1987).

1.3 *Some Norwegian experience:* Research conducted in Statistics Norway often utilises data from the LFS. Laake (1978, 1979) contained some early attempts at the synthetic estimator in combination with post-stratification. The study was carried further in Heldal, Swensen, and Thomsen (1987) in connection with Norwegian Census 1990. Spjøtvoll and Thomsen (1987) concentrated on the composite estimator and proposed an efficient empirical Bayes approach. Also Neural Network (e.g. Nordbotten, 1996) has been investigated in more recent methodological works.

2 While earlier methods often appeared ambiguous in this respect, the metodological developement on SMAREST has witnessed an increasing emphasis on modelling.

2.1 *(Sample) regression-symptomatic method:* The use of *symptomatic variables* have originated from the so-called *Symptomatic Accounting Technique (SAT)* (Marker, 1983), which is one of the oldest small area estimation methods. Basically, these should be known from various administration registers, and are correlated with the survey variable. (Sample) regression-symptomatic method (Ericksen, 1974; Purcell and Kish, 1979, 1980; Zidek, 1982; Marker, 1983) generalises the SAT under the multiple regression framework.

2.1.1 Denote by y the survey variable, and x the vector of symptomatic variables. Let a be the index of small area, and t the index of time. Let $p_{a,t}(y) = Y_{a,t}/\sum_a Y_{a,t}$, and $r_{a,t}(y) = p_{a,t}(y)/p_{a,t-1}(y)$. Similarly, let $p_{a,t}(x) = X_{a,t}/\sum_a X_{a,t}$, and $r_{a,t}(x) = p_{a,t}(x)/p_{a,t-1}(x)$. In particular, $(t-1,t)$ represent past census, so that both $r_{a,t}(x)$ and $r_{a,t}(y)$ are known for the entire population. The multiple regression gives us, at $t+1$,

$$r_{a,t}(y) = \beta_0 + r_{a,t}(x)^T\beta \qquad \Rightarrow \qquad \hat{r}_{a,t+1}(y) = \hat{\beta}_0 + r_{a,t+1}(x)^T\hat{\beta},$$

where $t+1$ can be any time after the last census at t, and $r_{a,t+1}$ are known from the updated registers, and $(\hat{\beta}_0, \hat{\beta})$ are based on census at $t-1$ and t.

2.1.2 The sample symptomatic regression method assumes that sample based estimates of $\hat{r}_{a,t+1}(y)$ are available for some, but not all, of the small areas. The multiple regression model is first fitted based on these areas at $(t, t+1)$, i.e. $\hat{r}_{a,t+1}(y) = \beta_0 + r_{a,t+1}(x)^T\beta$, and afterwards used on the rest small areas.

Remark *One needs to balance the cost and uncertainty associated with sample based $\hat{r}_{a,t+1}(y)$ against the bias of $(\hat{\beta}_0, \hat{\beta})$ based on past census.*

2.2 *SPREE:* Purcell and Kish (1980) summarized the *structure preserving estimation (SPREE)* method, developed in Freeman and Koch (1976), Chambers and Feeney (1977), and Purcell (1979). The SPREE adjusts past simultaneous distribution of the survey variable and the auxiliary variable over the small areas, according to the updated/present marginal distributions. In a way it can be viewed as a constrained raking method.

2.2.1 Assume categorical variables. Denote by $N_{axy}(t)$ the *a priori* simultaneous distribution of (X, Y) over small areas, which is called the *association structure*. (Typically, $N_{axy}(t)$ can obtained from the last census.) Denote by $m_{\cdot xy}(t+1)$ the updated marginal distribution (summed over all small areas), which is called the *allocation structure*. The estimator for the updated $N_{axy}(t+1)$, which is said to preserve the association structure while respecting the allocation structure, is defined as

$$\hat{N}_{axy}(t+1) = \{N_{axy}(t)/N_{\cdot xy}(t)\}m_{\cdot xy}(t+1).$$

2.2.2 Alternative association and allocation structures (Purcell and Kish, 1980), omitting index $(t, t+1)$, include (i) $\{N_{axy}, (m_{\cdot xy}, m_{a\cdot\cdot})\}$, (ii) $\{N_{axy}, (m_{\cdot xy}, m_{ax\cdot})\}$, (iii) $\{N_{ax\cdot}, m_{\cdot xy}\}$, (iv) $\{N_{ax\cdot}, (m_{\cdot xy}, m_{a\cdot\cdot})\}$, and (v) $\{N_{ax\cdot}, (m_{\cdot xy}, m_{ax\cdot})\}$. These sometimes lead to iterative proportional fitting (IPF), or raking, if the allocation structure contains more than one marginal distribution. Moreover, correspondence between these procedures and their log-linear model-representations implies e.g. that N_{axy} can often be reduced to lower-order interactions, without much loss of information.

Remark *Before applying the SPREE one should consider whether it is appropriate to preserve the (associate) structure in the first place.*

2.3 *Synthetic estimator:* According to Gonzalez (1973), "An unbiased esetimate is obtained from a sample survey for a large area; when this estimate is used to derive estimates for subareas under the assumption that the small area have the same characteristics as the large area, we identify these estimates as synthetic estimates."

2.3.1 In practice it is seldom to apply the mean of a large area directly to *all* the small areas. Synthetic estimates are often formed in combination with post-stratification, the assumption being that the post-stratum mean does not vary over some or all of the small areas, which usually cut across the post-strata (Laake, 1978; Heldal, Swensen, and Thomsen, 1987). Let h be the post-stratum index. A *synthetic estimator* is given as

$$\hat{Y}_a = \sum_h N_{ah}(\hat{Y}_{\cdot h}/N_{\cdot h}) \qquad N_{\cdot h} = \sum_a N_{ah} \quad \text{and} \quad Y_{\cdot h} = \sum_a Y_{ah}.$$

2.3.2 Holt and Smith (1979) noted that the "post-stratified" synthetic estimator can be evaluated under the *group-mean model*, i.e. for $i \neq j$ and $h \neq g$,

$$y_{i,ah} = \mu_h + \epsilon_{i,ah} \qquad E[\epsilon_{i,ah}] = 0 \quad \text{and} \quad Var(\epsilon_{i,ah}) = \sigma^2 \quad \text{and} \quad Cov(\epsilon_{i,ah}, \epsilon_{j,ag}) = 0.$$

Whereas Laake (1979) had a more general variance structure on $\epsilon_{i,ah}$, further extensions can e.g. be found in Lui and Cumberland (1991).

Remark *Holt and Smith (1979) studied sensitivity of the group-mean model under alternative models, such as (i)* $\mu_h = \mu$, *(ii)* $E[y_{i,ah}] = \mu_a$, *(iii)* $E[y_{i,ah}] = \mu_a + \mu_h$, *or (iv)* $E[y_{i,ah}] = \mu_{ah}$, *etc.*

2.4 *Composite estimator:* Composite estimator balances the potential bias of an indirect estimator against the instability of a direct/local estimator from the small area in question.

2.4.1 In the literature, the indirect estimator is often set to be a synthetic estimator, and the composite estimator, denoted by $\hat{Y}_a^C$, takes the simple linear form, i.e.

$$\hat{Y}_a^C = w_a \hat{Y}_a^D + (1 - w_a)\hat{Y}_a^I,$$

where $\hat{Y}_a^D$ is the direct estimator, and $\hat{Y}_a^I$ the indirect one.

2.4.2 In mean-square-error based approach, w_a is continuous on $(0, 1)$. The MSE of $\hat{Y}_a^C$ is minimized at

$$w_a = MSE(\hat{Y}_a^I)/\{MSE(\hat{Y}_a^I) + Var(\hat{Y}_a^D)\}.$$

In practice, though, it is often difficult to obtain stable estimates of w_a, and several remedies were suggested in Schaible (1978) and Purcell and Kish (1979).

2.4.3 Drew, Singh, and Choudhry (1982) proposed a *sample-dependent estimator*, where $w_a = 1$ provided $\hat{N}_a \geq \delta N_a$, and $w_a = \hat{N}_a/(\delta N_a)$ otherwise. In particular, $\hat{N}_a$ is the direct, unbiased estimator of small area population size N_a, and δ some preassigned constant. Whereas Särndal and Hidiroglou (1989) suggested $w_a = 1$ provided $\hat{N}_a \geq N_a$, and $w_a = (\hat{N}_a/N_a)^{\gamma-1}$ otherwise. Notice that the two estimators coincide in case $\delta = 1$ and $\gamma = 2$.

Remark *The difficulty of the sample-dependent estimator lies in the choice of the* 'cut-off' limit *for* w_a. *Unless the total sample size is sufficiently large,* $\hat{Y}_a^C$ *can fail to borrow strength from related small areas, even when* $E[n_a]$ *is actually not large enough to make* $\hat{Y}_a^D$ *reliable (Ghosh and Rao, 1994). In any case, due to the discontinuity caused by the 'cut-off' limit, the composite estimator may behave unreasonably in those small areas close to the chosen limit, depending on which side they happen to be.*

2.5 *Non-Bayesian predictive methods:* Unless small area estimation is solely based on direct estimators, modeling of the survey variable is necessary. This can be seen clearly once the model assumptions, which underline the various estimators so far discussed, are made explicit — see e.g. Marker (1983) on the (sample) regression-symptomatic methods, and Holt and Smith (1979) on the synthetic estimator. Both the *random area-effect model* (e.g. Fay and Herriot, 1979) and the *nested error regression model* (Battese, Harter, and Fuller, 1988) extend the group-mean model to incoorperate the between-area variation. While the former introduces a random error at the area-level, the latter remains at the individual level.

2.5.1 The random area-effect model adds to the group-mean model a random error at the small area level,

$$Y_a = X_a^T\beta + z_a e_a \qquad E[e_a] = 0 \quad \text{and} \quad Var(e_a) = \sigma^2.$$

Let $\hat{Y}_a$ be some unbiased direct estimator, we have

$$\hat{Y}_a = X_a^T\beta + z_a e_a + \epsilon_a \qquad E[\epsilon_a] = 0 \quad \text{and} \quad Var(\epsilon_a|Y_a) = \tau_a^2,$$

where τ_a^2 refers to the sampling error. Cressie (1992), Prasad and Rao (1990), and Ghosh and Rao (1994) studied the random effect model under the variance component approach.

2.5.2 The nested error regression model assumes individual auxiliary information, and

$$y_{i,a} = x_{i,a}^T\beta + e_a + \epsilon_{i,a},$$

where both e_a and $\epsilon_{i,a}$ are model effects, i.e. none of them depends on sampling. In particular, modeling at the individual level implies the predictive approach. The *best linear unbiased predictor (BLUP)* depends not only on the BLUE of β, but also the predicted e_a conditional to the realized sample. Battese, Harter, and Fuller (1988), Fuller and Harter (1987) (the multivariate version), Prasad and Rao (1990), and Stukel (1991) studied the model in details; whereas Holt and Moura (1993) extends it to allow for area-specific "slope".

Remark *The between-area variation is introduced at the area-level through* e_a *in both models, which is meaningful for a particular small area only if it is represented in the sample.*

2.6 *Empirical and hierarchical Bayesian methods:* Whereas empirical Bayes (EB) approach (Morris, 1983) does not require explicit form of the prior distribution of the parameters, hierarchical Bayes (HB) approach (Datta and Ghosh, 1991; Ghosh, Natarajan, Stroud, and Carlin, 1998), operates under full parameterization, where the posterior distribution is obtained using the Bayes theorem.

2.6.1 With the *empirical Bayes (EB)* approach, one first derives the posterior distribution of the survey variable, denoted by $p(Y_a|\hat{Y}_a, \theta)$, as if the model parameters θ were known. To base prediction on $p(Y_a|\hat{Y}_a, \hat{\theta})$ alone would obviously lead to underestimation of the posterior variance. Adjustments need to be made to account for the uncertainty in $\hat{\theta}$ (Laird and Louis, 1987; Kass and Steffey, 1989).

2.6.2 The past twenty years have witnessed enormous development in Markov Chain Monte Carlo (MCMC) methods, which made the HB approach more feasible than ever before. Whereas Ghosh and Larihi (1987), Raghunathan (1993) and DeSouza (1992) contained approximate, or modified, EB or HB methods.

2.6.3 Ghosh (1992) established, under rather general settings, that the Bayesian posterior estimates, denoted by $\hat{Y}_a^B$, contain less variation than that among the true Y_a. A general method was proposed (Ghosh, 1992), which leads to the *constrained hierarchical Bayes (CHB)* approach. Earlier the problem was dealt with by Louis (1984) and Spjøtvoll and Thomsen (1987) under alternative EB frameworks, i.e. the CEB approach.

Remark *Similar problems exist also in the non-Bayesian predictive approach, which call for similar development of constrained approach.*

3 To make use of indrect data, modelling of small area statistics must contain structural features which are common to the population; whereas to allow for between-area variation, it must also deal with area-specific deviation from these common, baseline, synthetic features.

3.1 Let μ_a be the *mean-parameter* of the survey variable from area a, i.e. $\mu_a = E[y_a]$. The synthetic features common to the population can often be summarized in the following manner, i.e.

$$h(\mu_a) = \xi_a = g(\xi, x_a)$$

Notice that $h()$, ξ and $g()$ are independent of a. The model is specified at the area-/domain-level if y_a is a scalar, in which case auxiliary x_a is a vector in general; whereas the model is specified at the individual-level if y_a is a vector, i.e. $y_a = (y_1, ..., y_{N_a})^T$, in which case x_a is a matrics in general.

Remark *Typically $g()$ is of the linear form which, through the link function $h()$, leads to generalized linear models. We have retained the general form which, among others, allow for non-parametric approach as well.*

Remark *The group-mean model (which motivates the synthetic estimator based on post-stratification) can be expressed in this way, where ξ stands for the post-stratum mean across the areas, and x_a the known post-stratum proportions within the relevant small area, and $g = x_a^T \xi$.*

3.2 The deviation part can similarly be summarized in another parameter, denoted by $\eta_a = \eta(z_a, e_a)$, where z_a contains relevant auxiliary information, and e_a are random errors with prior distribution $\pi(\psi)$ — though often $\pi(\psi)$ is only specified up to the first two moments of e_a.

Remark *In the random area-effect model we have $\eta_a = z_a e_a$, whereas in the nested error regression model, we have $\eta_a = e_a$ and $z_a = 1$.*

3.3 Combining synthetic feature with local deviation, we have, for ξ_a and η_a defined as above,

$$h(\mu_a) = \xi_a + \eta_a = g(\xi, x_a) + \eta(z_a, e_a).$$

We call ξ_a the *synthetic-parameter*, and η_a the *deviation-parameter*, and $\xi_a + \eta_a$ the linear predictor of the mean-parameter μ_a, which is obtained through a transformation defined by the link function $h()$.

Remark *Let $g = x_a^T \xi$, we obtain the standard linear model if $h(\mu_a) = \mu_a$ and $\eta_a = 0$; the variance-component model if $h(\mu_a) = \mu_a$ and $\eta_a = \eta(z_a, e_a)$; the generalized linear model if $h(\mu_a) = h(\mu_a)$ and $\eta_a = 0$; the generalized linear mixed model if $h(\mu_a) = h(\mu_a)$ and $\eta_a = \eta(z_a, e_a)$. Whereas in non-parametric, or semi-parametric, approach, $g()$ can be left unspecified.*

SMAREST (II): Models for continuous survey variables

1 In small area estimation, the finite population is divided into a number of sub-groups, i.e. domains.

1.1 Denote by U the population, which is divided into H domains, denoted by U_h, such that $U = \cup_{h=1}^{H} U_h$ where $U_g \cap U_h = \emptyset$ for $g \neq h$. Denote by s the sample, and by s_h the h-th domain in the sample, and so on.

Example In a business survey conducted by the Norwegian Fishing Directory in 1996, the sample contained 394 fish boats (from 1283 in the population). Classified according to (i) the length of the boat (4 classes), (ii) the type of lisence granted (22 types), and (iii) the county in which the boat was registered (9 counties), there were altogether 166 non-empty domains in the population, of which 109 were represented in the sample.

1.2 Denote by y the survey variable. Let $Y_h = \sum_{i \in U_h} y_i$, and $Y = \sum_h Y_h$. Let $y_h = \sum_{i \in s_h} y_i$, and $y_s = \sum_h y_h$. In particular, let $y_h = 0$ if $s_h = \emptyset$.

Example (cont'd) Let the amount of fished catched be y, such that $\bar{Y}_h$ is the mean Catch in domain h, *etc.*. The Catch is in fact known for all the fish boats in the population, so the various methods of prediction can be checked against the true vaules.

1.3 Denote by x the auxiliary variable, which may be (column) vector-valued. Let $X_h = \sum_{i \in U_h} x_i$, and $X = \sum_h X_h$. Let $x_h = \sum_{i \in s_h} x_i$, and $x_s = \sum_h x_h$. In particular, let $x_h = 0$ if $s_h = \emptyset$.

Example (cont'd) Based on monthly report to the Directory, a yearly fishing income, denoted by x, is available for each boat, which will be used as the auxiliary variable.

2 The random (domain-) effect model (Fay and Herriot, 1979; Cressie, 1992; Prasad and Rao, 1990; Ghosh and Rao, 1994) accounts for the between-domain variation at the domain level. Whereas the (one-fold) nested-error model (Battese, Harter, and Fuller, 1988; Prasad and Rao, 1990; Stukel, 1991) further introduces a random error at the individual level. Variance Components approach (Harville, 1977; Robinson, 1991) is applied in both cases.

2.1 Let v_h be the random domain-effect, such that

$$Y_h = X_h^T \beta + z_h v_h \qquad E[v_h] = 0 \quad \text{and} \quad Var(v_h) = \sigma_v^2 \quad \text{and} \quad Cov(v_g, v_h) = 0$$

for some domain-related constant z_h. Let $\hat{Y}_h$ be some unbiased direct estimator of Y_h, we have

$$\hat{Y}_h = X_h^T \beta + z_h v_h + e_h \qquad E[e_h] = 0 \quad \text{and} \quad Var(e_h | Y_h) = \tau_h^2.$$

Notice that while v_h is introduced by the model, e_h is the sampling error which is independent of the model.

2.1.1 Under the random effect model, $\hat{Y}_h$ of different domains are uncorrelated, and the variance of $\hat{Y}_h$ is given as $\psi_h^2 = z_h^2 \sigma_v^2 + \tau_h^2$. The transformation

$$\hat{Y}_h / \psi_h = (X_h / \psi_h)^T \beta + u_h$$

achieves constant variance in u_h. In other words, estimating β under the random effect model is the same as applying the ordinary least square (OLS) technique to the regression of $\hat{Y}_h / \psi_h$ on X_h / ψ_h, i.e.

$$\hat{\beta} = (\sum_h X_h X_h^T / \psi_h^2)^{-1} (\sum_h X_h \hat{Y}_h / \psi_h^2).$$

Remark *Given univariate auxiliary variable, this reduces to* $\hat{\beta} = (\sum_h X_h \hat{Y}_h / \psi_h^2) / (\sum_h X_h^2 / \psi_h^2)$.

2.1.2 The best linear unbiased predictor (BLUP) of Y_h is $X_h^T \beta + z_h \hat{v}_h$, where $X_h^T \hat{\beta}$ is the regression synthetic predictor, and $z_h \hat{v}_h = (\hat{Y}_h - X_h^T \hat{\beta}) \gamma_h$ the predictor of the domain-effect conditional to $\hat{Y}_h - X_h^T \hat{\beta}$ and $\gamma_h = (z_h^2 \sigma_v^2) / \psi_h^2$. The BLUP of Y_h is thus seen to be

$$\hat{Y}_h = X_h^T \hat{\beta} + (\hat{Y}_h - X_h^T \hat{\beta}) \gamma = \gamma \hat{Y}_h + (1 - \gamma) X_h^T \hat{\beta},$$

and it turns out to be the weighted sum of a direct estimator and a synthetic estimator.

Remark *Due to historical reasons, estimators of the random effects are referred to as predictors.*

2.1.3 The empirical best linear unbiased predictor (EBLUP) is obtained from replacing σ_v^2 in the BLUP with any asymptotically consistent estimator $\hat{\sigma}_v^2$. Since $E[\sum_h (Y_h - X_h^T \hat{\beta})^2/\psi_h^2] = H - \dim(\beta)$, a method of moments estimator, which does not depend on normality of v_h, is given by $\max(\hat{\sigma}_v^2, 0)$, where

$$\sum_h (\hat{Y}_h - X_h^T \hat{\beta})^2/(z_h^2 \hat{\sigma}_v^2 + \tau_h^2) = H - \dim(\beta).$$

Remark *The mean square error (MSE) of the EBLUP is often estimated by replacing σ_v^2 in the MSE of the BLUP with $\hat{\sigma}_v^2$. However this may leads to serious underestimation, and an additional term accounting for the uncertainty in $\hat{\sigma}_v^2$ will be needed (Prasad and Rao, 1990). The authors there also gave an alternative moment estimator of σ_v^2.*

2.2 Let $u_{h,i}$ be the random effect in y_i for $i \in U_h$. The (one-fold) nested-error model assumes decomposition $u_{h,i} = v_h + e_i$, where v_h is the random domain-effect as under the random effect model, so that, for $i \in U_h$,

$$y_i = x_i^T \beta + v_h + e_i \qquad E[e_i] = 0 \quad \text{and} \quad Var(e_i) = \sigma_e^2 \quad \text{and} \quad Cov(e_i, e_j) = 0.$$

Through e_i the random effect is now modeled directly at the individual level. Notice also that the nested-error model here contains no variance-inflating constants, such as z_h under the random effect model.

Remark *Fuller and Harter (1987) contains the multivariate extension of the nested-error model; whereas Holt and Moura (1993) discusses mixed models which allow for domain-specific β.*

2.2.1 The simple within-domain-deviation transformation gets rid of the domain effect v_h such that, for $i \in s_h$,

$$y_i - \bar{y}_h = (x_i - \bar{x}_h)^T \beta + (e_i - \bar{e}_h) \qquad Var(e_i - \bar{e}_h) = (1 - n_h^{-1})\sigma_e^2 \quad \text{and} \quad Cov(e_i - \bar{e}_h, e_j - \bar{e}_h) = 0.$$

Regressing the y-deviations on the x-deviations gives us, for $i \in s_h$ where $n_h > 1$,

$$\hat{e}_i = (y_i - \bar{y}_h) - (x_i - \bar{x}_h)^T \{[\sum_{h; n_h > 1} (x_i - \bar{x}_h)(x_i - \bar{x}_h)^T/(1 - n_h^{-1})]^{-1} [\sum_{h; n_h > 1} (x_i - \bar{x}_h)(y_i - \bar{y}_h)]\},$$

and $\hat{e}_i = 0$ for $i \in s_h$ where $n_h = 1$. The method of moments estimator of σ_e^2 is given as

$$\hat{\sigma}_e^2 = (\sum_{i \in s} \hat{e}_i^2)/\{\sum_h n_h - 1 - \dim(\beta)\}.$$

Remark *Applying simple OLS technique to obtain alternative estimates of $e_i - \bar{e}_h$ is equivalent to ignoring the variation in the sample domain size n_h.*

Example (cont'd) Among the 109 domains represented in the sample, 66 have more than one observation. Based on these, we obtain $\hat{\sigma}_e^2 = 6.456 \times 10^{11}$. Whereas the OLS technique gives us $\hat{\sigma}_e^2 = 6.449 \times 10^{11}$.

2.2.2 To estimate σ_v^2 the method of mements is often applied. Starting with the OLS regression of y on x, one may derive the second moment of the resulting residuals, denoted by $\tilde{u}$, as a weighted sum of σ_v^2 and σ_e^2. One such estimator, obtained from substituting σ_e^2 with $\hat{\sigma}_e^2$, is given as $\max(0, \hat{\sigma}_v^2)$, where

$$\hat{\sigma}_v^2 = \{(\sum_i \hat{\tilde{u}}_i^2) - [n - \dim(\beta)]\hat{\sigma}_e^2\}/\{n - \text{Tr}[(\sum_h x_i x_i^T)^{-1}(\sum_h x_h^T x_h)]\}.$$

Example (cont'd) With univariate auxiliary variable the denominator above reduces to $n - (\sum_h x_h^2)/(\sum_i x_i^2)$, and we obtain $\hat{\sigma}_v^2 = 6.316 \times 10^{12}$.

2.2.3 Given the variance components, the following transformation, i.e. for $i \in s_h$,

$$\tilde{y}_i = y_i - \alpha_h \bar{y}_h \quad \text{and} \quad \tilde{x}_i = x_i - \alpha_h \bar{x}_h \quad \text{and} \quad \tilde{e}_i = \tilde{y}_i - \tilde{x}_i^T \beta \quad \text{and} \quad \alpha_h = 1 - [\sigma_e^2/(n_h \sigma_v^2 + \sigma_e^2)]^{\frac{1}{2}}$$

makes $\tilde{e}_i$ uncorrelated with constant variance σ_e^2 (Fuller and Battese, 1973). In other words, estimating β under the nested-error model is the same as applying the OLS technique to the regression of $\tilde{y}$ on $\tilde{x}$, i.e.

$$\hat{\beta} = (\sum_i \tilde{x}_i \tilde{x}_i^T)^{-1} (\sum_i \tilde{x}_i \tilde{y}_i)$$

The EBLUP is obtained by replacing (σ_e^2, σ_v^2) with their respective estimates.

Example (cont'd) In this case $\hat{\beta}_{EBLUP} = 0.260$, whereas $\hat{\beta}_{OLS} = 0.272$ based directly on y and x. Also, $\tilde{R} = (\sum_i \tilde{y}_i)/(\sum_i \tilde{x}_i) = 0.306$, and $\hat{R} = (\sum_i y_i)/(\sum_i x_i) = 0.293$. In other words, the choice between ratio estimation and regression estimation appears to make a bigger difference than the assumptions about the variance structure.

2.2.4 The BLUP of $\bar{Y}_h$ is the sum of the regression synthetic estimator $\bar{X}_h^T \hat{\beta}$, and the conditional expected domain effect $E[v_h | \bar{u}_h]$, where $\bar{u}_h = \bar{y}_h - \bar{x}_h^T \hat{\beta}$, and is given as

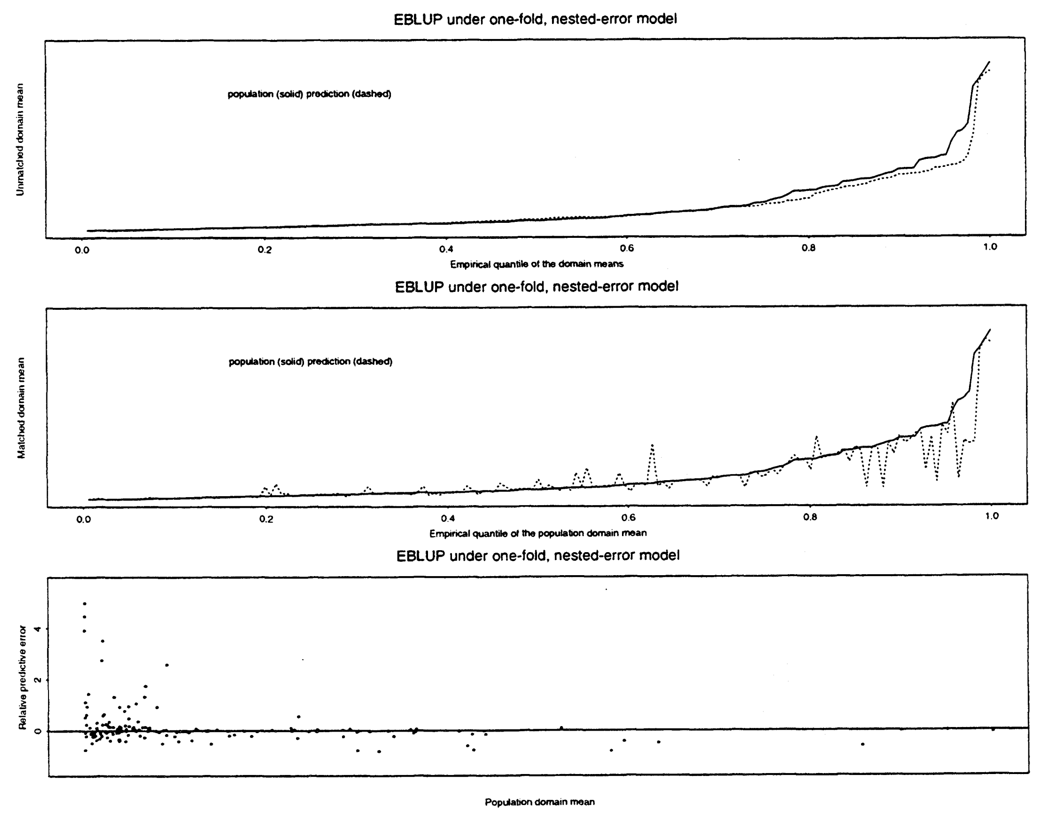
$$\hat{\bar{Y}}_h = \bar{X}_h^T \hat{\beta} + (\bar{y}_h - \bar{x}_h^T \hat{\beta}) \eta_h = \eta_h \bar{y}_h + (\bar{X}_h^T - \eta_h \bar{x}_h^T) \hat{\beta},$$

where $\eta_h = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2 / n_h)$. The EBLUP is obtained by replacing the variance components with their respective asymptotically consistent estimators.

Remark *In case the between-domain variation is much larger than the remaining individual variation, i.e. $\sigma_v^2 >> \sigma_e^2$ so that $\eta_h \approx 1$, the EBLUP becomes approximately the same as the survey regression predictor under the nested-error model, i.e.*

$$\bar{X}_h^T \hat{\beta}_{EBLUP} + (\bar{y}_h - \bar{x}_h^T \hat{\beta}_{EBLUP}) = \bar{y}_h + (\bar{X}_h^T - \bar{x}_h^T) \hat{\beta}_{EBLUP}.$$

Example (cont'd) The estimated random domain-effect is about 10 times as large as the remaining individual effect, the EBLUP is therefore practically the same as the survey regression predictor. Comparsion between the predicted mean Catch and true population vaules is given below for all the 166 domains.



Remark *In the predictive approach the EBLUP is only applied to the rest population outside of the sample.*

3 *Generalized regression model with LINearized AREa-effect (LINARE)* postulates a linear structure of the area-/domain-effect. Inference is based on the faimiliar generalized regression techinuques.

Remark *The random domain-effect of both the random effect and the nested-error model is trivial and can not be 'predicted', unless the relevant domain is represented in the sample. This causes often problems in unanticipated, or badly planned, production of small area statistics — "In general the approach involving components of variance has arisen from the need to take account of the between-small-area components of variance. However a much more rewarding approach is to seek to explain why small areas differ." (Holt and Moura, 1993).*

3.1 Consider the case with univariate auxiliary variable. The regression model which fully takes account of the between-domain effect would allow the parameter to differ from one domain to another, i.e. for $i \in U_h$,

$$y_i = x_i\beta_h + e_i \qquad E[e_i] = 0 \text{ and } Var(e_i) = v(x_i)\sigma^2.$$

The LINARE postulates a linear structure of $\underline{\beta} = (\beta_1, ..., \beta_H)^T$ through a constant *(design) structure-matrix* $B_{H\times p}$, and correspondingly a parameter vector $\underline{\xi}$ of p components, such that

$$\underline{\beta} = B\underline{\xi}.$$

The univariate within-domain regression model is thus replaced by a synthetic multiple regression with linearized domain-effect. (Obviously setting B to be the $H \times H$ identity matrix recovers the complete model.) In particular, the structure-matrix B can arise from dummy-indexing in the same way calibration arises from post-stratification.

Remark *Given the corresponding structure matrices, the LINARE can be extended to include domain-dependent intercepts, as well as multivariate auxiliary variables, while the linear predictor remains additive.*

Illustration Suppose that the domains arise from post-stratification according to Sex and Age with, say, 3 age-groups. Suppose calibration w.r.t. marginal totals of both variables. The dummy indices for the post-stratification, denoted by I, and the calibration, denoted by B, and the linearization of the domian effect, involving $\underline{\beta} = (\beta_1, ..., \beta_6)^T$ and $\underline{\xi} = (\xi_1, ..., \xi_5)^T$, can be expressed as

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \underline{\beta} = I\underline{\beta} = B\underline{\xi}.$$

3.2 Let $Q = (q_{ij})$ be the $n \times p$ *(sample) structure-matrix*, defined according to the design structure matrix, where the i-th row corresponds to $i \in s$. Let Q_x be such that its (i,j)-th element is given by $x_i q_{ij}$. The sample can thus be rewritten as

$$\underline{y} = Q_x\underline{\xi} + \underline{e} \qquad \underline{y} = (y_1, ..., y_n)^T \text{ and } \underline{e} = (e_1, ..., e_n)^T.$$

Standardize the data so that $\tilde{y}_i = y_i/\sqrt{v_i}$ and $\tilde{x}_i = x_i/\sqrt{v_i}$, and define $\tilde{Q}_x$ based on $\tilde{x}$ similiarly to Q_x on x. Generalized regression (GREG) of y on x is the same as the OLS regression of $\tilde{y}$ on $\tilde{x}$, i.e.

$$\hat{\underline{\xi}} = (\tilde{Q}_x^T\tilde{Q}_x)^{-1}(\tilde{Q}_x^T\tilde{\underline{y}}) = (Q_x^T V^{-1} Q_x)^{-1}(Q_x^T V^{-1}\underline{y}) \text{ and } \text{diag}(V) = (v_1, ..., v_n)^T.$$

Typically the variance inflating constant v_i takes the form $v_i = x_i^r$ for some fixed r. The standardized $\tilde{e}$ has constant variance σ^2, whose method of mements estimator is thus given as

$$\hat{\sigma}^2 = (\hat{\tilde{e}}^T\hat{\tilde{e}})/\{n - \text{rank}(\tilde{Q}_x)\}.$$

Remark *The sample structure-matrix Q_x is usually not of the full rank, in which case deletion of the redundant columns are necessary.*

3.3 If normality of e holds, the scaled log-likelihood and its derivatives are given as, let $\tau = \sigma^2$ and $e = \tilde{e}^T\tilde{e}$,

$$2l = -n\log\tau - e/\tau$$

$$\partial 2l/\partial\xi = 2\tilde{Q}_x^T\tilde{e}/\tau \qquad \partial 2l/\partial\tau = -n/\tau + e/\tau^2$$

$$\partial^2 l/\partial\xi^2 = -(\tilde{Q}_x^T\tilde{Q}_x)/\tau \qquad \partial^2 2l/\partial\tau^2 = n/\tau^2 - 2e/\tau^3.$$

Since $\partial e/\partial\xi \equiv 0$ whenever evaluated at $\xi = \hat{\xi}$, the variances of $\hat{\xi}$ and $\hat{\tau}$ can be estimated separately, i.e.

$$\hat{Var}(\hat{\xi}) = (\tilde{Q}_x^T\tilde{Q}_x)^{-1}\hat{\tau} \qquad \hat{Var}(\hat{\tau}) = 2\hat{\tau}^2/\{n - 2\text{rank}(\tilde{Q}_x)\}.$$

3.4 More important, however, is the estimation of the MSE of $\hat{Y}_h$. Under the model, the BLUP is unbiased. Let $\theta = (\xi, \sigma^2)$ and ignoring the sampling fraction, we have

$$E[\hat{Y}_h|\hat{\theta}] = X_h B_h\hat{\xi} \quad \text{and} \quad Var(\hat{Y}_h|\hat{\theta}) = \hat{\sigma}^2(\sum_{i\in U_h} v_i^2) = \hat{\sigma}^2 V_h^2,$$

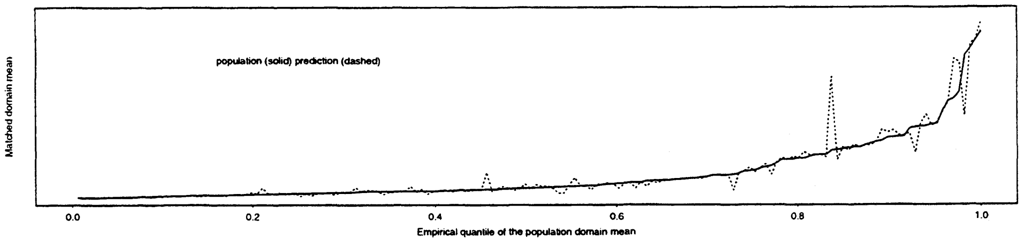
where B_h is the h-th row of the design structure-matrix which corresponds to domain h, i.e.

$$Var(\hat{Y}_h) = \sigma^2 V_h^2 + (X_h B_h)Var(\hat{\xi})(X_h B_h)^T.$$

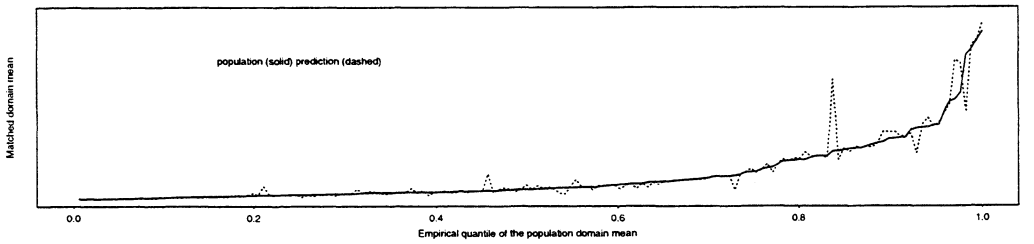
Remark *These variance estimates, as well as those derived under any other model, depends on the validity of the model, and should be treated with caution. For instance, it is probably unwise to base the choice of predictor on variances derived under their respective models alone. In fact one of the greatest challanges in small area estimation is to develop robust and sensible measures of error and uncertainty.*

Example (cont'd) In applying the LINARE to the present data, we first compared the choice of the variance inflating constant r.

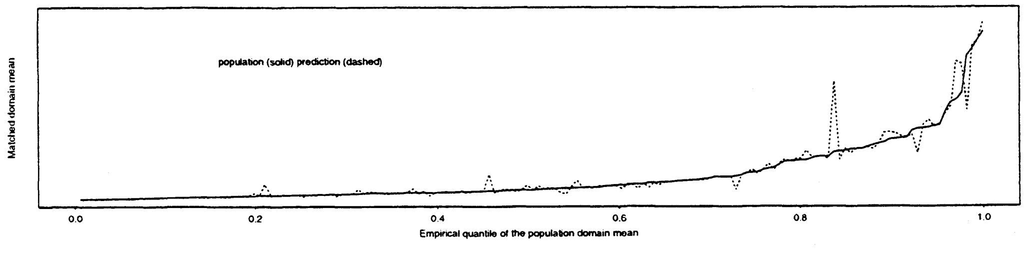
BLUP under LINARE (r=0) of domain-effect, without intercept



BLUP under LINARE (r=0.5) of domain-effect, without intercept

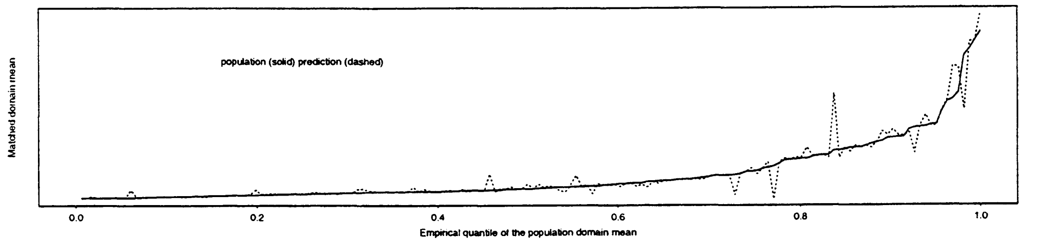


BLUP under LINARE (r=1) of domain-effect, without intercept

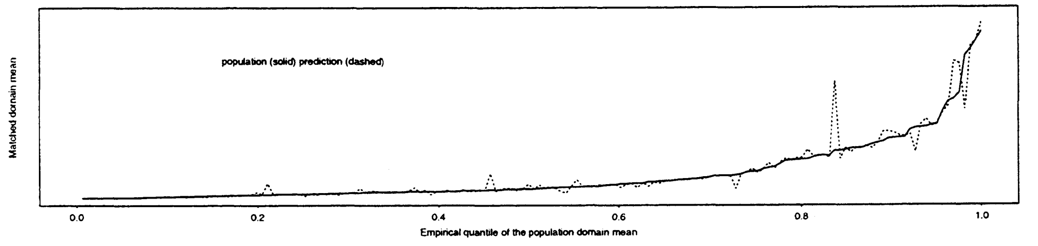


Next we investigated whether intercept, in constant or linearized form, is necessary.

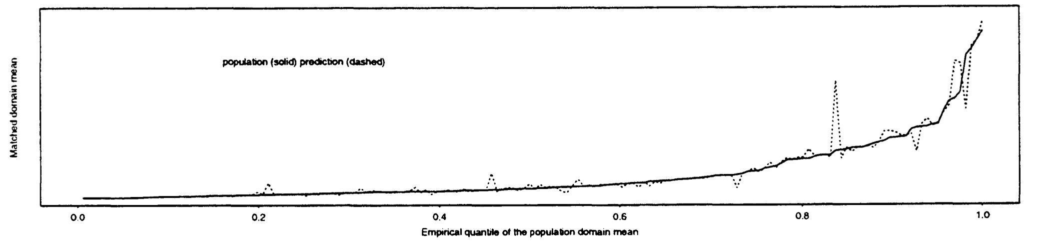
BLUP under LINARE (r=1) of domain-effect, with linearized intercept



BLUP under LINARE (r=1) of domain-effect, with constant intercept

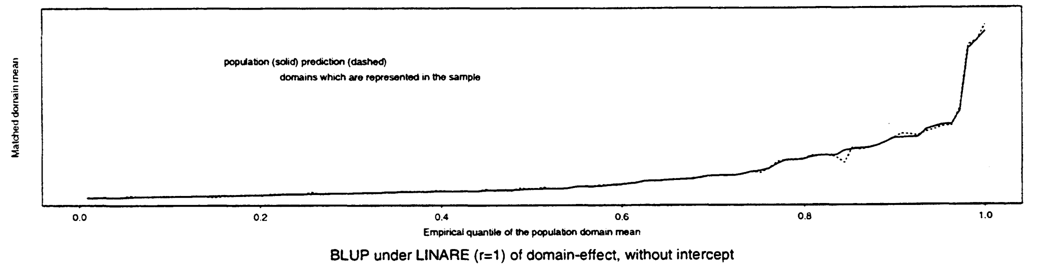


BLUP under LINARE (r=1) of domain-effect, without intercept

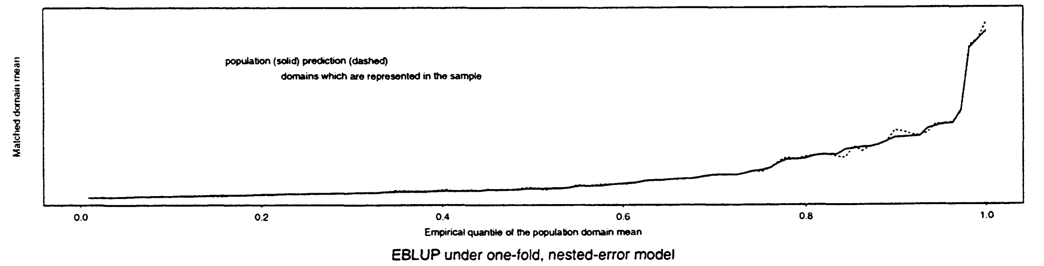


Fixing the LINARE at $r = 1$, we compare it with the nested-error model, first, for all the 109 domains which were represented in the sample — also shown is the direct within-domain ratio predictor.

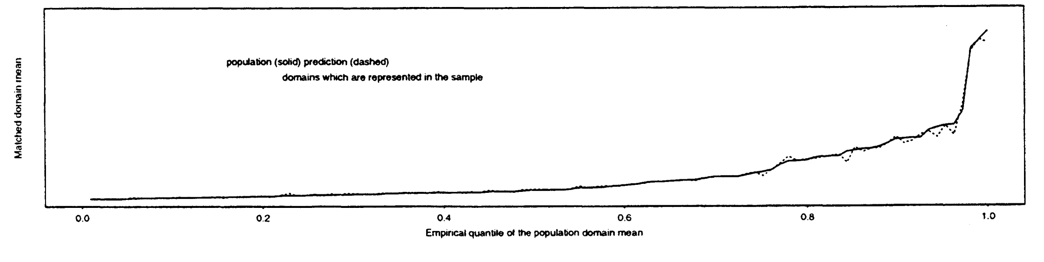
Direct within-domain ratio predictor



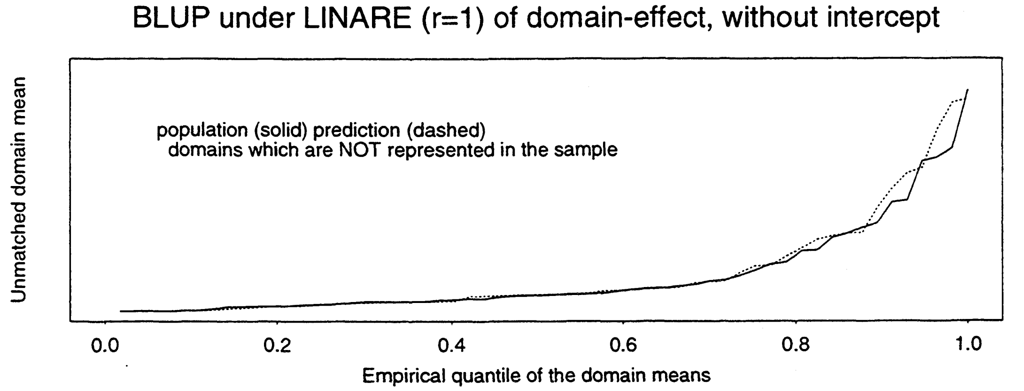
BLUP under LINARE (r=1) of domain-effect, without intercept

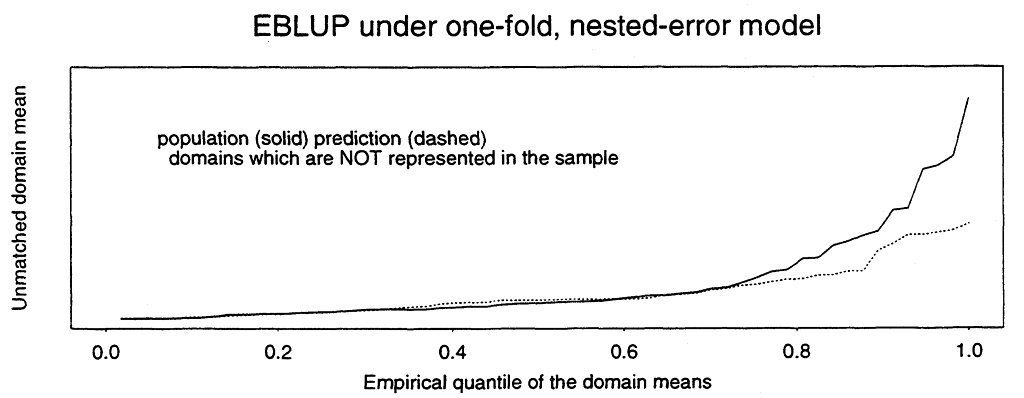


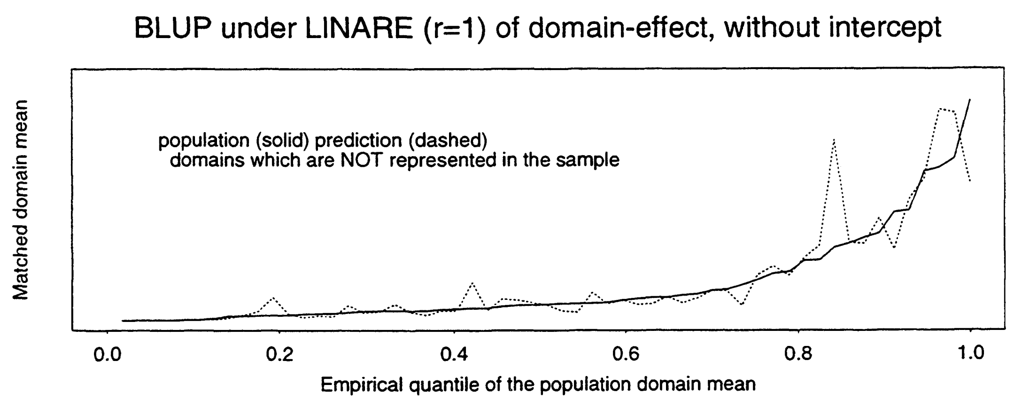
EBLUP under one-fold, nested-error model

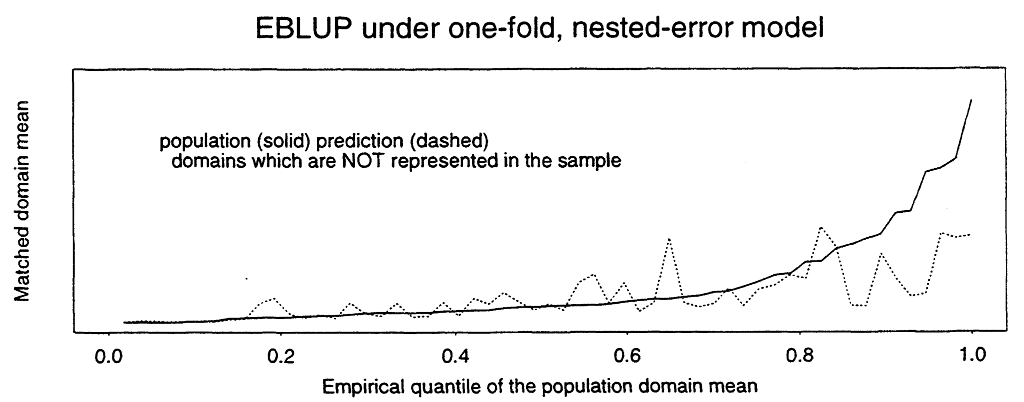


We next compare the two methods for the 57 domians which were not represented in the sample.





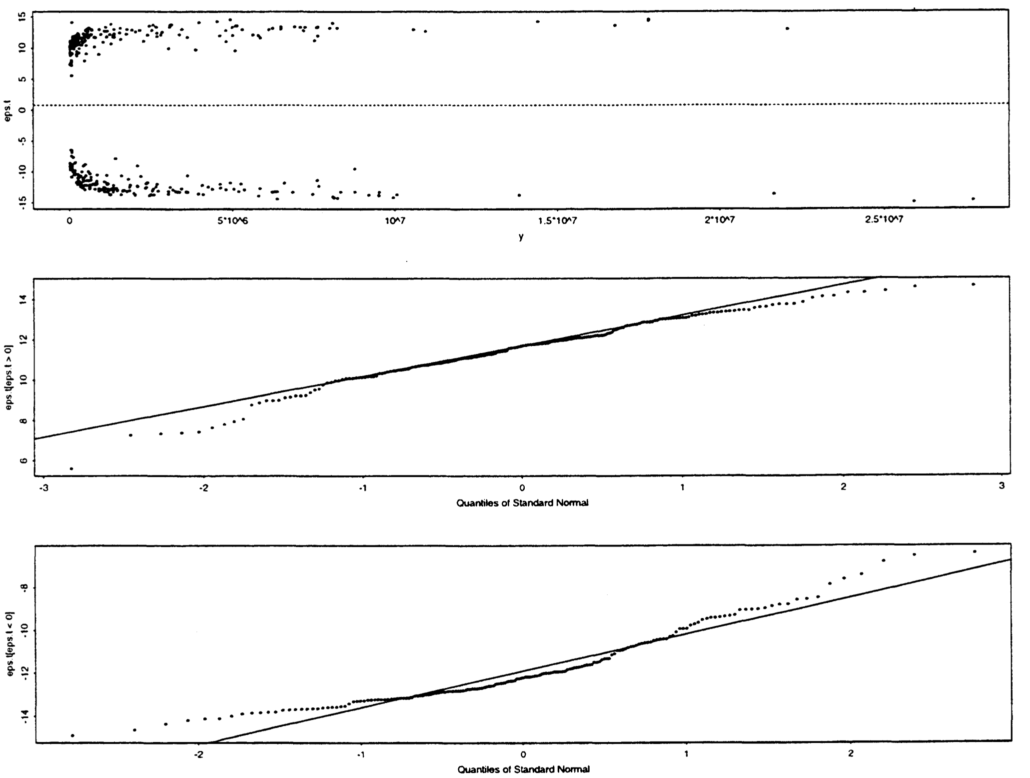




The conclusion seems clear: both the LINARE and the nested-error models account adequately for the between-domain variation for the domains which were represented in the sample; however the nested-error model is unsuitable for those which were not represented in the sample, as compared to the LINARE.

Remark *It is interesting to notice that the variance inflating constant has very little effect on the result. Now a key condition for the OLS estimator to be consistent is the constant variance of the residual. (Appropriate variance specification affects mostly the efficiency of the estimator.) Often this is questioned if one observes a fan-shaped plot of the estimated residuals against the predicted, or true, values of the response variable, which is easily confirmed by a further qq-plot against the standard Normal distribution. However, while the qq-plot requires normality of the residual, the fan-shapeness does not. Situations may arise in which a simple remodelling of the residuals, i.e. without the normality assumption, may reveal that the fan-shapeness is not strong enough to have much effect on the inference. In other words, approximate constancy of variance is maintained under some other non-normal distribution of the residual.*

Illustration Having fitted the LINARE with $r = 0$, i.e. constant variance of the residual, we transformed the estimated residuals to the log-scale, separately for the positive and the nagative ones. We then used the qq-plot to check separately the positive and the negative transformed residuals.



The approximate normality in both cases suggest that the residuals might be considered to approximately follow a mixed log-normal distribution, independent of the size of the auxiliary variable, under which the approximate constancy of the residual remains valid, which could be why the OLS estimator works! (The mean of the transformed residuals, i.e. the horisontal dashed line in the first plot, suggests symmetry of the residuals around zero.)

SMAREST (III): Models for binary survey variables

1 Apply post-stratification to the Labour Force Survey (LFS) based on auxiliary information of (i) Register-Status, (ii) Sex and (iii) Age. Let the LFS-Employment within each municipality be the interest of the survey.

1.1 Let h, where $1 \leq h \leq H$, be the post-stratum index. Let a, where $1 \leq a \leq A$, be the municipality index. Let Y_{ah} be the total LFS-Employment within munipicality a and post-stratum h. Let $p_{ah} = Y_{ah}/N_{ah}$ be the corresponding LFS-Employment Rate, where N_{ah} is the corresponding sub-population total. Similarly, let y_{ah} be the observed LFS-Employment in the corresponding sub-sample. Let $\bar{y}_{ah} = y_{ah}/n_{ah}$ provided $n_{ah} > 0$ and $\bar{y}_{ah} = 0$ otherwise, where n_{ah} is the size of the sub-sample.

1.2 The *complete group-mean model* allows $\{p_{ah}\}$, i.e. $A \times H$ of them, to be entirely free, i.e.

$$p_{ah} = p_h + e_a \qquad \text{where} \quad p_h = (\sum_a N_{ah} p_{ah})/(\sum_a N_{ah}). \tag{1}$$

2 While the synthetic estimator assumes that the post-stratum mean has null variation across the municipalities, empirical Bayes (EB) and generalized linear mixed models (GLMM) allow for random area-effects. Whereas EB leads towards a constrained approach (Spjøtvoll and Thomsen, 1987), estimation under the GLMM is based on the penalized quasi-likelihood (Green, 1987).

2.1 The synthetic, or post-stratified, estimator can be based on the following *synthetic model*, i.e.

$$p_{ah} = p_h \qquad \Leftrightarrow \qquad e_a = 0. \tag{2}$$

Illustration Observed within-post-stratum LFS-Employment Rate $\bar{y}_h = (\sum_a n_{ah} \bar{y}_{ah})/(\sum_a n_{ah})$.

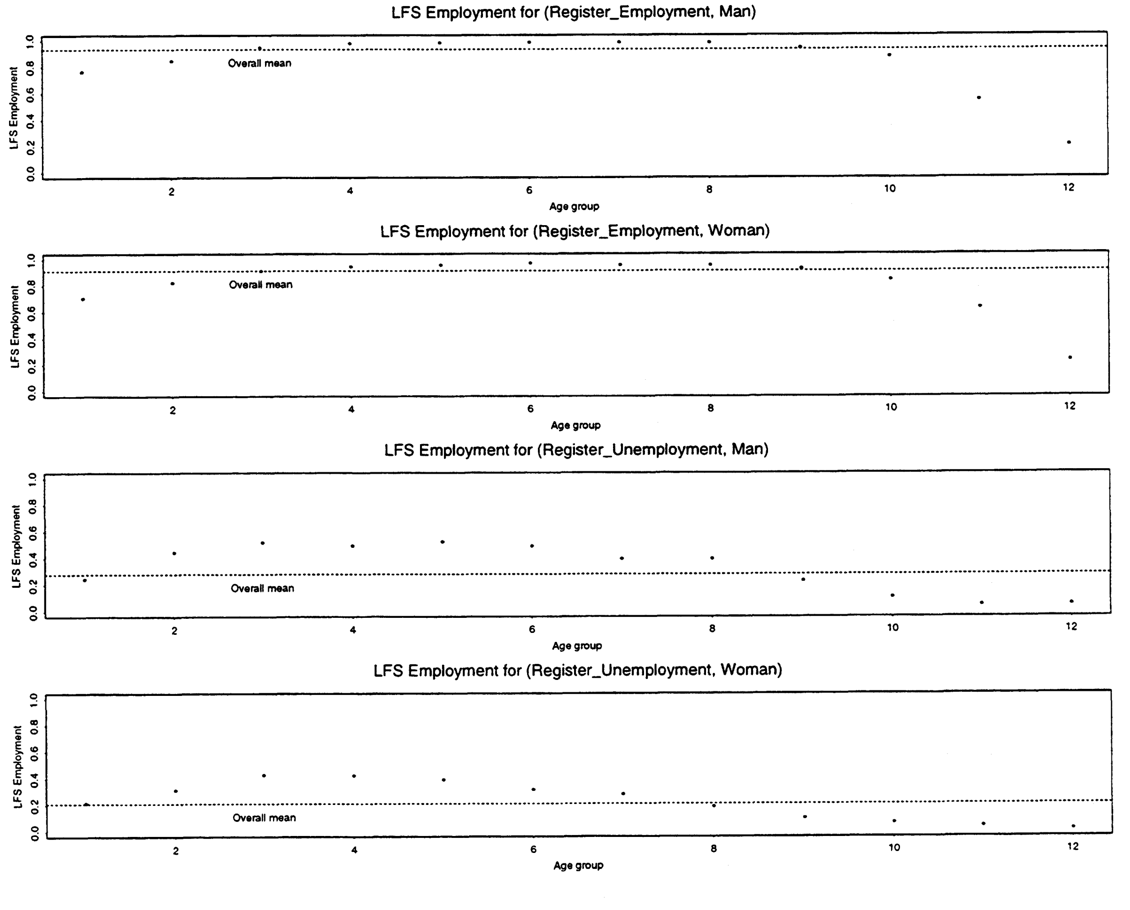
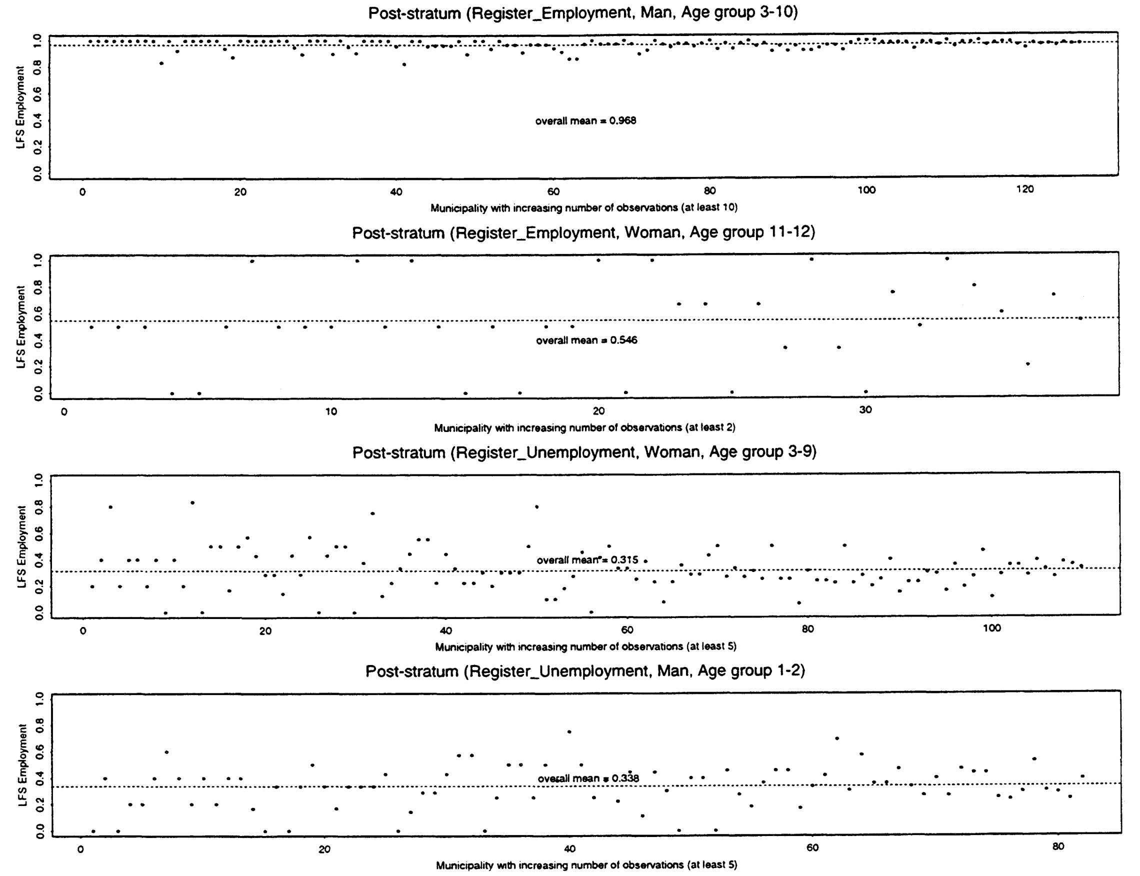


Illustration Observed variation over municipalities within some collapsed post-strata.



Remark *Notice that different post-strata have different dispersion.*

2.2 The *EB model* allows for random area-effect with zero mean and a constant variance *a priori*, i.e.

$$p_{ah} = p_h + e_{ah} \qquad \text{where} \quad E[e_{ah}] = 0 \quad \text{and} \quad Var(e_{ah}) = \sigma_h^2. \tag{3}$$

In other words, we allow σ_h^2 to vary over different post-strata.

2.2.1 Spjøtvoll and Thomsen (1987) considered estimators in the form of a linear adjustment of $\bar{y}_{ah}$. In particular, the so-called *optimal estimator*, which minimizes $E[(\hat{p}_{ah} - p_{ah})^2]$, is given as

$$\hat{p}_{ah} = w_{ah}\bar{y}_{ah} + (1 - w_{ah})p_h \qquad \text{where} \quad w_{ah} = n_{ah}\sigma_h^2/[(n_{ah} - 1)\sigma_h^2 + p_h(1 - p_h)].$$

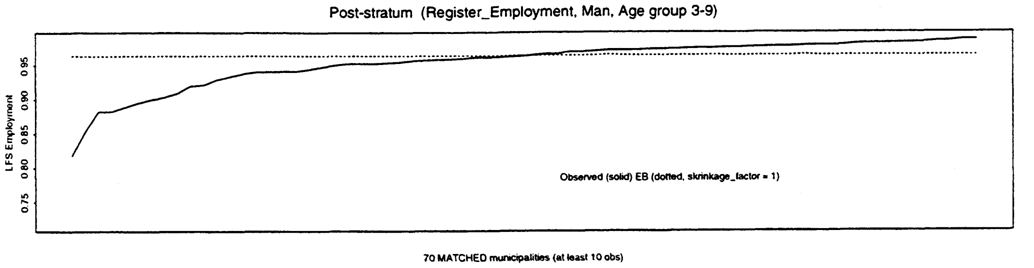
This in general results into *over-shrinkage* since $E[\sum_a(\hat{p}_{ah} - p_{ah})^2/A] \leq \sigma_h^2$ — see Morris (1983) and Ghosh (1992) for similar/general results. The *constrained EB estimator* is such that (i) $E[\hat{p}_{ah}] = p_h$, and (ii) $Var(\hat{p}_{ah}) = \sigma_h^2$. Within the class of linear estimators, this corresponds to choosing $w_{ah}^{1/2}$ and $1 - w_{ah}^{1/2}$ as the respective weights of $\bar{y}_{ah}$ and p_h.

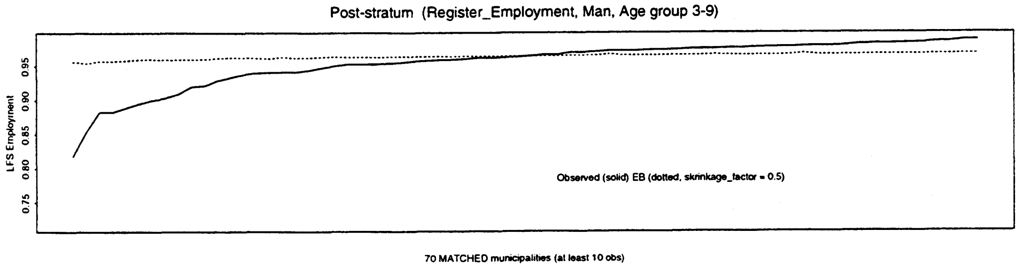
2.2.2 Often the variance *a priori* is unknown and needs to be estimated (Spjøtvoll and Thomsen, 1987). In general, however, this may again cause over-shrinkage of the constrained estimator. A general *shrinkage composite estimator* motivated by the EB approach can therefore be defined as

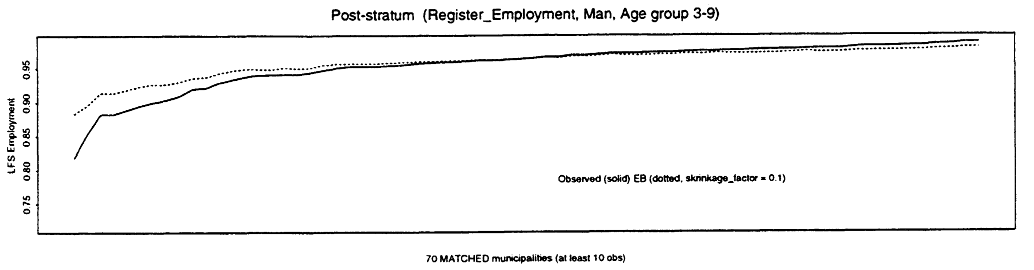
$$\hat{p}_{ah} = w_{ah}^{\gamma}\bar{y}_{ah} + (1 - w_{ah}^{\gamma})p_h \qquad \text{where} \quad w_{ah} = n_{ah}\sigma_h^2/[(n_{ah} - 1)\sigma_h^2 + p_h(1 - p_h)],$$

and γ, for $0 \leq \gamma \leq 1$, is the *shrinkage factor* which balances between $\bar{y}_{ah}$ and p_h. In particular, we recover the optimal weighting at $\gamma = 1$, and the constrained EB at $\gamma = 1/2$.

Illustration Over-shrinkage of the skrinkage composite estimator under the EB approach.







Remark *It is quite clear that the constrained EB, with shrinkage factor* $\gamma = 0.5$, *did not successfully adjust for over-shrinkage. We believe that this is because* σ_h^2 *was estimated based on the sample. Further investigation should therefore consider using known variance* σ_h^2 a priori, *for instance, based on earlier Census results.*

2.3 *Generalized linear mixed models (GLMM)* (Breslow and Clayton, 1993) are useful for accommodating overdispersion in binomial data — a general discussion of hierarchical generalized linear model can be found in Lee and Nelder (1996). Under the present setting, we have

$$\text{logit } p_{ah} = \log p_{ah} - \log(1 - p_{ah}) = \xi_h + e_{ah} \qquad \text{where} \quad E[e_{ah}] = 0 \quad \text{and} \quad Var(e_{ah}) = \sigma_h^2. \qquad (4)$$

2.3.1 Without restrictions on e_{ah}, the maximum likelihood estimator (mle) results into over-fitting. A penalized (quasi) log-likelihood (Green, 1987) can be defined here as

$$l_e(\xi_h, \sigma_h^2; y) = [\sum_a y_{ah}(\xi_h + e_{ah}) + n_{ah}\log(1 - p_{ah})] - \frac{1}{2}\sum_a e_{ah}^2,$$

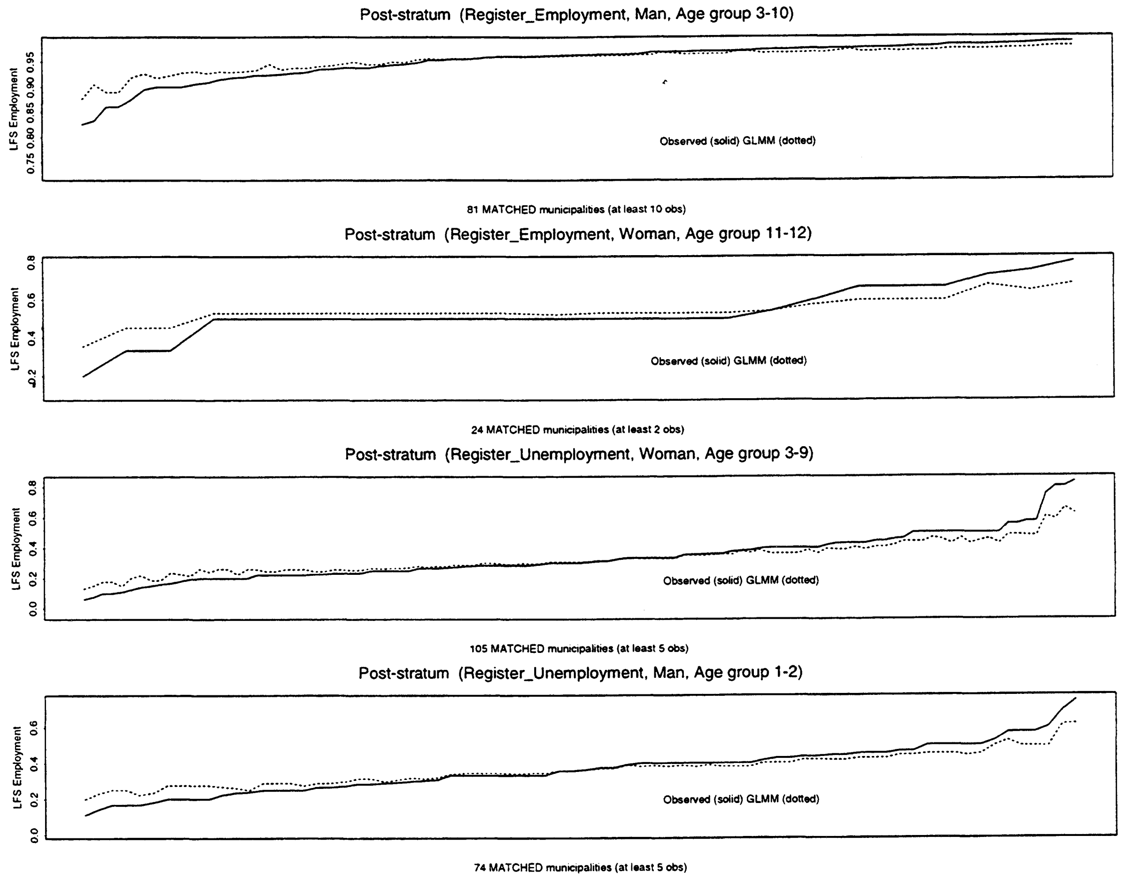
where the last term on the right-hand side is the penalty to be paid for non-zero random effects.

2.3.2 Estimation based on l_e employs Fisher scoring: For the present values of ξ_h and e_{ah}, calculate p_{ah} according to (4), and let $\eta_{ah} = y_{ah} - n_{ah}p_{ah}$ and $v_{ah} = n_{ah}p_{ah}(1 - p_{ah})$, and

$$u = \begin{pmatrix} \sum_a \eta_{ah} \\ \eta_{1h} - e_{1h} \\ \vdots \\ \eta_{Ah} - e_{Ah} \end{pmatrix} \quad \text{and} \quad j = \begin{pmatrix} \sum_a v_{ah} & v_{1h} & v_{2h} & \cdots & v_{Ah} \\ v_{1h} & 1 + v_{1h} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & 0 \\ v_{Ah} & 0 & \cdots & \cdots & 1 + v_{Ah} \end{pmatrix}.$$

Update $\theta = (\xi_h, e_{1h}, ..., e_{Ah})^T$ as $\theta + j^{-1}u$, and iterate till convergence, upon which estimate σ_h^2 based on the estimated e_{ah}.

Illustration Fitting GLMM to some collapsed post-strata.



3 The *random area-effect model* of the GLMM-type can be defined at the level of municipality. A linear structure can be introduced to accommodate area-specific post-stratification, which allows for better data usage. A similar linear structure can be introduced for the random area-effect to account for difference between the sample and population configurations.

3.1 Let p_a be the LFS-Employment Rate of municipality a, with auxiliary vector X_a which, typically, arises from calibrating post-stratification (Zhang, 1998). Based on the sample, we define, for $\bar{x}_a = x_a/n_a$,

$$\text{logit } p_a = \bar{x}_a^T \xi + e_a \qquad \text{where} \quad E[e_a] = 0 \quad \text{and} \quad Var(e_a) = \sigma^2. \tag{5}$$

In this way, within each municipality, there is only one random effect. Notice that, the usual GLM is obtained from setting $e_a \equiv 0$. Estimation is, as under(4), based on the penalized quasi log-likelihood, defined as

$$l_e(\xi, \sigma^2; y) = [\sum_a y_a(\bar{x}_a^T \xi + e_a) + n_a \log(1 - p_a)] - \frac{1}{2} \sum_a e_a^2.$$

Let $\eta_a = y_a - n_a p_a$ and $v_a = n_a p_a(1 - p_a)$. Let $e = (e_1, ..., e_A)^T$, and $\eta = (\eta_1, ..., \eta_A)^T$, and V the diagonal matrix with v_a as the ath element on the diagonal. Let $B = (b_{aj})$ be the $A \times q$ design matrix whose ath row is given by $\bar{x}_a^T$, and

$$u = \begin{pmatrix} B^T \eta \\ \eta - e \end{pmatrix} \quad \text{and} \quad j = \begin{pmatrix} B^T V B & B^T V \\ VB & I + V \end{pmatrix},$$

where I is the $A \times A$ identity matrix. Update $\theta = (\xi_1, ..., \xi_q, e_1, ..., e_A)^T$ as $\theta + j^{-1}u$, and iterate till convergence, upon which estimate σ^2 based on the estimated e_a.

Illustration Fitting GLM and GLMM with random area-effect to random sub-sample of the LFS (I).

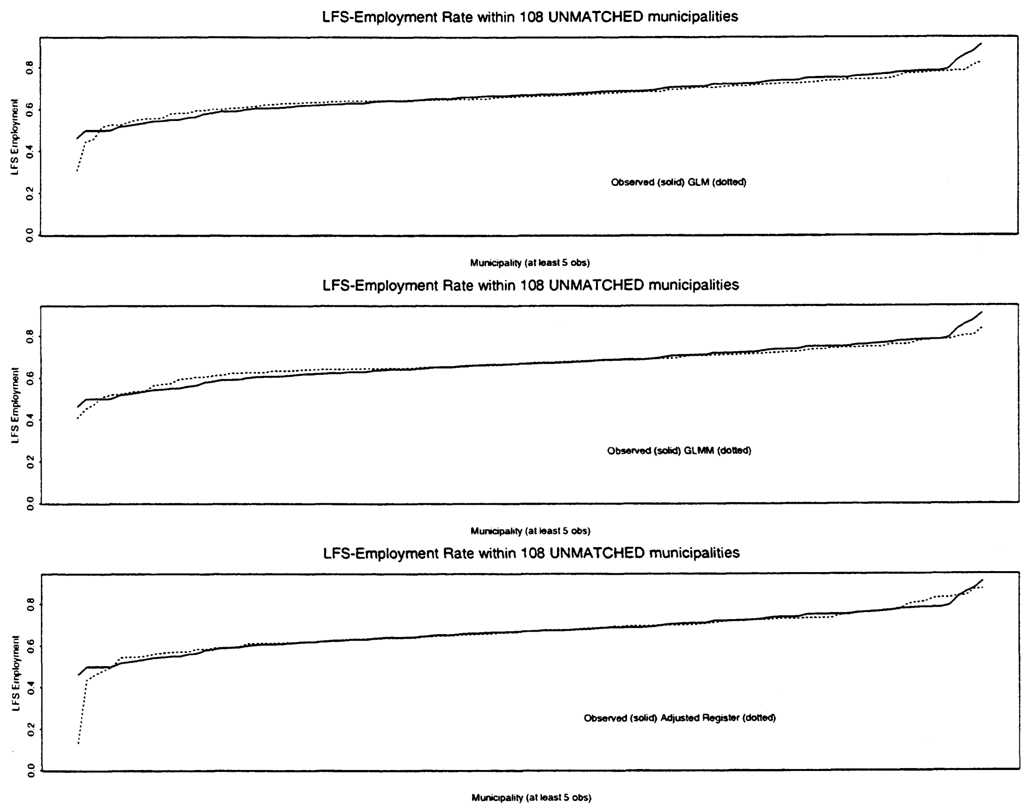


Illustration Fitting GLM and GLMM with random area-effect to random sub-sample of the LFS (II).

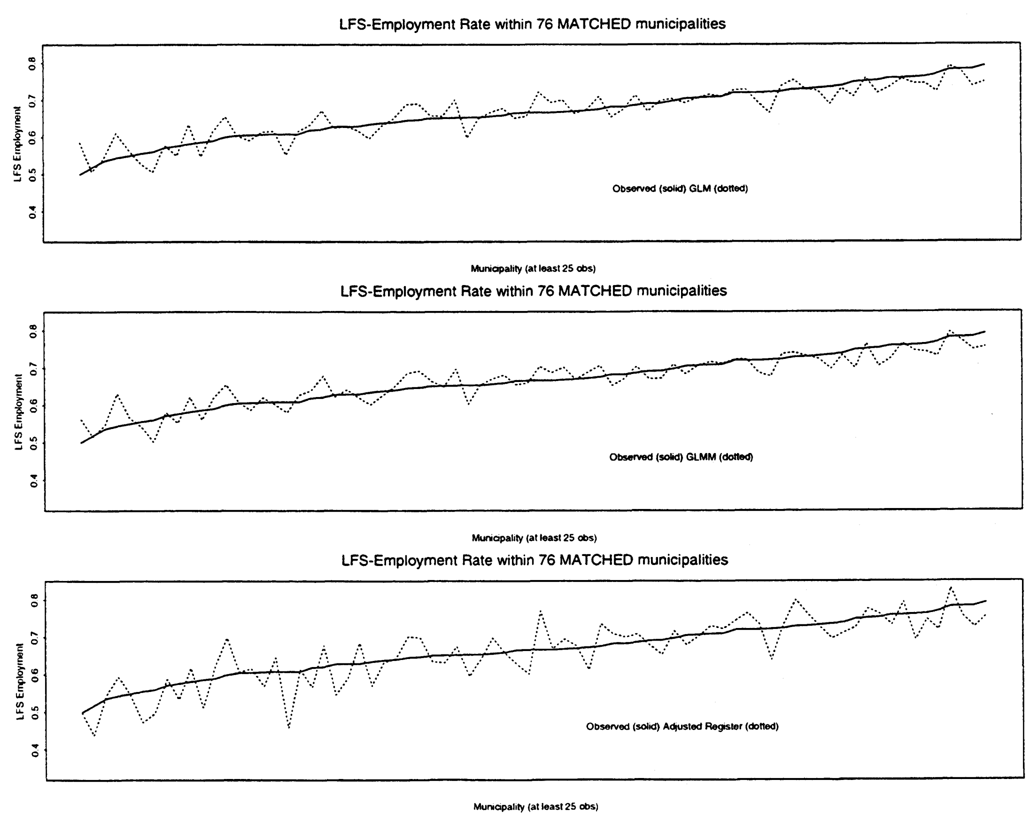
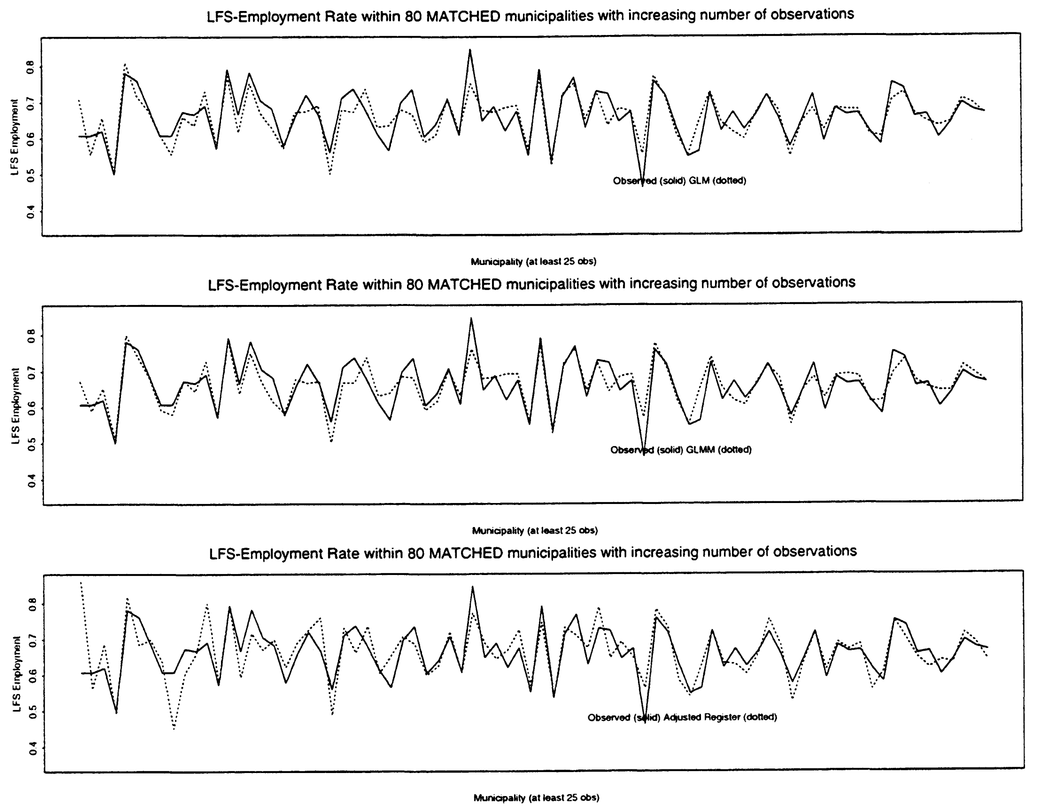


Illustration Fitting GLM and GLMM with random area-effect to random sub-sample of the LFS (III).



Remark *The Adjusted Register was obtained by multiplying the Register-Employment Rate, denoted by* $\bar{z}_a$*, with a factor so that the average Rate across the municaplities is made the same as observed in the chosen sub-sample, i.e.* $\sum_a n_a \bar{z}_a = \sum_a n_a \bar{y}_a$.

Remark *Model* (5) *can be useful w.r.t. overdispersion observed in the sample. However, it is likely inappropriate to apply the estimated* e_a *as such to the rest population, due to the difference between* $\bar{x}_a$ *and* $\bar{X}_a$*. As it was shown earlier, different post-strata have different degrees of dispersion.*

3.2 To account for the fact that different post-strata have different dispersion across the municipalities, we can introduce a linear structure to the vector of random effect of the same dimension as that of ξ, i.e.

$$\text{logit } p_a = \bar{x}_a^T \xi + \bar{x}_a^T \epsilon_a \qquad \text{where} \quad E[\epsilon_a] = 0 \quad \text{and} \quad Cov(\epsilon_{aj}, \epsilon_{ak}) = \delta_{jk} \sigma_j^2, \tag{6}$$

and $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ otherwise. The penalized quasi log-likelihood can be defined as

$$l_\epsilon(\xi, \sigma_k^2; y) = [\sum_a y_a(\bar{x}_a^T \xi + \bar{x}_a^T \epsilon_a) + n_a \log(1 - p_a)] - \frac{1}{2} \sum_a \epsilon_a^T \epsilon_a.$$

Fisher scoring can be formulated similarly as before, though the design matrix has become much larger now. Let η, V and B be defined as under (5), we have

$$u = \begin{pmatrix} B^t \eta \\ \bar{x}_1 \eta_1 - \epsilon_1 \\ \vdots \\ \bar{x}_A \eta_A - \epsilon_A \end{pmatrix} \qquad \text{and} \qquad j = \begin{pmatrix} B^T V B & v_1 \bar{x}_1 \bar{x}_1^T & \cdots & v_A \bar{x}_A \bar{x}_A^T \\ v_1 \bar{x}_1 \bar{x}_1^T & I + v_1 \bar{x}_1 \bar{x}_1^T & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ v_A \bar{x}_A \bar{x}_A^T & 0 & 0 & I + v_A \bar{x}_A \bar{x}_A^T \end{pmatrix}.$$

Notice that deletion of the zero-components in $\bar{x}_a$ is necessary for u and, likewise, deletion of the all-zero rows and columns in $\bar{x}_a \bar{x}_a^T$ is necessary for j.

References

Ansen, H., Hallen, S.-A., and Ylander, H. (1988). Statistics for regional and local planning in Sweden. *J. Off. Statist.*, **4**, 35–46.

Battese, G.A., Harter, R.M., and Fuller, W.A. (1988). A error-component model for prediction of county crop areas using survey and satelite data. *J. Amer. Statist. Assoc.*, **83**, 28–36.

Brackstone, G.J. (1987). Small area data: policy issues and technical challenges. In *Small Area Statistics*, ed. by R. Platek, J. Rao, C.-E. Särndal, and M. Singh, pp. 3–20. Wiley, New York.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.*, **88**, 9–25.

Chambers, R.L. and Feeney, G.A. (1977). Log linear models for small area estimation. *Unpublished paper, Australian Bureau of Statistics.*

Chaudhuri, A. (1992). Small domain statistics: A review. Technical Report ASC/92/2, Indian Statistical Institute, Calcutta.

Cressie, N. (1989). Empirical Bayes estimation of undercount in the decennial census. *J. Amer. Statist. Assoc.*, **84**, 1033–44.

Cressie, N. (1992). REML estimation in empirical Bayes smoothing of census undercount. *Survey Methodology*, **18**, 75–94.

Datta, G.S: and Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *Ann. Statist.*, **19**, 1748–70.

Decaudin, G. and Labat, J.-C. (1997). A synthetic, robust and efficient method of making small area population estimates in France. *Survey Methodology*, **23**, 91–8.

DeSouza, C.M. (1992). An approximate bivariate Bayesian method for analyzing small frequencies. *Biometrics*, **48**, 1113–30.

Drew, D., Singh, M.P., and Choudhry, G.H. (1982). Evaluation of small area estimation techniques for the Canadian Labour Force Survey. *Survey Methodology*, **8**, 17–47.

Ericksen, E.P. (1974). A regression method for estimating populations of local areas. *J. Amer. Statist. Assoc.*, **69**, 867–75.

Ericksen, E.P. and Kadane, J.B. (1985). Estimating the population in a census year (with discussion). *J. Amer. Statist. Assoc.*, **80**, 98–131.

Ericksen, E.P. and Kadane, J.B. (1987). Sensitivity analysis of local estimates of undercount in the 1980 U.S. Census. In *Small Area Statistics*, ed. by R. Platek, J. Rao, C.-E. Särndal, and M. Singh, pp. 23–45. Wiley, New York.

Ericksen, E.P., Kadane, J.B., and Tukey, J.W. (1989). Adjusting the 1981 census of population and housing. *J. Amer. Statist. Assoc.*, **84**, 927–44.

Falorsi, P.D., Falorsi, S., and Russo, A. (1994). Empirical comparison of small area estimation methods for the Italian Labour Force Survey. *Survey Methodology*, **20**, 171–76.

Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269–77.

Freedman, D.A. and Navidi, W.C. (1986). Regression models for adjusting the 1980 Census (with discussion). *Statistical Science*, **1**, 1–39.

Freedman, D.A. and Navidi, W.C. (1992). Should we have adjusted the U.S. Census of 1980? (with discussion). *Survey Methodology*, **18**, 3–74.

Freeman, D.H. and Koch, G.G. (1976). An asymptotic covariance structure for testing hypotheses on raked contingency tables from complex sample surveys. In *Proccedings of the Social Statistics Section*, pp. 330–5. Amer. Statist. Assoc., Washington, DC.

Fuller, W.A. and Battese, G.E. (1973). Transformations for estimation of linear models with nested-error structure. *J. Amer. Statist. Assoc.*, **68**, 626–32.

Fuller, W.A. and Harter, R.M. (1987). The multivariate components of variance model for small area estimation. In *Small Area Statistics*, ed. by R. Platek, J. Rao, C.-E. Särndal, and M. Singh, pp. 103–23. Wiley, New York.

Ghosh, M. (1992). Constrained Bayes estimation with applications. *J. Amer. Statist. Assoc.*, **87**, 533–40.

Ghosh, M. and Larihi, P. (1987). Robust empirical Bayes estimation of means from stratified samples. *J. Amer. Statist. Assoc.*, **82**, 1153–62.

Ghosh, M., Natarajan, K., Stroud, T.W.F., and Carlin, B.P. (1998). Generalized linear models for small-area estimation. *J. Amer. Statist. Assoc.*, **93**, 273–82.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal (with discussion). *Statistical Science*, **9**, 55–93.

Gonzalez, M.E. (1973). Use and evaluation of synthetic estimators. In *Proceedings of the Social Statistics Section*, pp. 33–6. Amer. Statist. Assoc., Washington, DC.

Green, P.J. (1987). Penalized likelihood for general semi-parametric regression models. *Int. Statist. Rev.*, **55**, 245–59.

Harville, D.A. (1977). Maximum likelihood approaches to Variance Component estimation and to related problems. *J. Amer. Statist. Assoc.*, **72**, 320–38.

Heldal, J., Swensen, A.R., and Thomsen, I. (1987). Census statistics through combined use of surveys and registers. *Statistical Journal of the United Nations ECE*, **5**, 43–51.

Holt, D. and Moura, F. (1993). Mixed models for making small area estimates. In *Small Area Statistics and Survey Designs*, ed. by G. Kalton, J. Kordos, and R. Platek, vol. **1**, pp. 221–31. Central Statistical Office, Warsaw.

Holt, D. and Smith, T.M.F. (1979). Post stratification. *J. Roy. Statist. Soc. A*, **142**, 33–46.

Isaki, C.T., Schultz, L.K., Smith, P.J., and Diffendal, D.J. (1987). Small area etimation research for census undercount - progress report. In *Small Area Statistics*, ed. by R. Platek, J. Rao, C.-E. Särndal, and M. Singh, pp. 219–38. Wiley, New York.

Kalton, G., Kordos, J., and Platek, R. (1993). *Small Area Statistics and Survey Designs Vol I: Invited Papers; Vol II: Contributed Papers and Panel Discussion*. Central Statisticcal Office, Warsaw.

Kass, R.E. and Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical modles (parametric empirical Bayes models). *J. Amer. Statist. Assoc.*, **84**, 717–26.

Laake, P. (1978). An evaluation of synthetic estimates of employment. *Scand. J. Statist*, **5**, 57–60.

Laake, P. (1979). A predictive approach to subdomain estimation in finite populations. *J. Amer. Statist. Assoc.*, **74**, 355–8.

Laird, N.M. and Louis, T.A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *J. Amer. Statist. Assoc.*, **82**, 739–50.

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models (with discussion). *J. Roy. Statist. Soc. B*, **58**, 619–78.

Louis, T. (1984). Estimating a population of parameter values using bayes and empirical Bayes methods. *J. Amer. Statist. Assoc.*, **79**, 393–8.

Lui, K.-J. and Cumberland, W.G. (1991). A model-based approach: composite estimators for small area estimation. *J. Off. Statist.*, **7**, 69–76.

Lundstrøm, S. (1987). An evaluation of small area estimation methods: The case of estimating the number of nonmarried cohabiting persons in Swedish munucipalities. In *Small Area Statistics*, ed. by R. Platek, J. Rao, C.-E. Särndal, and M. Singh, pp. 239–56. Wiley, New York.

Marker, D.A. (1983). Organization of small area estimators. In *Proceedings of Survey Research Methods Section*, pp. 409–14. Amer. Statist. Assoc., Washington, DC.

Morris, C. (1983). Parametric empirical Bayes inference: Theory and applications (with discussion). *J. Amer. Statist. Assoc.*, **78**, 47–65.

Morrison, P. (1971). Demographic information for cities: a manual for estimating and projecting local population characteristics. RAND report R-618-HUD.

Nordbotten, S. (1996). Neural Network imputation applied to the Norwegian 1990 population census data. *J. Off. Statist.*, **12**, 385–401.

Platek, R., Rao, J.N.K., Särndal, C.-E., and Singh, M.P. (1987). *Small Area Statistics.* Wiley, New York.

Platek, R. and Singh, M.P. (1986). *Small Area Statistics: Contributed Papers.* Laboratory for Research in Statistics and Probability, Carleten Univ.

Prasad, N.G. and Rao, J.N.K. (1990). The estimation of mean square errors of small area estimators. *J. Amer. Statist. Assoc.*, **85**, 163–71.

Purcell, N.J. (1979). Efficient Small Domain Estimation: A categorical data analysis approach. Ph.D. thesis, Univ. of Michigan, Ann Arbor, Michigan.

Purcell, N.J. and Kish, L. (1979). Estimation for small domain. *Biometrics*, **35**, 365–84.

Purcell, N.J. and Kish, L. (1980). Postcensal estimates for local areas (or domains). *Int. Statist. Rev.*, **48**, 3–18.

Raghunathan, T.E. (1993). A quasi-empirical Bayes method for small area estimation. *J. Amer. Statist. Assoc.*, **88**, 1444–8.

Rao, J.N.K (1986). Synthetic estimators, SPREE and best model based predictors. In *Proceedings of the Conference on Survey Research Methods in Agriculture*, pp. 1–16. U.S. Dept. Agriculture, Washington, DC.

Robinson, G.K. (1991). That BLUP is a good thing: the estimation of random effects (with discussion). *Statistical Science*, **6**, 15–51.

Särndal, C.-E. and Hidiroglou, M.A. (1989). Small domain estimation: a conditional analysis. *J. Amer. Statist. Assoc.*, **84**, 266–75.

Schaible, W.L. (1978). Choosing weights for composite estimators for small area statistics. In *Proceedings of Survey Research Methods Section*, pp. 741–6. Amer. Statst. Assoc., Washington DC.

Singh, M.P., Gambino, J., and Mantel, H. (1993). Issues and options in the provision of small area data. In *Small Area Statistics and Survey Designs*, ed. by G. Kalton, J. Kordos, and R. Platek, vol. **1**, pp. 37–75. Central Statistical Office, Warsaw.

Small Area Estimation Research Team (1983). A bibliography for small area estimation. *Survey Methodology*, **9**, 241–61.

Spjøtvoll, E. and Thomsen, I. (1987). Application of some empirical Bayes methods to small area statistics. *Bulletin of the International Statistical Institute*, **2**, 435–49.

Statistics Canada (1987). *Population Estimation Methods, Canada*. Catalogue 91-528E. Statistics Canada, Ottawa.

Stukel, D. (1991). *Small Area Estimation Under One and Two-fold Nested Error Regression Model*. Ph.D. thesis, Carleton Univ.

Zhang, L.-C. (1998). Post-stratification and calibration — A synthesis. *Statistics Norway (Discussion paper No. 216)*.

Zidek, J.V. (1982). A review of methods for estimating the populations of local areas. Technical Report 82-4, Univ. British Columbia, Vancouver.

Recent publications in the series Documents

97/9 H. Berby and Y. Bergstrøm: Development of a Demonstration Data Base for Business Register Management. An Example of a Statistical Business Register According to the Regulation and Recommendations of the European Union

97/10 E. Holmøy: Is there Something Rotten in this State of Benchmark? A Note on the Ability of Numerical Models to Capture Welfare Effects due to Existing Tax Wedges

97/11 S. Blom: Residential Consentration among Immigrants in Oslo

97/12 Ø. Hagen and H.K. Østereng: Inter-Baltic Working Group Meeting in Bodø 3-6 August 1997 Foreign Trade Statistics

97/13 B. Bye and E. Holmøy: Household Behaviour in the MSG-6 Model

97/14 E. Berg, E. Canon and Y. Smeers: Modelling Strategic Investment in the European Natural Gas Market

97/15 A. Bråten: Data Editing with Artificial Neural Networks

98/1 A. Laihonen, I. Thomsen, E. Vassenden and B. Laberg: Final Report from the Development Project in the EEA: Reducing Costs of Censuses through use of Administrative Registers

98/2 F. Brunvoll: A Review of the Report "Environment Statistics in China"

98/3: S. Holtskog: Residential Consumption of Bioenergy in China. A Literature Study

98/4 B.K. Wold: Supply Response in a Gender-Perspective, The Case of Structural Adjustments in Zambia. Tecnical Appendices

98/5 J. Epland: Towards a register-based income statistics. The construction of the Norwegian Income Register

98/6 R. Chodhury: The Selection Model of Saudi Arabia. Revised Version 1998

98/7 A.B. Dahle, J. Thomasen and H.K. Østereng (eds.): The Mirror Statistics Exercise between the Nordic Countries 1995

98/8 H. Berby: A Demonstration Data Base for Business Register Management. A data base covering Statistical Units according to the Regulation of the European Union and Units of Administrative Registers

98/9 R. Kjeldstad: Single Parents in the Norwegian Labour Market. A changing Scene?

98/10 H. Brüngger and S. Longva: International Principles Governing Official Statistics at the National Level: are they Relevant for the Statistical Work of International Organisations as well?

98/11 H.V. Sæbø and S. Longva: Guidelines for Statistical Metadata on the Internet

98/12 M. Rønsen: Fertility and Public Policies - Evidence from Norway and Finland

98/13 A. Bråten and T. L. Andersen: The Consumer Price Index of Mozambique. An analysis of current methodology – proposals for a new one. A short-term mission 16 April - 7 May 1998

98/14 S. Holtskog: Energy Use and Emmissions to Air in China: A Comparative Literature Study

98/15 J.K. Dagsvik: Probabilistic Models for Qualitative Choice Behavior: An introduction

98/16 H.M. Edvardsen: Norwegian Regional Accounts 1993: Results and methods

98/17 S. Glomsrød: Integrated Environmental-Economic Model of China: A paper for initial discussion

98/18 H.V. Sæbø and L. Rogstad: Dissemination of Statistics on Maps

98/19 N. Keilman and P.D. Quang: Predictive Intervals for Age-Specific Fertility

98/20 K.A. Brekke (Coauthor on appendix: Jon Gjerde): Hicksian Income from Stochastic Resource Rents

98/21 K.A.Brekke and Jon Gjerde: Optimal Environmental Preservation with Stochastic Environmental Benefits and Irreversible Extraction

99/1 E. Holmøy, B. Strøm and T. Åvitsland: Empirical characteristics of a static version of the MSG-6 model

99/2 K. Rypdal and B. Tornsjø: Testing the NOSE Manual for Industrial Discharges to Water in Norway

99/3 K. Rypdal: Nomenclature for Solvent Production and Use

99/4 K. Rypdal and B. Tornsjø: Construction of Environmental Pressure Information System (EPIS) for the Norwegian Offshore Oil and Gas Production

99/5 M. Søberg: Experimental Economics and the US Tradable SO_2 Permit Scheme: A Discussion of Parallelism

99/6 J. Epland: Longitudinal non-response: Evidence from the Norwegian Income Panel

Documents

PORTO BETALT
VED
INNLEVERINGA
P.P.
NORGE/NOREG

Tillatelse nr.
159 000/502

B

Returadresse:
Statistisk sentralbyrå
Postboks 8131 Dep.
N-0033 Oslo

Statistics Norway
P.O.B. 8131 Dep.
N-0033 Oslo

Tel: +47-22 86 45 00
Fax: +47-22 86 49 73

ISSN 0805-9411

Statistisk sentralbyrå
Statistics Norway