A Hierarchical Bayesian Approach for Addressing Multiple Objectives in Poverty Research for Small Areas

Stefano Marchetti [a*], Gaia Bertarelli [b**], Monica Pratesi [c*] and Partha Lahiri[d***]

[*]Department of Economics and Management, University of Pisa
[**]Department of XXX, Sant'Anna School XXX
[***]Department of Mathematics, University of Maryland

Abstract

Nowadays the information extracted from data should be the key to good policy, therefore, analysts must make the best possible use of all available information. However, data availability often is limited by cost or for other reasons. Consequently, there is the need to use data from different sources. Our goals are to develop hierarchical models and to demonstrate their ability to improve inferences about quantities for which there are meager data. When a hierarchical model can be found to represent the situation properly, analysis of that model often can be used to extract most or all of the relevant information and so provide the best possible estimates. The application considered will include small area estimation in the context of the EU Statistics on Income and Living Conditions. In developing the hierarchical model, we use together survey data and population registers. As for the implementation of the hierarchical model, we propose to use Bayesian methodology assisted by Monte Carlo Markov Chain.

Keywords— poverty mapping, population register, multilevel modelling

1 Introduction

The main goal of this work is to propose a hierarchical model that is able to use data at a different level of aggregation and that come from different sources in

[a]stefano.marchetti@unipi.it
[b]gaia.bertarelli@sssup.it
[c]monica.pratesi@unipi.it
[d]plahiri@umd.edu

order to make inference on the Foster et al. (1984) poverty measures (FGT) at small area level.

In particular we make use of survey data and administrative data, which are collected for many different porpoises, therefore they are available at different level of aggregation. Often, administrative data are aggregated following administrative subdivision of the country, Italy in our application. Italy is divided into 5 repartitions (NUTS 1 level according to the EU nomenclature), 20 regions (NUTS 2), 107 provinces (NUTS 3/LAU 1) and about 8000 municipalities (LAU 2). The goal of our application is the estimation of FGT indexes at the provincial level, in particular poverty incidence. Nevertheless, domains different from administrative boundaries are possible under the proposed framework, e.g. provinces by age class and gender.

The combined use of administrative data and survey data fits the new register-based paradigm of many national statistical offices, where administrative data play a central role in the production of official statistics. This new paradigm requires appropriate models to exploit all the data available in order to produce sound statistics at the small area level.

2 Target paramters and data

Our goal is to obtain FGT indexes at the province level in Italy. Let w_{ijkl} be a wealth variable in region $i = 1, \ldots, R$, province $j = 1, \ldots, D_i$, municipality $k = 1, \ldots, M_{ij}$ and household $l = 1, \ldots, N_{ijk}$ and t be a fixed national threshold that classified poor and non poor households. Then, FGT poverty measure at provincial level is defined as follows:

$$FGT(t)_{ij,\alpha} = \frac{1}{\sum_{k=1}^{M_{ij}} N_{ijk}} \sum_{k=1}^{M_{ij}} \sum_{l=1}^{N_{ijk}} \Big(\frac{t - w_{ijklt}}{t}\Big)^{\alpha} I(w_{ijkl} < t),$$

where $\alpha = \{0, 1, 2\}$ define poverty incidence, intensity and severity respectively. As wealth variable we use the equivalised household income, which is computed as the total available household income divided by the equivalised household size according to the OECD modified scale that assign weight 1 to the first adult, 0.5 to other adults and 0.3 for children (age less than 14).

The equivalised household income is available from the EU Statistics on Income and Living Conditions (SILC) survey, which is conducted yearly by Istat and represent the reference source in the EU for comparative statistics on income distribution and social exclusion. It surveys personal data, income, working status, housing, leisure activities. The 2017 Italian EU-SILC survey has a sample size of about 22 thousand households. It is a two-stage sample design stratified by region and type of municipality. The PSU are the municipalities and the SSU are the households. The survey use a rotating panel each 4 year. More details are available on the Istat website.

Municipalities level data are organised in administrative archives, which are integrated by Istat under the ARCHIMEDE project. The administrative sources used to build ARCHIMEDE microdata are: municipal population registers, tax return registers, central register of pensioners, social security and fiscal sources, social security benefit registers and the population census. Italian population counts about 60 million persons in about 24 million households.

Province level data can be obtained properly aggregating municipality level data, since municipalities are partitions of provinces. However, some data are available only at provincial level, such as the labour force data. This information come the labour force survey (LFS), which is a cross-sectional and longitudinal household sample survey. It provides information about main labour market indicators, broken down by socio-demographic variables. The LFS in Italy follows a rotating sample design where households participate for two consecutive quarters, then they exit for the next two quarters, and finally come back for other two quarters (2 in - 2 out - 2 in rotation). The 2017 LFS in Italy has a sample size of about 250 thousand households, about 600 thousand persons, which guarantee reliable estimates also at the provincial level for what concern annual estimates.

Region specific data can be available from other surveys, but are not considered at this stage in this work.

3 Proposed hierarchical Bayes multilevel model

Hierarchical Bayesian (HB) models have been extensively used in small area estimation, see for example Rao and Molina (2015) for a general review. They can accommodate very complex models based on very simple models as building blocks. Another great advantage of these models is about the estimation of the standard error of the small area HB estimators, which can be found exactly without using approximations. In this framework we can obtain credible intervals and useful summaries from the posterior distributions with practically no additional effort.

In order to obtain reliable estimates of poverty incidence $FGT(t)_{ij,0}$ at provincial level, we propose a three-level cross-sectional model. The three levels are household, municipality and province. Some parameters are defined region specific. The proposed method require to have a transformation $T(w)$ of the wealth variable such that $y = T(w)$ is approximately normal. The mode can be represented as follows:

$$\begin{aligned}
&\text{L.1 } y_{ijkl}|\theta_{ijk}, \boldsymbol{\beta}_i, \sigma_i^2 \sim N(\theta_{ijk} + \mathbf{a}_{ijkl}\boldsymbol{\beta}_i, \sigma_i^2)\\
&\text{L.2 } \theta_{ijk}|\eta_{ij}, \boldsymbol{\gamma}_i, \tau_i^2 \sim N(\eta_{ij} + \mathbf{b}_{ijk}\boldsymbol{\gamma}_i, \tau_i^2)\\
&\text{L.3 } \eta_{ij}|\xi, \boldsymbol{\lambda}, \delta^2 \sim N(\xi + \mathbf{c}_{ij}\boldsymbol{\lambda}, \delta^2),
\end{aligned}$$

where θ_{ijk} is the municipality random effect, $\mathbf{a}_{ijkl}$ are the household level covariates from EU-SILC, η_{ij} is the provincial level random effect, $\mathbf{b}_{ijk}$ are the municipality level covariates from ARCHIMEDE, ξ is a fixed effect and $\mathbf{c}_{ij}$ are the province level covariates from LFS. As a note, c_{ij} covariates are affected by sampling error, which is considered negligible in this work, and then they are treated as true values.

Following Gelman (2015) we use proper informative priors, half-Cauchy for δ, τ_i, σ_i, multivariate normal for $\boldsymbol{\gamma}_i, \boldsymbol{\beta}_i$ and normal for $\xi, \boldsymbol{\lambda}$.

The household level covariate we use is the household size groups: 1 member, 2 members, 3 members, 4 members, 5 or more members. The municipality level covariates are the proportion of persons in age classes (13 to 35, 36 to 65, 66 or more), proportion of male, proportion of persons in 3 type of work contract (dependent, independent, other), median of equivalised taxable income. Note, this last covariate is different from the median equivalised household income estimated from the EU-SILC survey because of a different taxonomy. The province level covariate is the unemployment rate, that is the proportion of persons who don't work while seeking for a job.

An estimate of the unknown quantity $FGT_{ij,\alpha}$ can be obtained as follows:

$$F\bar{G}T(t, \theta_{ijk}, \beta_i, \sigma_i)_{ij,\alpha} = \frac{1}{N_{ij}} \sum_{k=1}^{m_{ij}} \sum_{l=1}^{n_{ijk}} E[g_\alpha(w_{ijkl})|\theta_{ijk}, \boldsymbol{\beta}_i, \sigma_i^2]\omega_{ijkl},$$

where

$$g_\alpha(w_{ijkl}) = \Big(\frac{t - w_{ijkl}}{t}\Big)^\alpha I(w_{ijkl} < t),$$

m_{ij} are the sampled municipality in province j of region i, ω_{ijkl} is the survey weight for household l in municipality k in province j in region i, n_{ijk} is the sample size in municipality k in province j in region i.

For $\alpha = 0$,

$$E[g_\alpha(w_{ijkl})|\theta_{ijk}, \boldsymbol{\beta}_i, \sigma_i^2] = \int_{-\frac{\theta_{ijk}+\mathbf{a}_{ijk}\boldsymbol{\beta}_i}{\sigma_i}}^{\frac{\log t-(\theta_{ijk}+\mathbf{a}_{ijk}\boldsymbol{\beta}_i)}{\sigma_i}} \phi(z|\theta_{ijkl}, \sigma_i, \boldsymbol{\beta}_i)dz$$
$$= \Phi\Big(\frac{\log t - (\theta_{ijk} + \boldsymbol{a}_{ijk}\boldsymbol{\beta}_i)}{\sigma_i}\Big) - \Phi\Big(-\frac{\theta_{ijk} + \mathbf{a}_{ijk}\boldsymbol{\beta}_i}{\sigma_i}\Big),$$

where ϕ and Φ are respectively the density function and the distribution function of the standard normal distribution.

The model parameters are estimated using Gibbs sampling by Monte Carlo Markov Chain (MCMC). To obtain stable posterior distribution of model parameters we use a *lasso* penalty. Let H be the number of MCMC samples after burn-in. Let $\theta_{ijk,h}$, $\boldsymbol{\beta}_{i,h}$ and $\sigma_{i,h}$ denote the hth MCMC draw of θ_{ijk}, $\boldsymbol{\beta}_i$ and σ_i, respectively ($h = 1, \ldots, H$). We define the $D_i \times H$ matrix $\mathbf{F}_\alpha$, where the j, h entry is defined as $\mathbf{F}_{(j,h);\alpha} = F\bar{G}T(t, \theta_{ijk,h}, \beta_{i,h}, \sigma_{i,h}))_{ij,\alpha}$.

According to Lahiri and Suntornchost (2018) the matrix $\mathbf{F}_\alpha$ provides samples generated from the posterior distribution of $F\bar{G}T(t, \theta_{ijk}, \beta_i, \sigma_i)_{ij,\alpha}, j = 1 \ldots, D_i, i = 1, \ldots, R$ and so is adequate for solving a variety of inferential problems in a Bayesian way. Lahiri and Suntornchost (2018) suggest three different inferential problems: 1. estimates $\widehat{FGT}(t)_{ij,\alpha}$ of $FGT(t)_{ij,\alpha}$ obtained as the posterior mean of $\mathbf{F}_{(j,h);\alpha}, h = 1, \ldots, H$, and estimates $\widehat{MSE}(\widehat{FGT}(t)_{ij,\alpha})$ of $MSE(\widehat{FGT}(t)_{ij,\alpha})$ obtained as the posterior standard deviation of $\mathbf{F}_{(j,h);\alpha}, h = 1, \ldots, H$; 2. identification of provinces that are out of predefined bounds and 3. identify the worst and best provinces according to FGT indexes. In this work we focus on point 1 only, with $\alpha = 0$. As a remark inference on points 2. and 3. make use of $\mathbf{F}_\alpha$ taking advantage of a unique hierarchical Bayes framework.

4 Application Results

In this section we show the $FGT(t)_{ij,0}$ estimates for 27 provinces in three Italian regions, namely Lombardia, Tuscany and Campania. This choice is due to the availability of ARCHIMEDE data, which have been available to us under an agreement between ISTAT and University of Pisa.

We analyse the model parameters estimates through convergence plot.

We compare the direct estimates with the HB estimates (Figure 1 and Figure 2).

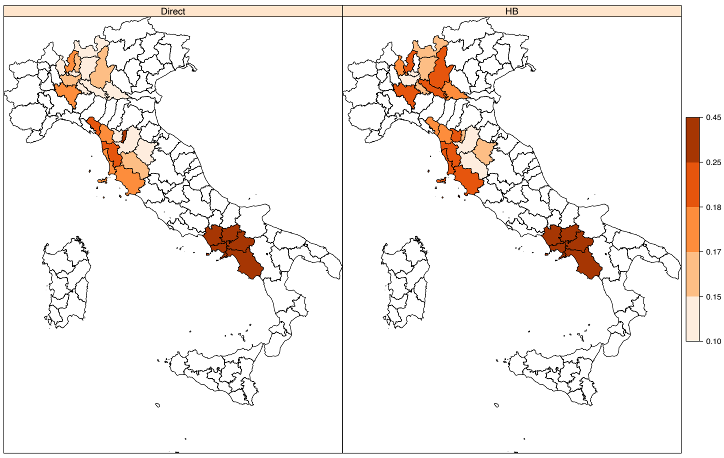

Figure 1: Direct and HB ARPR estimates computed using EU-SILC 2017 at provincial level (NUTS 3)

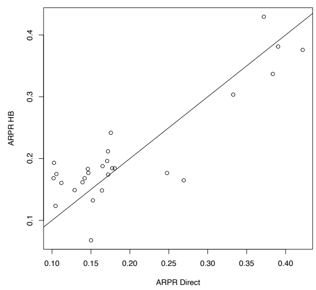

Figure 2: Direct and HB ARPR estimates computed using EU-SILC 2017 at provincial level (NUTS 3).

Model estimates are more reliable than direct ones, with a clear reduction in their coefficients of variation (Figure 3).

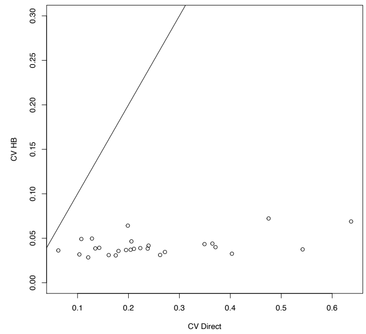

Figure 3: Direct and HB ARPR estimates CVs.

5 Conclusions

In this work we have successfully integrate administrative and survey data at different level of aggregation to obtain posteriors distribution of a multilevel model parameters, which allow different inferential goals. In particular we focus on the incidence of relative poverty, one of the main indicators used by policy makers and stakeholders.

In future works the hierarchical Bayes model can be improved by taking into account the measurement error of auxiliary variables coming from survey data, such as the unemployment rates coming from LFS. Furthermore, the model can be enriched by big data coming for example from google trends, twitter text analysis or supermarket scanner data (which collect price and quantity of retail chains spread across Italy).

References

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52:761–766.

Gelman, A. (2015). 3 new priors you can't do without, for coefficients and variance parameters in multilevel regression. Statistical Modeling, Causal Inference, and Social Science blog.

Lahiri, P. and Suntornchost, J. (2018). A general bayesian approach to meet different inferential goals in poverty research for small areas.

Rao, J. and Molina, I. (2015). *Small Area Estimation*. Wiley Series in Survey Methodology. Wiley.

On benchmarking small area estimators when the model is misspecified

Laura Marcis [a*], Maria Chiara Pagliarella [b**] and Renato Salvatore [c*]

[*]University of Cassino and Southern Lazio, Italy
[**]Italian National Institute for Public Policy Analysis

Abstract

One of the main motivations of concern when we apply a small area estimation model is to relate individual area estimates with some direct estimates in a larger area. External and internal benchmarked estimators provide adjusted model-based estimates, in order to agree with that aggregated results. The use of multiple calibration quantities in the benchmarking matrix suggests that the underlying "true" model is misspecified by the actual model equation. We examine the appropriateness of employing the benchmarking matrix to account for omitted variables in the model, through an additional regression term.

Keywords— Fay-Herriot model, benchmarking estimators, model misspecification, augmented model

1 Introduction

In the context of small area estimation, benchmarking is justified by the need for adjusting individual area level estimates to agree with direct estimates of a larger area. The Eblup estimators do not satisfy the benchmarking property, and thus, in the last years, many authors studied a variety of benchmarking techniques, in order to address this issue. In general, these methods rely on some modification of the Eblup by simple adjustments, as for the ratio and the difference benchmarking estimators (Steorts and Ghosh, 2013). Otherwise, an optimal benchmarking estimator that is model unbiased and at the same time satisfies the design-consistency property was obtained by Wang et al. (2008). Bell et al. (2013) give a general result for the optimal estimator in case of

[a]laura.marcis@unicas.it
[b]mc.pagliarella@inapp.org
[c]rsalvatore@unicas.it

multiple benchmarking constraints, by joining together external and internal benchmarking using a common relation. Under model misspecification, Wang et al. (2008) also proposed an augmented model, by inserting a sampling variance model covariate, adjusted by the proportion of units in the corresponding area. Simulation experiments has shown that the augmented model estimator performs well in case of model misspecification, when the omitted variable is correlated with the augmented covariate. Nevertheless, self-benchmarking as in the You and Rao (2002) approach generally ensures efficiency in terms of the MSE, when direct estimates for the larger area suggest a model failure. By a general approach with multiple benchmarking constraints, this paper introduces a benchmarking linear estimator, assuming a model misspecification by an omitted variable factor. The underlying assumption is that direct estimates for the larger area accounts for the true model. Then, we propose an augmented model that incorporates a tentative approach to the model failure. We show that misspecification is proportional to the orthogonal projection of the direct estimate in the subspace of the benchmarking constraints. An application study is reported, in order to introduce the validity of this approach.

2 Theory

Following the Bell et al. (2013) approach to benchmarking small area estimators, as regards the application of the Fay-Herriot model, we say that a) $\theta = X\beta + u$ represents the population area-level parameter model, b) $y = \theta + e$ the sampling model, and, c) $t = W'\theta + \eta$, the benchmarking model by an external random data vector. θ is the $m \times 1$ vector of area parameters, X is the $m \times p$ covariates design matrix, β the $p \times 1$ regression parameters, u is the regression error, y the $m \times 1$ vector of sampling estimates, e is the sampling error, with given $var(e) = R = diag(\psi_1, ..., \psi_m)$. Furthermore, t is the $q \times 1$ vector of benchmarking constraints ($q < m$) to which the area-level estimates must agree, with η the $q \times 1$ related sampling errors. W is a $m \times q$ "benchmarking" matrix, that contains the multiple constraints that links the small area parameters with t. The model variance for θ is $var(y) = Q = \Sigma_u + R$, $\Sigma_u = \sigma_u^2 I_m$. Finally, $cov(u, \eta) = cov(u, e) = 0$, with a chance of a non-zero covariance between sampling errors, i.e. $cov(e, \eta) = C$. Assuming model normality, together with standard Bayesian prior for β, we know that $\Sigma_y = \sigma_\beta^2 XX' + Q$, and, as $\sigma_\beta^2 \longrightarrow \infty$ and by matrix inversion rules, $\Sigma_y^{-1} = Q^{-1}(I - P_X)$. $P_X = X(X'Q^{-1}X)^{-1}X'Q^{-1}$ is the projection matrix onto the subspace by X in the metric of Q^{-1}.

Knowing σ_u^2, and denoting by $\widetilde{\theta}_y = E(\theta|y)$ the best unbiased linear predictor, we have that $\widetilde{\theta}_y = y - RPy$, and $mse(\widetilde{\theta}_y) = var(\widetilde{\theta}_y) = R - RPR$, with P is the projection matrix onto the complement of the column space of X in the metric of Q^{-1}. Assuming the benchmarking model for θ, we get (Bell et al., 2013) $\widetilde{\theta}_{y,t} = E(\theta|y, t)$ as the best linear "adjusted" predictor based on both data sources (y, t):

$$\widetilde{\theta}_{y,t} = E(\theta|y,t) = \widetilde{\theta}_y + cov(\theta, t|y) var(t|y)^{-1} \{t - E(t|y)\}, \quad (1)$$

$$mse(\widetilde{\theta}_{y,t}) = var(\widetilde{\theta}_y) - cov(\theta, t|y) var(t|y)^{-1} \{cov(\theta, t|y)\}'. \quad (2)$$

When $var(\eta) = \Sigma_\eta \longrightarrow 0$, $\widetilde{\theta}_{y,t}$ becomes the "externally" benchmarked predictor $\widetilde{\theta}_E = \widetilde{\theta}_y + var(\widetilde{\theta}_y)W[W'var(\widetilde{\theta}_y)W]^{-1}[t - W'\widetilde{\theta}_y]$. Considering $W'\theta = t$ as the projection of the parameter vector θ onto the subspace of W, when we have no external data t and that projection is that relating to y, i.e. $W'y$, the (1) becomes the "internal" benchmarked predictor $\widetilde{\theta}_I = \widetilde{\theta}_y + var(\widetilde{\theta}_y)W[W'var(\widetilde{\theta}_y)W]^{-1}W'(y - \widetilde{\theta}_y)$. In both cases, $\widetilde{\theta}_E$ and $\widetilde{\theta}_I$ verify the benchmarking property $W'\widetilde{\theta}_E = t$ and $W\prime\widetilde{\theta}_I = W'y$, respectively.

In the standard regression theory it is well-known that if for the model in a) we get $E(u|x) \neq 0$, the covariates are said endogeneous in the linear model, i.e. almost one of the explanatory variables is correlated with the regression error u. One of the most important endogeneity problem arises when model misspecification is due to some omitted variables in the equation model. This situation leads in general to the "omitted variable bias" of the fixed-effects estimator, together with an overestimation of the error variance. When important regressors are ignored and the correlation between included and omitted regressors is relevant, the correlation between the covariates and the model error u increases. Conversely, it matters to delete from the model "unimportant" regressors, because they may increase the sampling variance. In large samples, the bias of estimates becomes the major issue (Davidson et al., 2004). Although the standard area level model is mixed linear model, ignoring omitted variables in the regression component of the model, and the consequent unseemly random-area effect variance estimation, may adversely affect the linear predictor. For example, it can be shown that if the true mixed linear model with known model variance Q is $y = X_1\beta_1 + X_2\beta_2 + \pi + Zv + e$, with π as the unobservable omitted vector of fixed effects, and v and e, the random effects and the residual error, respectively, the bias for the fixed-effects estimates of β_1 is $B(\widehat{\beta}_{1,gls}) = A'(I - X_2B')\pi$, $A = Q^{-1}X_1(X_1'Q^{-1}X_1)^{-1}$, $B = PX_2(X_2'PX_2)^{-1}$.

Generally, the presence of omitted variables in the structural linear model significantly correlated with the regression error may be expressed by a model error u, composed of two parts. Denoting by $\theta^m = X\beta + u$ the assumed but incorrect population model, and q the omitted regressor, then $u = q\gamma + v$, where v is the "true" structural regression error. Given the true model θ, we have:

$$\theta = \theta^m + (\theta - \theta^m) = X\beta + q\gamma + v. \quad (3)$$

As q is the unobservable factor in the model for θ, v is uncorrelated with all the covariates $x_1, ..., x_p$, and q. Further, other relations are given, since by the normalization due by the model q has zero mean: $E(u|x) \neq E(u)$, $E(u|q) \neq 0$,$cov(u,q) \neq 0$, $E(v|x,q) = 0$, $E_v(\theta - \theta^m) = q\gamma$. With the true model for θ, and given the benchmarking model c), it is straightforward that $E(t) = W'\theta \neq W'\theta^m$. In the same way, taking the subspace spanned by the columns of W, we observe that $BW'\theta^m \neq BW'\theta = Bt$, $B = W(W'W)^{-1}$, as BW' is the orthogonal projection matrix for W. A proxy-variable solution, say z, for the unobserved factor q, can be assumed as dependent on the difference $(W'\theta - W'\theta^m)$, i.e. $z \propto E(BW'\theta - BW'\theta^m)$. As z becomes a linear regressor for q, we may have in general:

$$q = z\lambda + r = \lambda_0 + \lambda_1 z + r, \tag{4}$$

$E(r|z) = 0, E(r|x,z) = 0$, and, as requested by standard "redundancy" conditions about proxy variables, we can easily check for z that $E(\theta|x,q,z) = E(\theta|x,q)$. Furthermore, given the linear projection $L(q|x,z)$, due to the circumstance that $cov(x,r) = 0$ we may observe that $L(q|x,z) = L(q|z)$. By putting (4) into (3), we get:

$$\theta = X\beta + \gamma\lambda_0 1 + \lambda_1 BW'(\theta - X\beta)\gamma + \gamma r + v,$$
$$(I - \gamma\lambda_1 BW')(\theta - X\beta) = \gamma\lambda_0 1 + \gamma r + v.$$

Thus:

$$\theta_M = X\overline{\beta} + (I - b_1 BW')^{-1}\epsilon, \tag{5}$$

with $\overline{\beta} = \beta_0 + (I - b_1 BW')^{-1} b_0 1$, $\epsilon = \gamma r + v$, $b_0 = \gamma\lambda_0$, $b_1 = \gamma\lambda_1$. By mitigating the omitted factor in the assumed model, equation (5) defines a new model for θ, due to the availability of a proxy variable z. The last by "mirroring" differences in the parameter θ by the projection onto the subspace defined by the benchmarking matrix W. It is straightforward to see for the model (5) that $var(\theta_M) = \sigma^2_\epsilon (I - b_1 BW')^{-1}[(I - b_1 BW')^{-1}]'$, and $cov(\epsilon, \theta_M) = \Sigma_\epsilon (I - b_1 BW')^{-1}$. Although the benchmarking property is always verified for the adjusted predictor (1), whatever the estimate $\widetilde{\theta}_y$ or $\widetilde{\theta}_M$, is certainly interesting to investigate how different models may change estimate of $mse(\theta)$. By the decomposition of the $mse(\widetilde{\theta}) = g_1 + g_2 + g_3$ for the Fay-Herriot model, with the leading term $g_1(\sigma^2_u) = R - RQ^{-1}R = diag(\frac{\sigma^2_u\psi_1}{\sigma^2_u+\psi_1}, ..., \frac{\sigma^2_u\psi_m}{\sigma^2_u+\psi_m})$, it is straightforward to note that $g_{1,i}(\sigma^2_v) < g_{1,i}(\sigma^2_u)$, $\forall i$, $i = 1, ..., m$, when $\sigma^2_v < \sigma^2_u$. Further, with $Q = diag(\sigma^2_v + \psi_1, ..., \sigma^2_v + \psi_m)$, $Q_m = diag(\sigma^2_u + \psi_1, ..., \sigma^2_u + \psi_m)$, and following standard matrix inversion rules, $Q_m^{-1} = Q^{-1} - S^{-1}$, $S = diag[\frac{(\sigma^2_v+\psi_1)(\sigma^2_u+\psi_1)}{\sigma^2_u-\sigma^2_v}, ..., \frac{(\sigma^2_v+\psi_m)(\sigma^2_u+\psi_m)}{\sigma^2_u-\sigma^2_v})]$, it can be shown that:

$$\begin{aligned} tr[mse(\widetilde{\theta}_M)] &= tr\left\{var[\widetilde{\theta}_y(\sigma^2_v)] + R(Q_m^{-1} + S^{-1})P_{(I-P_X)q}R\right\} \\ &< tr\left\{var[\widetilde{\theta}_y(\sigma^2_u)] + R(Q_m^{-1} + S^{-1})P_{(I-P_X)q}R\right\} = tr[mse(\widetilde{\theta}_y)]. \end{aligned}$$

Here the projection matrix of the model (3) is partitioned by the decomposition $P_{X,q} = P_X + P_{(I-P_X)q}$, where the matrix $P_{(I-P_X)q}$ projects vectors onto the space spanned by the columns of $(q - P_X q)$.