Comments on: Single and two-stage cross-sectional and time series benchmarking procedures for small area estimation

Domingo Morales

Published online: 21 October 2014
© Sociedad de Estadística e Investigación Operativa 2014

This paper gives a very complete and comprehensive account of the literature on benchmarking procedures for small area estimation. The paper reviews and empirically compares cross-sectional and time series benchmarked predictors and presents some interesting conclusions. The paper introduces a two-stage cross-sectional benchmarking of time series estimators and gives an illustrative example. The authors have made a major contribution to the literature of small area estimation on a topic that is far from being simple. They also have found the way of presenting results in a friendly and rigorously way.

The paper starts by discussing the single-stage cross-sectional benchmarking under the frequentist approach. In that setup, let us consider a finite population U partitioned in D domains $U_1, \dots, U_D$. For a variable y measured in all the units of the population, let $Y = \sum_{j \in U} y_j$ and $Y_d = \sum_{j \in U_d} y_j$ be the totals for U and U_d respectively, so that

$$Y = \sum_{d=1}^D Y_d. \quad (1)$$

In a very simple form, the benchmarking problem can be described as follows. For a size n sample $s \subset U$ partitioned on subsamples $s_d \subset U_d$, $d = 1, \dots, D$, let $\hat{Y}^{\text{dir}} = \sum_{j \in s} w_j y_j$ and $\hat{Y}_d^{\text{dir}} = \sum_{j \in s_d} w_j y_j$ be direct estimators of Y and Y_d respectively, where the w_j s are known sampling weights. It holds that

---

Supported by the grant MTM2012-37077-C02-01.

This comment refers to the invited paper available at: doi:[10.1007/s11749-014-0398-y](https://doi.org/10.1007/s11749-014-0398-y).

D. Morales
Centro de Investigación Operativa, Universidad Miguel Hernández de Elche, Elche, Spain
e-mail: d.morales@umh.es

$$\hat{Y}^{\text{dir}} = \sum_{d=1}^D \hat{Y}_d^{\text{dir}}. \quad (2)$$

Therefore, the consistency relation (1) also holds for the direct estimators of Y and Y_d , $d = 1, \dots, D$. If the sample size n is large enough, then $\hat{Y}^{\text{dir}}$ is a reliable estimator of Y and it is widely used in the data publications of the Statistical Offices. This is to say, $\hat{Y}^{\text{dir}}$ gives in many cases the official estimate of Y and the property (2) is desired and sometimes officially demanded for estimators of domain totals.

In general, (2) does not hold for model-based estimators, $\hat{Y}_d^{\text{mod}}$, of Y_d . This is why statisticians have studied the problem of finding benchmarked estimators, $\hat{Y}_d^{\text{bmk}}$, as close to $\hat{Y}_d^{\text{mod}}$ as possible and fulfilling (2). More formally, we might be interested in finding the solution in $\hat{Y}_d^{\text{bmk}}$, $d = 1, \dots, D$, of the minimizing problem

$$\text{Min} \Psi = \sum_{d=1}^D \hat{Y}_d^{\text{mod}} \phi(\hat{Y}_d^{\text{bmk}} / \hat{Y}_d^{\text{mod}}) \text{ subject to } \sum_{d=1}^D \hat{Y}_d^{\text{bmk}} = \hat{Y}^{\text{dir}}, \quad (3)$$

where ϕ is a convex function defined on $[0, \infty)$ and satisfying $\phi(1) = \phi'(1) = 0$. Ψ is the divergence measure introduced independently by Csiszár (1963) and Ali and Silvey (1966). Some frequently used functions are $\phi(x) = (x - 1)^2/2$ (chi-squared) or $\phi(x) = x \log x$ (Kullback-Leibler), $x \in R$. For the minimization problem (3), the Lagrangian function is

$$L = L(\hat{Y}_1^{\text{bmk}}, \dots, \hat{Y}_D^{\text{bmk}}; \lambda) = \sum_{d=1}^D \hat{Y}_d^{\text{mod}} \phi(\hat{Y}_d^{\text{bmk}} / \hat{Y}_d^{\text{mod}}) + \lambda \left(\hat{Y}^{\text{dir}} - \sum_{d=1}^D \hat{Y}_d^{\text{bmk}} \right).$$

By taking derivatives and equating to zero, we have

$$0 = \frac{\partial L}{\partial \hat{Y}_d^{\text{bmk}}} = \phi'(\hat{Y}_d^{\text{bmk}} / \hat{Y}_d^{\text{mod}}) - \lambda, \quad d = 1, \dots, D; \quad 0 = \frac{\partial L}{\partial \lambda} = \hat{Y}^{\text{dir}} - \sum_{d=1}^D \hat{Y}_d^{\text{bmk}}.$$

For $\phi(x) = (x - 1)^2/2$, $\phi'(x) = x - 1$, we have

$$0 = \hat{Y}_d^{\text{bmk}} - \hat{Y}_d^{\text{mod}} - \lambda \hat{Y}_d^{\text{mod}}, \quad \hat{Y}_d^{\text{bmk}} = (1 + \lambda) \hat{Y}_d^{\text{mod}}, \quad d = 1, \dots, D.$$

By summing over $k = 1, \dots, D$, we get

$$1 + \lambda = \frac{\hat{Y}^{\text{dir}}}{\sum_{k=1}^D \hat{Y}_k^{\text{mod}}}, \quad \hat{Y}_{d,R}^{\text{bmk}} = \frac{\hat{Y}^{\text{dir}}}{\sum_{k=1}^D \hat{Y}_k^{\text{mod}}} \hat{Y}_d^{\text{mod}}, \quad d = 1, \dots, D,$$

which is the ratio benchmarked estimator described by the authors in his formula (2). The mean squared error (MSE) of $\hat{Y}_{d,R}^{\text{bmk}}$ with respect to the model distribution can be calculated by the bootstrap approaches described by González-Manteiga et al. (2008a) and (2008b) for area-level and unit-level models respectively.

The estimator $\hat{Y}_{d,R}^{\text{bmk}}$ is a very simple, universal and widely used procedure that has the drawbacks described in the paper of Pfeffermann, Sikov and Tiller. For this reason, many authors have investigated alternative benchmarking procedures with better statistical properties and adapted to the assumed data generating model. Depending on the paradigm (frequentist or Bayesian), the modelling level (unit or area), the type of target variable (continuous, binary, counts, ...) and the conditional distribution of the target variable, we can find many or few works on benchmarking.

Under the frequentist setup, there are several benchmarking procedures derived for empirical best linear predictors (EBLUPs) based on area-level linear mixed models, and more concretely for the Fay-Herriot model. Many of these procedures consist in minimizing a weighted model MSE subject to the restriction appearing in (3). In comparison, less effort has been put on unit-level model-based estimators, see e.g. Ugarte et al. (2009), and almost no effort in estimators based on generalized linear mixed models or in multivariate benchmarking with or without domain restrictions.

To illustrate a problem of multivariate benchmarking with domain constraints, let us consider the totals, Y_1 , Y_2 and Y_3 , of employed, unemployed and inactive people, such that $Y_1 + Y_2 + Y_3 = N$ for a known population size N (people over 16 years old). Let N_d 's be the corresponding domain sizes such that $N = \sum_{d=1}^D N_d$. The domain totals, Y_{1d} , Y_{2d} and Y_{3d} fulfil the relations

$$Y_k = \sum_{d=1}^D Y_{kd}, \quad k = 1, 2, 3; \quad Y_{1d} + Y_{2d} + Y_{3d} = N_d, \quad d = 1, \dots, D. \quad (4)$$

and the same do the direct estimator, i.e

$$\hat{Y}_k^{\text{dir}} = \sum_{d=1}^D \hat{Y}_{kd}^{\text{dir}}, \quad k = 1, 2, 3; \quad \hat{Y}_{1d}^{\text{dir}} + \hat{Y}_{2d}^{\text{dir}} + \hat{Y}_{3d}^{\text{dir}} = N_d, \quad d = 1, \dots, D.$$

The target is thus how to obtain benchmarked model-based estimators fulfilling (4). López-Vizcaíno et al. (2013) gave a simple solution to this problem using multinomial logit models, by doing ratio benchmarking on the components of employed and unemployed people and by deriving $\hat{Y}_{3d}^{\text{bmk}}$ from the relations $\hat{Y}_{3d}^{\text{bmk}} = N_d - \hat{Y}_{1d}^{\text{bmk}} - \hat{Y}_{2d}^{\text{bmk}}$, $d = 1, \dots, D$. Further research is needed for introducing model-adapted benchmarking procedures for this set up.

The paper also treats the problem of single-stage cross-sectional benchmarking of time series estimators and points out the drawbacks of applying a cross-sectional procedure at each time point separately. The paper describes the approach of Pfeffermann and Burck (1990) and Pfeffermann and Tiller (2006) that added the monthly benchmark constraints defined in (3) to the observation equations of a state-space time series model fitted jointly to the direct estimates in all the areas. It would be quite interesting to investigate a similar problem under some other time models. For example, let us consider the Rao and Yu (1994) model

$$y_{dt} = \mathbf{x}_{dt}\boldsymbol{\beta} + u_{1,d} + u_{2,dt} + e_{dt}, \quad d = 1, \dots, D, \quad t = 1, \dots, T, \quad (5)$$

where y_{dt} is a direct estimator of the mean $\bar{Y}_{dt} = Y_{dt}/N_{dt}$ of area d and time instant t , $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables,

the random effects $u_{1,d}$'s are i.i.d. $N(0, \sigma_1^2)$, the random vectors $(u_{2,d1}, \dots, u_{2,dT})$, $d = 1, \dots, D$, are i.i.d. AR(1) with common variance and auto-correlation parameters σ_2^2 and ρ respectively, the errors e_{dt} 's are independent $N(0, \sigma_{dt}^2)$ with known σ_{dt}^2 's, and the $u_{1,d}$'s, the $(u_{2,d1}, \dots, u_{2,dT})$'s and the e_{dt} 's are independent. In matrix notation, the model (5) is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (6)$$

where $\mathbf{y} = \text{col}_{1 \leq d \leq D}(\mathbf{y}_d)$, $\mathbf{y}_d = \text{col}_{1 \leq t \leq T}(\mathbf{y}_{dt})$, $\mathbf{u}_1 = \text{col}_{1 \leq d \leq D}(u_{1,d})$, $\mathbf{u}_2 = \text{col}_{1 \leq d \leq D}(\mathbf{u}_{2,d})$, $\mathbf{u}_{2,d} = \text{col}_{1 \leq t \leq T}(u_{2,dt})$, $\mathbf{u} = (\mathbf{u}'_1, \mathbf{u}'_2)'$, $\mathbf{e} = \text{col}_{1 \leq d \leq D}(\mathbf{e}_d)$, $\mathbf{e}_d = \text{col}_{1 \leq t \leq T}(e_{dt})$, $\mathbf{X} = \text{col}_{1 \leq d \leq D}(\mathbf{X}_d)$, $\mathbf{X}_d = \text{col}_{1 \leq t \leq T}(\mathbf{x}_{dt})$, $\mathbf{x}_{dt} = \text{col}'_{1 \leq j \leq p}(\mathbf{x}_{dtj})$, $\boldsymbol{\beta} = \text{col}_{1 \leq j \leq p}(\beta_j)$, $\mathbf{Z}_1 = \text{diag}_{1 \leq d \leq D}(\mathbf{1}_T)$, $\mathbf{Z}_2 = \mathbf{I}_M$, $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, $M = \sum_{d=1}^D T$. This model assumes that $\mathbf{u}_1 \sim N(\mathbf{0}, \mathbf{V}_{u_1})$, $\mathbf{u}_2 \sim N(\mathbf{0}, \mathbf{V}_{u_2})$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$ are independent with known error variances σ_{dt}^2 's and covariance matrices $\mathbf{V}_u = \text{diag}(\mathbf{V}_{u_1}, \mathbf{V}_{u_2})$, $\mathbf{V}_{u_1} = \sigma_1^2 \mathbf{I}_D$, $\mathbf{V}_{u_2} = \sigma_2^2 \Omega(\rho)$, $\Omega(\rho) = \text{diag}_{1 \leq d \leq D}(\Omega_d(\rho))$, $\mathbf{V}_e = \text{diag}_{1 \leq d \leq D}(\mathbf{V}_{ed})$, $\mathbf{V}_{ed} = \text{diag}_{1 \leq t \leq T}(\sigma_{dt}^2)$,

$$\Omega_d = \Omega_d(\rho) = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \ddots & & \rho^{T-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{T-2} & & \ddots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \dots & \rho & 1 \end{pmatrix}_{T \times T}.$$

The BLU estimators and predictors of $\boldsymbol{\beta}$ and $\mathbf{u}$ and the BLUP of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ are

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \quad \hat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \quad \hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}, \quad (7)$$

where $\boldsymbol{\mu} = \text{col}_{1 \leq d \leq D}(\text{col}_{1 \leq t \leq T}(\hat{\mu}_{dt}))$, $\hat{\mu}_{dt} = \mathbf{x}_{dt}\hat{\boldsymbol{\beta}} + \hat{u}_{1,d} + \hat{u}_{2,dt}$ and

$$\begin{aligned} \mathbf{V} = \text{var}(\mathbf{y}) &= \sigma_1^2 \mathbf{Z}_1 \mathbf{Z}_1' + \sigma_2^2 \text{diag}_{1 \leq d \leq D}(\Omega_d(\rho)) + \mathbf{V}_e \\ &= \text{diag}_{1 \leq d \leq D} \left(\sigma_1^2 \mathbf{1}_T \mathbf{1}_T' + \sigma_2^2 \Omega_d(\rho) + \mathbf{V}_{ed} \right) = \text{diag}_{1 \leq d \leq D}(\mathbf{V}_d). \end{aligned}$$

Esteban et al. (2012) fitted model (6) by applying the residual maximum likelihood (REML) method. The REML log likelihood is

$$\begin{aligned} l_{reml}(\sigma_1^2, \sigma_2^2, \rho) &= -\frac{M-p}{2} \log 2\pi + \frac{1}{2} \log |\mathbf{X}'\mathbf{X}| - \frac{1}{2} \log |\mathbf{V}| \\ &\quad - \frac{1}{2} \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2} \mathbf{y}'\mathbf{P}\mathbf{y}, \end{aligned}$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (\sigma_1^2, \sigma_2^2, \rho)$, $\mathbf{V}_1 = \frac{\partial \mathbf{V}}{\partial \sigma_1^2} = \text{diag}_{1 \leq d \leq D}(\mathbf{1}_T \mathbf{1}_T')$, $\mathbf{V}_2 = \frac{\partial \mathbf{V}}{\partial \sigma_2^2} = \text{diag}_{1 \leq d \leq D}(\Omega_d(\rho))$ and $\mathbf{V}_3 = \frac{\partial \mathbf{V}}{\partial \rho} =$

$\sigma_2^2 \text{diag}_{1 \leq d \leq D} (\dot{\Omega}_d(\rho))$, $\dot{\Omega}_d(\rho) = \frac{\partial \Omega_d(\rho)}{\partial \rho}$. The likelihood equations are obtained by equating the partial derivatives of l_{reml} (scores) to zero, i.e.

$$0 = S_a = \frac{\partial l_{reml}}{\partial \theta_a} = -\frac{1}{2} \text{tr}(\mathbf{P} \mathbf{V}_a) + \frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{V}_a \mathbf{P} \mathbf{y}, \quad a = 1, 2, 3. \quad (8)$$

If we introduce the time-dependent benchmarking parameters γ_t , $t = 1, \dots, T$, and define $\hat{\mu}_{dt}^{\text{bmk}} = \hat{\mu}_{dt} + \gamma_t$, then the benchmarking equations for the $\hat{\mu}_{dt}^{\text{bmk}}$'s are

$$\sum_{d=1}^D N_{dt} \hat{\mu}_{dt}^{\text{bmk}} = \sum_{d=1}^D \hat{Y}_{dt}^{\text{dir}}, \quad t = 1, \dots, T,$$

which can be written in the matrix form

$$\mathbf{T} \mathbf{N} (\mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{Z} \hat{\mathbf{u}} + \boldsymbol{\gamma}) = \mathbf{y}_T^{\text{dir}}, \quad (9)$$

with matrices $\mathbf{T} = \text{col}_{1 \leq t \leq T} (\text{col}'_{1 \leq d \leq D} (\text{col}'_{1 \leq j \leq T} (\delta_{jt})))$, $\mathbf{N} = \text{diag}_{1 \leq d \leq D} (\text{diag}_{1 \leq t \leq T} (N_{dt}))$ and vectors $\boldsymbol{\gamma} = \text{col}_{1 \leq d \leq D} ((\text{col}_{1 \leq t \leq T} (\gamma_{dt})))$, $\mathbf{y}_T^{\text{dir}} = \text{col}_{1 \leq t \leq T} (\sum_{d=1}^D \hat{Y}_{dt}^{\text{dir}})$. By solving the system of Eqs. (8) and (9) in σ_1^2 , σ_2^2 , ρ and γ_t , $t = 1, \dots, T$, the benchmarked estimator $\hat{\mu}_{dt}^{\text{bmk}}$ can be calculated.

The paper finally deals with the very new problem of two-stage cross-sectional benchmarking, where less literature is available. We first note that EBLUP estimators under hierarchical linear mixed model are auto-benchmarked; this is to say, the EBLUP estimates of subdomain totals sum up the EBLUP estimate of the domain. Therefore, a simple solution to the problem of benchmarking estimates of totals at both levels, subdomain and domain, can be obtained by working only at the subdomain level. First, we calculate the benchmarked subdomain estimates $\hat{Y}_{ds}^{\text{bmk}}$, $d = 1, \dots, D$, $s = 1, \dots, S_{(d)}$, and second we sum them for obtaining the benchmarked domain estimates $\hat{Y}_d^{\text{bmk}}$, $d = 1, \dots, D$. The following equations are then fulfilled.

$$\sum_{d=1}^D \sum_{s=1}^{S_{(d)}} \hat{Y}_{ds}^{\text{bmk}} = \hat{Y}^{\text{dir}}, \quad \hat{Y}_d^{\text{bmk}} = \sum_{s=1}^{S_{(d)}} \hat{Y}_{ds}^{\text{bmk}}, \quad d = 1, \dots, D,$$

This approach might reduce the bias of the EBLUP estimators at the subdomain level but not at the domain level. For this reason, Ghosh and Steorts 2013 introduced benchmarked estimators by solving a constrained minimization problem that takes into account the estimators at the domain and at the subdomain level simultaneously.

The paper finally introduces for the first time a two-stage cross-sectional benchmarking for time series estimators, which is appropriate when the small areas of interest define a hierarchical structure. The authors also develop a corresponding two-stage cross-sectional time series benchmarking procedure. The simulations presented in the section 4 deal with empirical comparisons between cross-sectional and time series benchmarked predictors using simulated series. Some very remarkable findings and

conclusions are obtained from the simulations and they are summarized in the section 8 together with some future possible developments.

I congratulate the authors for preparing an insightful and informative overview paper with original contributions on the important topic of benchmarking in small area estimation. Finally I would like to thank the editors of TEST for giving me the opportunity to read and discuss on this paper.

References

- Ali SM, Silvey SD (1966) A general class of coefficient of divergence of one distribution from another. *J R Stat Soc Ser B* 286:131–142
- Csiszár I (1963) Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ Math Inst Hung Acad Sci Ser A* 8:85–108
- Esteban MD, Morales D, Pérez A, Santamaría L (2012) Small area estimation of poverty proportions under area-level time models. *Comput Stat Data Anal* 56:2840–2855
- Ghosh M, Steorts RC (2013) Two-stage bayesian benchmarking as applied to small area estimation. *Test* 22:670–687
- González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2008a) Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. *Comput Stat Data Anal* 52:5242–5252
- González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2008b) Bootstrap mean squared error of small-area EBLUP. *J Stat Comput Simul* 78:443–462
- López-Vizcaíno E, Lombardía MJ, Morales D (2013) Multinomial-based small area estimation of labour force indicators. *Stat Model* 13:153–178
- Pfeffermann D, Burck L (1990) Robust small area estimation combining time series and cross-sectional data. *Surv Methodol* 16:217–237
- Pfeffermann D, Tiller RL (2006) Small area estimation with state-space models subject to benchmark constraints. *J Am Stat Assoc* 101:1387–1397
- Rao JNK, Yu M (1994) Small area estimation by combining time series and cross sectional data. *Can J Stat* 22:511–528
- Ugarte MD, Militino AF, Goicoa T (2009) Benchmarked estimates in small areas using linear mixed models with restrictions. *Test* 18:342–364