Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 66(1) 2012 187-201

Labour Force Status Estimates under a Bivariate Random Components Model

Ayoub Saei[1*] and Alan Taylor[2]
[1]*Centre for Infections, Health Protection Agency, Statistics Unit, Statistics, Modelling and Bioinformatics Department, 61 Colindale Avenue, NW9 5EQ, London, UK*
[2]*Office for National Statistics, U.K.*

Received 30 April 2011; Revised 17 August 2011; Accepted 17 August 2011

SUMMARY

Models for a multi-category response data, labour force status, based on a generalized linear model specification typically assume that the regression coefficients to be varied with the response category. However, they can be extended to random components by allowing the area random effects to be depended on response category. In this paper, we describe a multinomial linear mixed model with a bivariate random component in estimating totals of the inactive, unemployed and employed people at Local Authority District (LAD) level. The random effects are assumed to follow a bivariate normal distribution. The model parameters including variance components and correlation coefficient are estimated by maximum and residual maximum likelihood methods. The estimated parameters and predicted values of the LAD (area) random effects are then used in calculating the empirical best linear unbiased-type estimates. The mean squared error estimates are obtained by using an analytical approximation approach. The application is the UK LFS data in Molina *et al.* (2007) and estimates are compared with the results in that paper. A simulation study demonstrates a good performance of the proposed model.

Keywords : Bivariate, Category-specific, LFS, Maximum likelihood, Multinomial, Random component, REML.

1. INTRODUCTION

A reliable estimate at both national and sub-regional levels of the number of individuals in all three labour force categories is a high profile in Great Britain (GB). In particular, category unemployed is one of the most important indicators of the health of an area. Sample unemployment information is collected in the Labour Force Survey (LFS) and it is the largest regular household survey in GB. The LFS is the major source of data in estimating number of unemployed individuals at both national and Local Authority District (LAD) levels. The LFS produces reliable estimates of labour force status for national and some LADs with large sample sizes. However, application of design-based estimation theory to LFS data leads to poor estimates of the labour force status for LADs with small sample size. An alternative approach that is now widely used in small area estimation is the so-called indirect or model-based approach. This uses auxiliary information for the small areas of interest and has been characterized in the statistical literature as "borrowing strength" from the relationship between the values of the response variables and the auxiliary information.

A flexible and popular way of borrowing strength is based on the application of linear mixed models with area specific random effects (Rao 2003), with estimation and inferences typically carried out using empirical best linear unbiased prediction (EBLUP), see Prasad and Rao (1990), Longford (2007) and Saei and Chambers (2003). For recent review, see Jian and Lahiri (2006). However, LFS data, which motivates this study, the survey variable of interest is not normally

*Corresponding author : Ayoub Saei
E-mail address : ayoub.saei@hpa.org.uk or axs96@soton.ac.uk

distributed. In particular, data are counts from a multi-category (ploytomous) response variable, labour force status (inactive, unemployed and employed). The application of standard methods (see Rao 2003) for small area estimation based on linear models fails in this situation not only because the response variable is polytomous but also the distribution of small counts does not approach normality.

For discrete response data in particular polytomous data, Zhang and Chambers (2004) introduced generalized linear structure mixed model (GLSMM) and Saei *et al*. (2005) developed a generalized structure preserving estimation equation (GSPREE) approach. Saei and Chambers (2003) extended the empirical best linear unbiased estimation to a binary/binomial data under a logistic model with area random effects. This method was used in the EURAREA Consortium (2004), Curtis and Saei (2005) and Cruddas and Saei (2006) to model the proportion of unemployed population. The model provides estimated totals of unemployed individuals, which are then combined with the direct estimates of the totals of employed individuals to derive the rates of unemployment (Hastings *et al.* 2003). While the unemployment estimates are calibrated to the survey estimates at regional government office region (GOR), socio-economic classification (Cluster) and age-sex groups, there are inconsistencies at LAD levels. This means that the sum of the modelled unemployed and employed levels will not match a survey estimate of total economically active population and that the sum of all three labour force categories will not match the LAD total population estimate. Molina *et al.* (2007) provides a consistent unemployment totals estimate at LAD level under a multinomial logistic model. However, the underlying model is unable to discriminate between response categories in term of area random effect. In the absence of the auxiliary information (covariates), the model produces the same estimates for both categories (unemployed and inactive) in the model. This assumption is hardly justified in many applications, especially in LFS. Thus, small area model for a multi-category response data needs to allow a different random area effect for different response categories. This paper extends the logistic linear mixed model for binomial response data and allows the area random effect to be varied by response category (i.e. we have a category-specific area random effects). The two sets of random effects are sharing some commonality because they originate from the same LAD. The model assumes that the LAD random effects for two categories unemployed and inactive are correlated. In particular, bivariate random effects are embedded into the multinomial logistic linear model, where area random effects in both unemployed and inactive are assumed to follow a bivariate normal distribution. Consequently the model differs from the standard mixed model used in small area estimation in that it contains extra unknown parameters corresponding to covariance between the area random effects as well as variances.

Section 2 defines the category-specific correlated random effects and associated notations. Assuming the variance components of this model are known, we develop the corresponding best linear unbiased predictor-type (BLUP-type) estimates of totals of inactive, unemployed and employed people in Section 3. The empirical best linear unbiased predictor-type (EBLUP-type) estimates of totals in all three categories, inactive, unemployed and employed, are developed in Section 4. Results from a simulation study of the performance of the new model and method are then provided in Section 5. The application of the method to UK LFS data is developed in Section 6. Section 7 concludes the paper with a discussion of potential avenues for further research.

2. MODEL AND NOTATION

Let vectors $\{y_{sdij}\}$ and $\{y_{rdij}\}$ denote the sample and non-sample population values of the survey variable total number of people in three LFS categories (inactive, unemployed and employed). The subscripts d, i and j ($d = 1, 2,\ldots, D$; $i = 1, 2,\ldots, I$; $j = 1, 2, 3$) represent area, group (for example age-sex class/combination) and response category respectively. The sampled and non-sampled are presented by s and r in subscripts. The objective is to estimate/predict the number of people in all three LFS categories (inactive, unemployed and employed),

$$\theta = \{\theta_{dj}\} = \left\{ \sum_{i=1} y_{sdij} + \sum_{i=1} y_{rdij} \right\}.$$

The vector $\mathbf{y}$ can be partitioned as $\mathbf{y} = [\mathbf{y}_s', \mathbf{y}_r']'$ after sample is observed. Similarly, a known matrix $\mathbf{A}$ is partitioned conformably as $\mathbf{A} = [\mathbf{A}_s, \mathbf{A}_r]$. In matrix format, the parameters of interest are then

$$\theta = \mathbf{A}_s\mathbf{y}_s + \mathbf{A}_r\mathbf{y}_r.$$

The first term depends only on the sample values and is known after the sample is observed. The second term, which depends on the non-sample values, is unknown. The estimate or predicted value of θ, say $\hat{\theta}$, is then obtained by replacing $\mathbf{y}_r$ with its predicted value and it is,

$$\hat{\theta} = \mathbf{A}_s\mathbf{y}_s + \mathbf{A}_r\mathbf{y}_r \tag{1}$$

where a "hat" denotes an estimate of an unknown quantity.

Let p_{dij} be probability that an individual in group (age-sex) i from area d belong to LFS category j (j = 1 = inactive, j = 2 = unemployed) and let n_{di} be total number of individuals (active and inactive) in group i from area d in the sample. Similarly, let $\mathbf{x}_{sdij}$ be the vector of sample values of the auxiliary information/ covariates.

We consider two different forms of the linear predictor for the multinomial logistic linear mixed model. In the first, the predictor is assumed to be a linear function of a vector $\mathbf{x}_{sdij}$ of p covariates as well as an area (LAD) random effect u to account for variation not explained by the values in $\mathbf{x}_{sdij}$. That is,

$$\tau_{sdij} = \alpha_j + \mathbf{x}'_{sdij}\beta_j + u_d \tag{2}$$

where α_j is the category-specific effect, β_j is an unknown vector of regression coefficients for the response group j (j = 1 = inactive, j = 2 = unemployed). The random effects u_d are assumed to be realisations of independent N(0, ϕ) random variables. The linear predictor (2) is connected to multinomial response by a generalized logit function, i.e.

$$m\ \text{logit}(p_{sdij}) = \log\left(\frac{p_{sdij}}{p_{sdi3}}\right) = \tau_{sdij} \text{ for } j = 1, 2. \tag{3}$$

Molina *et al*. (2007) used model (3) in estimating labour force status. Although it overcomes the inconsistency estimate problems in Hastings *et al.* (2003), it is unable to discriminate between response categories in term of area random effects. As a result, it produces a single estimate/predicted value for all response categories for a given area in the absence of the covariates/auxiliary information.

In applications the random effect assumption of model (3) often is highly questionable and hardly justified. Our second (new) model assumes the area random effects are depended on the response category (category-specific random effect). It allows a possible change in variance and pattern of association according to response categories. This is consistent with the idea that the random area effects are used to account for variation not explained by the auxiliary variables $\mathbf{x}_{dij}$. The model is

$$\tau_{sdij} = \alpha_j + \mathbf{x}'_{sdij}\beta_j + u_{dj} \text{ for } j = 1, 2. \tag{4}$$

The $\mathbf{u}_d = (u_{d1}, u_{d2})'$s are independent bivariate normal variables with zero mean vector and variances of ϕ_1, ϕ_2 and covariance between them of ϕ_3 . The u_{d1} and u_{d2} are the area random effect for the first (inactive) and second (unemployed) response categories for a multinomial variable with three categories. A model with independent category-specific random effects is special case of model (4) where $\phi_3 = 0$.

Let Δ represents an indicator matrix for the category-specific effects α_j's and **X** denotes matrix of regression variables. Let $\mathbf{u}_1 = (u_{11}, u_{21}, ..., u_{D1})$, $\mathbf{u}_2 = (u_{12}, u_{22}, ..., u_{D2})$ and $\mathbf{X}^* = [\Delta\mathbf{X}_s]$. Similarly, let $\mathbf{Z}_1$ and $\mathbf{Z}_2$ denote the incidence matrices for the random effect vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ respectively. In general the right hand side of the (4) can be written as

$$\tau_s = \mathbf{X}^*\beta + \mathbf{Z}\mathbf{u} \tag{5}$$

where $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2]$ and $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$. The random vectors **u** is distributed as multivariate normal with zero mean vector and variance-covariance matrices given by

$$\mathbf{\Phi} = \mathbf{I}_D \otimes \begin{bmatrix} \phi_1 & \phi_3 \\ \phi_3 & \phi_2 \end{bmatrix}$$

where $\mathbf{I}_D$ is identity matrix of D (D = 406 in UK LFS application) order and ⊗ denotes direct product.

3. BEST LINEAR UNBIASED PREDICTOR-TYPE

A widely used method for defining the estimates in (1) is via substitution of the corresponding best linear unbiased estimators (for unknown regression coefficients) and best linear unbiased predictors (for unknown realisations of random variables). Under generalized linear mixed model (GLMM), the distribution of the vector **y** of sample values of the variable of interest is assumed to depend on a vector quantity τ_s that is related to regression covariates and random components through the equation (5). The linear predictor τ_s is connected to response variable **y** via a known function h, defined by $E(y_s \mid \mathbf{u}) = h(\tau_s)$. Here $h(\tau_s)$ is a vector of the elements of

$$n_{di} \frac{\exp(\tau_{sdij})}{1+\sum_{j=1}^{2}\exp(\tau_{sdij})} \tag{6}$$

where n_{di} is the number of sampled people in group i for the area d.

Henderson (1963, 1973a, and 1975) develops best linear unbiased predictors for linear mixed models. McGilchrist (1994) extended this approach to generalized linear mixed models. Below we outline the extension of this approach to multinomial logistic linear mixed model with category-specific area random components.

Let l_1 be the log-likelihood function of the multinomial observations vector $\mathbf{y}_s$ conditional on the value of the random component vector **u** and let l_2 be the logarithm of the probability density function of **u**. For the model (4) the functions l_1 and l_2 are

$$l_1 = \ln(f_1(\mathbf{y}_s \mid \mathbf{u})) = \sum_{d=1}^{D}\sum_{i=1}^{I}\sum_{j=1}^{J-1}\left[y_{sdij}\tau_{sdij} - n_{di}\ln\left(1+\sum_{j=1}^{J-1}\exp(\tau_{sdij})\right)\right]$$

$$l_2 = \ln(f_2(\mathbf{u})) = -(½)\,[\text{Const.} + \ln|\Phi| + \mathbf{u}'\Phi^{-1}\mathbf{u}]$$

where s indicates sample, $J = 3$ and $\tau_{sdij} = \alpha_j + \eta_{sdij}$. The α, β and **u** values that jointly maximise $l = l_1 + l_2$ are called the penalised quasi likelihood (PQL) estimates; $\tilde{\alpha}, \tilde{\beta}$, and $\tilde{\mathbf{u}}$. Put $\tau_s = \alpha + \mathbf{X}_s\beta + \mathbf{Z}\mathbf{u}$ and $\mathbf{X}^* = [\Delta\ \mathbf{X}_s\ \mathbf{Z}]$ (here $\mathbf{Z} = \mathbf{Z}_s = \mathbf{Z}_r$). The iterative procedure used to obtain the PL estimators can be specified as follows:

(a) Starting from initial values α_0, β_0, $\mathbf{u}_0$ and ϕ_{j0} (hence Φ_0) successive iterations are obtained by finding changes $\Delta\alpha$, $\Delta\beta$ and $\Delta\mathbf{u}$ to the current estimates from the equations

$$\mathbf{V}\begin{bmatrix}\Delta\alpha \\ \Delta\beta \\ \Delta\mathbf{u}\end{bmatrix} = X^{*T}\frac{\partial l_1}{\partial \tau_0} - \begin{bmatrix}\mathbf{0} \\ \Phi_0^{-1}\mathbf{u}_0\end{bmatrix}$$

where $\mathbf{V} = X^{*T}\left(-\frac{\partial^2 l_1}{\partial\tau_{s0}\partial\tau_{s0}^T}\right)X^* + \begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Phi_0^{-1}\end{bmatrix}$ and $\frac{\partial l_1}{\partial\tau_{s0}}$ and $\frac{\partial^2 l_1}{\partial\tau_{s0}\partial\tau_{s0}^T}$ are first and second order derivatives of l_1 with respect to τ_s and evaluated at initial value τ_{s0}.

At converge values $\tilde{\alpha}, \tilde{\beta}$ and $\tilde{\mathbf{u}}$, best linear unbiased predict-type (BLUP-type) estimate of θ is then

$$\tilde{\theta} = A_s y_s + A_r \tilde{y}_r = A_s y_s + A_r\left\{(N_{di} - n_{di})\exp(\hat{\tau}_{rdij})\left(1+\sum_{j=1}^{2}\exp(\tilde{\tau}_{rdij})\right)^{-1}\right\}$$

where N_{di} is the total population (active and inactive) in age-sex group i for LAD d and $\tilde{\tau}_{rdij} = \tilde{\alpha}_j + x'_{rdij}\tilde{\beta}_j + \tilde{u}_{dj}$ for j = 1, 2. The new estimator is called BLUP-type because it is neither linear nor unbiased.

It is often the case that variance components parameters defining the matrix **Φ** are unknown and have to be estimated from the sample data. A further step could follow the step (a) to obtain penalised likelihood estimates of the variance components. However, these estimates of are negatively biased and they are not recommended in the practice; see McGilchrist (1994). The maximum likelihood (ML) and residual/restricted maximum likelihood (REML) are two important approaches in estimating variance components. The biases in estimating variance components are very small by REML method.

4. EMPIRICAL BEST LINEAR UNBIASED PREDICTOR-TYPE

For normal error model, the interrelationship between BLUP (PL) with maximum likelihood (ML) and restricted or residual maximum likelihood (REML) estimators was developed in Harville (1977) and investigated further in Thompson (1980), Fellner (1986, 1987) and Speed (1991). McGilchrist (1994) extends this approach to generalized linear mixed models. This method has elements in common with Schall (1991), Breslow and Clayton (1993), Wolfinger (1993), Nelder and Lee (1996) and Saei and McGilchrist (1998). Lee and Nelder (2001a, 2001b) further extended the work of Nelder and Lee (1996) to correlated non-normal data. Below we outline the extension of this approach to the multinomial logistic linear mixed model with category-specific area random components. In this case, the PL estimates and predictors $\tilde{\alpha}, \tilde{\beta}$ and $\tilde{\mathbf{u}}$ are used as an initial step in finding ML and REML estimates of ϕ_j via Anderson (1973) and Henderson (1973b) algorithm. The iterative procedure used to obtain the ML and REML estimators and their approximate variance-covariance matrices can be specified as follows:

(b) (a) Starting from initial values α_0, β_0, $\mathbf{u}_0$ and ϕ_{j0} (hence Φ_0) successive iterations are obtained by finding changes $\Delta\alpha$, $\Delta\beta$ and $\Delta\mathbf{u}$ to the current estimates from the equations

$$\mathbf{V}\begin{bmatrix}\Delta\alpha\\ \Delta\beta\\ \Delta\mathbf{u}\end{bmatrix} = X^{*T}\frac{\partial l_1}{\partial\tau_0} - \begin{bmatrix}\mathbf{0}\\ \Phi_0^{-1}\mathbf{u}_0\end{bmatrix}$$

This is the PL estimation equation and its elements are described in previous section.

(c) Once iterations of (a) have converged to $\tilde{\alpha}, \tilde{\beta}$ and $\tilde{\mathbf{u}}$, let $\tilde{\tau} = \Delta\tilde{\alpha} + \mathbf{X}\tilde{\beta} + \mathbf{Z}\tilde{\mathbf{u}}$, $\mathbf{B}_s = -\partial^2 l_1/\partial\tilde{\tau}_s\partial\tilde{\tau}_s^T$, $a_1 = \mathrm{tr}(\mathbf{T}_{11}^*) + \tilde{\mathbf{u}}_1^T\tilde{\mathbf{u}}_1$, $a_2 = \mathrm{tr}(\mathbf{T}_{22}^*) + \tilde{\mathbf{u}}_2^T\tilde{\mathbf{u}}_2$ and $a_3 = 2(\mathrm{tr}(\mathbf{T}_{12}^*) + \tilde{\mathbf{u}}_1^T\tilde{\mathbf{u}}_2)$ and

$\mathbf{T}^* = [\,\mathbf{T}_{jk}^*\,] = [\,\Phi_0^{-1} + \mathbf{Z}^{\mathrm{T}}\mathbf{B}_s\mathbf{Z}]^{-1}$.

The estimates of ϕ_j are then given by

$$\hat{\phi}_1 = \frac{\left[a_1 + \left(\frac{\phi_{30}}{\phi_{20}}\right)^2 a_2 + \left(\frac{D\phi_{30}^2}{\phi_{20}}\right) - \left(\frac{\phi_{30}}{\phi_{20}}\right)a_3\right]}{D}$$

$$\hat{\phi}_2 = \frac{\left[a_2 + \left(\frac{\phi_{30}}{\phi_{10}}\right)^2 a_1 + \left(\frac{D\phi_{30}^2}{\phi_{10}}\right) - \left(\frac{\phi_{30}}{\phi_{20}}\right)a_3\right]}{D} \quad (7)$$

$$\hat{\phi}_3 = \frac{\phi_{10}\phi_{20}a_3}{D\phi_{30}^2 - a_3\phi_{30} + \phi_{20}a_1 + \phi_{10}a_2 - \phi_{10}\phi_{20}D}$$

(d) The preceding two (a) and (b) steps are then repeated, with initial values set to $\tilde{\alpha}, \tilde{\beta}$ $\tilde{\mathbf{u}}$, and $\hat{\phi}_j$.

At convergence, $\hat{\alpha}, \hat{\beta}$ and $\hat{\phi}_j$ are the ML estimates of fixed parameters α, β and variance components ϕ_j. At this stage the ML-based empirical best linear unbiased predict-type (EBLUP-type) of θ is

$$\hat{\theta} = A_s y_s + A_r\hat{y}_r \quad (8)$$

$$= A_s y_s + A_r\{(N_{di} - n_{di})\exp(\hat{\tau}_{rdij})(1 + \sum_{j=1}^{2}\exp(\tilde{\tau}_{rdij}))^{-1}\}$$

where $\hat{\tau}_{rdij} = \hat{\alpha}_j + x'_{rdij}\hat{\beta}_j + \hat{u}_{dj}$ for j = 1, 2. The $\hat{\alpha}_j$, $\hat{\beta}_j$ and $\hat{u}_{dj}$ are the final values of the α_j, β_j and u_{dj} in the iteration processes (a) and (b).

Remember that estimate values for the last category (employed) can be obtained by subtraction of sum of the inactive and unemployed from total N_{di}.

The ML estimators $\hat{\phi}_{(\mathrm{ML})} = (\hat{\phi}_{1(\mathrm{ML})}, \hat{\phi}_{2(\mathrm{ML})}, \hat{\phi}_{3(\mathrm{ML})})$ have asymptotic variance-covariance matrix

$$\mathrm{var}(\hat{\phi}_{(\mathrm{ML})}) = 2\begin{bmatrix} r_{111} + r_{211} - 2r_{311} & r_{112} + r_{212} - 2r_{312} & r_{113} + r_{213} - 2r_{313} \\ & r_{122} + r_{222} - 2r_{322} & r_{123} + r_{223} - 2r_{323} \\ & & r_{133} + r_{233} - 2r_{333} \end{bmatrix}^{-1} \quad (9)$$

where

$$r_{1jj'} = \mathrm{tr}\left[\mathbf{\Phi}(\partial\mathbf{\Phi}^{-1}/\partial\phi_j)\,\mathbf{\Phi}(\partial\mathbf{\Phi}^{-1}/\partial\phi_{j'})\right]$$

$$r_{2jj'} = \mathrm{tr}\left[\mathbf{T}^*(\partial\mathbf{\Phi}^{-1}/\partial\phi_j)\,\mathbf{T}^*(\partial\mathbf{\Phi}^{-1}/\partial\phi_{j'})\right]$$

$$r_{3jj'} = \mathrm{tr}\left[\mathbf{\Phi}(\partial\mathbf{\Phi}^{-1}/\partial\phi_j)\,\mathbf{T}^*(\partial\mathbf{\Phi}^{-1}/\partial\phi_{j'})\right]$$

for j, j' = 1, 2, 3

where tr( ) denotes trace of the matrix inside the bracket. The derivatives of the inverse of variance-covariance matrix $\mathbf{\Phi}$ are given in Appendix A.

Let $\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$ and $\mathbf{V}^{-1} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix}$ denote the partitions of the matrix $\mathbf{V}$ and its inverse corresponding to the dimensions of fixed effects (α, β) and random effect $\mathbf{u}$. Replacing $\mathbf{T}^*$ by $\mathbf{T}_{22}$ in (7) yields the REML estimates $\hat{\alpha}, \hat{\beta}$ and $\hat{\phi}_{j(\text{REML})}$ of α, β and ϕ_j. This leads REML-based empirical best linear unbiased predict-type (EBLUP-type) estimates in (8) and variance-covariance for the REML estimators $\hat{\phi}_{(\text{REML})} = (\hat{\phi}_{1(\text{REML})}, \hat{\phi}_{2(\text{REML})}, \hat{\phi}_{3(\text{REML})},)$ in (9).

An important measure of the "quality" of a statistical estimator is its mean square error (MSE), and so it is important that any small area estimate be accompanied by an estimate of its MSE. Here the prediction error of $\hat{\theta}$ is $\hat{\theta} - \theta = \mathbf{a}_r(\hat{\mathbf{y}}_r - \mathbf{y}_r) = \mathbf{a}_r(h(\hat{\tau}_r) - \mathbf{y}_r)$. Replacement of $\hat{\tau}_s$ by $\hat{\tau}_r$ and n_{di} by $N_{di} - n_{di}$ in (6) yields $h(\hat{\tau}_r)$. Let $\tilde{\tau}_r$ represents BLUP-type counterpart of the EBLUP-type $\hat{\tau}_r$. The prediction error of $\hat{\theta}$ can be further simplified to $\hat{\theta} - \theta = \mathbf{a}_r(h(\hat{\tau}_r) - \mathbf{y}_r) = \mathbf{a}_r(h(\hat{\tau}_r) - h(\tilde{\tau}_r) + h(\tilde{\tau}_r) - \mathbf{y}_r)$. The mean cross-product error matrix is therefore MCPE $(\ddot{\boldsymbol{\theta}})$ $= E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})']$.

For an identity link function (linear mixed model), following the work of Kackar and Harville (1984), Prasad and Rao (1990) introduced an analytical expression for the MSE estimate. Datta and Lahiri (2000) and Saei and Chambers (2003) provided MSE components under ML and REML estimates of variance components. Singh *et al.* (2005) extended MSE estimates to the correlated area random effects models. Saei and Chambers (2003) gave an approximate MSE estimate of the small area estimators under a generalized linear mixed model. A similar approach was used in estimating MSE of the LFS estimates by Molina *et al.* (2007).

While the above approaches are concentrated on deriving an analytical expression for the MSE estimate, an alternative method has been introduced by using resampling techniques. Jiang *et al.* (2002) proposed a jackknife methodology for estimation under generalized linear mixed models. Pfeffermann and Tiller (2005) proposed a parametric and a non-parametric bootstrap estimator of mean prediction errors under state space models. Hall and Maiti (2006) proposed a double-bootstrap approach for bias correction, which is applicable for constructing bias-corrected estimators of the mean squared error and for computing prediction regions under general settings. Molina *et al.* (2007) also used a bootstrap technique under a multinomial logistic mixed model. In this paper, we report mean square estimates by using an analytical approximation approach.

5. SIMULATION RESULTS

A limited simulation study was undertaken to examine the performance of the model and method. The survey variable is a multinomial with three categories (trinomial), $J = 3$, and focus is the estimation of small area totals for all three categories. The values $\mathbf{u} = [u_{11}, u_{21}, ..., u_{D1}, u_{D2}]'$ were generated from a multivariate normal distribution with zero mean vector and variance-covariance matrix

$$\boldsymbol{\Phi} = \mathbf{I}_D \otimes \begin{bmatrix} \phi_1 & \phi_3 \\ \phi_3 & \phi_2 \end{bmatrix}$$

The x_{di} values were randomly assigned to values of 0 and 1 and were also kept fixed throughout the simulations. These $\mathbf{u}$ and x_{di} along with α_j and β_j values were used to generate $\tau_{dij} = \alpha_j + x_{di}\beta_j + u_{dj}$. The values τ_{dij} were used to generate population values using model (3) for $D = 40$ areas. The generated population values were 1, 2 and 3 according to first, second and third response categories. Random samples of size n_d, from N_d, were taken with n_d increasing with d. The sample size varied from 4 to 61. These sample sizes were according to sampling ratios range of 0.0005 to 0.0035. The sample size was kept fixed for all 40 areas throughout the simulations. The y_{rdij} and y_{sdij} are then obtained by aggregating non-sample and sample values over combination of the area, $x(0, 1)$ and response category ($J = 3$). The sample data were used by two models, new model and Molina *et al.* (2007), in estimating the model parameters via REML. The estimated parameters by two models were then used to obtain two different sets of estimates. For the last response category, the small area total was calculated by subtraction from total population for given area. The totals for first two categories were obtained by

$$\hat{\theta} = A_s y_s + A_r \hat{y}_r$$
$$= A_s y_s + A_r \{(N_{di} - n_{di}) \exp(\hat{\tau}_{rdij})(1 + \sum_{j=1}^{2} \exp(\hat{\tau}_{rdij}))^{-1}\}$$

For a given area, the $\hat{\theta}$ is a vector of the estimated values for the first and second category totals. The process of generating population and sample data, estimation of model parameters and calculation of $\hat{\theta}$ was independently replicated 1000 times. For each set of estimates $\hat{\theta}$ and for each small area d, we then calculated the actual and average estimated mean squared errors

$$ActMSE_d = \text{diag}_d\left(\sum_{k=1}^{1000} (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k)' (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k)/1000\right)$$

$$EstMSE_d = \text{diag}_d\left(\sum_{k=1}^{1000} \widehat{MCPE}(\hat{\boldsymbol{\theta}}_k)/1000\right)$$

where diag_d (**X**) denotes the d^{th} element of the main diagonal of **X**. Because of limited space, the results were reported for just last category in model, category 2. The actual coefficient of variation

$$ActCV_d = 100 \times \frac{\sqrt{ActMSE_d}}{\sum_{k=1}^{1000} \boldsymbol{\theta}_{d2k}/1000}$$

and the estimated coefficient of variation

$$EstCV_d = 100 \times \frac{\sqrt{EstMSE_d}}{\sum_{k=1}^{1000} \hat{\boldsymbol{\theta}}_{d2k}/1000}$$

were then calculated, as was the average coverage of the area d total by the nominal 95% confidence intervals. The 2 in denominators indicates the second category in the above equations.

We have also generated population data from the model in Molina *et al.* (2007), i.e. a multinomial logistic linear mixed model, with linear predictor was defined by

$$\tau_{dij} = \alpha_j + x_{di}\beta_j + u_d.$$

This model was called a single random effect model. The values $\mathbf{u} = [u_1, u_2, ..., u_D]'$ were generated from a multivariate normal distribution with zero mean vector and variance-covariance matrix $\boldsymbol{\Phi} = \varphi \mathbf{I}_D$. The whole processes of generating, drawing samples and calculating area estimates in previous paragraph were repeated except that the estimation was carried out only by single random effect model.

Four and three different combinations of variance components were used in the simulations under a bivariate and single random effect models respectively. The variance components (ϕ_1, ϕ_2, ϕ_3) were (1, 0.25, 0.4), (0.25, 1, 0.4), (1, 1, 0.4) and (0.25, 0.25, 0.1) for bivariate random component model. For single random effect model, three different simulation values of φ were 0.5, 1.5 and 1 respectively. A single set of regression coefficients was used in both bivariate and single random effect models, i.e. $\alpha_1 = 1$, $\alpha_2 = 0.5$, $\beta_1 = 0.5$ and $\beta_2 = -0.5$. The simulation results are presented from a single random effect first.

Fig. 1 shows the average values of both the actual coefficient variation (*ActCV*), estimated coefficient of variation (*EstCV*) and estimated 95% coverage in estimating totals for the second category. The horizontal axis is sorted in ascending order of the area sample sizes. The results show that estimated CVs are in very close agreement with their actual values. They decrease as area sample sizes increases. The estimated 95% coverages are very close to 95% nominal value for all 40 areas regardless of their sample sizes. A similar conclusion was obtained in the simulation study by Molina *et al.* (2007). This does show that the adopted

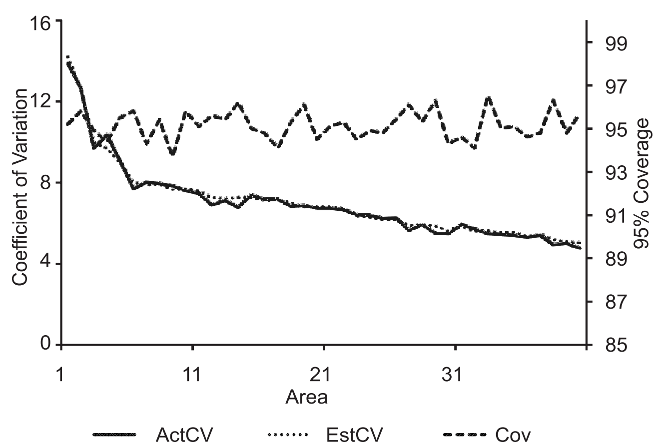

Fig. 1. Coverage, actual and estimated coefficient of variations, *ActCV* and *EstCV*.

mean square estimation approach works reasonably well.

Fig. 2 shows the coefficient of variations for the total estimators by application of Molina *et al.* (2007) (a single random effect model) to data from model (10). These show that for large variance components in particular, estimated CVs are far from their actual values, irrespective of the area sample size. Although the difference between *ActCV* and *EstCV* is reduced for small variance components, the differences are not small and *EstCV*s are even less than 50% of the *ActCV*s. This problem persists, albeit in a somewhat reduced form, with the total estimators of the two other categories, first and last (not reported here).

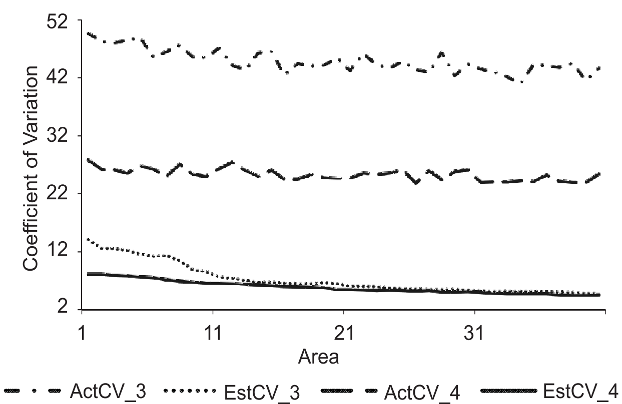

Fig. 2. Actual and estimated coefficient of variations, *ActCV* and *EstCV,* X_i denotes X for the simulation data i.

The bivariate area random components model (proposed model in this paper) performs much better in this regard, with estimated and actual CVs for the all variance component combinations being very close. These, *ActCV* and *EstCV* values, are shown in Fig. 3. As expected, CV decreases by increasing area sample size.

The results, not reported here, show that there is also a gain in efficiency from modelling correlation between bivariate area random effects. These gains are more pronounced for areas with small size. Irrespective of potential increases in efficiency, an important gain by a bivariate area random component is better estimation of mean squared error. Molina *et al.* (2007) generally leads to severe underestimate MSE estimation. An implication of this is a poor coverage and records coverages very far from the nominal 95% level. In contrast, a bivariate area random component coverages are very close to the 95% level. Furthermore,

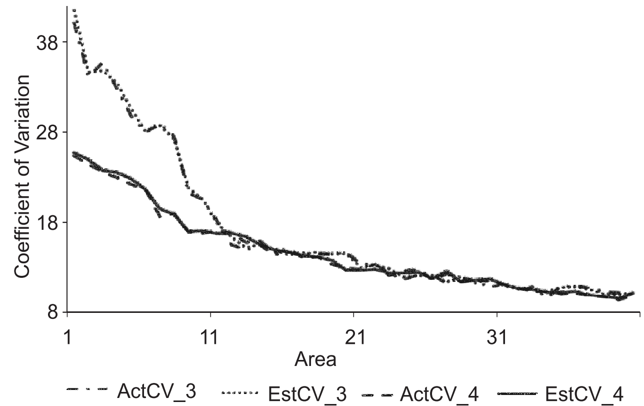

Fig. 3. Actual and estimated coefficient of variations, *ActCV* and *EstCV,* X_i denotes X for the simulation data set i.

this overall good performance holds across all sets of the parameter values investigated. Fig. 4 shows 95% coverage by both bivariate area random component and single random effect.

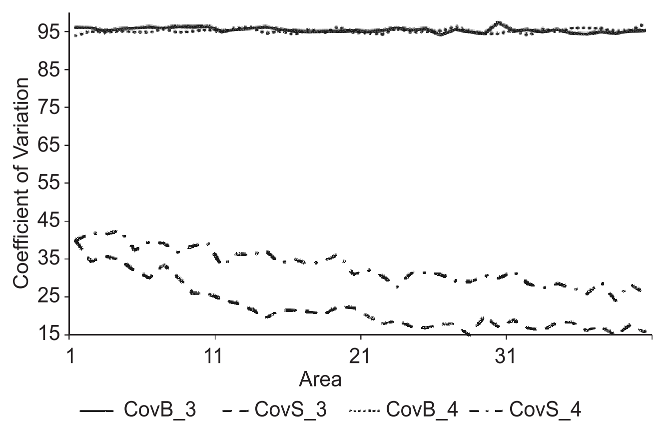

Fig. 4. Coverage of nominal 95% confidence intervals (*95% Coverage*) generated by Bivariate (B) and Single (S) random effect model

6. APPLICATION TO UK LFS DATA

In this section we illustrate the preceding model and method to the same UK LFS data in Molina *et al.* (2007). The LFS is the largest regular household survey in Great Britian (GB) and it is the main source of information on unemployment. The focus of this application is estimating unemployment, inactive and employee for the primary areas for UK resource allocation, Unitary Authorities and Local Authority Districts (UA/LADs). The data are available at six age and sex (agesex) categories for each of 406 LADs. There are 12 government regional offices (GOR) and 7 clusters. The cluster is a socio-economic classification and seven clusters are rural areas, urban fringe, coast and services, prosperous England, mining, manufacturing and industry, education centres and outer London, and inner London. The number of registered unemployed is the other variable that we are going to use in modelling. As we expected, this is an important predictor for unemployment and Fig. 5 confirms that by showing a positive correlation between logit of proportion of unemployed and logit of registered unemployed. The number of registered unemployed is given at six age and sex combinations/groups for each LAD areas. The sample size varies from 2 to 633 for LAD-age-sex combination. There is also a single LAD-age-sex combination of zero sample size. The sample size ranges from 52 to 2702 at LAD level.

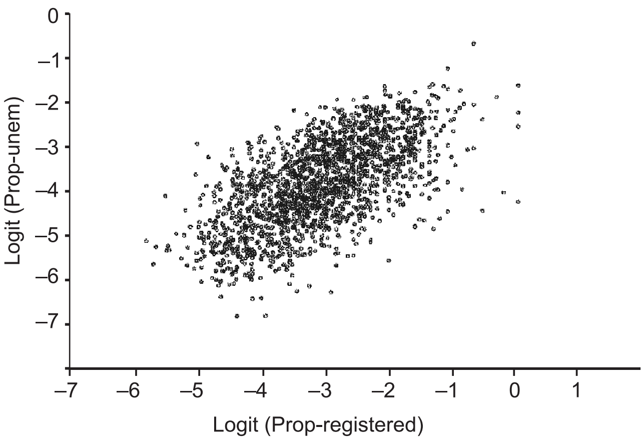

Fig. 5. Scatter graph of logit of the proportion of unemployed vs logit of the proportion of registered unemployed.

Table 1 sets out the parameter estimates and associated standard errors for two multinomial logistic linear models fitted to these data, with the response variable in each case corresponding to the number of inactive, unemployed and employed people observed on each LAD-agesex combination. Model 1 and 2 are multinomial logistic linear models with bivariate and single LAD random effect respectively. A third model relaxes the dependence assumption in Model 1 and assumes LAD effects for inactive and unemployed are independent. The parameter estimates are not reported here for this model. All three models are fitted using REML.

Examination of the results for Model 1 show that there is statistically significant variation between LADs for inactive category ($\hat{\varphi}_1$ = 0.02 with an estimated standard error of 0.003). This between LADs variation is also significant for unemployment category ($\hat{\varphi}_2$ = 0.016 with an estimated standard error of 0.006). Furthermore, the two LAD random effects are correlated with estimated covariance of $\hat{\varphi}_3$ = 0.008 and estimated standard error of 0.003. The estimated correlation between two LAD random effects is $\hat{\rho}$ = 0.45. Both other models, 2 and 3, support the conclusions by Model 1 on LAD effect that is a statistically significant variation between LADs. In term of the fixed effects/regression coefficients, there is not much choice between three models. The differences between regression coefficient estimates are negligible or very small. The gender, both covariates associated with the registered unemployed, registered unemployed and registered unemployed total, are statistically significant. Under Model 1, the estimated values are 0.053 and 0.383 for registered unemployed and registered unemployed total respectively. The estimates are both positive, and this backs up the relationship between number of unemployed and registered unemployed in Fig. 5. In contrast, there is no evidence of an interaction between age and sex, with Wald statistics $(\hat{\boldsymbol{\beta}}^T [\mathrm{var}(\hat{\boldsymbol{\beta}})]^{-1} \hat{\boldsymbol{\beta}})$ of 3.5 and 3.7 for the inactive and unemployed categories respectively. A part from the government office region (GOR) for unemployed category, all other covariates in the model are playing important role on the inactive and unemployed total variations. For the category inactive, Wald statistics are 91.2, 39.5, 147 and 39 for age, GOR, interaction between age and registered unemployed and cluster (Socio-economic classification) respectively. Similar Wad statistics are 7.7, 12.2, 18.3 and 38 for the category unemployed. These conclusions are also supported by other two models. Under Model 2, the Wald statistics for inactive and unemployed categories are 88.8, 10.1, 41.9, 150 and 8.54, 11.2, 16.4, 36.9.

The estimated coefficients and predicted values of area (LAD) random effects are used to calculate EBLUP-type estimate of totals number of inactive and unemployed people for each 406 LAD in UK. This is a composite estimate, a combination of observed values in sample and predicted values of non-sample. The EBLUP-type estimates of the reference category, employed, are obtained by substruction from LAD total population. The performances of three models are compared in estimating of the totals inactive and unemployed. Although the results in previous paragraph supported a category-specific LAD random effect, the small variance component estimates did not make a big changes in estimating regression coefficients and associated standard errors by Model 1 from other two

Table 1. REML parameter estimates (est), standard errors (se) and t (t = est/se) for bivariate area random component and single random effect models fitted to LFS data.

		Model 1						Model 2					
		Inactive			Unemployment			Inactive			Unemployment		
		Est	SE	Est/SE	Est	SE	Est/SE	Est	SE	Est/SE	Est	SE	Est/SE
	ϕ_1	0.020	0.003	7.873				0.018	0.002	7.824			
	ϕ_2	0.016	0.006	2.718									
	ϕ_3	0.008	0.003	2.789									
Inact	α_1	1.131	0.163	6.936				1.133	0.158	7.164			
Unem	α_2	0.391	0.386	1.013				0.316	0.383	0.826			
Sex	β_1	–0.054	0.171	–0.318	–1.834	0.750	–2.445	–0.084	0.170	–0.492	–1.729	0.749	–2.309
Age	β_2	–1.020	0.155	–6.586	–0.348	0.380	–0.916	–0.995	0.154	–6.444	–0.382	0.379	–1.009
	β_3	–1.436	0.151	–9.541	–0.773	0.361	–2.142	–1.410	0.150	–9.397	–0.824	0.359	–2.293
Age*Sex	β_4	0.266	0.236	1.127	1.430	0.797	1.793	0.286	0.236	1.213	1.342	0.796	1.687
	β_5	–0.167	0.222	–0.754	0.938	0.795	1.180	–0.159	0.222	–0.716	0.873	0.794	1.100
	β_6	0.006	0.043	0.144	–0.035	0.079	–0.436	0.006	0.041	0.145	–0.028	0.079	–0.352
	β_7	0.071	0.038	1.858	–0.016	0.073	–0.216	0.071	0.037	1.924	–0.008	0.073	–0.110
	β_8	0.094	0.062	1.510	0.157	0.098	1.599	0.096	0.060	1.604	0.160	0.099	1.625
	β_9	0.217	0.084	2.577	–0.013	0.118	–0.113	0.210	0.080	2.631	0.022	0.118	0.186
	β_{10}	0.049	0.058	0.856	0.131	0.091	1.442	0.046	0.055	0.830	0.155	0.090	1.725
GOR	β_{11}	0.031	0.043	0.719	0.006	0.075	0.082	0.028	0.041	0.675	0.018	0.074	0.247
	β_{12}	–0.060	0.050	–1.212	0.035	0.080	0.442	–0.060	0.047	–1.259	0.031	0.079	0.394
	β_{13}	–0.025	0.041	–0.607	0.035	0.076	0.461	–0.027	0.040	–0.668	0.043	0.076	0.573
	β_{14}	0.173	0.052	3.356	0.088	0.086	1.030	0.168	0.049	3.396	0.107	0.086	1.247
	β_{15}	–0.043	0.045	–0.962	0.075	0.077	0.967	–0.037	0.043	–0.874	0.057	0.076	0.749
	β_{16}	–0.023	0.051	–0.440	–0.066	0.082	–0.814	–0.028	0.049	–0.568	–0.033	0.080	–0.409
	β_{17}	0.102	0.035	2.936	0.005	0.064	0.079	0.103	0.033	3.095	0	0.064	–0.003
	β_{18}	0.105	0.038	2.748	0.118	0.064	1.835	0.107	0.036	2.938	0.118	0.064	1.849
Cluster	β_{19}	–0.025	0.035	–0.695	0.072	0.066	1.088	–0.022	0.034	–0.660	0.058	0.066	0.884
	β_{20}	0.202	0.036	5.621	0.213	0.058	3.670	0.205	0.034	5.964	0.203	0.058	3.518
	β_{21}	0.016	0.060	0.269	–0.017	0.092	–0.186	0.016	0.058	0.272	–0.011	0.092	–0.119
	β_{22}	0.154	0.085	1.818	0.069	0.125	0.553	0.152	0.081	1.879	0.092	0.125	0.733
Reg	β_{23}	–0.184	0.039	–4.725	0.503	0.106	4.753	–0.194	0.039	–5.003	0.517	0.105	4.909
Reg*Age-	β_{24}	–0.083	0.034	–2.464	–0.368	0.150	–2.450	–0.086	0.034	–2.564	–0.351	0.150	–2.345
Sex	β_{25}	–0.003	0.043	–0.069	–0.307	0.099	–3.102	0.001	0.043	0.027	–0.312	0.099	–3.167
	β_{26}	0.428	0.043	10.020	–0.068	0.092	–0.736	0.433	0.043	10.160	–0.080	0.092	–0.866
	β_{27}	0.104	0.061	1.695	0.275	0.174	1.585	0.106	0.061	1.729	0.262	0.173	1.510
	β_{28}	–0.364	0.055	–6.661	0.047	0.167	0.280	–0.364	0.055	–6.665	0.039	0.166	0.232
Total-Reg	β_{29}	0.437	0.044	10.040	0.383	0.083	4.584	0.448	0.042	10.570	0.349	0.082	4.236

models. Results on Table 1 indicate that classical model test is unable to discriminate between three models if the aim is the inference on regression coefficients. In this case, a bivariate random component model gains a little over a simple single random effect model, Molina, *et al.* (2007). Moreover, the gain in efficiency would be negligible or small by a bivariate random component model, Model 1, if the predicted values of LAD random effects are excluded in estimation small area totals. However, in the simulation study in previous section, the small area estimates did not behave very well when a single random effect model, Molina *et al.* (2007), was applied to the data generated by a bivariate random component models. This poor performance was evident in all different variance component combinations in particular for large variance components. There was also evident of more gain in efficiency by modelling the correlation between category-specific area random effects. Under Model 2, the standard errors, in particular for the category unemployed, are severely underestimated. The average estimated the mean squared errors are 2788604.23 and 162737.80 for the totals inactive and unemployed estimators respectively. Model with independent category-specific area random effects overestimates the mean squared errors with the estimated average values of 3323591.44 and 403398.80 for the totals of inactive and unemployed estimators. The averaged ratios between two mean squared error estimates by Model 2 and Model 1 are 0.87 and 0.54 for categories inactive and unemployed respectively. These values are 1.023 and 1.274 in comparing Model 1 with the third model (model with independent category-specific area random effects). Fig. 6 presents mean squared error ratios in estimating the totals of inactive and unemployed for all 406 LADs. The horizontal axis is sorted by LAD total population. Fig. 6(a) shows the ratios between two mean squared errors by Model 2, Molina *et al.* (2007), and Model 1. While these ratios are almost all over 80% for the inactive category, they are less than 50% for a significant number of LADs for the unemployed category. Under Model 2, the underestimation of mean square errors is more pronounced for large LADs. The impact of modelling the correlation between categoryspecific area random effects on the mean square error estimates are depicted in Fig. 6(b). These

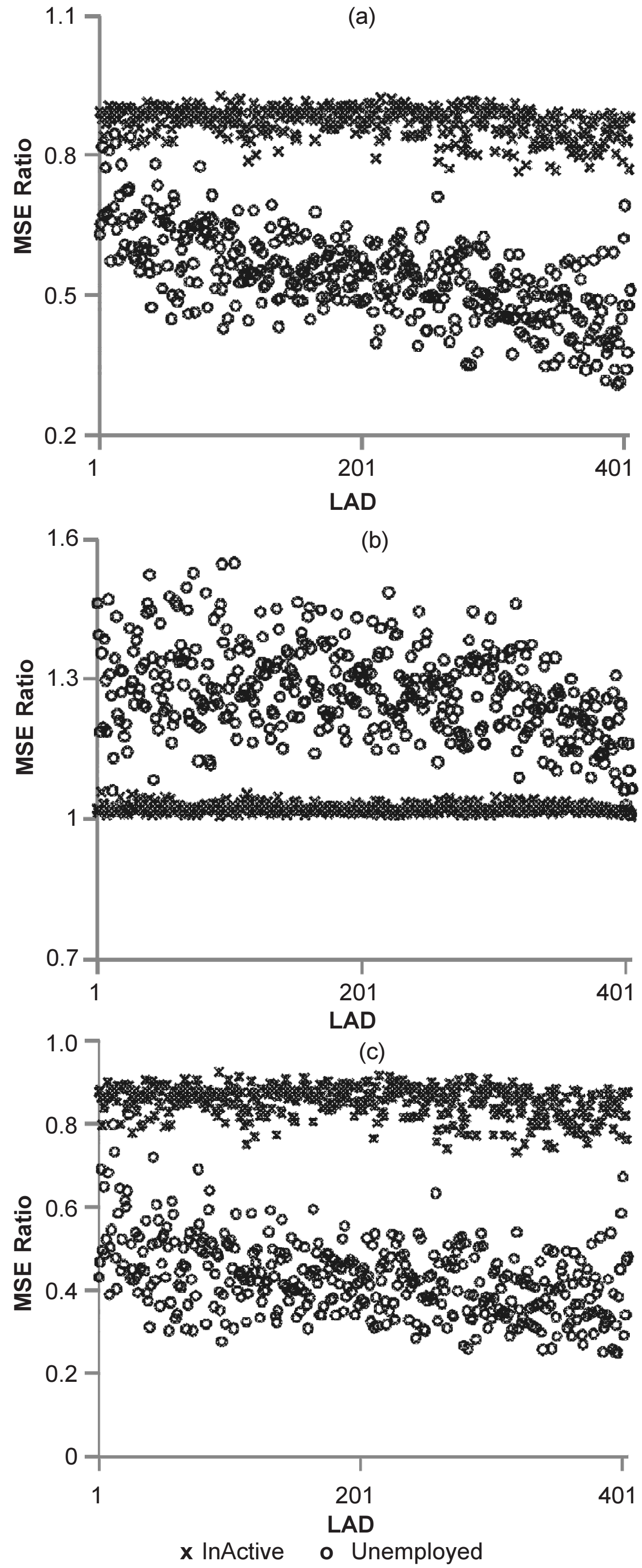

Fig. 6. Ratio of MSE estimates (a) Model 2 over Model 1 (b) independent category-specific area random effects model over Model 1 (c) Model 2 over independent category-specific area random effects model.

are ratios between two mean squared errors by independent category-specific area random effects and Model 1. For the inactive category, category with large values, the MSE ratios pattern is similar to the pattern in Fig. 6(a). The LAD total population has a little or no influence on the MSE ratios and the pattern is almost flat. The maximum gain in efficiency is 6% by modelling the correlation between category-specific area random effects. As in Fig. 6(a), the performance of a bivariate random components model more pronounced in estimating total of category unemployed, category with small values. However, the pattern is more promising and it is different to that in Fig. 6(a). The MSE ratios are decreasing by increasing total LAD population and are approached to 1 for large of the LADs. A large number of the LADs gain more than 30% in efficiency by modelling the correlation between category-specific area random effects. The maximum gain is 55%, and there are 25 LADs with the gain more than 45% in efficiency.

Although not presented here, in the simulation study, the performance of independent category-specific area random effect model was much better than a single random effect model when they applied to the multinomial data that were generated by a bivariate random components model. For the category inactive, the ratios between two mean squared error estimates, independent category-specific area random effect model and Model 1, follows the same pattern as in Fig. 6(a) with the values slightly smaller. The category unemployed shows a major change. The MSE ratios are not sharply decreasing by increasing total LAD population for the category unemployed. Fig. 6 about here

7. CONCLUSIONS

In this paper we introduce a multinomial logistic linear mixed model with a bivariate random component for estimating totals of inactive, unemployed and employed people at subregional level. The model allows the area random component to be varied by the response category (category-specific area random effects). It also assumes that the area random effects for two categories unemployed and inactive are correlated. In particular, bivariate random effects are embedded into the multinomial logistic linear model, where area random effects in both unemployed and inactive are assumed to follow a bivariate normal distribution. This model is compared with two other models, independent category-specific random effect model and a single random component model, Molina *et al.* (2007). In application to UK LFS data, the bivariate random components model indicates the existence of statistically significant LAD heterogeneity for both inactive and unemployed categories. The two random components are also significantly correlated. Although the estimated variance components are small, the gain in estimation efficiency in particular for the category unemployed is not ignorable. The proposed model benefits from both category-specific LAD random effect and correlation between them. If both of them are ignored, Molina *et al.* (2007), the mean square errors especially for the category unemployed, a category with small values, are severely underestimated. The interesting is that the poor performance increases by increasing LAD total population.

Results from a small scale simulation support this conclusion, in the sense that the mean squared error estimates of the total estimators, in particular for categories with small values, are underestimated if the category-specific area random components are ignored when modelling data that includes such effects. Our proposed model has the potential to lead to substantial increases in prediction efficiency of total estimators for small areas when there is considerable variation associated with category-specific area random effects. The results also show that the estimates of mean square error calculated under the model with correlated category-specific area random effects are much more accurate than those based on the model with the independent category-specific area random effects. As a consequence, confidence intervals based on these estimates of mean square error tend to be more accurate, in the sense of achieving their nominal level of coverage.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to the referee for carefully reading and making very useful comments for improving the paper. This work was supported by the ONS via S3RI under the Small area estimation for labour market statistics, SAE1, project.

REFERENCES

Anderson, T.W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.,* **1**, 135-141.

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

Cruddas, M. and Saei, A. (2006). Developing model-based estimates of unemployment for local authority districts. *The Annual Conference of the Royal Statistical Society, Queen's University of Belfast*.

Curtis, D. and Saei, A. (2005). EBLUP-type estimation of local authority unemployment. *ASC2005, Maximizing Data Value Use and Re-Use, Newland Park.*

Datta, G.S. and Lahiri, P. (2000). A unified measure of uncertainty of estimates for best linear unbiased predictors in small area estimation problem. *Statist. Sinica*, **10**, 613-627.

Fellner, W.H. (1986). Robust estimation of variance components. *Technometrics*, **28**, 51-60.

Fellner, W.H. (1987). Spare matrices, and the estimation of variance components by likelihood methods. *Comm. Statist. Simulation,* **16**, 439- 463.

EURAREA Consortium (2004). EURAREA Project IST-2000-26290. (Available from http://www.statistics.gov.uk/eurarea.)

Hall, P. and Maiti, T. (2006). On parametric bootstrap methods for small area prediction. *J. Roy. Statist. Soc.* **B68**, 221-238.

Hastings, D., Maine, N., Brown, G. and Crudas, M. (2003). Development of improved estimation methods for local area unemployment levels and rates. In: *Technical Report, Labour Market Trends*, 37-43. London: Office for National Statistics.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.*, **72**, 320-340.

Henderson, C.R. (1963). Selection index and expected genetic advance. In: *Statistical Genetics and Plant Breeding* (W.D. Hanson and H.F. Robinson, eds.), 141-163. National Academy of Sciences and National Research Council Publication No. 982, Washington, D.C.

Henderson, C.R. (1973a). Sire evaluation and genetic trends. In: *Proceedings of the Animal Breeding and Genetics Symposium in Honour of Dr. Jay L. Lush* 10-41. Amer. Soc. Animal Sci.-Amer. Dairy Sci. Assn.-Poultry Sci. Assn. Champaigne, Illinois.

Henderson, C.R. (1973b). Maximum likelihood estimation of variance components. Unpublished manuscript.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under selection model. *Biometrics*, **31**, 423-447.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, **15**, 1-96.

Jiang, J., Lahiri, P. and Wan, S. (2002). Aunified jackknife theory for empirical best prediction with *M*-estimation. *Ann. Statist.*, **30**, 1782-1810.

Kacker, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimations of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.*, **79**, 853-862.

Lee, Y. and Nelder, J.A. (2001a). Hierarchical generalized linear models: A synthesis of generalised linear models, random-effect model and structured dispersions. *Biometrika*, **88**, 987-1006.

Lee, Y. and Nelder, J.A. (2001b). Modelling and analysing correlated non-normal data. *Statist. Model.*, **1**, 3-16

Longford, N.T. (2007). On standard errors of model-based small area estimators. *Survey Methodology*, **33**, 69-79.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*. Second Ed. Chapman and Hall, London.

McGilchrist, C.A. (1994). Estimation in generalized mixed models. *J. Roy. Statist. Soc.*, **B56**, 61-69.

Molina, I., Saei, A. and Lombardía, M.J. (2007). Small area estimates of labour force participation under a multinomial logit mixed model. *J. Roy. Statist. Soc.,* **A**, 975–1000.

Nelder, J.A. and Lee, Y. (1996). Hierarchical generalized linear models. *J. Roy. Statist. Soc.*, **B58**, 619-678.

Pfeffermann, D. and Tiller, R. (2005). Bootstrap approximation to prediction MSE for statespace models with estimated parameters. *J. Time Series Anal.*, **26**, 893-916.

Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of smallarea estimators. *J. Amer. Statist. Assoc.*, **85**, 163-171.

Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, New York.

Saei, A. and McGilchrist, C. (1998). Longitudinal threshold models with random components. *The Statistician* (*J. Roy. Statist. Soc.*), **D47**, 365-375.

Saei, A. and Chambers, R. (2003). Small area estimation under linear and generalized linear mixed models with time and area effects.*Working Paper M03/15*. Southampton Statistical Sciences Research Institute, University of Southampton, Southampton.

Saei, A., Zhang, L.C. and Chambers, R.L. (2005). Generalized structure preserving estimation models for small areas. *Statist. Trans.*, **7(3)**, 685-696.

Schall, R. (1991). Estimation in generalised linear models with random effects. *Biometrika,* **78**, 719-727.

Singh, B.B., Shukla, G.K. and Kundu, D. (2005). Spatio-temporal models in small area estimation. *Survey Methodology*, **31**, 183-195.

Speed, T. (1991). Comment on robinson: Estimation of random effect. *Statist. Sci.,* **6**, 42-44.

Thompson, R. (1981). Maximum likelihood estimation of variance components. *Math. Operforch. Statist. Ser. Statist*., **11**, 125-131.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika,* **80**, 791-796.

Zhang, L.C. and Chambers, R.L. (2004). Small area estimates for cross-classifications. *J. Roy. Statist. Soc.,* **B66**, 479-496.

APPENDIX A

The variance-covariance matrix in equation (9) of the Section 3 is completed by determining its derivative elements. Let $\Delta = \phi_1\phi_2 - \phi_3^2$, derivatives of the inverse of variance-covariance matrix $\mathbf{\Phi} = \mathbf{I}_D \otimes \begin{bmatrix} \phi_1 & \phi_3 \\ \phi_3 & \phi_2 \end{bmatrix}$ with respect to φ_j are then

$$\frac{\partial \mathbf{\Phi}^{-1}}{\partial \phi_1} = \partial \frac{\left(\Delta^{-1}\mathbf{I}_D \otimes \begin{bmatrix} \phi_1 & -\phi_3 \\ -\phi_3 & \phi_2 \end{bmatrix}\right)}{\partial \phi_1}$$

$$= \Delta^{-2}\mathbf{I}_D \otimes \begin{bmatrix} -\phi_2^2 & \phi_3\phi_2 \\ \phi_3\phi_2 & -\phi_3^2 \end{bmatrix}$$

$$\frac{\partial \mathbf{\Phi}^{-1}}{\partial \phi_2} = \partial \frac{\left(\Delta^{-1}\mathbf{I}_D \otimes \begin{bmatrix} \phi_2 & -\phi_3 \\ -\phi_3 & \phi_1 \end{bmatrix}\right)}{\partial \phi_2}$$

$$= \Delta^{-2}\mathbf{I}_D \otimes \begin{bmatrix} -\phi_3^2 & \phi_3\phi_1 \\ \phi_3\phi_1 & -\phi_2^2 \end{bmatrix}$$

$$\frac{\partial \mathbf{\Phi}^{-1}}{\partial \phi_3} = \partial \frac{\left(\Delta^{-1}\mathbf{I}_D \otimes \begin{bmatrix} \phi_2 & -\phi_3 \\ -\phi_3 & \phi_1 \end{bmatrix}\right)}{\partial \phi_3}$$

$$= \Delta^{-2}\mathbf{I}_D \otimes \begin{bmatrix} 2\phi_3\phi_2 & -(\phi_1\phi_2 + \phi_3^2) \\ -(\phi_1\phi_2 + \phi_3^2) & 2\phi_1\phi_3 \end{bmatrix}$$

Replacement of the above derivatives into (9) in Section 3 yields asymptotic variance covariance for ML estimators $\hat{\phi}_{(ML)} = (\hat{\phi}_{1(ML)}, \hat{\phi}_{2(ML)}, \hat{\phi}_{3(ML)})^T$.