

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

William R. Bell

Published online: 21 October 2014

© Sociedad de Estadística e Investigación Operativa (outside the USA) 2014

Pfeffermann, Sikov, and Tiller (hereafter PST) have provided a nice review of the literature on benchmarking in small area estimation, and also presented a new two-stage benchmarking procedure for small area time series models. The latter further develops previous work of Pfeffermann and Burck (1990) and Pfeffermann and Tiller (2006). My remarks here will focus on some general points they surface related to the differing rationales for external benchmarking (as is commonly practiced with time series data from repeated business surveys) versus internal benchmarking (as is commonly practiced in small area estimation).

PST (Remark 3) note that external benchmarking lowers variances of estimates, “when the benchmark is a known constant . . . and the benchmark constraint is correct.” In fact, even when the external benchmark data are survey estimates with sampling error, benchmarking may lower variances relative to the variances of predictors that make no use of the benchmark data (this being subject to a condition given by Bell et al. (2012) noted by PST in Remark 3). Thus, variance reduction provides one reason for doing external benchmarking. Note, however, that when the benchmark data contain sampling error, maximum variance reduction results from optimal predictors that do not force exact agreement with the “benchmarks”, though they pull the predictions towards them (Hillmer and Trabelsi 1987). One may question whether such optimal prediction should be called “benchmarking”? Doing so may draw useful connections with exact benchmarking, since exact benchmarking results, in the limit, when the variances of the benchmarks go to zero.

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).

W. R. Bell (✉)
U.S. Census Bureau, Washington, USA
e-mail: William.r.bell@census.gov

In practice, external benchmarking of time series of estimates from monthly or quarterly business surveys typically does force exact agreement with the benchmarks. See Dagum and Cholette (2006) for detailed review of many approaches to achieving this result, and Trabelsi and Hillmer (1990) for adapting time series model-based estimation to this problem. While, as noted, this can reduce the variances of direct survey estimates for small areas, often external benchmarking also aims to reduce nonsampling error. One might thus think of external benchmarking as attempting to reduce the mean squared error of the estimates through balanced reductions both of variance and of bias due to nonsampling error, though rarely would the problem be approached at such a level of statistical formality. The potential reduction of nonsampling error arises with monthly or quarterly business surveys because respondents to such surveys generally have readily available annual figures to report (these provide the benchmark data) but some, especially small businesses, may not have monthly or quarterly figures readily available, and so may report approximate figures.

Internal benchmarking, in contrast, cannot be motivated by variance reduction since, as PST note (again Remark 3), internal benchmarking actually raises prediction error variances by making suboptimal use of the data under the model. In addition, since internal benchmarks come from the same survey data that produced the direct estimates being modeled, the benchmarks are subject to the same sorts of nonsampling errors as the direct small area estimates, and this leads to nonsampling errors in the resulting benchmarked small area predictions. One should thus not expect internal benchmarking to reduce nonsampling error. The failure of these two rationales may lead to the oft-cited rationale for internal benchmarking as providing potential protection against model failure. PST (Remark 4), however, question this third rationale, noting that some approaches to cross-sectional internal benchmarking either increase all the estimates or decrease all of them, and they note that “it is hard to surmise a cross-sectional model misspecification under which the resulting (cross-sectional) model-dependent estimators will all either systematically underestimate or overestimate the true targets.” Indeed, such a model misspecification seems mathematically impossible, since virtually all statistical model-fitting procedures produce fitted values that, in some sense, must pass through the center of the data.

PST illustrate that internal cross-sectional benchmarking of time series estimators from repeated surveys does not suffer from this limitation. What seems to happen is this. Time series signal extraction can lead to “oversmoothing” when the series being modeled move rapidly up or down, at least more rapidly than is accounted for by the slow evolution typically assumed by time series models.¹ If the rapid movement is common to all the time series being modeled, then incorporating cross-sectional

¹ Some time series models provide exceptions to “slow evolution”. For example, regression covariates in time series models can account for rapid movements, such as when substantive covariates are being driven by the same underlying forces as the observed series (common trends). Alternatively, when rapid movements are observed they can be empirically modeled by level shift or other types of intervention effects. In either of these cases the time series smoothing would then be of regression residuals after removal of the rapid movements. The estimated regression effects would then be added to the predicted residuals, so the resulting predictions would reflect the rapid movements.

benchmarking constraints into the process can correct the oversmoothing bias. If the rapid movement is particular to a limited set of areas, this may not be the case.²

External time series benchmarking (to annual totals or to end-of-year values from an annual survey or census) would, at best, only partly address the oversmoothing bias of time series estimators. On the other hand, imposing external time series benchmarking constraints on a sequence of cross-sectional small area predictors could correct for misspecification of a small area regression model *if* some areas were consistently under-predicted while others were consistently over-predicted.

The general conclusion may be that, in the right circumstances, imposing benchmarking constraints of one type (internal cross-sectional or external time series) can correct bias due to model misspecification of the other type.

Reference

Trabelsi A, Hillmer SC (1990) Benchmarking time series with reliable benchmarks. *Appl Stat* 39:367–379

² Note that the amount of smoothing done by time series estimators increases with the sampling variances of the direct survey estimators. Hence, oversmoothing may be most pronounced in those areas with the smallest samples.