HIERARCHICAL BAYESIAN MODELS FOR THE ESTIMATION OF UNEMPLOYMENT RATES IN SMALL DOMAINS OF THE ITALIAN LABOUR FORCE SURVEY

Enrico Fabrizi

1. INTRODUCTION

The Italian National Institute of Statistics (ISTAT) publishes quarterly estimates of key Labour Market aggregates on the basis of a repeated sample survey. Estimates are published for the whole nation, for regions, for provinces and by sex and age groups within territorial domains. Provinces are sub-regional administrative units, about 100 in Italy and from 5 to 10 in most of the regions. For the smallest domains (province and sub-province populations) the published estimates are characterized by very large variances. The aim of this work is to obtain shrinkage estimators with smaller variances for aggregates related to these domains.

The repetitive nature of the Survey suggests the proposal of models based on time series. In particular, we propose exchangeable small area models in which the process of borrowing information is based both on "neighbouring areas" from the same repetition of the survey and on previous ones. This may be very useful as past estimates can provide valuable information on the current value of the parameter of interest. Moreover, time series allows us both to estimate province specific intercepts measuring the peculiarities of local economies and to take account for cyclic trends, year effects and seasonality.

For building these small area models we do not dispose of any auxiliary information apart from the demographic statistics already employed in ISTAT estimators, since no independent source of information (for instance from Unemployment Insurance systems records) is available for small domains of the Italian Labour Force Survey (Falorsi *et al.*, 1994). For this reason we propose simple time series models in which the direct province estimates are related to those obtained for larger areas of the same survey. This approach differs from most of the small area literature applied to similar problems (Cronkhite 1987, Singh *et al.*, 1991, Rao and Yu 1994, Datta *et al.* 1999) that usually makes use of auxiliary information external to Labour Force surveys data.

Since even domain estimates from which we borrow information are exposed to sampling variability we model them as random in order to obtain reliable and not optimistically biased measures of variability for the small area estimators.

In this work we focus on the estimation of the Unemployment rate for the age class 15-24 within provinces. In the Italian Labour Force Survey the Unemployment rate is the only relevant Labour Market indicator published for age classes within provinces. The 15-24 age class is the smallest among the three considered by the ISTAT (the other are 15-29 and 30-64); moreover, in this age class, a large proportion of the population is still out of the Labor Force with the consequence of particularly small samples and unstable estimates.

We propose models in the class of Linear Mixed Models and adopt the normal hypothesis. This assumption allows us to take account for the variability of the estimates being modelled in a very simple way. We consider both additive random effects models and mixed linear models.

Predictive Frequentist or Empirical Bayes methods are often used for the estimation of such models. Although simple in theory, these methods need complex estimation algorithms when variance components are not known, as is the case in our problem and in most situations of practical relevance.

We adopt a full Hierarchical Bayesian approach to estimation. This allows for a straightforward specification even when complicated structure of random effects are introduced and allows us to manage the sampling variability of estimates from the Labour Force Survey we use as covariates. The solution of the models is based on Markov Chain Monte Carlo integrations and computation are implemented with the software BUGS (Spiegelhalter *et al.*, 1995). The introduction of simulation based methods makes the solutions of models relatively fast to obtain. The Hierarchical Bayes modelling is therefore a feasible approach for the analysis of large scale surveys. We show that simple hierarchical models based on the principle of borrowing strength lead to shrinkage estimators characterized by a significantly lower variance than the design based published estimates.

The paper is organized as follows: in section 2 the basic features of the data set is outlined along with the method employed for the estimation of the variance of the published estimates; in section 3 a brief comparative review of the methods employed in model based small area estimation is conducted and the choice of the Hierarchical Bayes approach is motivated; in section 4 the proposed methodology and models are discussed, while the results of estimation are discussed in section 5.

2. ESTIMATION OF THE UNEMPLOYMENT RATE IN THE ITALIAN LABOUR FORCE SURVEY

The sample of the Italian quarterly Labour Force Survey has an approximate size of 73,000 households collected from 1,200 different municipalities. Municipalities are sampled according to a stratified design where the resident population is the stratification variable.

As regards the sampling of households the Labour Force Survey uses a 2-2-2 rotating sampling design based on simple random sampling of the households within municipalities; each sampled units stays in the sample for two consecutive quarters, stays out for the two succeeding repetitions and is included again for other two quarters.

the estimation of Labour Market aggregates, the ISTAT employs a two stage post-stratified ratio estimator (POS). The auxiliary information on which the ratio estimator is based is given by statistics on population size by age and sex, obtained by the demographic registers held by each municipality. Post stratification is applied since classification by sex and age is not used in the design of the sample.

Unemployment rate is defined as a percentage of Unemployed people on the Labour Force. Estimates for these ratios are obtained as ratios of estimated numbers:

$$ {}_{POS}UR_d = \frac{{}_{POS}U_d}{{}_{POS}LF_d} \tag{1} $$

where d indicates the domain to which the estimate is referred and POS indicates the method of estimation. It should be noted that the panel structure of the sample is not directly exploited in the estimation process.

Measures of variability are published only for regional estimates. For smaller domains, approximated measures of variability can be obtained by means of Generalized Variance Function models (Wolter, 1985; Tiller, 1992). In particular the following model for the coefficient of variation (CV) of both the number of Unemployed and the Labour Force is proposed:

$$ CV^2(X) = \alpha + \beta \ln(X) \tag{2} $$

As regards the CV of the Unemployment rate, independence between the ratio $R = X / Y$ and its denominator Y is assumed, according to a suggestion contained in Wolter (1985) and the following model introduced:

$$ CV\left(\frac{X}{Y}\right) = \sqrt{CV^2(X)} - \sqrt{CV^2(X)} \tag{3} $$

Estimates of the parameters of these models have been kindly provided to us by the ISTAT Research Bureau.

The data set considered in this work is constituted by the series of provincial estimates of the Unemployment rate endowed with the measures of variability just described. In particular, we focus on the 15-24 age class and restrict our study to the nine provinces of the Emilia-Romagna region and to the Survey replications from January 1995 up to October 1999 (20 replications). This restriction allows us to avoid the problems involved by the modifications in the Survey's sampling design and re-definition of aggregates (occurred in 1992 and 1993) and by the establishment of new provinces by the Italian Parliament in 1994.

3. DIFFERENT APPROACHES TO THE SOLUTION OF MIXED LINEAR MODELS IN SMALL AREA ESTIMATION

Let's describe Linear Mixed Models as follows:

$$\underset{m\times 1}{y} = \underset{m\times 1}{\eta} + \underset{m\times 1}{e}$$
$$\underset{m\times 1}{\eta} = \underset{m\times k}{X}\ \underset{k\times 1}{\beta} + \underset{m\times p}{Z}\ \underset{p\times 1}{\upsilon} \tag{4}$$

The first equation in (4) expresses the vector of observations as the sum of an underlying unobservable vector η and an error term *e*. The second equation in (4) introduces a linear model for η. The matrix *X* includes real valued covariates, while *Z* is a 0-1 design matrix. According to the frequentist terminology, β is referred to as the vector of ''fixed effects'', while the components of υ are called ''random effects'' (while this distinction is immaterial in a Bayesian perspective).

The following assumptions on *e* and υ are usually introduced in the standard analysis of Mixed Linear Models:

$$E\begin{pmatrix} e \\ \upsilon \end{pmatrix} = 0 \qquad V\begin{pmatrix} e \\ \upsilon \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix} \tag{5}$$

The vector β is treated as fixed but unknown under the frequentist approach, and usually assumed a priori independent in the Bayesian literature. In this latter case β is assumed a priori independent of *v* and *e*. In the context of small area estimation it is usual to define a linear combination of fixed and mixed effects

$$\eta = \lambda'\beta + \delta'\upsilon \tag{6}$$

as the object of inference. Since (6) is a function of mixed effects υ the term prediction is often used in the place of estimation (Henderson, 1975); nonetheless the parameter β in (4) is to be considered a superpopulation parameter. Several approaches for inference have been considered in literature for the prediction of (6). In particular, if we set $\lambda = X$ and $\delta = Z$ the frequentist BLUP predictor is:

$$\hat{\eta}^{BLUP} = X' b_{GLS} + ZGZ' R^{-1}\left(y - X' b_{GLS}\right) \tag{7}$$

(Goldberger 1962, Henderson 1975). Provided that *G* and *R* are known, the estimator (7) can be obtained without explicit distributive assumptions (but only supposing the finiteness of the first two moments); moreover if posterior linearity holds it agrees exactly with the posterior mean when an improper uniform prior on R^k is assumed for β (Fay and Herriot, 1979). The solution is more complicated when *G* is not completely known. Usually it assumed it is known up to an unknown parameter θ (variance component). The traditional solution in this case is to "plug in" sampling estimates of θ into (7), obtaining an Empirical BLUP (EBLUP) estimator (Ghosh and Rao, 1994). Unfortunately, unbiasedeness and optimality with respect to quadratic loss no longer hold for the EBLUP, and normality the of υ is needed to derive Taylor series based approximations of the MSE (Kackar and Harville 1984, Prasad and Rao, 1990). These approximations are based on the unconditional approach, that is, on removing the conditioning on observed data typical of the "model based" frequentist sampling theory. There is no universally accepted method for conducting inference about random effects even within this framework (Datta and Lahiri, 2000). Moreover a conditional approach seems to be more appropriate for non normal and non linear models (Booth and Hobert, 1998).

Empirical Bayes methods have also been applied (see for instance Ghosh and Lahiri 1987, Farrell *et al.* 1997, Datta *et al.*, 1997). Even though they have a very different logical basis, Empirical Bayes methods are rather similar to the predictive frequentist ones just discussed; basically both methods require sample based estimation of the variance components and apply the "plug in" principle. Since most of the problems in BLUP theory are caused by the estimation of these variance components Empirical Bayes does not represent a better solution with respect to BLUP under this respect.

Under normality a perfect agreement between the two methods for the estimation of (6) is achieved (Fay and Herriot, 1979, Arora *et al.*, 1997).

In both Frequentist and Empirical Bayes approaches the derivation of measures of variability is usually based on the decomposition of variance obtained by conditioning on θ:

$$V(u) = E(V(u \mid \theta)) + V(E(u \mid \theta)) \tag{8}$$

where $u = (\hat{\eta} \mid y)$ in the Empirical Bayes case and $u = (\hat{\eta} \mid \eta)$ in the Frequentist one. The definition of $E(V(u \mid \theta))$ is different under the two approaches, while $V(E(u \mid \theta))$ is, under both approaches, a frequentist variance, so problems raised by its estimation are met also in the Empirical Bayes estimation.

The Hierarchical Bayes approach has been seldom used in the past because of the difficulty in obtaining exact posterior distributions even for rather standard models. The recent dramatic development of computer technology favoured the research on simulation techniques and nowadays complex and large Hierarchical models can be easily solved by means of computational methods. In particular Markov Chain Monte Carlo have developed fast and nowadays a number of efficient and well-tested algorithm are available for the handling of a large variety of problems. The Hierarchical Bayes approach, since suitable solution tools are now available, has many advantages since it enjoys both the logical consistency and richness of the Bayesian theory and the flexibility of computation based methods. In particular the specification of large, realistically complex models is easy and the generalization to non normal and non linear models straightforward (Ghosh *et al.*, 1998).

4. THE PROPOSED MODELS

For the estimation of the Unemployment rate in provincial domains, we propose shrinkage estimators based on borrowing strength from both the current survey and past repetitions. We base our shrinkage estimators on simple Linear Mixed Models.

Using an area level notation (which is slightly different from that of (4), let's introduce first a "sampling equation", relating the published estimates with the underlying unobserved value of the Unemployment Rate:

$$y_{it} = \eta_{it} + e_{it} \tag{9}$$

where y_{it} is the design based POS estimate for the area *i* in time *t*, η_{it} is the underlying value of the unemployment rate for the same domain and e_{it} is the sampling error. For this error term we suppose,

$$e_{it} \overset{ind}{\sim} N\left(0, \sigma_{it}^2\right) \tag{10}$$

that is, we define *R* in (5) as $diag\left(\sigma_{it}^2\right)$.

As anticipated, we introduce normality because it allows simple modelling of data consisting in estimates characterized by sampling variability. Moreover normality is typical of literature on Linear Mixed models (Fay and Herriot 1979, Datta *et al.* 1999). The assumption of serial uncorrelation of e_{it} is questionable since a consistent overlapping is present in successive samples of Labour Force Survey. Nonetheless the identification and estimation of a time series model is difficult, because of the decreasing sample sizes as growing temporal lags are considered (see Cocchi and Castellini, 1988). Moreover, we note that the panel structure of the sample is not directly exploited in the POS estimator, and this mitigates the effect of overlapping. We consider different models for η_{it} that can be grouped in two broad classes: Random effects additive "ANOVA type" models and linear mixed models.

In the first case, we link the unemployment rates estimated for provinces to their regional averages; the discrepancy between the provincial and regional estimates is modelled by means of time dependent random effects. In summary:

$$\eta_{it} = \mu_t^* + \text{random effects} \tag{11}$$

This implies imposing $X = 0$ in (4) and basing our model exclusively on the design matrix *Z*. Moreover we assume that

$$\mu_t^* \sim N\left(\mu_t, V\left(\mu_t\right)\right) \tag{12}$$

that is, the published regional estimate of the unemployment rate for the age class 15-24 is supposed to be normally distributed with variance calculated according to the GVF model described in previous section 2.

As regards the second class of models we regress the 15-24 provincial unemployment rate on the provincial rate for all age classes, that is:

$$\eta_{it} = X^*_{it}\beta_i + \text{random effects} \tag{13}$$

The provincial rate is modelled as a random effect, assuming normality with variance calculated according to the described GVF model:

$$X^*_{it} \sim N(X_{it}, V(X_{it})) \tag{14}$$

The regression parameters β_i are province-specific, that is, their estimation is based on time series of rates computed for the same province. The two classes of models summarize alternative hypotheses on the age-specific provincial rate. The first is based on its relationship with the same rate in neighbouring provinces, while the second relies on intra-province relationships among the rates for different age classes. For both classes of models we consider an overall province intercept (α_i), quarter (γ_{is}) and year (δ_{ja}) effects along with an auto-correlated "residual" term ε_{it} accounting for the slow changes in Labour Market conditions and the sampling autocorrelation not modelled in (10). All time dependent effects are introduced as area specific; in fact overall season or year effects are naturally included in the time variation of the regional average μ^*_t in (11) or of the provincial rate X^*_{it} in (13).

For both the considered classes of models we can write the full (saturated) model as:

$$\eta_{it} = K_{it} + a_i + \gamma_{is} + \delta_{ia} + \varepsilon_{it} \tag{15}$$

with $K_{it} = \mu^*_t$ in additive models and $K_{it} = X^*_{it}\beta_i$ in linear mixed models.

As regards prior assumptions on random effects, we select conjugate priors in most of the cases, since they simplify and speed up Markov Chains computations. The parameters of priors are chosen in order to let they have very large prior variances and therefore diffuse distributions. As noted in Natarajan and McCullogh (1998) parameters about which there is little information may be sensitive even to the choice of a diffuse prior, but this is not a concern in our case since we focus mostly on the estimation of the η_{it} that are first level parameters directly indexing the likelihood. In particular, for the regression parameter in the linear mixed models and the province intercept we choose the priors:

$$\alpha_i \overset{ind}{\sim} N(0,100) \tag{16}$$

$$\beta_i \mid A \overset{ind}{\sim} N(0, A) \tag{17}$$

with

$$A^{-1} \sim Gamma(\omega, \omega) \tag{18}$$

Similarly to the priors just discussed, hierarchical priors with non area-specific variance components are assumed for the seasonality and year effects:

$$\begin{aligned} &\gamma_{is} \mid A \overset{ind}{\sim} N(0, B) \qquad s = 1,\dots,4 \\ &B_s^{-1} \overset{ind}{\sim} Gamma\ (\omega, \omega) \end{aligned} \tag{19}$$

and

$$\begin{aligned} &\delta_{ia} \mid B \overset{ind}{\sim} N(0, C) \qquad a = 1,\dots,5 \\ &C^{-1} \sim Gamma(\omega, \omega) \end{aligned} \tag{20}$$

This hierarchical structure links the provincial effects to a common distribution from which they can be thought to be sampled, thus reducing the actual number of parameters in the model. For the provincial intercept α_i we do not propose a hierarchical prior since it is reasonable that they are very different from province to province. We have assumed season-specific variances B_s for season effects, that can be supposed to be sampled from ''seasonal distributions''

over the years, but a unique, common variance C for years effects on which we have scant information. As regards the choice of ω since

$$\begin{aligned} &E\left(A^{-1}\right) = E\left(B_s^{-1}\right) = E\left(C^{-1}\right) = 1 \\ &V\left(A^{-1}\right) = V\left(B_s^{-1}\right) = V\left(C^{-1}\right) = \omega^{-1} \end{aligned}$$

very small values (as 100^{-1} or 1000^{-1}) imply a highly positively skewed distribution (skewness being equal to $2\omega^{-1/2}$). We note that a $Gamma(\alpha, \beta)$ prior on the precision is equivalent to an $InvGamma(\alpha, \beta)$ on the variance component. Since $\sigma^2 \sim InvGamma(\alpha, \beta)$ we have

$$\begin{aligned} &E\left(\sigma^2\right) = \frac{\beta}{\alpha - 1} \quad \text{(provided } \alpha > 1\text{)} \\ &\mathrm{V}\left(\sigma^2\right) = \frac{\beta^2}{(\alpha - 1)^2 (\alpha - 2)} \end{aligned}$$

As a consequence, a "small parameter" $Gamma(\omega,\omega), \omega<1$ corresponds to an $InvGamma(\alpha,\beta)$ with infinite expected value and variance. Sampling form diffuse distributions as a $Gamma(\omega,\omega)$ with very small ω may be accompanied by computational difficulty when using the Gibbs Sampler (Natarajan and McCullogh, 1998). We then propose to set $\omega=0.1$, that is a moderately small value.

As regards the "residual" ε_{it}, we consider two alternative hypotheses: i) non stationary random walk (as in Datta *et al.*, 1999) and ii) auto-correlated AR(1) process (as in Rao and Yu, 1994). In this latter case we assume, a *Uniform* prior over the interval (-1,1) for the autoregression parameter ρ.

We also consider the "classical" James Stein and Fay-Herriot models for the additive random effects and linear mixed classes of models as benchmarks. In these models, the borrowing strength is based only on the current repetition of the survey and we therefore refer to them as "longitudinal". The linking model can be described by the following assumption:

$$\eta_{it}=K_{it}+\alpha_{it} \tag{21}$$

As in (15) we have $K_{it}=\mu_t^*$ for additive models and $K_{it}=X_{it}^*\beta_i$ for linear mixed models. For both James-Stein and Fay-Herriot models the intercepts α_{it} are thought to be sampled from common longitudinal distributions:

$$\begin{aligned}&\alpha_{it}\mid C_t \overset{ind}{\sim} N(0,C_t)\\&C_t^{-1}\overset{ind}{\sim} Gamma(0.1,0.1)\end{aligned} \tag{22}$$

4.1 *Model assessment and model selection*

The adequacy of the proposed models is evaluated by means of their average performances, that is, by measuring how they behave on the whole set of domains studied. It is in fact likely that model performances varies across areas due to different area sizes. An average evaluation integrates out this size effect. We remark also that average evaluation is consistent with the practice of many applied research works (see for instance Falorsi *et al.*, 1994 and Ghosh *et al.*, 1996).

The evaluation we propose relies on two main features of model performances: (1) the fit on the values of published estimates and (2) the gains in efficiency. Model selection is based on two simple measures of the model features just mentioned. It is worth to note that improvements in the precision of the estimates is considered not only in the model selection but also in the check of model adequacy. Among possible measures of gain in efficiency we propose the average reduction of the Coefficient of Variation realized by the model estimates with respect to the published POS values:

$$CVR = \frac{1}{180}\sum_{i=1}^{9}\sum_{t=1}^{20}\left(\frac{sd(\hat{\eta}_{it})}{\hat{\eta}_{it}}\right)/\left(\frac{sd({}_{POS}UR_{it})}{{}_{POS}UR_{it}}\right) \tag{23}$$

while as regards the "fit" on published estimates, we propose the simple mean of absolute differences:

$$AFIT = \frac{1}{180}\sum_{i=1}^{9}\sum_{t=1}^{20}\left|\hat{\eta}_{it} - {}_{POS}UR_{it}\right| \tag{24}$$

a measure considered also by Ghosh *et al.* (1996).

A trade off between improvement of fit and variability reduction is to be expected. Intuitively, if we replace the provincial estimates with their regional analogues we obtain a dramatic reduction in CVs at the price of neglecting what direct estimates tell us about single domains. The more flexible models become, the more they are likely to show a better fit, but at the price of introducing further sources of variation and of inflating the variance of model based estimators. The model allowing for the best compromise between variability reduction on the published estimates has then selected, provided it can be judged adequate according to the procedure described below.

Model adequacy is evaluated using a posterior predictive approach, that is performing a self consistency check between hypothetical future observations generated from the posterior predictive distribution

$$f(y_{new} \mid \theta) = \int f(y_{new} \mid \theta)p(\theta \mid y)d\theta$$

and "observations" (in our case the published estimates). The general goal is then to assess whether actual data are plausible under the assumed model. The posterior predictive approach is usually based on building the "discrepancy measures" $D(y,\theta)$ and successively evaluating the probability $P(D(y_{new},\theta) > D(y,\theta) \mid y)$ that indicates a lack of fit of the model when its value is far from 0.5. Following this approach, Datta *et al.* (1999) propose a sum of standardized residuals, close to the usual "omnibus goodness of fit" measure described in Carlin and Louis (1996, p 57).

Since the goal of our modelling is to provide a basis for shrinkage, reduced variance estimators, we are willing to accept a somewhat poor fit if it allows a remarkable variance reduction. For each area we propose to evaluate the consistency of y_{new} with the published estimates by means of:

$$P(y_{new,it} > y_{it} \mid y_{it}, \eta_{it}) \tag{25}$$

We fix then a threshold ($v = 0.1$ or $v = 0.25$ for instance) and define the following indicator variable:

$$T_{it} = \begin{cases} 1 & \text{if } \left(P\left(y_{new,it} > y_{it} \mid y_{it}, \eta_{it}\right) < v\right) \text{or} \\ & \left(P\left(y_{new,it} > y_{it} \mid y_{it}, \eta_{it}\right) > 1 - v\right) \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

We then compute a summary measure averaging over all the domains considered:

$$OUT = \frac{1}{180} \sum_{i=1}^{9} \sum_{t=1}^{20} T_{it} \tag{27}$$

that can be interpreted as the proportion of domains for which a given model shrinks "too much", according to the fixed threshold v. A model is then to be considered as adequate if the associated value of *OUT* is 0 or very low.

5. DATA ANALYSIS AND DISCUSSION OF MODEL RESULTS

All models described in the previous section are solved by means of a Gibbs sampler algorithm implemented in the software WinBUGS (Spiegelhalter *et al.*, 1995) For all models a run of 15,000 iterations for three parallels chains has been carried out.

For the assessment of convergence the visual inspection of chain paths together with autocorrelation diagrams is considered along with the modified Gelman and Rubin statistic (Brooks and Gelman, 1998). Despite the fast reach of convergence shown by all models, a conservative burn in of 5,000 iteration is chosen for all chains; consequently an overall sample of size 30,000 is drawn from each posterior distribution.

Summary statistics from posterior distributions for η_{it} in all models highlight unimodality and approximate symmetry or moderate positive asymmetry. We note also that, the performances in terms of *CVR* are generally similar in all provinces, while the fit as expressed by *AFIT* is better for provinces whose rates are characterized by lower estimated variances, since we have in those cases a better identification of random effects.

Results from Random effects additive "ANOVA type" models are summarized in figure 1. The average adequacy measure OUT is calculated for $v = 0.1$ and $v = 0.25$.

TABLE 1

Performances of various "anova type" proposed models

Linking model	CVR(%)	AFIT(%)	OUT(%) $\nu = 0.25$	OUT(%) $\nu = 0.1$
$\mu_t + \alpha_i$	29.6	4	42.8	17.2
$\mu_t + \alpha_i + \varepsilon_{it}$ (RW)	48.7	3.5	35.5	10.4
$\mu_t + \alpha_i + \delta_{ia}$	49.3	3.3	31.1	7.8
$\mu_t + \alpha_i + \gamma_{is}$	53.5	3.4	30.1	7.7
$\mu_t + \alpha_i + \delta_{ia} + \gamma_{is}$	65.7	2.8	23.3	3.1
$\mu_t + \alpha_i + \varepsilon_{it}$ (AR(1))	70.4	1.8	28	0
$\mu_t + \alpha_{it}$ (longitudinal)	84.1	2	0	0
$\mu_t + \alpha_i + \delta_{ia} + \gamma_{is} + \varepsilon_{it}$ (AR(1))	96.5	0.25	0	0

The model considering the only province effect besides the regional mean shows a serious lack of fit, and the same do, even though to a lesser degree, models including also season or year effects. The full model (15) tends to overfit data and does not provide a remarkable reduction in the variability of estimates; similarly the longitudinal James-Stein model fits quite well the series of published estimates but has limited impact over the CV reduction.

The model including province intercepts and stationary autocorrelated residual seems to offer the better compromise between fit and in terms of CV reduction. Moreover it can be considered good according to the guideline defined in previous section, while all the models characterized by lower levels of *CVR* show large values of *OUT*.

We note that this model outperforms definitely its analogous characterized by the assumption of non stationary Random Walk residual; in fact the posterior mean of the autocorrelation parameter is approximately 0.4 and the assumption of non stationariety for ε_{it} is therefore inadequate.

As regards mixed linear models, results of the estimation of mixed models are summarized in table 2. Thresholds $\nu = 0.1$ and $\nu = 0.25$ in the computation of *OUT* are considered in this case as well.

TABLE 2

Performances of various "mixed linear" proposed models

Linking model	CVR(%)	AFIT(%)	OUT(%) $\nu = 0.25$	OUT(%) $\nu = 0.1$
$X_{it}\beta_i + \alpha_i$	48.2	2.8	24.4	3.3
$X_{it}\beta_i + \alpha_i + \gamma_{is}$	54.5	2.4	20.1	1.4
$X_{it}\beta_i + \alpha_i + \delta_{ia}$	56.8	2.6	20.5	1.2
$X_{it}\beta_i + \alpha_i + \varepsilon_{it}$ $(AR(1))$	73.6	2.0	4.0	0.6
$X_{it}\beta_i + \alpha_i + \gamma_{is} + \delta_{ia}$	82.4	1.8	0	0
$X_{it}\beta_i + \alpha_{it}$ (longitudinal)	89.2	0.7	0	0
$X_{it}\beta_i + \alpha_i + \gamma_{is} + \delta_{ia} + \varepsilon_{it}$ $(AR(1))$	98	0.4	0	0

The results are comparable to those found for additive random effects models. We may note, in general, a better fit and lower reduction of variability. This is consistent with the use of provincial estimates of the Unemployment rate X_{it}^* as auxiliary information: they are local estimates as the age specific rates being estimated but are characterized by a quite large variance. In this view we note that the longitudinal Fay-Herriot model is not very effective in variance reduction. In considering the fit of the alternative models it is worth to note that $Corr(y_{it}, X_{it}^*)$ varies substantially among the nine provinces, ranging form 0.3 to 0.8. As for the additive random effects models just discussed, the fit is influenced by the size the variances associated to the published estimates; in the case of mixed models we have that for similar levels of average design variance, the statistic *AFIT* is smaller in provinces characterized by higher correlations.

As regards the comparison of the different models, there is evidence that the model including autocorrelated stationary residuals offer the best compromise between the reduction of the CV and the fit on the series of estimated data. Moreover it can be judged good according to the value of the *OUT* statistic.

6. CONCLUSIONS

In this paper we introduced shrinkage estimators based on Hierarchical Bayes linear mixed models for the estimation of the Unemployment rate in small domains of the Italian Labour Force Surveys. In particular we have considered the 15-24 age class within provinces; since a large proportion of the population of this age class is still out of the Labour Force estimates associated to these domains are particularly unreliable.

For the building of small area models auxiliary information like Unemployment Insurance records or other administrative sources were not available. The need to overcome this problem, along with the repetitive nature of Labour Force Survey suggest the application of models combining both longitudinal and time series "borrowing strength" features (Rao and Yu 1994, Datta *et al.*, 1999). We propose models borrowing information from estimates obtained from larger areas of the same survey. In this context Hierarchical Bayes approach allows us to take account for all sources of uncertainty in a very simple way.

Computations have been made using MCMC integration algorithms, as implemented in the dedicated software BUGS (Spiegelhalter *et al.*, 1995). The availability of such computational methods makes the Hierarchical Bayes approach a feasible instrument for the analysis of large scale survey which is also simpler when compared to the other approaches traditionally applied to the analysis of linear mixed models in the small area context.

Results show that, even simple Hierarchical Bayes models allows for significant reduction of the variance of published estimates and also that models based on both time series and cross sectional borrowing of information a clearly outperform models based exclusively on a single repetition of the survey.

Dipartimento di Scienze statitistiche "Paolo Fortunati" ENRICO FABRIZI
Università di Bologna

ACKNOWLEDGEMENTS

I thank Stefano Falorsi from the ISTAT research staff for his invaluable support in preparing the basic data set and discussing various aspects of the research. I thank also Daniela Cocchi and Carlo Trivisano for their continuous scientific and human support during this research.

REFERENCES

V. ARORA, P. LAHIRI, K. MUKHERJEE (1997) *Empirical Bayes estimation of finite Population Means form Complex Surveys*, "Journal of the American Statistical Association", 92, pp. 1555-1561.

J.G. BOOTH AND J.P. HOBERT (1998) *Standard errors of predictors in generalized linear mixed models*, "Journal of the American Statistical Association", 93, pp. 362-372.

S.P. BROOKS AND A. GELMAN (1998) *Alternative methods for monitoring convergence of iterative simulations*, "Journal of Computational and Graphical Statistics", 7, pp. 434-455.

F.R. CRONKHITE (1987) *Use of Regression Techniques for developing State and Area Employment and Unemployment Estimates*, in R. PLATEK , J.N.K. RAO, C.E SÄRNDAL., M.P. SINGH (eds.), Small Area Statistics, John Wiley, New York.

G.S. DATTA, P. LAHIRI, T. MAITI (1997) *Empirical Bayes Estimation of Median Income of four-person families by state using time series and cross-sectional data*, to appear in "Journal of Statistical Planning and Inference".

G.S. DATTA, P. LAHIRI, T. MAITI, K.L. LU (1999) *Hierarchical Bayes Estimation of Unemployement Rates for the U.S.*, "Journal of the American Statistical Association", 94, pp. 1074-1082.

G.S. DATTA AND P. LAHIRI (2000) *A unified measure of uncertainty of estimated Best Linear Unbiased predictor in Small Area estimation problems*, to appear in "Statistica Sinica".

P.D. FALORSI, S. FALORSI, A. RUSSO (1994) *Empirical comparison of small area estimation methods for the Italian Labour Force Survey*, "Survey Methodology", 20, pp. 171-176.

R.E. FAY, R.A. HERRIOT (1979) *Estimates of income for small places: an application of James-Stein procedures to census data*, "Journal of the American Statistical Association", 74, pp. 269-277.

P.J. FARRELL, B. MCGIBBON, T.J. TOMBERLIN (1997) *Empirical Bayes small area estimation using logistic regression models and summary statistics*, "Journal of Business and Economic Statistics", 15, pp. 101-108.

M. GHOSH, P. LAHIRI (1987) *Robust empirical Bayes estimation of means from stratified samples*, "Journal of the American Statistical Association", 82, pp. 1153-1162.

M. GHOSH, J.N.K. RAO (1994) *Small area estimation: an appraisal*, "Statistical Science", 9, pp. 55-93.

M. GHOSH, N. NANGIA, D. HO KIM (1996) *Estimation of income of four person families: a Bayesian time series approach*, "Journal of the American Statistical Association", 91, pp. 1423-1431.

A. GOLDBERGER (1962) *Best Linear Unbiased Prediction in the Generalized Linear Regression Model*, "Journal of the American Statistical Association", 57, pp. 369-375.

M. GOLDSTEIN (1975) *Approximate Bayes solutions to some nonparametric problems*, "Annals of Statistics", 3, pp. 512-517.

C.R. HENDERSON (1975) *Best Linear Unbiased estimation and prediction under a selection model*, "Biometrics", 31, pp. 423-447.

R. N. KACKAR AND D. A. HARVILLE (1984) *Approximation for standard error of estimators of fixed and random effects in mixed linear models*, "Journal of the American Statistical Association", 79, pp. 853-862.

R. NATARAJAN, C.E. MC CULLOGH (1998) *Gibbs sampling with diffuse proper priors: a valid approach to data-driven inference?*, "Journal of Computational and Graphical Statistics", 7, pp. 267-277.

D. PFEFFERMANN, L. BURCK (1990) *Robust small area estimation combining time series and cross-sectional data*, "Survey Methodology", 16, pp. 217-237.

N.G.N. PRASAD AND J.N.K. RAO (1990) *The estimation of mean squared error of small area estimators*, "Journal of the American Statistical Association", 85, pp. 163-171.

J.N.K. RAO AND M. YU (1994) *Small area estimation by combining time series and cross sectional data*, "Canadian Journal of Statistics", 22, pp. 511-528.

J.N.K. RAO (1999) *Recent advances in Small Area Estimation*, "Survey Methodology", 25, pp. 48-57.

A. C. SINGH, H.J. MANTEL, B.W. THOMAS (1994) *Time series EBLUPs for small areas using survey data*, "Survey metholology", 20, pp. 33-43.

D.J. SPIEGELHALTER, A. THOMAS, N. BEST , W.R. GILKS (1995) *BUGS: Bayesian Inference using Gibbs sampling, version 0.50*, Techinal Report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.

R. B. TILLER (1992) Time series modeling of sample survey data from the U.S. Current Population Survey, "Journal of Official Statistics", 8, pp. 149-166.

K.M. WOLTER (1985) *Introduction to variance estimation*, Springer Verlag.

RIASSUNTO

Modelli bayesiani gerarchici per la stima dei tassi di disoccupazione in piccoli domini di studio dell'indagine sulle Forze di Lavoro

In questo lavoro vengono introdotti stimatori a *shrinkage* per la stima del tasso di disoccupazione in piccoli domini di studio dell'Indagine sulle Forze di Lavoro. Questi stimatori sono basati su modelli lineari misti gerarchici che attuano principio del prestito di informazione (*borrowing strength*) sia rispetto ad altre stime locali per la stessa ripetizione dell'indagine che in serie storica. Non vengono invece considerate informazioni ausiliarie esterne all'indagine. Viene adottato un approccio bayesiano gerarchico in cui i modelli vengono risolti attraverso algoritmi computazionali di tipo MCMC. Ciò garantisce di poter misurare la variabilità degli stimatori considerando adeguatamente tutti i fattori di incertezza legati al loro calcolo. I risultati evidenziano come semplici modelli bayesiani gerarchici consentano una riduzione notevole della varianza rispetto alle stime pubblicate; inoltre si evidenzia come i modelli che considerano la dimensione temporale accanto a quella spaziale offrano prestazioni decisamente migliori.

SUMMARY

Hierarchical Bayesian models for the estimation of unemployment rates in small domains of the italian labour force survey

In this paper we introduce shrinkage estimators for the estimation of the Unemployment rate in small domains of the Italian Labour Force Survey. The proposed estimators are based on Hierarchical Linear Mixed Models and on the borrowing strength on both time series and cross section. Auxiliary information from sources external to the Labour Force Survey is not considered. A Hierarchical Bayesian approach is adopted, in which models are solved by means of MCMC sampling algorithms. This allows to measure variability associated to estimators accounting, in a simple way, for all the uncertainty sources. Results highlight how, simple hierarchical models allows for remarkable gain in efficiency with respect to published estimates, and that models with a time series component perform better than those based exclusively on data from the same repetition of the survey.