Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 66(1) 2012 15-30

Small Area Estimation under Additivity Constraints to Published Direct Survey Estimates

Daniel Elazar
Australian Bureau of Statistics
Received 15 February 2011; Revised 31 August 2011; Accepted 01 September 2011

SUMMARY

Users of small area estimates (SAEs) produced by national statistical offices often require that published SAEs are coherent with survey estimates or benchmark totals published at national or state levels. In some countries, this requirement is mandated by legislation. Ensuring the coherence of SAEs with published statistics not only helps reassure users about the quality of the SAEs, it can also help correct for model misspecification bias. In previous small area applications carried out at the Australian Bureau of Statistics (ABS), coherence with broader level published estimates was achieved after the small area estimation process by using a technique such as iterative proportional fitting. However, the disadvantages of this approach were firstly, the lack of integration with the small area estimation process itself, and secondly, the fact that the mean square error estimates did not take proper account of the constraints imposed.

In this paper we use the Lagrange mulitplier method to adapt the penalised quasi likelihood (PQL) approach applied to random effects logistic models, to take account of the additivity contraints placed on SAEs. We trial four levels of survey estimates as benchmark constraints to examine the impact on estimates and relative root MSEs (RRMSEs). We find that survey benchmark estimates with low levels of sampling error have a very small impact on estimates and RRMSEs, whereas finer level benchmarks with high sampling error, result in increased and highly volatile estimates of RRMSEs.

Keywords : Small area estimation, Logistic binomial, Random effects, Benchmarks, Constrained regression, Lagrange multipliers.

1. INTRODUCTION

Consistency and coherence are two important statistical quality dimensions covered in the "7 Dimensions of Quality". Users of small area data based on official statistics expect that the small area estimates provided to them are both consistent and coherent (ABS 2009), with respect to the small area data package and other published official statistics. Small area estimates that are not coherent with the published survey estimates on which they are based, will struggle to gain credibility with users. Pfeffermann and Barnard (1991) developed a methodology for benchmarking SAEs in the context of a multilevel linear model. Currently there is considerable research work going into developing methodologies for producing benchmarked SAEs for more conventional SAE models. You *et al.* (2002) and Datta *et al.* (2010) looked at the problem of benchmarking SAEs in the context of Bayesian model estimation. Ugarte *et al.* (2008) derive a small area estimator with restrictions applied, for the linear random effects model. In the terminology of Datta *et al.* (2010), benchmarking may be internal or external. Internal benchmarks are usually the regional estimates from the survey used to fit the small area model. External benchmarks are totals obtained from some administrative data source not used as auxiliary

E-mail address : daniel.elazar@abs.gov.au

information in the small area modelling process. In this paper we will be focusing on an application involving internal benchmarking, although the methodology can be readily applied to external benchmarking.

The ABS has investigated the feasibility of producing SAEs for all three labour force statuses (employed, unemployed and not in the labour force) using logistic random effects models. The purpose of this paper is to derive an adapted maximum penalized quasi-likelihood (MPQL) approach for estimating a GLMM with additivity constraints on the SAEs predicted from the model.

2. THE CONSTRAINT

Before we discuss the constraints in detail, we briefly present the model under consideration. Let **y** be a $C \times 1$ vector of counts of a random variable Y with distribution Binomial (**n**; p), where C is the total number of binomial classes, **n** is the corresponding vector of sample sizes for each binomial class and p are the corresponding binomial probabilities. Given **X**, a $C \times P$ matrix of covariates, **Z**, a known $C \times D$ design matrix (where D is the number of small areas) and $\mathbf{u} \sim N(0, \mathbf{\Omega} = \text{Diagonal}(\phi))$, a $D \times 1$ vector of small area level effects, then the population model is given by

$$\text{logit}(p) = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} \tag{2.1}$$

Let $\hat{\theta}$ be a $D \times 1$ vector of the estimates for all D small areas (including those out-of-sample) predicted from the small area model and let $\hat{\mathbf{E}}$ be an $R \times 1$ vector of known direct survey estimates at some benchmark level. In our case, benchmark levels may be either Australia, State/Territory, State/Territory by Capital City/Non-Capital City (e.g. Sydney and New South Wales without Sydney) and labour force dissemination region. It is assumed that each small area, the statistical local area (SLA), belongs to one and only one benchmark level region. Then the constraint equation that forces SAEs $\hat{\theta}$ to add to the corresponding dissemination region estimates can be expressed as:

$$\alpha\hat{\theta} = \hat{\mathbf{E}} \tag{2.2}$$

where α is known $R \times D$ matrix, consisting of 1s and 0s, that effectively sums the model predicted estimates for small areas up to their respective benchmark region.

In the Australian context, there are around 70 dissemination regions. In an hypothetical case of 9 small areas within 3 dissemination regions, α might look like:

$$\alpha = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & \vdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \vdots & 1 & 1 \end{pmatrix}$$

Note that as we list out-of-sample areas after the in-sample areas, out-of-sample areas appear in the right hand sub-matrix of α. So in the example above, the first two dissemination regions have no out-of-sample areas while the third dissemination region has one in-sample small area and two out-of-sample areas.

Now using the nomenclature of Saei and Chambers (2003), we can express the small area estimator $\hat{\theta}$ as

$$\hat{\theta} = \mathbf{a}(\mathbf{y}_{sr} + \hat{\mathbf{y}}_r) \tag{2.3}$$

where

a: is a known $D \times C$ matrix (with $C = \sum_{d=1}^{D} C_d$ and C_d the number of binomial classes in small area d) of ones and zeros with ones in the d'th row corresponding to the binomial classes belonging to the d'th small area,

$\mathbf{y}_{sr}$**:** is a C-vector of observed sample response variable counts (with zeros for un-sampled binomial classes within in-sample areas and those within out-of-sample areas), $\hat{\mathbf{y}}_r = [(N_{di} - n_{di})\ \hat{p}_{di}]$, $d = 1, \ldots, C$, $i = 1, \ldots, C_d$ is a C-vector of model predictions for the sample complement in the ith age/sex cell within the dth small area, and

$$\hat{p}_{di} = \frac{\exp\left(\mathbf{x}_{di}\hat{\beta} + \hat{u}_d\right)}{1 + \exp\left(\mathbf{x}_{di}\hat{\beta} + \hat{u}_d\right)} \tag{2.4}$$

for all $d = 1, \ldots, D$, $i = 1, \ldots, C_d$ and where $\mathbf{x}_{di}$ is the $(1 \times P)$ row vector of covariate values and $\hat{\beta}$ is a $(P \times 1)$ column vector of model parameter estimates. Note that the above formula is calculated for all small areas d, including out-of-sample areas, for which the random effects u_d, are assigned the value zero.

In matrix notation, $\hat{\mathbf{y}}_r = (\mathbf{N} - \mathbf{n})\,\#\,\hat{\mathbf{p}}$ where # designates element-wise vector multiplication and

$$\hat{\mathbf{p}} = \frac{\exp\left(\mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\mathbf{u}}\right)}{1 + \exp\left(\mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\mathbf{u}}\right)} \tag{2.5}$$

where **X** is the matrix of population covariates consisting of rows $\mathbf{x}_{di}$, $d = 1,\ldots, D$, $i = 1,\ldots, C_d$ and **Z** is the design matrix defining which binomial classes belong to which small areas.

The constraint (2.2) can be re-expressed in the form:

$$\mathbf{A}\hat{\mathbf{y}}_r = \mathbf{K} \tag{2.6}$$

where

$$\mathbf{A} = \alpha\mathbf{a}$$

$$\mathbf{K} = \hat{\mathbf{E}} - \mathbf{A}\mathbf{y}_{sr} \tag{2.7}$$

Note that the matrix **A** is a matrix of ones and zeros designating membership of small area by age by sex cells to dissemination regions (benchmark regions). **A** has dimensions $(R \times D)$ $(D \times C) = (R \times C)$ and **K** is an $(R \times 1)$ column vector.

3. MPQL SUBJECT TO AN ADDITIVITY CONSTRAINT

The objective is to reformulate the GLMM model (2.1), subject to the constraint equation (2.6). Note that from the previous section, the constraint **K** from equation (2.7) depends on $\hat{\mathbf{E}}$, the vector of direct survey estimates at the benchmark level. **K** is therefore subject to sampling error, and this increases as you go from the Australian level down to dissemination region level. In fact, dissemination region estimates can have sampling variances of over 25%. When benchmarking small area estimates, practitioners commonly use methods such as iterative proportional fitting. However, the disadvantages of this approach are that firstly, it is usually done after the small area estimation process and is therefore not necessarily model consistant. Secondly, it does not provide a way of estimating the relative root MSEs for the SAEs.

To estimate the constrained GLMM, we need to use the Langrange multiplier method, which involves introducing an $R \times 1$ vector λ of constraint parameters (Lagrange multipliers). Following Saei and Chambers (2003) for the unconstrained problem, let $f_1(\mathbf{y}_s \mid \mathbf{u}_s)$ be the probability density function of $\mathbf{y}_s$ conditional on $\mathbf{u}_s$ and let $f_2(\mathbf{u}_s)$ be the probability density function of $\mathbf{u}_s$.

Then the extended loglikelihood, as defined by Pawitan (2001), is given by:

$$l = l_1 + l_2 + \lambda^T (\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K}) \tag{3.1}$$

where

$$l_1 = \ln (f_1 (\mathbf{y}_s \mid \mathbf{u}_s))$$

$$l_2 = -\frac{1}{2}\left(\text{Const.} + \ln|\Omega| + \mathbf{u}_s^T\Omega^{-1}\mathbf{u}_s\right) \tag{3.2}$$

Let $\zeta = \left(\beta^T, \mathbf{u}_s^T, \lambda^T\right)^T$ be a $(P + D_s + R) \times 1$ vector of parameters to be optimised, where $\mathbf{u}_s$ is a $D_s \times 1$ vector of random effects for only those areas that are in-sample and D_s is the number of in-sample small areas, with $D_s \le D$. Also let

$$\begin{aligned}\eta &= \mathbf{X}_{ss}\beta + \mathbf{Z}_{ss}\mathbf{u}_s \\ &= \left(\mathbf{X}_{ss}\mathbf{Z}_{ss}\mathbf{0}_{C_{ss}\times R}\right)\zeta\end{aligned} \tag{3.3}$$

be a $C_{ss} \times 1$ vector, where $C_{ss} = \sum_{d=1}^{D_s} C_{ds}$ and C_{ds} is the number of age sex cells, i, in small area d that do have sample, i.e. for which $n_{ij} > 0$, and $\mathbf{0}_{C_{ss}\times R}$ is a zero matrix of dimensions $C_{ss} \times R$. Also note that η is of dimension C_{ss} as the parameters β and $\mathbf{u}_s$ need to be estimated for the model fitted to the observed sample data, i.e. not including age-sex cells with no sample that fall within in-sample areas. The constraint space, however, is spanned by all age-sex cells in all small areas, regardless of whether they have sample or not because predictions for all cells need to be accounted for in ensuring coherence with the broad level publication estimates we wish to constrain our small area estimates to.

Using Newton's method, the $(k + 1)$'th iteration is given in terms of the previous iteration, k, by:

$$\zeta^{(k+1)} = \zeta^{(k)} - \left(\left.\frac{\partial^2 l}{\partial\zeta\,\partial\zeta^T}\right|_{\zeta=\zeta^{(k)}}\right)^{-1} \left.\frac{\partial l}{\partial\zeta}\right|_{\zeta=\zeta^{(k)}} \tag{3.4}$$

where

$$\frac{\partial l}{\partial\zeta} = \begin{pmatrix}\mathbf{X}_{ss}^T \\ \mathbf{Z}_{ss}^T \\ \mathbf{0}\end{pmatrix}\frac{\partial l_1}{\partial\eta} - \begin{pmatrix}\mathbf{0} \\ \Omega^{-1}\mathbf{u}_s \\ \mathbf{0}\end{pmatrix} + \begin{pmatrix}\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda \\ \mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda \\ \mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K}\end{pmatrix} \tag{3.5}$$

and

$$\frac{\partial^2 l}{\partial\zeta\partial\zeta^T} = \begin{pmatrix} \mathbf{X}_{ss}^T \\ \mathbf{Z}_{ss}^T \\ \mathbf{0} \end{pmatrix} \left(-\frac{\partial^2 l_1}{\partial\eta\partial\eta^T} \right) (\mathbf{X}_{ss}\mathbf{Z}_{ss}\,\mathbf{0}) + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

$$-\begin{pmatrix} \mathbf{X}^T\mathbf{F}\mathbf{X} & \mathbf{X}_s^T\mathbf{F}_s\mathbf{Z}_s & \mathbf{X}^T\mathbf{J}\mathbf{A}^T \\ \mathbf{Z}_s^T\mathbf{F}_s\mathbf{X}_s & \mathbf{Z}_s^T\mathbf{F}_s\mathbf{Z}_s & \mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T \\ \mathbf{A}\mathbf{J}\mathbf{X} & \mathbf{A}_s\mathbf{J}_s\mathbf{Z}_s & \mathbf{0}_{R\times R} \end{pmatrix} \tag{3.6}$$

Proofs for equations (3.5) and (3.6) are shown in Results (B.9) and (B.10), respectively, found in the Appendix.

The inverse of $\frac{\partial^2 l}{\partial\zeta\partial\zeta^T}$ can be calculated using this formula, however it requires inverting a $(P + D + R)^2$ matrix where $P = 36$, $D = 420$ and $R <= 70$. While this is not too bad for estimation of SAEs, it would be preferable to make this calculation more effcient for parameteric bootstrap simulation of the MSE estimates.

Let $\mathbf{H} = -\frac{\partial^2 l_1}{\partial\eta\,\partial\eta^T}$. Then

$$\frac{\partial^2 l}{\partial\zeta\,\partial\zeta^T}$$

$$= \begin{pmatrix} \mathbf{X}_{ss}^T\mathbf{H}\mathbf{X}_{ss} - \mathbf{X}^T\mathbf{F}\mathbf{X} & \mathbf{X}_{ss}^T\mathbf{H}\mathbf{Z}_{ss} - \mathbf{Z}_s^T\mathbf{F}_s\mathbf{Z}_s & -\mathbf{X}^T\mathbf{J}\mathbf{A}^T \\ \mathbf{Z}_{ss}^T\mathbf{H}\mathbf{X}_{ss} - \mathbf{Z}_s^T\mathbf{F}_s\mathbf{X}_s & \mathbf{Z}_{ss}^T\mathbf{H}\mathbf{Z}_{ss} + \Omega^{-1} - \mathbf{Z}_s^T\mathbf{F}_s\mathbf{Z}_s & -\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T \\ -\mathbf{A}\mathbf{J}\mathbf{X} & -\mathbf{A}_s\mathbf{J}_s\mathbf{Z}_s & \mathbf{0}_{R\times R} \end{pmatrix}$$

We can readily see that the matrix expression for $\frac{\partial^2 l}{\partial\zeta\,\partial\zeta^T}$ takes the form $\begin{pmatrix} \mathcal{A} & \mathcal{B}^T \\ \mathcal{B} & 0 \end{pmatrix}$ where

$$\mathcal{A} = \begin{pmatrix} \mathbf{X}_{ss}^T\mathbf{H}\mathbf{X}_{ss} - \mathbf{X}^T\mathbf{F}\mathbf{X} & \mathbf{X}_{ss}^T\mathbf{H}\mathbf{Z}_{ss} - \mathbf{X}_s^T\mathbf{F}_s\mathbf{Z}_s \\ \mathbf{Z}_{ss}^T\mathbf{H}\mathbf{X}_{ss} - \mathbf{Z}_s^T\mathbf{F}_s\mathbf{X}_s & \mathbf{Z}_{ss}^T\mathbf{H}\mathbf{Z}_{ss} + \Omega^{-1} - \mathbf{Z}_s^T\mathbf{F}_s\mathbf{Z}_s \end{pmatrix}$$

and $\mathcal{B} = (-\mathbf{A}\mathbf{J}\mathbf{X} - \mathbf{A}_s\mathbf{J}_s\mathbf{Z}_s)$.

Then

$$\left(\frac{\partial^2 l}{\partial\zeta\partial\zeta^T}\right)^{-1} = \begin{pmatrix} \mathcal{A} & \mathcal{B}^T \\ \mathcal{B} & 0 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \mathcal{A}^{-1}\left(\mathbf{I} - \mathcal{B}^T\left(\mathcal{B}\mathcal{A}^{-1}\mathcal{B}^T\right)^{-1}\mathcal{B}\mathcal{A}^{-1}\right) & [\mathcal{A}^{-1}\mathcal{B}^T\left(\mathcal{B}\mathcal{A}^{-1}\mathcal{B}^T\right)^{-1}] \\ (\mathcal{B}\mathcal{A}^{-1}\mathcal{B}^T)^{-1}\mathcal{B}\mathcal{A}^{-1} & -(\mathcal{B}\mathcal{A}^{-1}\mathcal{B}^T)^{-1} \end{pmatrix}$$

Now we only have to contend with finding $\mathcal{A}^{-1}$.

Let $\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{21}^T \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{pmatrix}$.

Using a well known partitioned matrix identity given in Henderson and Searle (1981),

$$\mathcal{A}^{-1} = \begin{pmatrix} \frac{1}{\left(\mathcal{A}_{11} - \mathcal{A}_{21}^T\mathcal{A}_{22}^{-1}\mathcal{A}_{21}\right)} & \frac{(\mathcal{A}_{21}^T\mathcal{A}_{22}^{-1})}{-(\mathcal{A}_{11} - \mathcal{A}_{21}^T\mathcal{A}_{22}^{-1}\mathcal{A}_{21})} \\ \frac{(-\mathcal{A}_{22}^{-1}\mathcal{A}_{21})}{(\mathcal{A}_{11} - \mathcal{A}_{21}^T\mathcal{A}_{22}^{-1}\mathcal{A}_{21})} & \frac{\mathcal{A}_{22}^{-1} + (\mathcal{A}_{22}^{-1}\mathcal{A}_{21})(\mathcal{A}_{21}^T\mathcal{A}_{22}^{-1})}{(\mathcal{A}_{11} - \mathcal{A}_{21}^T\mathcal{A}_{22}^{-1}\mathcal{A}_{21})} \end{pmatrix}$$

4. RESULTS

In the previous section we extended the methodology involving the PQL algorithm with maximum likelihood/REML, to estimate model parameters and small area predictions subject to the constraints that the sum of the small area estimates within a given region, equal some known benchmark value for that region.

In this section we present and discuss the results from applying this methodology to small area labour force estimates using internal benchmarking at four levels:

- Australia,
- state,
- state by capital city / non-capital city, and
- labour force dissemination region.

The 'state' level refers to the eight states and territories of Australia. For statistical output purposes these are often split by capitial city / non-capital city, also known as part-of-state (POS). Dissemination regions are currently the finest output level from the monthly Labour Force Survey. These are broad regions with population sizes of around 250,000 to 300,000 persons that have been constructed to give satisfactory, but not necessarily reliable, sampling errors. These dissemination regions do not necessarily reflect homogeneous labour market regions.

Now, it goes without saying that these internal benchmarks are themselves subject to sampling error, which will be inversely proportional to the size of the region. Thus, for example, the sampling error will be condiderably larger for dissemination regions than it will be for the national level. In this paper, we do not take explicit account of the sampling error when estimating the benchmarked model nor in MSE estimation. However it can be argued that implicitly, the additional statistical noise in the benchmarks will increase model error and should therefore increase the estimated MSEs. However this may not account for all of the impact of the uncertainty in the benchmarks, upon the SAEs and their estimated MSEs. This, we hope, will be the subject of future work.

The benchmarked logistic random effects models were fitted to labour force data to produce small area predictions for the three labour force statuses. The models for each status were benchmarked to each of the four benchmark levels listed above. The main objectives were firstly to, discover how SAEs produced from the benchmarked models compare with the equivalent unconstrained SAEs, and secondly to see what the effect of different benchmarking levels has on the SAEs and their estimated RRMSEs.

4.1 Benchmarked Small Area Estimates

To test whether the methodology is feasible and coherent, we set the benchmark constraints to be the sum of the unconstrained model SAEs at the given benchmark level. For brevity, we will refer to these benchmarks as the "pseudo unconstrained benchmarks". Heuristically, this is like setting a constraint on the model that does not unduly shift model estimation away from its natural, unconstrained course. It is in effect akin to a non-constraint, constraint. If the method is plausible, when comparing the SAEs under the pseudo unconstrained benchmarks to the original unconstrained SAEs, we would expect to observe the data points sitting along a unit line.

Fig. 1 shows this plot for employment with pseudo unconstrained benchmarks at the Australian level. It can be easily seen that the regression line and the unit line are very close.

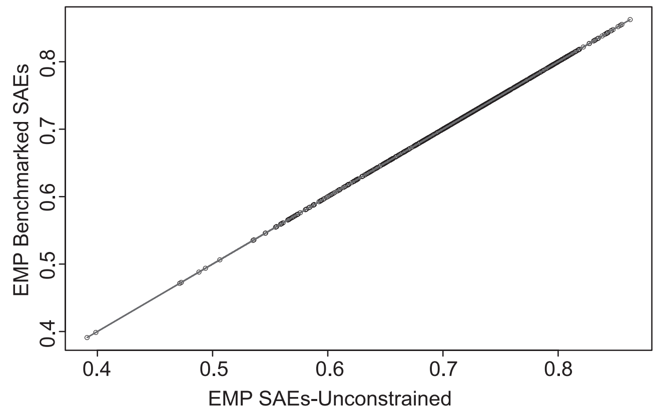

Fig. 1. Employment SAEs benchmarked to the sum of unconstrained SAEs at Australian level

The plot for unemployment, Fig. 2, shows a small amount of volatility but is still close to the unit line.

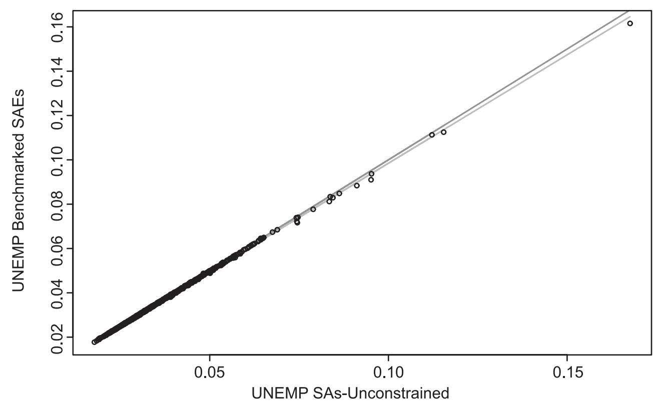

Fig. 2. Unemployment SAEs benchmarked to the sum of unconstrained SAEs at Australian level

Employment and unemployment SAEs constrained to pseudo unconstrained benchmarks at the finest level, dissemination regions, Figs. 3 and 4, show a slight amount of noise but essentially follow the unit line.

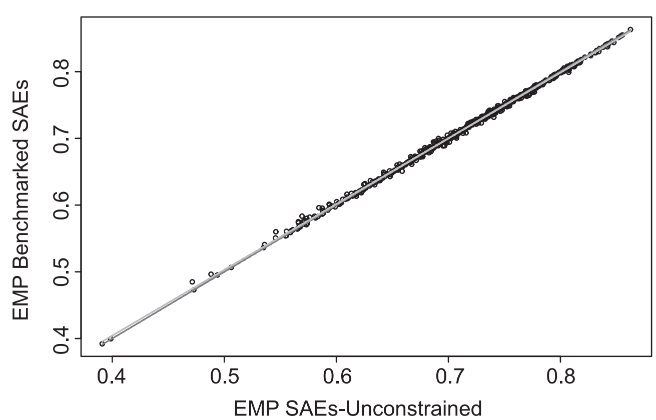

Fig. 3. Employment SAEs benchmarked to the sum of unconstrained SAEs at REGNDISM level

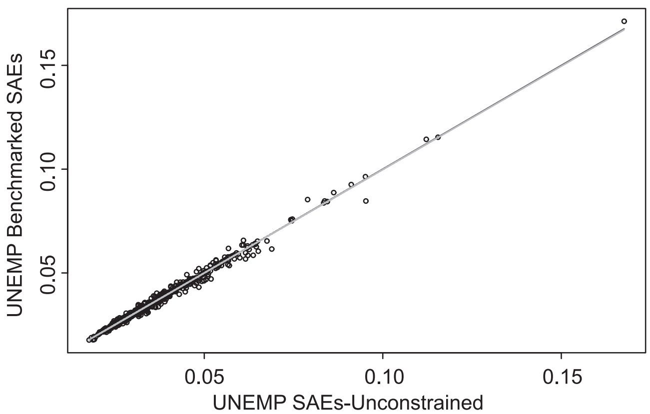

Fig. 4. Unemployment SAEs benchmarked to the sum of unconstrained SAEs at REGNDISM level

We now turn to a comparison between benchmarked SAEs and unconstrained SAEs. Fig. 5 shows when benchmarked at the Australian level, the benchmarked SAEs are slightly lower than their unconstrained SAE counterparts. The regression line (light line) through the scatter plot is highly parallel to the unit line (dark), indicating that the relative bias is constant irrespective of the size of the estimate. A separate simulation study has shown that the unconstrained SAEs are likely to be biased due to not taking full account of the survey design when estimating the small area model, and that this bias was larger than that due to model misspecification. It appears that the benchmarking has, at least to some extent, corrected for this bias.

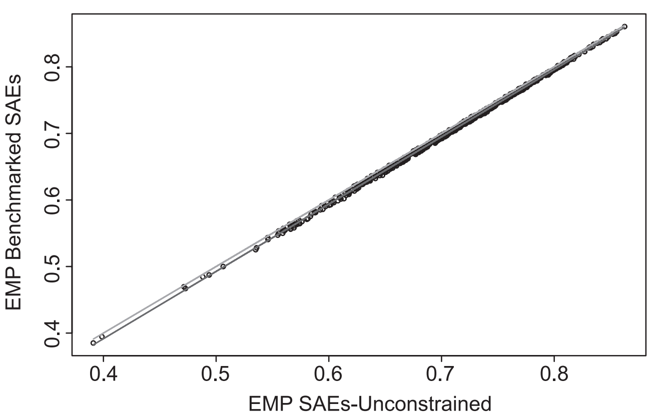

Fig. 5. Employment: Benchmarked SAEs compared to unconstrained SAEs at the Australian level

When it comes to benchmarking to the state level (Fig. 6), the regression line (slope = 0.99988) and the unit line are again highly parallel with a vey small bias (intercept of regression line = –0.0052). While most points fall close to the regression and unit lines, there are a few points that sit in a line below these lines. All these estimates are for small areas in the Australian Capital Territory (ACT) and Northern Territory (NT). Earlier investigations into our small area model for labour force showed that the most influential points occured in binomial classes in these territories. The

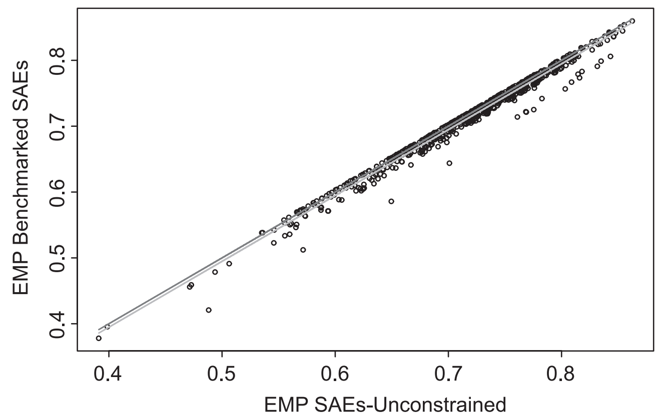

Fig. 6. Employment: Benchmarked SAEs compared to unconstrained SAEs at the State level

benchmarking estimation process is reducing the estimates for these areas to a considerably larger extent than all other areas.

Benchmarking at the state by capital city/non-capital city (aka part of state (POS)) introduces more benchmark sampling error into the benchmarked estimates. The satay stick appearance of Fig. 7 is due to this increased volatility.

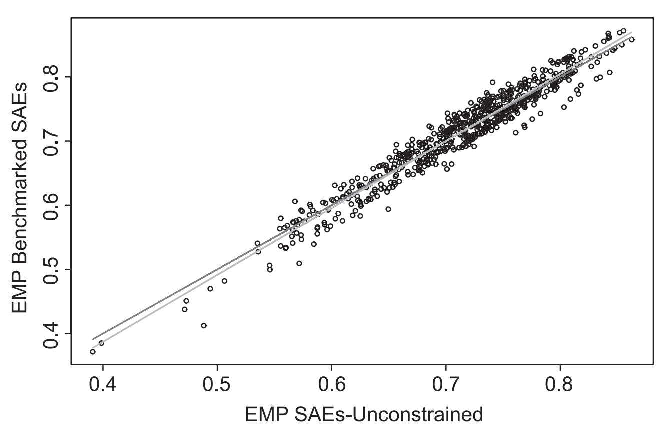

Fig. 7. Employment: Benchmarked SAEs compared to unconstrained SAEs at the State by POS level

And even more so when the benchmarking is done at the dissemination region level as shown in Fig. 8.

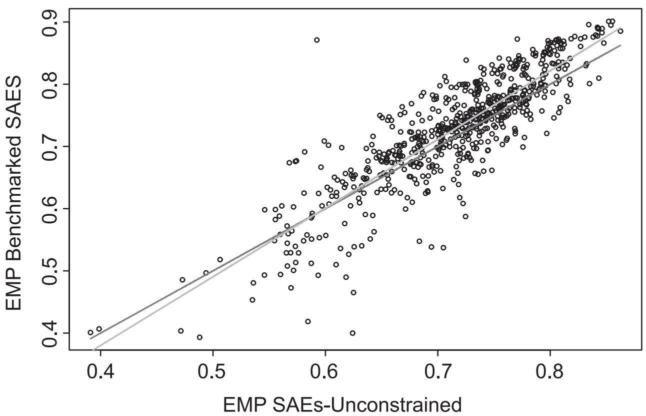

Fig. 8. Employment: Benchmarked SAEs compared to unconstrained SAEs at Dissemination

4.2 Relative Root Mean Squared Errors

In terms of the accuracy of the benchmarked SAEs, there are three estimates of relative root MSE (RRMSE) that can be compared. These are the:

- RRMSES of the original unconstrained SAEs
- naive RRMSEs for the benchmarked SAEs
- bootstrap RRMSEs for the benchmarked SAEs

The naive RRMSEs for the benchmarked SAEs are obtained simply by plugging the estimated values for the parameters $\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}$ and $\hat{\boldsymbol{\phi}}$ obtained from the benchmarked model, into the usual analytic formula for MSE estmation as per the unconstrained model estimation. We would expect the naive RRMSEs to be biased downwards as they do not include the additional (negative) correlation between SAEs, arising from the enforced additivity to benchmarks. Looking at it another way, the constraints will pull the regression away from its natural unconstrained fit, resulting in larger residuals.

The bootstrap RRMSEs on the other hand should be an unbiased estimate of the true RRMSEs. They were calculated using a similar approach to that of Scealy (2010) and Molina (2007), which starts with the estimated logistic binomial random intercept model fitted to the labour force data. This model is assumed to be the working model from which bootstrap estimates will be generated. Assign the estimated parameters from this model as follows: $\beta = \hat{\beta}, f = \hat{\phi}$. The bootstrap algorithm proceeds as follows:

1. randomly generate random effects $u_i^{(b)}$: $i = 1, \ldots, D$ from the normal distribution N(0; ϕ). (note that this includes random effects for the out-of-sample areas)
2. conditional on $u_i^{(b)}$, calculate the binomial proportions $p_{ij}^{(b)} = \dfrac{\exp(\mathbf{X}_{ij}\beta + u_i^{(b)})}{1+\exp(\mathbf{X}_{ij}\beta + u_i^{(b)})}$
3. conditional on $p_{ij}^{(b)}$, generate the sample data $y_{ij}^{(s)(b)}$ from the binomial distribution $Bin(n_{ij},\ p_{ij}^{(b)})$ for all areas i
4. also generate the non-sample data $y_{ij}^{(s)(b)}$ from the binomial distribution $Bin(N_{ij} - n_{ij};\ p_{ij}^{(b)})$ again for all areas $i = 1, \ldots, D$.
5. calculate the population totals for area i as $Y_i^{(b)} = \sum_{j=1}^{10} Y_{ij}^{(b)}$ where $Y_{ij}^{(b)} = y_{ij}^{(s)(b)} + y_{ij}^{(r)(b)}$
6. re-estimate the model parameters by fitting the logistic random effects model to the current sample data $\{y_{ij}^{(ss)(b)}\} = \{ y_{ij}^{(s)(b)} : n_{ij} > 0\}$. Refer to these estimated parameters as $\hat{\beta}^{(b)}, \hat{u}_s^{(b)}$ and $\hat{\phi}^{(b)}$. Note that $\hat{u}_s^{(b)}$ will contain zero values for out-of-sample areas.
7. calculate the binomial proportion estimates $$\hat{p}_{ij}^{(b)} = \frac{\exp(\mathbf{X}_{ij}\hat{\beta}^{(b)} + \hat{u}_{si}^{(b)})}{1+\exp(\mathbf{X}_{ij}\hat{\beta}^{(b)} + \hat{u}_{si}^{(b)})}$$
8. estimate the small area estimates $$\hat{y}_i^{(b)} = \sum_{j=1}^{10}\left(y_{ij}^{(s)(b)} + (N_{ij} - n_{ij})\hat{p}_{ij}^{(b)}\right)$$
9. calculate the bootstrap estimate of MSE for the small area estimate $\hat{y}_i$ as $$M\hat{S}E^{(b)}(\hat{y}_i) = \frac{1}{B}\sum_{b=1}^{B}(\hat{y}_i^{(b)} - Y_i^{(b)})^2 \qquad (4.1)$$

Calculating the bootstrap RRMSEs is computationally highly intensive, taking up to eight days. For this reason they were only calculated for unemployment. While the intention was to run 4,000 bootstrap iterations, unfortunate server crashes and reboots, meant the realised number of iterations ranged from 2,000 to 4,000 iterations.

Fig. 9 shows the relationship between the bootstrap RRMSEs for unemployment SAEs constrained at the Australian level and the analytic RRMSEs for the unconstrained SAEs. The plot shows that there is pretty much a one to one correspondence between these two sets of RRMSE estimates.

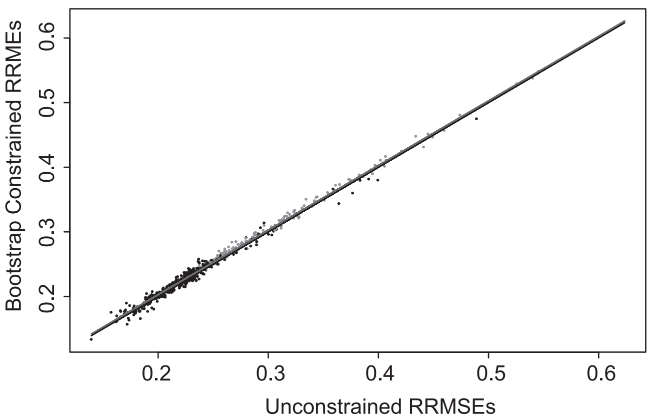

Fig. 9. Unemployment: Bootstrap constrained versus unconstrained RRMSEs at the Australian level

The bootstrap RRMSEs also compared very favourably with the naive RRMSEs for SAEs benchmarked at the Australian level (see Fig. 10). In other words, when benchmarking at a level with low sampling error, the naive RRMSEs are a reliable approximation to the bootstrap RRMSEs.

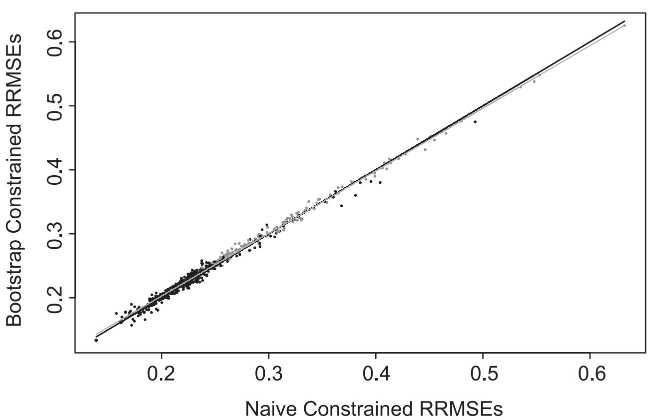

Fig. 10. Unemployment: Bootstrap versus naive constrained RRMSEs at the Australian

At the other extreme, Fig. 11 shows that when the benchmarking occurs at the finest dissemination region level the bootstrap RRMSEs are considerably higher than the RRMSEs for the unconstrained SAEs. The more volatile dissemination region estimates perturb the model estimation resulting in RRMSEs that are about 30% higher. The high volatilty points in Fig. 11 are probably due to a lack of convergence in the MPQL estimation. A limit of 180 iterations was set, however for around 7% of bootstrap samples, model estimation was terminated at this limit. Hence the estimates used had not fully converged. Another possible cause related to this is that benchmark constraints at the dissemination region level can induce ripples into the constrained loglikelihood surface, thereby leading to possibly numerous local optima. A grid search using a number of initial values should overcome this problem, but at the expense of computational speed.

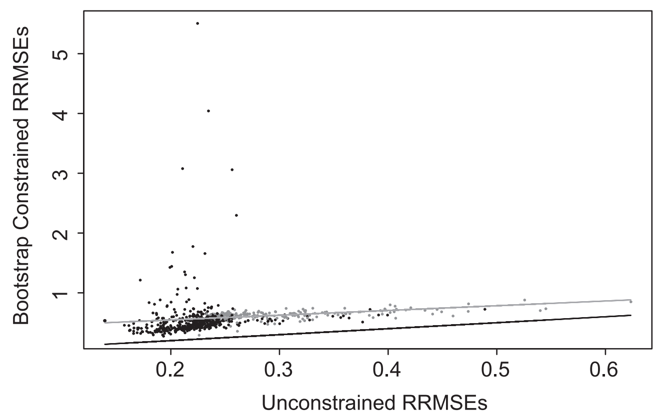

Fig. 11. Unemployment: Bootstrap constrained versus unconstrained RRMSEs at the dissemination region level

Fig. 12 confirms that the RRMSEs calculated using the naive approach are not sufficiently reliable.

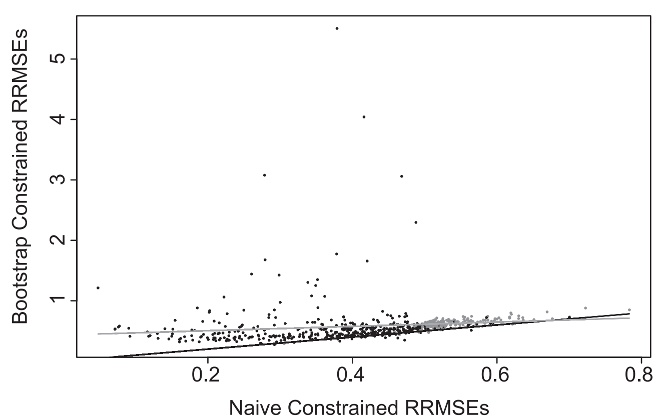

Fig. 12. Unemployment: Bootstrap versus naive constrained RRMSEs at the dissemination region level

ACKNOWLEDGMENTS

The author wishes to thank Professor Ray Chambers for advice and encouragement given during the course of this work and Professor Peter Hall for advice on the use of the parametric bootstrap under a constrained model.

REFERENCES

ABS Data Quality Framework (2009). Cat. 1520.0 website: http: /www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1520.0Main%20Features7May%202009? opendocument & tabname =Summary &prodno =1520.0 &issue =May %202009 &num =&view=

Datta, G., Ghosh, M., Steorts, R. and Maples, J. (2010). Bayesian benchmarking with applications to small area estimation. *TEST,* Published online: http://www.springerlink.com/content/y629x83058337762/

Henderson, H.V. and Searle, S.R. (1981). On deriving the inverse of a sum of matrices. *SIAM Rev.,* **23,** 53-60.

Molina, I., Saei, A. and Lombarda, M.J. (2007). Small area estimates of labour force participation under a multinomial logit mixed model. *J. Roy. Statist. Soc.,* **A170(4),** 975-1000.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Clarendon Press, Oxford.

Pfeffermann, D. and Barnard, C.H. (1991). Some new estimators for small-area means with application to the assessment of farmland values. *J. Busi. Eco. Statist.*, **9(1)**, 73-83.

Saei, A. and Chambers, R. (2003). Small Area Estimation Under Linear and Generalized Linear Mixed Models With Time and Area Eeffcts, Southampton Statistical Sciences Research Institute, S3RI Methodology Working Papers, M03/15, website: http://eprints.soton.ac.uk/8165/1/8165-01.pdf

Ugarte, M.D., Militino, A.F. and Goicoa, T. (2008). Benchmarked estimates in small areas using linear mixed models with restrictions. *TEST*, **18(2)** 342-364.

You, Y., Rao, J.N.K. and Dick, P. (2002). Benchmarking Hierarchical Bayes Small Area Estimators with application in census undercoverage estimation. SSC Annual Meeting, *Proceedings of the Survey Methods Section.*

A. ADDITIONAL NOTATION

$$\mathbf{J}_{di} = (N_{di} - n_{di})\, p_{di}\,(1 - p_{di})$$

$$\kappa_{di} = (N_{di} - n_{di})\, p_{di}\,(1 - p_{di})\,(1 - 2p_{di})$$

$$\mathbf{J} = \text{Diagonal } [J_{di}],\ d = 1, \ldots, D;\ i = 1, \ldots, C_d$$

$$\mathbf{J}_s = \text{Diagonal } [J_{di}],\ d = 1, \ldots, D_s;\ i = 1, \ldots, C_d$$

$$\mathbf{J}_s^* = [J_{di}]\ ;\ d = 1, \ldots, D;\ i = 1, \ldots, C_d$$

$$\mathbf{J}_d = (J_{d1}, \ldots,\ J_{dC_d})$$

$$\mathcal{K} = \text{Diagonal } [\mathcal{K}_{di}]\ ;\ d = 1, \ldots, D;\ i = 1, \ldots, C_d$$

$$\mathcal{K}_s = \text{Diagonal } [\mathcal{K}_{di}]\ ;\ d = 1, \ldots, D_s;\ i = 1, \ldots, C_d$$

$$\mathbf{Z}_{so} = \begin{pmatrix} \mathbf{Z}_s \\ \mathbf{0}_{(C - C_s) \times D_s} \end{pmatrix}$$

$$\mathbf{J}_{so} = \begin{pmatrix} \mathbf{J}_s & \mathbf{0}_{C_s \times (C - C_s)} \\ \mathbf{J}_{(C - C_s) \times C_s} & \mathbf{0}_{(C - C_s) \times (C - C_s)} \end{pmatrix}$$

$$\mathbf{F} = \text{Diagonal } [\mathbf{A}^T \lambda]\mathcal{K}$$

$$\mathbf{F}_s = \text{Diagonal } [\mathbf{A}_s^T \lambda]\mathcal{K}_s$$

B. APPENDIX

Definition B.1. Let $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ be a $n \times 1$ vector with each element y_k a function of a $p \times 1$ vector $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$. Then the partial derivative of $\mathbf{y}$ with respect to β is defined as:

$$\frac{\partial \mathbf{y}}{\partial \beta} = \begin{pmatrix} \frac{\partial y_1}{\partial \beta_1} & \cdots & \frac{\partial y_n}{\partial \beta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial \beta_p} & \cdots & \frac{\partial y_n}{\partial \beta_p} \end{pmatrix}$$

and

$$\frac{\partial \mathbf{y}}{\partial \beta^T} = \begin{pmatrix} \frac{\partial y_1}{\partial \beta_1} & \cdots & \frac{\partial y_1}{\partial \beta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial \beta_1} & \cdots & \frac{\partial y_n}{\partial \beta_p} \end{pmatrix} = \left(\frac{\partial \mathbf{y}}{\partial \beta} \right)^T$$

Lemma B.1 :

$$\frac{dp_{di}}{d\beta_p} = x_{di}^{(p)}\, p_{di}(1 - p_{di})$$

where $x_{di}^{(p)}$ is the element of $\mathbf{X}$ located in the row indexed by di and column p.

Proof. From equation (2.4)

$$\frac{dp_{di}}{d\beta_p} = \frac{d}{d\beta_p}\left(\frac{\exp(x_{di}\beta + u_d)}{1 + \exp(x_{di}\beta + u_d)} \right)$$

$$= \frac{x_{di}^{(p)} \exp(x_{di}\beta + u_d)}{1 + \exp(x_{di}\beta + u_d)}$$

$$- \exp(x_{di}\beta + u_d)\ \frac{x_{di}^{(p)} \exp(x_{di}\beta + u_d)}{(1 + \exp(x_{di}\beta + u_d))^2}$$

$$= x_{di}^{(p)}\, p_{di}(1 - p_{di})$$

Lemma B.2 :

$$\frac{d}{d\beta_p}(p_{di}(1 - p_{di})) = x_{di}^{(p)}\, p_{di}(1 - p_{di})(1 - 2p_{di})$$

Proof. By applying the product rule for differentiation and using Lemma B.1 we obtain

$$\frac{d}{d\beta_p}(p_{di}(1 - p_{di})) = \frac{dp_{di}}{d\beta_p}(1 - p_{di}) - p_{di}\frac{dp_{di}}{d\beta_p}$$

$$= x_{di}^{(p)}\, p_{di}\,(1 - p_{di})\,(1 - 2p_{di})$$

Lemma B.3 :

$$\frac{dp_{di}}{du_{d'}} = \begin{cases} p_{di}\,(1 - p_{di} & \text{if} \quad d = d', \\ 0 & \text{if} \quad d \neq d' \end{cases}$$

Proof. In the case of $d = d'$, from equation (2.4)

$$\frac{dp_{di}}{du_d} = \frac{d}{du_d}\left(\frac{\exp(x_{di}\beta + u_d)}{1 + \exp(x_{di}\beta + u_d)} \right)$$

$$= \frac{\exp(x_{di}\beta + u_d)}{1 + \exp(x_{di}\beta + u_d)}$$

$$- \exp(x_{di}\beta + u_d)\ \frac{\exp(x_{di}\beta + u_d)}{(1 + \exp(x_{di}\beta + u_d))^2}$$

$$= p_{di}(1 - p_{di})$$

Clearly, in the case of $d \neq d'$ then $\frac{dp_{di}}{du_{d'}} = 0.$

Lemma B.4 :

$$\frac{d}{du_{sd'}}(p_{di}(1-p_{di}))$$

$$= \begin{cases} p_{di}(1-p_{di})\,(1-2p_{di}) & \text{if } d = d' \\ 0 & \text{if } d \neq d'' \end{cases}$$

Proof. Applying the product rule for differentiation to Lemma B.3 in the case where $d = d'$, we obtain:

$$\frac{d}{du_{sd'}}(p_{di}(1-p_{di})) = \frac{dp_{di}}{du_{sd'}}(1-p_{di}) - p_{di}\frac{dp_{di}}{du_{sd'}}$$
$$= p_{di}(1-p_{di})\,(1-2p_{di})$$

Clearly, in the case of $d \neq d'$ then $\frac{dp_{di}}{du_{d'}} = 0$.

Lemma B.5 : Let **X** be an $n \times m$ matrix of constants. Let **Y** be an $n \times n$ diagonal matrix of elements y_k, $k = 1, ..., n$ with $\mathbf{y} = (y_1, y_2, ..., y_n)^T$ being the vector formed from the diagonal elements of **Y**. Also assume that the y_k are functions of a $p \times 1$ vector $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^T$. Let **b** be an $n \times 1$ vector of constants and let $\mathbf{B}$ = Diagonal [**b**] be the diagonal matrix formed from the elements of **b**. Then the partial derivative of the m-vector $\mathbf{X}^T\mathbf{Yb}$ with respect to $\boldsymbol{\beta}$ is

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\mathbf{X}^T\,\mathbf{Yb}) = \frac{\partial \mathbf{y}}{\partial \boldsymbol{\beta}}\mathbf{BX}$$

Proof. It can easily be shown that:

$$\mathbf{X}^T\mathbf{Yb} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1m} & x_{2m} & \cdots & x_{nm} \end{pmatrix} \begin{pmatrix} y_1 & & 0 \\ & \ddots & \\ 0 & & y_n \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

$$= \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1m} & x_{2m} & \cdots & x_{nm} \end{pmatrix} \begin{pmatrix} b_1 y_1 \\ \vdots \\ b_n y_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1}^{n} x_{k1} b_k y_k \\ \vdots \\ \sum_{k=1}^{n} x_{km} b_k y_k \end{pmatrix}$$

The partial derivative of $\mathbf{X}^T\mathbf{Yb}$ with respect to $\boldsymbol{\beta}$ is therefore:

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\mathbf{X}^T\,\mathbf{Yb}) = \begin{pmatrix} \sum_{k=1}^{n} x_{k1} b_k \frac{\partial y_k}{\partial \beta_1} & \cdots & \sum_{k=1}^{n} x_{km} b_k \frac{\partial y_k}{\partial \beta_1} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} x_{k1} b_k \frac{\partial y_k}{\partial \beta_p} & \cdots & \sum_{k=1}^{n} x_{km} b_k \frac{\partial y_k}{\partial \beta_p} \end{pmatrix}$$

$$= \begin{pmatrix} b_1 \frac{\partial y_1}{\partial \beta_1} & \cdots & b_n \frac{\partial y_n}{\partial \beta_1} \\ \vdots & \ddots & \vdots \\ b_1 \frac{\partial y_1}{\partial \beta_p} & \cdots & b_n \frac{\partial y_n}{\partial \beta_p} \end{pmatrix} \mathbf{X}$$

$$= \begin{pmatrix} \frac{\partial y_1}{\partial \beta_1} & \cdots & \frac{\partial y_n}{\partial \beta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial \beta_p} & \cdots & \frac{\partial y_n}{\partial \beta_p} \end{pmatrix} \begin{pmatrix} b_1 & 0 & \cdots & 0 & 0 \\ & b_2 & & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \cdots & b_{n-1} & 0 \\ 0 & 0 & & 0 & b_n \end{pmatrix} \mathbf{X}$$

$$= \frac{\partial \mathbf{y}}{\partial \boldsymbol{\beta}}\mathbf{BX} \qquad \text{(B.1)}$$

The left hand side of (B.1) must be of dimension $p \times m$ and the right hand side is of dimension $(p \times n)$ $(n \times n)$ $(n \times m)$, which are conformable.

Result B.1

$$\frac{\partial \mathbf{y}_r}{\partial \boldsymbol{\beta}} = \mathbf{X}^T\mathbf{J}$$

Proof.

$$\frac{\partial \hat{\mathbf{y}}_r}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial y_{r11}}{\partial \beta_1} & \cdots & \frac{\partial y_r DC_D}{\partial \beta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_{r11}}{\partial \beta_P} & \cdots & \frac{\partial y_r DC_D}{\partial \beta_P} \end{pmatrix}$$

$$= \begin{pmatrix} (N_{11}-n_{11})\frac{\partial p_{11}}{\partial \beta_1} & \cdots & (N_{DC_D}-n_{DC_D})\frac{\partial p_D C_D}{\partial \beta_1} \\ \vdots & \ddots & \vdots \\ (N_{11}-n_{11})\frac{\partial p_{11}}{\partial \beta_P} & \cdots & (N_{DC_D}-n_{DC_D})\frac{\partial p_D C_D}{\partial \beta_P} \end{pmatrix}$$

which by Lemma B.1

$$= \begin{pmatrix} [(N_{11}-n_{11})x_{11}^{(1)} & \cdots & [(N_{DC_D}-n_{DC_D})x_{DC_D}^{(1)} \\ p_{11}(1-p_{11})] & & p_{DC_D}(1-p_{DC_D})] \\ \vdots & \ddots & \vdots \\ [(N_{11}-n_{11})x_{11}^{(P)} & \cdots & [(N_{DC_D}-n_{DC_D})x_{DC_D}^{(P)} \\ p_{11}(1-p_{11})] & & P_D C_D(1-p_{DC_D})] \end{pmatrix}$$

$$= \mathbf{X}^T\mathbf{J}$$

Result B.2 :

$$\frac{\partial \hat{\mathbf{y}}_r}{\partial \mathbf{u}_s} = \mathbf{Z}_{so}^T\mathbf{J}_{so}$$

Proof.

$$\frac{\partial \hat{\mathbf{y}}_r}{\partial \mathbf{u}_s} = \begin{pmatrix} \frac{\partial \hat{y}_{r11}}{\partial u_{s1}} & \cdots & \frac{\partial \hat{y}_{r1C_1}}{\partial u_{s1}} & \cdots & \cdots & \cdots & \frac{\partial \hat{y}_{rD_1}}{\partial u_{s1}} & \cdots & \frac{\partial \hat{y}_{rDC_D}}{\partial u_{s1}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \hat{y}_{r11}}{\partial u_{sD_s}} & \cdots & \frac{\partial \hat{y}_{r1C_1}}{\partial u_{sD_s}} & \cdots & \cdots & \cdots & \frac{\partial \hat{y}_{rD_1}}{\partial u_{sD_s}} & \cdots & \frac{\partial \hat{y}_{rDC_D}}{\partial u_{sD_s}} \end{pmatrix}$$

For all $i = 1, \ldots, C_d$ within each small area $d = 1, \ldots, D$, it can easily be shown from Lemma B.3 that:

$$\frac{\partial \hat{y}_{rd_i}}{\partial u_{sd'}} = \begin{cases} J_{di} & \text{if} \quad d = d', \\ 0 & \text{if} \quad d \neq d' \end{cases}$$

Now if every small area was in-sample then $D_s = D$ and

$$\frac{\partial \hat{\mathbf{y}}_r}{\partial \mathbf{u}_s} = \begin{pmatrix} \mathbf{J}_1 & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{J}_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{J}_{D-1} & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{J}_D \end{pmatrix}$$

otherwise, if $D_s < D$, then

$$\frac{\partial \hat{\mathbf{y}}_r}{\partial \mathbf{u}_s} = \begin{pmatrix} \mathbf{J}_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \mathbf{J}_2 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathbf{J}_{D_s-1} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{J}_{D_s} & 0 & \cdots & 0 \end{pmatrix}$$

$$= \mathbf{Z}_{so}^T\mathbf{J}_{so}$$

Result B.3 :

$$\frac{\partial}{\partial \zeta}(\lambda^T(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K})) = \begin{pmatrix} \mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda \\ \mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda \\ \mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K} \end{pmatrix}$$

Proof.

$$\frac{\partial}{\partial \zeta}(\lambda^T(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K})) = \begin{pmatrix} \frac{\partial}{\partial \beta}(\lambda^T\mathbf{A}\hat{\mathbf{y}}_r) \\ \frac{\partial}{\partial u_s}(\lambda^T\mathbf{A}\hat{\mathbf{y}}_r) \\ \frac{\partial}{\partial \lambda}(\lambda^T(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K})) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial \hat{y}_r}{\partial \beta}\mathbf{A}^T\lambda \\ \frac{\partial \hat{y}_r}{\partial u_s}\mathbf{A}^T\lambda \\ \mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K} \end{pmatrix}$$

Now using Results B.1 and B.2 this gives

$$= \begin{pmatrix} \mathbf{Z}^T\mathbf{J}\mathbf{A}^T\lambda \\ \mathbf{Z}_{so}^T\mathbf{J}_{so}\mathbf{A}^T\lambda \\ \mathbf{A}\hat{y}_r - \mathbf{K} \end{pmatrix}$$

However the second element of this vector can be simplied a little further. Using the denitions of $\mathbf{Z}_{so}^T$ so and $\mathbf{J}_{so}$ introduced in Result B.2 (see also Appendix A), we can show that

$$\mathbf{Z}_{so}^T\mathbf{J}_{so}\mathbf{A}^T = \begin{pmatrix} \mathbf{Z}_s^T & \mathbf{0}_{D_s \times (C - C_s)} \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{J}_s & \mathbf{0}_{C_s \times (C - C_s)} \\ \mathbf{0}_{(C - C_s) \times C_s} & \mathbf{0}_{(C - C_s) \times (C - C_s)} \end{pmatrix} \begin{pmatrix} \mathbf{A}_s^T \\ \mathbf{A}_0^T \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{Z}_s^T \mathbf{J}_s & \mathbf{0}_{D_s \times (C - C_s)} \end{pmatrix} \begin{pmatrix} \mathbf{A}_s^T \\ \mathbf{A}_0^T \end{pmatrix}$$

$$= \mathbf{Z}_s^T \mathbf{J}_s \mathbf{A}_s^T$$

Therefore

$$\frac{\partial}{\partial \zeta} (\lambda^T (\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K})) = \begin{pmatrix} \mathbf{X}^T \mathbf{J} \mathbf{A}^T \lambda \\ \mathbf{A}_s^T \mathbf{J}_s \mathbf{A}_s^T \lambda \\ \mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K} \end{pmatrix}$$

Result B.4 :

$$\frac{\partial}{\partial \beta^T} (\mathbf{X}^T \mathbf{J} \mathbf{A}^T \lambda) = \mathbf{X}^T \mathbf{F} \mathbf{X}$$

Proof. Using Lemma B.5 and noting that we are taking the partial derivative with respect to β^T, not β, we have

$$\frac{\partial}{\partial \beta^T} (\mathbf{X}^T \mathbf{J} \mathbf{A}^{\mathrm{T}} \lambda) = \left(\frac{\partial}{\partial \beta} (\mathbf{X}^T \mathbf{J} \mathbf{A}^T \lambda) \right)^T$$

$$= \left(\frac{\partial \mathbf{J}^*}{\partial \beta} \text{Diagonal}\,[\mathbf{A}^T \lambda]\, \mathbf{X} \right)^T$$

$$= \mathbf{X}^T \text{Diagonal}\,[\mathbf{A}^T \lambda] \frac{\partial \mathbf{J}^*}{\partial \beta^T}$$

Using Lemma B.2, we arrive at the following.

$$\frac{\partial \mathbf{J}^*}{\partial \beta^T} = \begin{pmatrix} \frac{\partial J_{11}}{\partial \beta_1} & \cdots & \frac{\partial J_{11}}{\partial \beta_P} \\ \vdots & \ddots & \vdots \\ \frac{\partial J_{DC_D}}{\partial \beta_1} & \cdots & \frac{\partial J_{DC_D}}{\partial \beta_P} \end{pmatrix}$$

$$= \begin{pmatrix} x_{11}^{(1)} \mathcal{K}_{11} & \cdots & x_{11}^{(P)} \mathcal{K}_{11} \\ \vdots & \ddots & \vdots \\ x_{DC_D}^{(1)} \mathcal{K}_{DC_D} & \cdots & x_{DC_D}^{(P)} \mathcal{K}_{DC_D} \end{pmatrix}$$

$$= \mathcal{K}\mathbf{X}$$

Therefore

$$\frac{\partial}{\partial \beta^T} (\mathbf{X}^T \mathbf{J} \mathbf{A}^T \lambda) = \mathbf{X}^T\, \text{Diagonal}[\mathbf{A}^T \lambda]\, \mathcal{K}\mathbf{X}$$

$$= \mathbf{X}^T \mathbf{F} \mathbf{X}$$

where $\mathbf{F}$ = Diagonal $[\mathbf{A}^T\lambda]\ \mathcal{K}$.

Result B.5 : Define $\mathbf{F}_s$ = Diagonal $\left[\mathbf{A}_s^T \lambda\right] \mathcal{K}_s$.

Then

$$\frac{\partial}{\partial \beta^T} \left(\mathbf{Z}_s^T \mathbf{J}_s \mathbf{A}_s^T \lambda \right) = \mathbf{Z}_s^T \mathbf{F}_s \mathbf{X}_s$$

Proof. To prove this result, we essentially use Lemma B.5. First of all note that we are taking the partial derivative with respect to β^T, not β. Secondly, note that we make the following variable substitutions in order to use Result B.5: $\mathbf{X} \rightarrow \mathbf{Z}_{so}^T$, $\mathbf{Y} \rightarrow \mathbf{J}_{so}$, $\mathbf{b} \rightarrow \mathbf{A}^T \lambda$.

Then we can show that,

$$\frac{\partial}{\partial \beta^T} \left(\mathbf{Z}_s^T \mathbf{J}_s \mathbf{A}_s^T \lambda \right) = \left(\frac{\partial}{\partial \beta} \left(\mathbf{Z}_s^T \mathbf{J}_s \mathbf{A}_s^T \lambda \right) \right)^T$$

$$= \left(\frac{\partial \mathbf{J}_s^*}{\partial \beta} \text{Diagonal} \left[\mathbf{A}_s^T \lambda \right] \mathbf{Z}_s \right)^T$$

$$= \mathbf{Z}_s^T\ \text{Diagonal} \left[\mathbf{A}_s^T \lambda \right] \frac{\partial \mathbf{J}_s^*}{\partial \beta^T}$$

where $\mathbf{J}_s^*$ is a vector of the diagonal elements of the diagonal matrix $\mathbf{J}_s$.

$$\frac{\partial \mathbf{J}_s^*}{\partial \beta^T} = \begin{pmatrix} \frac{\partial J_{11}}{\partial \beta_1} & \cdots & \frac{\partial J_{11}}{\partial \beta_P} \\ \vdots & \cdots & \vdots \\ \frac{\partial J_{1C_1}}{\partial \beta_1} & & \frac{\partial J_{1C_1}}{\partial \beta_P} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \frac{\partial J_{D_s 1}}{\partial \beta_1} & & \frac{\partial J_{D_s 1}}{\partial \beta_P} \\ \vdots & \cdots & \vdots \\ \frac{\partial J_{D_s C_{D_s}}}{\partial \beta_1} & \cdots & \frac{\partial J_{D_s C_{D_s}}}{\partial \beta_P} \end{pmatrix}$$

$$= \begin{pmatrix} x_{11}^{(1)}\mathcal{K}_{11} & \cdots & x_{11}^{(P)}\mathcal{K}_{11} \\ \vdots & \cdots & \vdots \\ x_{1C_1}^{(1)}\mathcal{K}_{1C_1} & & x_{1C_1}^{(P)}\mathcal{K}_{1C_1} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ x_{D_s1}^{(1)}\mathcal{K}_{D_s1} & & x_{D_s1}^{(P)}\mathcal{K}_{D_s1} \\ \vdots & \cdots & \vdots \\ x_{D_sC_{D_s}}^{(1)}\mathcal{K}_{D_sC_{D_s}} & \cdots & x_{D_sC_{D_s}}^{(P)}\mathcal{K}_{D_sC_{D_s}} \end{pmatrix}$$

$$= \begin{pmatrix} \mathcal{K}_1^* & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathcal{K}_{D_s}^* \end{pmatrix} \begin{pmatrix} x_{11}^{(1)} & \cdots & \cdots & x_{11}^{(P)} \\ \vdots & & & \vdots \\ x_{1C_1}^{(1)} & \cdots & \cdots & x_{1C_1}^{(P)} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ x_{D_s1}^{(1)} & \cdots & \cdots & x_{D_s1}^{(P)} \\ \vdots & & & \vdots \\ x_{D_sC_{D_s}}^{(1)} & \cdots & \cdots & x_{D_sC_{D_s}}^{(P)} \end{pmatrix}$$

$$= \boldsymbol{K}_s\boldsymbol{X}_s$$

where

$$\mathcal{K}_d^* = \begin{pmatrix} \mathcal{K}_{d1} & 0 & \cdots & 0 \\ 0 & \mathcal{K}_{d2} & & 0 \\ \vdots & & \ddots & 0 \\ 0 & 0 & 0 & \mathcal{K}_{dC_d} \end{pmatrix}$$

Therefore

$$\frac{\partial\left(\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda\right)}{\partial\beta^T} = \mathbf{Z}_s^T\text{Diagonal}\left[\mathbf{A}_s^T\lambda\right]\mathcal{K}_s\mathbf{X}_s$$

$$= \mathbf{Z}_s^T\mathbf{F}_s\mathbf{X}_s$$

Result B.6.

$$\frac{\partial\mathbf{J}^*}{\partial\mathbf{u}_s^T} = \mathcal{K}_{so}\mathbf{Z}_{so}$$

$$\frac{\partial\mathbf{J}_s^*}{\partial\mathbf{u}_s^T} = \mathcal{K}_s\mathbf{Z}_s$$

Proof.

$$\frac{\partial\mathbf{J}^*}{\partial\mathbf{u}_s^T} = \begin{pmatrix} \frac{\partial J_{11}}{\partial u_{s1}} & \cdots & \frac{\partial J_{11}}{\partial u_{sD_s}} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \frac{\partial J_{DC_D}}{\partial u_{s1}} & & \frac{\partial J_{DC_D}}{\partial u_{sD_s}} \end{pmatrix}$$

From Lemma B.4

$$\frac{\partial J_{di}}{\partial u_{d'}} = \begin{cases} \mathcal{K}_{di} & \text{if} \quad d = d', \\ 0 & \text{if} \quad d \neq d' \end{cases}$$

So

$$\frac{\partial J^*}{\partial u_s^T} = \begin{pmatrix} \mathcal{K}_{11} & 0 & \cdots & 0 \\ \mathcal{K}_{12} & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ \mathcal{K}_{1C_1} & 0 & \cdots & 0 \\ 0 & \mathcal{K}_{21} & \cdots & 0 \\ 0 & \mathcal{K}_{22} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & \mathcal{K}_{2C_2} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \mathcal{K}_{D_s1} \\ 0 & 0 & \cdots & \mathcal{K}_{D_s2} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathcal{K}_{D_sC_{D_s}} \\ 0 & 0 & \cdots & 0 \\ \vdots & & \cdots & \vdots \\ \vdots & & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$= \mathcal{K}_{so}\mathbf{Z}_{so}$$

From the derivation above, it can easily be seen that for the case when $\mathbf{J}^*$ is restricted to the space of sampled small areas (including all age by sex cells within these areas)

$$\frac{\partial\mathbf{J}_s^*}{\partial\mathbf{u}_s^T} = \mathcal{K}_s\mathbf{Z}_s$$

Result B.7.

$$\frac{\partial}{\partial \mathbf{u}_s^T}(\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda) = \mathbf{X}_s^T\mathbf{F}_s\mathbf{Z}_s$$

Proof.

$$\begin{aligned}\frac{\partial}{\partial \mathbf{u}_s^T}(\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda) &= \left(\frac{\partial}{\partial \mathbf{u}_s}\left(\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda\right)\right)^T \\ &= \left(\frac{\partial \mathbf{J}^*}{\partial \mathbf{u}_s}\text{Diagonal}\,[\mathbf{A}^T\lambda]\,\mathbf{X}\right)^T \\ &= \mathbf{X}^T\text{Diagonal}\,[\mathbf{A}^T\lambda]\frac{\partial \mathbf{J}^*}{\partial \mathbf{u}_s^T} \\ &= \mathbf{X}^T\text{Diagonal}\,[\mathbf{A}^T\lambda]\,\mathcal{K}_{so}\mathbf{Z}_{so} \\ &= \mathbf{X}^T\mathbf{F}_{so}\mathbf{Z}_{so} \\ &= \begin{pmatrix}\mathbf{X}_s^T & \mathbf{X}_0^T\end{pmatrix} \begin{pmatrix}\mathbf{F}_s & \mathbf{0}_{C_s\times(C-C_s)} \\ \mathbf{0}_{(C-C_s)\times C_s} & \mathbf{0}_{(C-C_s)\times(C-C_s)}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_s \\ \mathbf{Z}_0\end{pmatrix} \\ &= \begin{pmatrix}\mathbf{X}_s^T\mathbf{F}_s & \mathbf{0}_{P\times(C-C_s)}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_s \\ \mathbf{Z}_0\end{pmatrix} \\ &= \mathbf{X}_s^T\mathbf{F}_s\mathbf{Z}_s\end{aligned}$$

Result B.8.

$$\frac{\partial}{\partial \mathbf{u}_s^T}\left(\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda\right) = \mathbf{Z}_s^T\mathbf{F}_s\mathbf{Z}_s$$

Proof.

$$\begin{aligned}\frac{\partial}{\partial \mathbf{u}_s^T}\left(\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda\right) &= \left(\frac{\partial}{\partial \mathbf{u}_s}\left(\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda\right)\right)^T \\ &= \left(\frac{\partial \mathbf{J}_s^*}{\partial \mathbf{u}_s}\text{Diagonal}\,[\mathbf{A}_s^T\lambda]\,\mathbf{Z}_s\right)^T \\ &= \mathbf{Z}_s^T\text{Diagonal}\left[\mathbf{A}_s^T\lambda\right]\frac{\partial \mathbf{J}_s^*}{\partial \mathbf{u}_s^T}\end{aligned}$$

where $\mathbf{J}_s^*$ is a vector of the diagonal elements of the diagonal matrix $\mathbf{J}_s$. By Result B.6 this becomes:

$$\begin{aligned}&= \mathbf{Z}_s^T\text{Diagonal}\left[\mathbf{A}_s^T\lambda\right]\mathcal{K}_s\mathbf{Z}_s \\ &= \mathbf{Z}_s^T\mathbf{F}_s\mathbf{Z}_s\end{aligned}$$

Result B.9.

$$\frac{\partial l}{\partial \zeta} = \begin{pmatrix}\mathbf{X}_{ss}^T \\ \mathbf{Z}_{ss}^T \\ \mathbf{0}\end{pmatrix}\frac{\partial l_1}{\partial \eta} + \begin{pmatrix}\mathbf{0} \\ -\mathbf{\Omega}^{-1}u_s \\ \mathbf{0}\end{pmatrix} + \begin{pmatrix}\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda \\ \mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda \\ \mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K}\end{pmatrix} \quad \text{(B.3)}$$

Proof. Differentiating (3.1) with respect to ζ we obtain:

$$\begin{aligned}\frac{\partial l}{\partial \zeta} &= \frac{\partial l_1}{\partial \zeta} + \frac{\partial l_2}{\partial \zeta} + \frac{\partial}{\partial \zeta}\left(\lambda^T(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K})\right) \\ &= \frac{\partial \eta}{\partial \zeta}\frac{\partial l_1}{\partial \eta} + \frac{\partial l_2}{\partial \zeta} + \frac{\partial}{\partial \zeta}\left(\lambda^T(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K})\right) \quad \text{(B.4)}\end{aligned}$$

It follows from (3.3) that

$$\frac{\partial \eta}{\partial \zeta} = \begin{pmatrix}\mathbf{X}_{ss}^T \\ \mathbf{Z}_{ss}^T \\ \mathbf{0}_{R\times C_{ss}}\end{pmatrix} \quad \text{(B.5)}$$

Of the model parameters $\boldsymbol{\beta}$, $\mathbf{u}_s$ and λ that we wish to estimate, l_2 only involves the parameter $\mathbf{u}_s$. Therefore,

$$\frac{\partial l_2}{\partial \zeta} = \begin{pmatrix}\mathbf{0}_{P\times 1} \\ -\mathbf{\Omega}^{-1}\mathbf{u}_s \\ \mathbf{0}_{R\times 1}\end{pmatrix} \quad \text{(B.6)}$$

The third term from (B.4) is as follows. Given that **A** and **K** are constants with respect to the parameters $\boldsymbol{\beta}$, $\mathbf{u}_s$ and λ, using the chain rule for matrix calculus, Result B.3 of the Appendix gives:

$$\frac{\partial}{\partial \zeta}\left(\lambda^T(\mathbf{A}\hat{y}_r - \mathbf{K})\right) = \begin{pmatrix}\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda \\ \mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda \\ \mathbf{A}\hat{y}_r - \mathbf{K}\end{pmatrix} \quad \text{(B.7)}$$

where

$\mathbf{J}$ = Diagonal $[(N_{di} - n_{di})\,p_{di}\,(1 - p_{di})]$,
$d = 1,\ldots,D;\ i = 1,\ldots,C_d$

$\mathbf{J}_s$ = Diagonal $[(N_{di} - n_{di})\,p_{di}\,(1 - p_{di})]$,
$d = 1,\ldots,D_s$, i $= 1,\ldots,C_d$

and $\mathbf{A}_s$ is the submatrix of **A** obtained by taking the first $C_s = \sum_{d=1}^{D_s} C_d$ columns.

By putting equations (B.5), (B.6) and (B.7) into equation (B.4) we arrive at the required result.

Result B.10.

$$\frac{\partial^2 l}{\partial \zeta\,\partial \zeta^T} = \begin{pmatrix}\mathbf{X}_{ss}^T \\ \mathbf{Z}_{ss}^T \\ \mathbf{0}\end{pmatrix}\frac{\partial^2 l_1}{\partial \eta\,\partial \eta^T}(\mathbf{X}_{ss}\mathbf{Z}_{ss}\,\mathbf{0}) - \begin{pmatrix}\mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{pmatrix} + \begin{pmatrix}X^TFX & X_s^TF_sZ_s & X^TJA^T \\ Z_s^TF_sX_s & Z_s^TF_sZ_s & Z_s^TJ_sA_s^T \\ AJX & A_sJ_sZ_s & 0_{R\times R}\end{pmatrix} \quad \text{(B.8)}$$

Proof. The second partial derivative of the loglikelihood is given by:

$$\frac{\partial^2 l}{\partial\zeta\,\partial\zeta^T} = \frac{\partial^2 l_1}{\partial\zeta\,\partial\zeta^T} + \frac{\partial^2 l_2}{\partial\zeta\,\partial\zeta^T} + \frac{\partial^2}{\partial\zeta\,\partial\zeta^T}\left(\lambda^T(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K})\right) \qquad \text{(B.9)}$$

Now using the definition of the second partial derivative with respect to a vector and the chain rule for matrix differential calculus we have

$$\begin{aligned}
\frac{\partial^2 l_1}{\partial\zeta\,\partial\zeta^T} &= \frac{\partial}{\partial\zeta^T}\left(\frac{\partial l_1}{\partial\zeta}\right) \\
&= \frac{\partial}{\partial\zeta^T}\left(\frac{\partial\eta}{\partial\zeta}\frac{\partial l_1}{\partial\eta}\right) \\
&= \frac{\partial\eta^T}{\partial\zeta^T}\frac{\partial}{\partial\eta^T}\left(\frac{\partial\eta}{\partial\zeta}\frac{\partial l_1}{\partial\eta}\right) \\
&= \frac{\partial\eta}{\partial\zeta}\frac{\partial}{\partial\eta^T}\left(\frac{\partial\eta}{\partial\zeta}\frac{\partial l_1}{\partial\eta}\right) \\
&= \begin{pmatrix}\mathbf{X}_{ss}^T \\ \mathbf{Z}_{ss}^T \\ \mathbf{0}\end{pmatrix}\frac{\partial}{\partial\eta^T}\left(\begin{pmatrix}\mathbf{X}_{ss}^T \\ \mathbf{Z}_{ss}^T \\ \mathbf{0}\end{pmatrix}\frac{\partial l_1}{\partial\eta}\right) \\
&= \begin{pmatrix}\mathbf{X}_{ss}^T \\ \mathbf{Z}_{ss}^T \\ \mathbf{0}\end{pmatrix}\frac{\partial^2 l_1}{\partial\eta\,\partial\eta^T}\begin{pmatrix}\mathbf{X}_{ss} & \mathbf{Z}_{ss} & \mathbf{0}\end{pmatrix} \qquad \text{(B.10)}
\end{aligned}$$

and

$$\frac{\partial^2 l_2}{\partial\zeta\,\partial\zeta^T} = \frac{\partial}{\partial\zeta^T}\begin{pmatrix}\mathbf{0} \\ -\boldsymbol{\Omega}^{-1}\mathbf{u}_s \\ \mathbf{0}\end{pmatrix} = \begin{pmatrix}\mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\Omega}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{pmatrix} \qquad \text{(B.11)}$$

given that $\boldsymbol{\Omega}^{-1}$ is symmetric.

Using Result (B.3) and Definition B.1, the third term on the right hand side of (B.9) is

$$\frac{\partial^2}{\partial\zeta\,\partial\zeta^T}\left(\lambda^T(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K})\right) = \frac{\partial}{\partial\zeta^T}\begin{pmatrix}\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda \\ \mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda \\ \mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K}\end{pmatrix}$$

$$= \begin{pmatrix}
\frac{\partial}{\partial\beta^T}(\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda) & \frac{\partial}{\partial u_s^T}(\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda) & \frac{\partial}{\partial\lambda^T}(\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda) \\
\frac{\partial}{\partial\beta^T}(\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda) & \frac{\partial}{\partial u_s^T}(\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda) & \frac{\partial}{\partial\lambda^T}(\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda) \\
\frac{\partial}{\partial\beta^T}(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K}) & \frac{\partial}{\partial u_s^T}(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K}) & \frac{\partial}{\partial\lambda^T}(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K})
\end{pmatrix} \qquad \text{(B.12)}$$

Let $\mathbf{J}^*$ be the column vector consisting of the diagonal elements of $\mathbf{J}$. Then, from Result B.4

$$\frac{\partial}{\partial\beta^T}(\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda) = \mathbf{X}^T\mathbf{F}\mathbf{X} \qquad \text{(B.13)}$$

From Appendix B, we also have the following results:

$$\frac{\partial}{\partial\beta^T}(\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda) = \mathbf{Z}_s^T\mathbf{F}_s\mathbf{X}_s \qquad \text{(Result B.5)} \qquad \text{(B.14)}$$

$$\frac{\partial}{\partial\mathbf{u}_s^T}(\mathbf{X}^T\mathbf{J}\mathbf{Z}^T\lambda) = \mathbf{X}_s^T\mathbf{F}_s\mathbf{Z}_s \qquad \text{(Result B.7)} \qquad \text{(B.15)}$$

$$\frac{\partial}{\partial\mathbf{u}_s^T}(\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda) = \mathbf{Z}_s^T\mathbf{F}_s\mathbf{Z}_s \qquad \text{(Result B.8)} \qquad \text{(B.16)}$$

$$\frac{\partial}{\partial\beta^T}(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K}) = \mathbf{A}\mathbf{J}\mathbf{X} \qquad \text{(Result B.1)} \qquad \text{(B.17)}$$

$$\frac{\partial}{\partial\mathbf{u}_s^T}(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K}) = \mathbf{A}_s\mathbf{J}_s\mathbf{Z}_s \qquad \text{(Result B.2)} \qquad \text{(B.18)}$$

$$\frac{\partial}{\partial\lambda^T}(\mathbf{X}^T\mathbf{J}\mathbf{A}^T\lambda) = \mathbf{X}^T\mathbf{J}\mathbf{A}^T \qquad \text{(B.19)}$$

$$\frac{\partial}{\partial\lambda^T}(\mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T\lambda) = \mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T \qquad \text{(B.20)}$$

$$\frac{\partial}{\partial\lambda^T}(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K}) = \mathbf{0} \qquad \text{(B.21)}$$

Therefore, putting equations (B.13) to (B.21) into (B.12) we obtain:

$$\frac{\partial^2}{\partial\zeta\,\partial\zeta^T}\left(\lambda^T(\mathbf{A}\hat{\mathbf{y}}_r - \mathbf{K})\right) = \begin{pmatrix}
\mathbf{X}^T\mathbf{F}\mathbf{X} & \mathbf{X}_s^T\mathbf{F}_s\mathbf{Z}_s & \mathbf{X}^T\mathbf{J}\mathbf{A}^T \\
\mathbf{Z}_s^T\mathbf{F}_s\mathbf{X}_s & \mathbf{Z}_s^T\mathbf{F}_s\mathbf{Z}_s & \mathbf{Z}_s^T\mathbf{J}_s\mathbf{A}_s^T \\
\mathbf{A}\mathbf{J}\mathbf{X} & \mathbf{A}_s\mathbf{J}_s\mathbf{Z}_s & \mathbf{0}_{R\times R}
\end{pmatrix} \qquad \text{(B.22)}$$

which turns out to be a symmetric matrix.

Putting (B.10), (B.11) and (B.22) into (B.9), we obtain the required result.