RESEARCH REPORT SERIES
*(Statistics #2012-12)*

# Benchmarking Small Area Estimators

William R. Bell
Gauri S. Datta[1]
Malay Ghosh[2]

[1] U.S. Census Bureau and University of Georgia
[2] University of Florida

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: September 12, 2012

**Abstract**

This paper considers benchmarking issues in the context of small area estimation. We find optimal estimators within the class of benchmarked linear estimators under either external or internal benchmark constraints. This extends existing results for both external and internal benchmarking, and also provides some links between the two. In addition, necessary and sufficient conditions for self-benchmarking are found for an augmented model. Most results of this paper are found using ideas of orthogonal projection.

**Keywords:** Augmented model; Best linear unbiased; External; Internal; Optimal; Orthogonal projection.

**Acknowledgements:** M. Ghosh thanks the American Statistical Association/National Science Foundation/Census Bureau Fellowship Program for financial support. His research was also partially supported by a U.S. National Science Foundation grant, and by a grant from the U.S. National Security Agency. Datta's research was partially supported by a National Security Agency grant. We thank anonymous reviewers for helpful comments on a previous version of the paper.

## 1 Introduction

Small area estimation has become a topic of growing importance in recent years. Model-based small area estimates, when aggregated, need not correspond to the direct survey estimate for a larger area, e.g., national. This may be a cause for concern if $(i)$ the sample size for the larger area is sufficiently large that the direct estimate is regarded as reliable, and $(ii)$ the direct estimate for the larger area has any sort of official status. Substantial deviation of an aggregation of model-based small area estimates from the corresponding direct estimate for a large area may also suggest model failure. These considerations motivate benchmarking, which is some form of calibration that adjusts individual area level estimates so they aggregate to a direct estimate for a large area.

We can distinguish two general approaches to benchmarking depending on the source of the benchmark estimates. In external benchmarking the benchmark estimates come from an additional data source, typically, another survey or a census. In internal benchmarking the benchmark estimates come from the same survey data that produced the estimates we wish to benchmark.

External benchmarking has a long history in time series obtained from repeated economic surveys. The standard economic time series benchmarking problem is that a statistical agency has a monthly or quarterly economic survey whose estimates do not agree with corresponding estimates from an annual survey or economic census. The benchmarking task is to modify the monthly or quarterly estimates to force agreement with the annual survey or census estimates, which are the benchmarks. To force agreement with these benchmarks is to satisfy the benchmark constraints. For a detailed discussion of this topic, see Dagum and Cholette (2006).

While time series benchmarking is often done using ad-hoc optimization criteria, Hillmer and Trabelsi (1987) and Trabelsi and Hillmer (1990) showed how time series benchmarking could be accomplished using statistical time series models. Although Hillmer and Trabelsi dealt with time series benchmarking, their basic theoretical results were stated in a general way, so that they would apply in other contexts such as small area estimation. Durbin and Quenneville (1997) provided further developments to model-based time series benchmarking, including showing how one can deal with the nonlinear benchmark constraints that result when logs are taken of the data as part of the modeling, while the benchmark constraints reflect aggregation on the original scale.

Model-based external time series benchmarking can improve the monthly or quarterly economic survey estimates in two ways. First, the use of the additional data will, if used in the optimal predictor under the assumed model, lower the variance of the estimates. Second, the benchmarking can reduce nonsampling error in the estimates. This is because respondents to economic surveys generally have available annual figures that they can report in an annual survey or census, whereas some businesses, particularly small businesses, may not have monthly or quarterly figures readily available. This includes those businesses that keep their financial records on some basis other than calendar months. The expected reduction in nonsampling error provides a justification for forcing exact agreement with benchmark values from an annual sample survey, as in Trabelsi and Hillmer (1990), even though standard best linear prediction results, as in Hillmer and Trabelsi (1987), shrink estimates toward the benchmark values but do not force exact agreement if the benchmarks are themselves survey estimates and not census values.

Recent interest in benchmarking in small area estimation has focused on internal benchmarking, as can be seen in papers by Pfeffermann and Barnard (1991), You and Rao (2000), Wang et al. (2008), Datta et al. (2011), and Pfeffermann and Tiller (2006). One common thread in all these small area benchmarking papers is that one begins with best linear unbiased predictors of small area means, and then modifies these estimators to achieve the desired higher level agreement. The motivations for internal benchmarking are thus different from those for external time series benchmarking, since enforcing internal benchmark constraints pushes estimates away from the best linear unbiased predictors, thus leading to increased variances under the model. This can be seen in results of Pfeffermann and Barnard (1991), Wang et al. (2008), Datta et al. (2011), and in our results in Sections 3 and 4. Internal benchmarking also cannot be expected to reduce nonsampling error in the estimates since the benchmarks are subject to the same sources of nonsampling error as the small area data used for the modeling. Rather, the motivations for internal benchmarking are the practical considerations noted in the first paragraph, including the provision of some protection against possible model failure.

While the cited references cover several cases, the approaches used, which vary, all tend to have some limitations such as assuming means are known, assuming covariance matrices are diagonal, or allowing only a single benchmark constraint. Also, some of the approaches seem somewhat ad-hoc, so their rationale is not completely clear. In the present paper, we attempt to provide a more comprehensive treatment of benchmarking that extends existing results for both external and internal benchmarking, and clarifies the rationale in some cases.

## 2 EXTERNAL BENCHMARKING

The usual random effects area level model is

$$y = \theta + e, \qquad \theta = X\beta + u, \tag{1}$$

where $y$ is the vector of direct survey estimates for $m$ small areas, $\theta$ is the vector of corresponding population quantities being estimated, and $e$ is the vector of sampling errors in the estimates $y$. We assume that $E(e) = E(u) = 0$. Let $\text{var}(e) = \Sigma_e, \text{var}(u) = \Sigma_u$, and $\text{cov}(u, e) = 0$. The population quantities $\theta$ follow the regression model given in (1) with covariates given by the columns of $X$, and with $\beta$ the $p \times 1$ vector of regression parameters. We assume that $\text{rank}(X) = p < m$. In the small area context, Fay and Herriot (1979), see also Pfeffermann and Nathan (1981), considered the special case when $\Sigma_e$ is diagonal and $\Sigma_u = \sigma_u^2 I_m$.

For the results given here and in Sections 3 and 4, we treat $\Sigma_e$ and $\Sigma_u$ as known. In practice, $\Sigma_e$ will be estimated using survey microdata, incorporating any known independence restrictions. If samples are independently selected in the different areas, then $\Sigma_e$ is diagonal. The matrix $\Sigma_u$ will generally depend on some set of unknown parameters $\psi$ estimated in fitting the model given by (1). If, as is common, $\Sigma_u = \sigma_u^2 I_m$, then $\psi$ is just the single variance $\sigma_u^2$.

Under the model (1), the best linear unbiased predictor of $\theta$, and its mean squared error, are (Rao, 2003, pp. 96, 99)

$$\tilde{\theta} = y - \Sigma_e Q^{-1}(I - P_X)y, \tag{2}$$

$$V(\tilde{\theta}) \equiv \text{var}(\theta - \tilde{\theta}) = \Sigma_e - \Sigma_e Q^{-1}(I - P_X)\Sigma_e, \tag{3}$$

where $\text{var}(y) = Q \equiv \Sigma_e + \Sigma_u$ and $P_X \equiv X(X^T Q^{-1} X)^{-1} X^T Q^{-1}$.

In this section we consider, in addition to (1), an external source of data $t$ modeled as

$$t = W^T\theta + \eta,\ E(\eta) = 0,\ \text{var}(\eta) = \Sigma_\eta,\ \text{cov}(e, \eta) = C,\ \text{cov}(u, \eta) = 0. \tag{4}$$

Here $t$ is a vector of external estimates of the $q < m$ values $W^T\theta$, with sampling errors $\eta$. If $t$ comes from a census, then $\Sigma_\eta$ and $C$ would both be zero, while if $t$ and $y$ come from independent sample surveys, then $C = 0$ while $\Sigma_\eta$ would be nonzero. In the latter case, $\Sigma_\eta$ could have some zero elements, e.g., it could be diagonal. As with $\Sigma_e$, when $\Sigma_\eta$ and $C$ are nonzero, we assume they have been estimated using survey microdata.

The model defined by (1) and (4) generalizes that of Hillmer and Trabelsi (1987), who assumed that $E(\theta)$ was known in (1) and that $C = 0$. The first generalization, $E(\theta) = X\beta$ with $\beta$ to be estimated, is relevant since models of economic time series often use regression mean functions to account for such things as calendar variation (Bell and Hillmer, 1983). The second generalization, $C \neq 0$, occurs when the samples producing $y$ and $t$ are not independent. This can arise when benchmarking economic time series, since often the sample for the monthly or quarterly survey that provides $y$ is a subsample of the sample used for the annual survey that provides $t$.

For the model given by (1) and (4), Theorem 1 below gives expressions for the best linear unbiased predictor of $\theta$ based on both $y$ and $t$, which we denote as $\tilde{\theta}_{y,t}$, and its mean squared error. An alternative approach to doing this would be to set up a joint model for $[y^T, t^T]^T$ and apply the formulas analogous to (2) and (3). Instead, Theorem 1 shows how the best linear unbiased predictor based on just $y$ can be adjusted to produce the best linear unbiased predictor based on $y$ and $t$, and how a corresponding expression for its mean squared error can be obtained. We start with the following simple lemma. Proofs of this lemma and other results are provided in the Appendix.

**Lemma 1.** *Let $x$, $y$, $z$ be zero mean random vectors with $(x^T, y^T, z^T)^T$ having a finite and positive definite covariance matrix. Let $P(\ \mid\ )$ be general notation for linear projection so that, e.g., $P(x \mid y)$ is the linear projection of $x$ on $y$. Let $r = z - P(z \mid y)$. Then*

$$\begin{aligned} P(x \mid y, z) &= P(x \mid y, r) = P(x \mid y) + P(x \mid r), \\ var\{x - P(x \mid y, z)\} &= var\{x - P(x \mid y)\} - var\{P(x \mid r)\}. \end{aligned}$$

Durbin and Koopman (2001, p. 37) prove an equivalent result for the Gaussian case using standard expressions for Gaussian conditional expectations and variances.

To use Lemma 1 to obtain $\tilde{\theta}_{y,t}$, we assume normality in (1) and (4), and use a Bayesian argument with the prior $\beta \sim N(0, \sigma_\beta^2 I)$ with $\sigma_\beta^2 \to \infty$. This is convenient since it is well-known (Robinson, 1991, Sec. 4.2) that the best linear unbiased predictor of $\theta$ is the same as the posterior mean of $\theta$ under a uniform prior on $\beta$. The prediction mean squared error is then the same as the posterior variance of $\theta$. Notice that with $\beta \sim N(0, \sigma_\beta^2 I)$, we now have $\text{var}(y) \equiv \Sigma_y = X(\sigma_\beta^2 I)X^T + Q$ and, using a matrix inversion result (Rao, 1973, p. 33),

$$\Sigma_y^{-1} = Q^{-1} - Q^{-1}X(\sigma_\beta^{-2} I + X^T Q^{-1} X)^{-1} X^T Q^{-1} \to Q^{-1}(I - P_X) \tag{5}$$

as $\sigma_\beta^2 \to \infty$. Thus, $E(\theta \mid y) = y - E(e \mid y) = y - \Sigma_e \Sigma_y^{-1} y \to y - \Sigma_e Q^{-1}(I - P_X)y = \tilde{\theta}$, and $\text{var}(\theta \mid y) = \Sigma_e - \Sigma_e \Sigma_y^{-1} \Sigma_e \to \Sigma_e - \Sigma_e Q^{-1}(I - P_X)\Sigma_e = V(\tilde{\theta})$, the best linear unbiased predictor and its mean squared error given by (2) and (3). We note that $V(\tilde{\theta})$ is positive definite.

Using this device, we can apply Lemma 1 to prove our main theorem.

**Theorem 1.** *Under the model given by (1) and (4),*

$$\begin{aligned} \tilde{\theta}_{y,t} &= \tilde{\theta} + cov(\theta - \tilde{\theta}, t - \tilde{t}) V(\tilde{t})^{-1} (t - \tilde{t}), \\ \text{MSE}(\tilde{\theta}_{y,t}) = var(\theta - \tilde{\theta}_{y,t}) &= V(\tilde{\theta}) - cov(\theta - \tilde{\theta}, t - \tilde{t}) V(\tilde{t})^{-1} cov(\theta - \tilde{\theta}, t - \tilde{t})^T, \end{aligned}$$

*where*

$$\begin{aligned} \tilde{t} &= W^T \tilde{\theta} + C^T Q^{-1} (I - P_X) y, \\ cov(\theta - \tilde{\theta}, t - \tilde{t}) &= V(\tilde{\theta}) W - \{I - \Sigma_e Q^{-1} (I - P_X)\} C, \\ V(\tilde{t}) &= W^T V(\tilde{\theta}) W + V(\tilde{\eta}) - W^T \{I - \Sigma_e Q^{-1} (I - P_X)\} C \\ &\quad - C^T \{I - \Sigma_e Q^{-1} (I - P_X)\}^T W, \\ V(\tilde{\eta}) &= \Sigma_\eta - C^T Q^{-1} (I - P_X) C. \end{aligned}$$

We now obtain the externally benchmarked predictor, $\hat{\theta}_{\text{ext}}$, and its mean squared error, via two corollaries to Theorem 1.

**Corollary 1.** *For $C = 0$,*

$$\tilde{\theta}_{y,t} = \tilde{\theta} + V(\tilde{\theta}) W \{W^T V(\tilde{\theta}) W + \Sigma_\eta\}^{-1} \{t - W^T \tilde{\theta}\}, \tag{6}$$

$$\text{MSE}(\tilde{\theta}_{y,t}) = V(\tilde{\theta}) - V(\tilde{\theta}) W \{W^T V(\tilde{\theta}) W + \Sigma_\eta\}^{-1} W^T V(\tilde{\theta}). \tag{7}$$

Hillmer and Trabelsi (1987, eqs. (2.16)–(2.17) and (2.9)) and Durbin and Quenneville (1997, p. 28) obtained results analogous to (6) and (7) for the case where the means, $E(\theta) = \mu$, are known. The difference is that, in their expressions, $\tilde{\theta}$ and $V(\tilde{\theta})$ are replaced by $\bar{\theta} = y - \Sigma_e Q^{-1}(y - \mu)$ and $V(\bar{\theta}) = \Sigma_e - \Sigma_e Q^{-1} \Sigma_e$, respectively. Rao (2003, p. 98) terms

$\bar{\theta}$ the best linear predictor based on $y$; it is also the best predictor, i.e., the conditional expectation given $y$, under normality. Corollary 1 thus shows that expressions of the same general form as those previously given for adjusting the best linear predictor based on $y$ to get the best linear predictor based on $(y, t)$, also hold for adjusting the best linear unbiased predictor based on $y$ to get the best linear unbiased predictor based on $(y, t)$. Thus, (6) and (7) accommodate the generalized least squares estimation of the regression parameters $\beta$ by implicitly updating the generalized least squares estimation of $\beta$ based on just $y$ to the generalized least squares estimation of $\beta$ based on $(y, t)$. If the known means $\mu$ in Hillmer and Trabelsi's result were replaced by $X\hat{\beta}$, where $\hat{\beta}$ is obtained by generalized least squares based on $(y, t)$, this would indeed yield $\tilde{\theta}_{y,t}$. Hillmer and Trabelsi's expression (2.9) for the mean squared error, however, would need an additional term to account for the error in estimating $\beta$.

Hillmer and Trabelsi's result would not produce the best linear unbiased predictor based on $(y, t)$ if $\mu$ were replaced by $X\hat{\beta}$ with $\hat{\beta}$ obtained by generalized least squares based on just $y$. This relates to a point made by Durbin and Quenneville (1997, p. 24) about the desirability of using both the original and benchmark data in model fitting. Durbin and Quenneville thus provided another approach in which they merged the benchmark data $t$ with the original time series data $y$, and then developed a modified state-space model for this merged series whose treatment by the Kalman filter and smoother would produce the best linear unbiased predictor. This is the time series version of setting up a joint model for $y$ and $t$.

**Corollary 2.** *As $\Sigma_\eta \to 0$, which implies also $C \to 0$,*

$$\tilde{\theta}_{y,t} \to \hat{\theta}_{\text{ext}} \equiv \tilde{\theta} + V(\tilde{\theta})W\{W^T V(\tilde{\theta})W\}^{-1}\{t - W^T\tilde{\theta}\}, \tag{8}$$

$$\text{MSE}(\tilde{\theta}_{y,t}) \to \text{MSE}(\hat{\theta}_{\text{ext}}) \to V(\tilde{\theta}) - V(\tilde{\theta})W\{W^T V(\tilde{\theta})W\}^{-1}W^T V(\tilde{\theta}). \tag{9}$$

It is easy to check that $W^T\hat{\theta}_{\text{ext}} = t$. Trabelsi and Hillmer (1990, eqs. (2.3)–(2.5)) gave the analogous result to (8) for the case where means are known.

The expression (9) gives the mean squared error of $\hat{\theta}_{\text{ext}}$ only in the limit as $\Sigma_\eta \to 0$, i.e., only when $\hat{\theta}_{\text{ext}}$ actually is the best linear unbiased predictor. If $\hat{\theta}_{\text{ext}}$ is used when $\Sigma_\eta > 0$, then $\hat{\theta}_{\text{ext}} \neq \tilde{\theta}_{y,t}$ and its mean squared error will be larger. From standard results, $\text{MSE}(\hat{\theta}_{\text{ext}}) = \text{MSE}(\tilde{\theta}_{y,t}) + \text{var}(\hat{\theta}_{\text{ext}} - \tilde{\theta}_{y,t})$, the term $\text{var}(\hat{\theta}_{\text{ext}} - \tilde{\theta}_{y,t})$ being the increase in mean squared error due to benchmarking.

It is of interest to consider how the mean squared error of $\hat{\theta}_{\text{ext}}$ compares to that of $\tilde{\theta}$, the best linear unbiased predictor using only the data $y$, when $\Sigma_\eta > 0$ so that $\hat{\theta}_{\text{ext}}$ is not optimal. For the case of $C = 0$, it can be shown that

$$\begin{aligned} \text{MSE}(\hat{\theta}_{\text{ext}}) - V(\tilde{\theta}) &= V(\tilde{\theta})W\{W^T V(\tilde{\theta})W\}^{-1} \times \\ &\quad \{\Sigma_\eta - W^T V(\tilde{\theta})W\}\{W^T V(\tilde{\theta})W\}^{-1} W^T V(\tilde{\theta}). \end{aligned} \tag{10}$$

Consider the term, $\Sigma_\eta - W^T V(\tilde{\theta})W$. Notice that $\Sigma_\eta$ is the error variance in predicting the benchmark targets by the benchmark data, i.e., $\text{var}(W^T\theta - t) = \text{var}(\eta)$. The subtracted term, $W^T V(\tilde{\theta})W$, is the variance of the error in predicting the benchmark targets using the best linear unbiased predictor based on $y$, i.e., $\text{var}(W^T\theta - W^T\tilde{\theta})$. Whether benchmarking using the data $t$ provides an improvement thus depends on whether $t$ is a better predictor of $W^T\theta$ than is $W^T\tilde{\theta}$. With one benchmark constraint $W$ is a vector, and benchmarking provides an improvement if $\text{var}(\eta) < \text{var}(W^T\theta - W^T\tilde{\theta})$; if the reverse is true, benchmarking does worse. Since the rank of (10) is at most $q$, there are $m - q$ independent linear combinations of $\theta$ whose prediction mean squared errors are unaffected by the benchmarking, because benchmarking does not alter their best linear unbiased predictors. Of course, these statements assume that the model given by (1) and (4) is true. In reality, as noted in the Introduction, the sub-optimal $\hat{\theta}_{\text{ext}}$ may be used to reduce nonsampling error in the data $y$, or to protect against possible model failure.

# 3 INTERNAL BENCHMARKING

Often in small area estimation problems, additional data $t$ are not available, and internal benchmarking is done based on linear functions of $y$. We write the internal benchmark constraints for an estimator $\hat{\theta}$ as $W^T\hat{\theta} = W^T y$. We will see now how a modification of $\tilde{\theta}$ given in (2) can lead to a benchmarked estimator that satisfies multiple internal benchmark constraints. This generalizes the work of Pfeffermann and Barnard (1991) and Wang et al. (2008) who considered a single internal benchmark constraint. Pfeffermann and Barnard showed that, under their model, there is no overall best linear unbiased predictor among the set of predictors that satisfy the benchmark constraint. To deal with this non-existence of an overall best linear unbiased predictor, Wang, et al. found optimal linear unbiased benchmarked estimators that solve a quadratic optimization problem with a single linear constraint. Here we generalize this approach to allow for multiple benchmark constraints, also using a more general quadratic loss function than did Wang, et al.

Consider the general quadratic loss $L_\Omega(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^T \Omega (\hat{\theta} - \theta)$ for estimating $\theta$ by any linear predictor $\hat{\theta}$ of $\theta$, where the weight matrix $\Omega$ is assumed to be known and positive definite. We have the following theorem giving the predictor that minimizes the expected loss subject to the benchmark constraints $W^T\hat{\theta} = W^T y$. We denote this predictor as $\hat{\theta}_{\text{QL}}$.

**Theorem 2.** *Consider the class of linear unbiased predictors $\hat{\theta} = Ky$ of $\theta$ which satisfy, in addition, $W^T K y = W^T y$ with probability 1, i.e., $W^T K = W^T$. Then, under the quadratic loss $L_\Omega(\theta, Ky)$, $E\{L_\Omega(\theta, Ky)\} \geq E\{L_\Omega(\theta, \hat{\theta}_{\text{QL}})\}$, with equality holding if and only if $Ky = \hat{\theta}_{\text{QL}}$ with probability 1, where*

$$\hat{\theta}_{\text{QL}} \equiv \tilde{\theta} + \Omega^{-1} W (W^T \Omega^{-1} W)^{-1} W^T (y - \tilde{\theta}). \tag{11}$$

*Remark 1.* For a vector $W$ and diagonal $\Omega$, $\hat{\theta}_{\text{QL}}$ reduces to the predictor obtained in Wang et al. (2008).

We now provide expressions for the overall mean squared error of $\hat{\theta}_{\text{QL}}$ and for its risk, $E\{L_\Omega(\theta, \hat{\theta}_{\text{QL}})\}$. Let $P_{\Omega,W} = W(W^T\Omega^{-1}W)^{-1}W^T\Omega^{-1}$. One can show that

$$\text{MSE}(\hat{\theta}_{\text{QL}}) = V(\tilde{\theta}) + P^T_{\Omega,W}\Sigma_e Q^{-1}(I - P_X)\Sigma_e P_{\Omega,W}. \tag{12}$$

The term $P^T_{\Omega,W}\Sigma_e Q^{-1}(I - P_X)\Sigma_e P_{\Omega,W}$ represents the increase in mean squared error due to benchmarking. The quadratic risk of $\hat{\theta}_{\text{QL}}$ is then, after some simplification,

$$\begin{aligned} E\{L_\Omega(\theta, \hat{\theta}_{\text{QL}})\} &= \text{tr}[\Omega\{\text{MSE}(\hat{\theta}_{\text{QL}})\}] \\ &= \text{tr}\{\Omega V(\tilde{\theta})\} + \text{tr}\{W(W^T\Omega^{-1}W)^{-1}W^T\Sigma_e Q^{-1}(I - P_X)\Sigma_e\}. \end{aligned}$$

*Remark 2 (Using external benchmarking results for internal benchmarking).* As an alternative to the quadratic loss approach, we can formally use the external benchmark predictor of Section 2 for internal benchmarking by setting $t = W^T y$ in (8). We denote this predictor by $\hat{\theta}_{\text{int}}$; it is given explicitly by

$$\hat{\theta}_{\text{int}} = \tilde{\theta} + V(\tilde{\theta})W\{W^T V(\tilde{\theta})W\}^{-1}W^T(y - \tilde{\theta}). \tag{13}$$

Comparing equations (13) and (11) shows that

$$\hat{\theta}_{\text{int}} = \hat{\theta}_{\text{QL}} \qquad \text{if} \quad \Omega^{-1} \propto V(\tilde{\theta}). \tag{14}$$

Wang et al. (2008) note that Battese et al. (1988) implicitly used weights corresponding to $\Omega = [\text{diag}\{V(\tilde{\theta})\}]^{-1}$ in benchmarking, where $\text{diag}\{V(\tilde{\theta})\}$ denotes the diagonal matrix with entries given by the diagonal elements of $V(\tilde{\theta})$. From (3), this will generally satisfy (14) only if $\Sigma_u$ and $\Sigma_e$ are diagonal, and if the model contains no regression variables so that $P_X$ disappears from $V(\tilde{\theta})$.

While the condition in (14) is not mathematically necessary for equality of the respective benchmarked estimators, special circumstances would be needed for the benchmarked estimators to agree without this condition holding. For example, if $\Omega$ and $V(\tilde{\theta})$ are diagonal and there is just one benchmark constraint defined by $W = (w_1, \ldots, w_m)^T$, the most obvious special circumstance would be to have some of the $w_i$ be zero, so that the estimators of the corresponding $\theta_i$ would be unconstrained.

When (14) holds, the reported mean squared errors of $\hat{\theta}_{\text{int}}$ and $\hat{\theta}_{\text{QL}}$ will also be the same if they are computed using the same model, namely, model (1). There is no obvious alternative for $\hat{\theta}_{\text{QL}}$, whose mean squared error is given by (12) with $\Omega^{-1} = V(\tilde{\theta})$, since any proportionality constant in $\Omega$ drops out of $P_{\Omega,W}$. For $\hat{\theta}_{\text{int}}$, setting $t = W^T y$ in (8) means that we implicitly assume that $\text{var}(W^T e) \to 0$ to get the point predictor, but we then avoid this assumption when computing its mean squared error under model (1). This approach to computing the mean squared error of an internally benchmarked predictor was suggested by Pfeffermann and Tiller (2006). It also leads to (12) with $\Omega^{-1} = V(\tilde{\theta})$. Comparing this with (9), the mean squared error of the predictor $\hat{\theta}_{\text{ext}}$ with external benchmarks equal to the truth, one can show that

$$\text{MSE}(\hat{\theta}_{\text{int}}) = \text{MSE}(\hat{\theta}_{\text{ext}}) + V(\tilde{\theta})W\{W^T V(\tilde{\theta})W\}^{-1}W^T \Sigma_e W\{W^T V(\tilde{\theta})W\}^{-1}W^T V(\tilde{\theta}).$$

The addition to $\text{MSE}(\hat{\theta}_{\text{ext}})$ is thus the cost of using internal benchmarks, rather than an externally provided truth.

Pfeffermann and Barnard (1991) obtained the predictor (13) for the case of a single benchmark constraint and a unit level model, but they reached it in an entirely different way. They derived unconstrained best linear unbiased predictors for their model through a generalized least squares calculation (Pfeffermann, 1984), and then added the benchmark constraint to produce predictors via generalized least squares regression with a linear restriction. This yields (13). Although Pfeffermann and Barnard applied this approach to internal benchmarking, the actual calculations could also be applied to external benchmarking to obtain $\hat{\theta}_{\text{ext}}$ in (8) in this alternative way.

# 4 INTERNAL BENCHMARKING VIA AN AUGMENTED MODEL/DIFFUSE PRIOR

Consider now the model (1) augmented with additional regression variables as follows:

$$y = \theta + e, \qquad \theta = X\beta + G\delta + u, \tag{15}$$

where $G$ is an $m \times q$ matrix of full rank $q$ with associated regression parameters $\delta$, and where $[X \mid G]$ is of full rank $p + q < m$. We later discuss how to construct $G$. Analogous to equations (2) and (3), the best linear unbiased predictor of $\theta$ and its mean squared error for model (15) are

$$\hat{\theta}_G = y - \Sigma_e Q^{-1}(I - P_{[X|G]})y, \tag{16}$$

$$\text{MSE}(\hat{\theta}_G) \equiv \text{var}(\theta - \hat{\theta}_G) = \Sigma_e - \Sigma_e Q^{-1}(I - P_{[X|G]})\Sigma_e, \tag{17}$$

where

$$P_{[X|G]} = [X \mid G]\left([X \mid G]^T Q^{-1}[X \mid G]\right)^{-1}[X \mid G]^T Q^{-1} \tag{18}$$

projects onto $\mathcal{L}[X \mid G]$, the vector space spanned by the columns of $[X \mid G]$, under the inner product $< a, b >= a^T Q^{-1} b$. We now ask if there exists a $G$ such that the best linear unbiased predictor (16) satisfies the internal benchmark constraints, $W^T\hat{\theta}_G = W^T y$. The answer turns out to be yes, as is established by the following theorem.

**Theorem 3.** *In model (15) assume that $[X \mid G]$ and $[X \mid \Sigma_e W]$ both have full rank. Then $\hat{\theta}_G$ given by (16) satisfies the internal benchmark constraints, $W^T\hat{\theta}_G = W^T y$, if and only if $G = \Sigma_e W R_1 + X R_2$ for some $q \times q$ nonsingular matrix $R_1$ and $p \times q$ matrix $R_2$. Furthermore, for such a $G$, $\hat{\theta}_G$ is invariant to alternative choices of $R_1$ and $R_2$.*

Model (15) with $G = \Sigma_e W$ generalizes the augmented model of Wang et al. (2008, Section 3), who considered (15) for the particular case of a single benchmark constraint and diagonal $\Sigma_e$. Wang, et al. note that such a model is self-calibrated, in that the resulting best predictor automatically satisfies the benchmark constraint. You et al. (2012) further consider this predictor and compare it to a different self-benchmarked predictor of You and Rao (2002) that was originally developed for a unit level model. Theorem 3 both extends Wang, et al.'s self-calibration result to the case of multiple benchmark constraints, and generalizes it to necessary and sufficient conditions for the model to be self-calibrated.

It can be shown that $Q^{-1}(I - P_{[X|G]}) = Q^{-1}(I - P_X - P_{(G-\tilde{G})})$ where $\tilde{G} = P_X G$ and $P_{(G-\tilde{G})}$ projects onto $\mathcal{L}[G - \tilde{G}]$, the vector space spanned by the columns of $G - \tilde{G} = (I - P_X)G$. The mean squared error of $\hat{\theta}_G$ from (17) can then be written as

$$\begin{aligned} \text{MSE}(\hat{\theta}_G) &= \Sigma_e - \Sigma_e Q^{-1}(I - P_X - P_{(G-\tilde{G})})\Sigma_e \\ &= V(\tilde{\theta}) + \Sigma_e Q^{-1} P_{(G-\tilde{G})}\Sigma_e. \end{aligned} \tag{19}$$

It can also be shown that $P_{(G-\tilde{G})}$ is invariant to the choice of $R_1$ and $R_2$ in $G = \Sigma_e W R_1 + X R_2$. The increase in mean squared error from benchmarking, given by the second term in (19), can thus be written as $\Sigma_e Q^{-1} M (M^T Q^{-1} M)^{-1} M^T Q^{-1} \Sigma_e$ by setting $G = \Sigma_e W$ and $M = G - \tilde{G} = (I - P_X)\Sigma_e W$.

*Remark 3.* Since $\hat{\theta}_G$ is invariant to alternative choices of $G = \Sigma_e W R_1 + X R_2$, to satisfy the benchmark constraints we might make the simplest choice of setting $R_1 = I$ and $R_2 = 0$ so that $G = \Sigma_e W$. Alternatively, we could choose $R_1 = I$ and $R_2 = -(X^T Q^{-1} X)^{-1} X^T Q^{-1} \Sigma_e W$, so that $G = (I - P_X)\Sigma_e W$, in which case $X$ and $G$ would be orthogonal. In the case of a single benchmark constraint, $\Sigma_e W$ and $G$ are just vectors, and the alternative choices of $G$ amount to multiplying $\Sigma_e W$ by a nonzero scalar and adding to this a linear combination of the columns of $X$. Clearly these actions will not affect the regression predictions, nor the best linear unbiased predictions, from model (15).

*Remark 4.* While $\hat{\theta}_G$ is invariant to alternative choices of $G$, the generalized least squares estimates of $\beta$ and of $\delta$ are generally affected by the choice of $G$ in (15). The Appendix discusses this point further.

*Remark 5.* It is worth considering how $[X \mid \Sigma_e W]$ might fail to have full rank, and what we could do in response. Consider the simplest possible example. Suppose that $\Sigma_e = \sigma_e^2 I$ with $\sigma_e^2 > 0$, and that the lone benchmark constraint is that the estimate of the total over all areas equals the direct estimate, so that $W = (1, \ldots, 1)^T$. If $X$ also contains a column of ones to include an intercept term in the regression, then $[X \mid \Sigma_e W]$ will not have full rank. In this case, there is a redundancy between the benchmark constraint and the regression intercept term. There is thus no need to impose this particular benchmark constraint; it will be satisfied automatically by having the intercept term in the regression. In general, if $[X \mid \Sigma_e W]$ does not have full rank, we can drop from $\Sigma_e W$ any columns contained in $\mathcal{L}[X]$.

We can alter the interpretation of the augmented model (15) by assuming that $\delta$ is a vector of random effects with zero mean, with $\text{var}(\delta) = \tau^2 I$, and with $\text{cov}(\delta, u) = \text{cov}(\delta, e) = 0$. Then, as for results (2) and (3) of Section 2, the best linear unbiased predictor and its mean squared error for this model are

$$\hat{\theta}_\tau = y - \hat{e}_\tau = y - \Sigma_e \Sigma_\tau^{-1}(I - P_{X,\tau})y, \tag{20}$$

$$\text{var}(\theta - \hat{\theta}_\tau) = \Sigma_e - \Sigma_e \Sigma_\tau^{-1}(I - P_{X,\tau})\Sigma_e, \tag{21}$$

where $\Sigma_\tau \equiv \text{var}(y) = Q + \tau^2 G G^T$ and $P_{X,\tau} = I - X(X^T \Sigma_\tau^{-1} X)^{-1} X^T \Sigma_\tau^{-1}$. The Appendix shows that $\Sigma_\tau^{-1}(I - P_{X,\tau}) \to Q^{-1}(I - P_{[X|G]})$ as $\tau^2 \to \infty$, implying that (20) and (21) then converge to (16) and (17). With an appropriate choice of $G$ from Theorem 3, we thus have another model whose best linear unbiased predictors automatically satisfy, in the limit as $\tau^2 \to \infty$, the benchmark constraints. This provides the diffuse prior interpretation of (15).

*Remark 6.* For the case where the regression parameters $\beta$ are known, the predictor $\hat{\theta}_\infty = \lim_{\tau\to\infty}\hat{\theta}_\tau$ can also be shown to be that of the transformation approach proposed by Ansley and Kohn (1985) to deal with initial conditions for nonstationary time series models. For model (15) with $\text{var}(\delta) = \tau^2 I$, this provides the best predictor among linear predictors whose errors depend only on $u$ and $e$, not on $\delta$. See the Appendix for further discussion.

*Remark 7.* We can compare $\hat{\theta}_{\Sigma_e W}$ to our other two benchmarked predictors by writing it as follows:

$$\hat{\theta}_{\Sigma_e W} = \tilde{\theta} + \Sigma_e Q^{-1}(I - P_X)\Sigma_e W\{W^T\Sigma_e Q^{-1}(I - P_X)\Sigma_e W\}^{-1}W^T(y - \tilde{\theta}). \qquad (22)$$

Since $I - P_X$ has rank $m - p < m$, $\Sigma_e Q^{-1}(I - P_X)\Sigma_e$ is singular. Comparing (22) with (13) and (11) then shows that, as long as there are regression variables in the model, $\hat{\theta}_{\Sigma_e W} \neq \hat{\theta}_{\text{QL}}$, in general, because we cannot set $\Omega \propto \{\Sigma_e Q^{-1}(I - P_X)\Sigma_e\}^{-1}$. Analogous considerations show that also $\hat{\theta}_{\Sigma_e W} \neq \hat{\theta}_{\text{int}}$, in general. For the case of no regression variables, or for known $\beta$, we drop $(I - P_X)$ from (22), in which case $\hat{\theta}_{\text{int}} = \hat{\theta}_{\Sigma_e W}$ if $V(\tilde{\theta}) \equiv \Sigma_e Q^{-1}\Sigma_u \propto \Sigma_e Q^{-1}\Sigma_e$, and $\hat{\theta}_{\text{QL}} = \hat{\theta}_{\Sigma_e W}$ if $\Omega \propto \{\Sigma_e Q^{-1}\Sigma_e\}^{-1}$. The latter condition can be chosen to hold, but the former condition is unlikely to hold. If $\Sigma_e$ is nonsingular, it requires $\Sigma_e \propto \Sigma_u$, something that would generally hold in practice only when $\Sigma_e$ and $\Sigma_u$ are both proportional to the identity matrix.

## 5 ON ESTIMATING THE MEAN SQUARED ERROR OF BENCHMARKED PREDICTORS

The results we have presented on the mean squared errors of the various benchmarked predictors have assumed that the covariance matrices $\Sigma_e$, $\Sigma_\eta$, and $C$ of the sampling errors $e$ and $\eta$, and $\Sigma_u$ of the model errors $u$, are known. As noted in Section 2, $\Sigma_e$, $\Sigma_\eta$, and $C$ will generally be estimated using survey microdata, while unknown parameters $\psi$ that determine $\Sigma_u$ will be estimated in fitting the model. Wang and Fuller (2003) and Rivest and Vandal (2003) provide results on accounting for error in estimating sampling variances when estimating the mean squared error of small area predictors. Considerably more attention has been given to accounting for the estimation error in the variance parameters $\psi$. Rao (2003, pp. 103-110) discusses this for models of a fairly general form, drawing on results of Prasad and Rao (1990) and Datta and Lahiri (2000).

The results of Rao (2003) just cited cover our predictors $\tilde{\theta}_{y,t}$ and $\hat{\theta}_{\Sigma_e W}$, as these are best linear unbiased under their respective models. In addition, Steorts and Ghosh, in a 2012 University of Florida technical report, consider the benchmarked predictor $\hat{\theta}_{\text{QL}}$ for the Fay-Herriot model when $\Omega = I$, providing results on the mean squared prediction error accounting for the estimation of $\psi$. Analogous results have not, to our knowledge, been developed for either the benchmarked predictor $\hat{\theta}_{\text{ext}}$ in the case when it is not optimal, i.e., when $\Sigma_\eta \neq 0$, or for the benchmarked predictor $\hat{\theta}_{\text{int}}$.

Our results can be used for developing predictors and their measures of uncertainty under a Bayesian approach. From standard results, the unbenchmarked Bayes estimator under model (1) is $E(\theta \mid y) = E_{\psi|y}(\tilde{\theta})$, and its measure of uncertainty is the posterior variance

$$\text{var}(\theta \mid y) = E_{\psi|y}\{\text{var}(\theta \mid y, \psi)\} + \text{var}_{\psi|y}\{E(\theta \mid y, \psi)\} \tag{23}$$

where $\text{var}(\theta \mid y, \psi) = V(\tilde{\theta})$ and $E(\theta \mid y, \psi) = \tilde{\theta}$. Given simulations of $\psi$ from its posterior distribution, the conditional expectation and variance on the right hand side of (23) can be obtained by appropriately averaging over these simulations. If $\psi$ is just $\sigma_u^2$, numerical integration can readily be used to approximate these terms. In the same way then, Bayes estimators and posterior variances can be developed for the model given by (1) and (4), and for the augmented model (15), whose posterior mean is the Bayesian self-benchmarked predictor, $E_{\psi|y}(\hat{\theta}_{\Sigma_e W})$. While our other benchmarked predictors do not lead to posterior means in a Bayesian treatment, e.g., $E_{\psi|y}(\hat{\theta}_{\text{int}}) \neq E(\theta \mid y)$, the corresponding posterior mean squared error matrix provides an uncertainty measure. For any such predictor $\hat{\theta}$ of $\theta$ based on $y$, including $\hat{\theta}_{\text{int}}$ and $\hat{\theta}_{\text{QL}}$ obtained using an estimated value of $\psi$, the mean squared error matrix is

$$E\{(\theta - \hat{\theta})(\theta - \hat{\theta})^T \mid y\} = \text{var}(\theta \mid y) + \{\hat{\theta} - E(\theta \mid y)\}\{\hat{\theta} - E(\theta \mid y)\}^T.$$

The posterior mean squared error matrix of $\hat{\theta}_{\text{ext}}$ would follow similarly using results from Theorem 1 or, when $C = 0$, from Corollary 1.

# 6 EXAMPLE

To illustrate the results of this paper, we present an example using a model and data from the U.S. Census Bureau's Small Area Income and Poverty Estimates program. We use the model for poverty rates of school-age children for states of the U.S. for the year 1998. For this year, data for the direct survey estimates $y_i$ came from the U.S. Current Population Survey's Annual Social and Economic Supplement. The model is of the form of (1) with $\Sigma_e$ diagonal and with $\Sigma_u = \sigma_u^2 I$. In addition to an intercept term, regressors in $X$ include two variables obtained from tabulations of U.S. federal income tax data: a tax data analog to state child poverty rates, and a measure of state tax nonfiler rates. A final regressor is the residuals obtained from regressing Census 2000 long-form state age 5–17 poverty rates on the other regressors just mentioned, but with the latter defined for 1999, the reference year for Census 2000 long form income questions. Estimation of the $\text{var}(e_i)$ is via a sampling error model with a generalized variance function. The model for $y_i$ is then given a Bayesian treatment, with flat priors for $\sigma_u^2$ and $\beta$. Modeling details, as well as additional information on the data sources, can be found on the Small Area Income and Poverty Estimates program web site at www.census.gov/did/www/saipe/index.html. General information on the Current Population Survey poverty estimates, including information on sampling and nonsampling errors in the data, is available at www.census.gov/cps/.

We consider here the estimators $\hat{\theta}_{\mathrm{QL}}$, $\hat{\theta}_{\mathrm{int}}$, and $\hat{\theta}_{\Sigma_e W}$ as given by (11), (13), and (22), replacing $\sigma_u^2$ by its posterior mean when calculating $\tilde{\theta}$ and $V(\tilde{\theta})$ from (2) and (3). For $\hat{\theta}_{\mathrm{QL}}$, we examine results for three choices of the weighting matrix $\Omega$: the $51 \times 51$ identity matrix; the diagonal matrix whose entries are the 5–17 state populations, denoted as $\mathrm{diag}(\mathrm{Pop}_i)$; and $\Omega = \mathrm{diag}(\mathrm{Pop}_i^2)$. With $\Omega = I$, the quadratic loss function $L_\Omega(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^T \Omega (\hat{\theta} - \theta)$ becomes the sum of squared errors in the estimated state 5–17 poverty rates, while with $\Omega = \mathrm{diag}(\mathrm{Pop}_i^2)$, it becomes, to a very good approximation, the sum of squared errors in the estimated state numbers of persons age 5–17 in poverty. The choice $\Omega = \mathrm{diag}(\mathrm{Pop}_i)$ provides an alternative in between these two.

For purposes of illustration, we examine results from benchmarking estimates so they reproduce the direct Current Population Survey estimates of the 5–17 poverty rates for the four regions of the U.S.: Northeast, Midwest, South, and West. A map showing which states belong to which regions is available on the Census Bureau web site at www.census.gov/geo/www/us_regdiv.pdf. In this case, $W$ is a $51 \times 4$ matrix, with the four columns corresponding to the four regions, and with the 51 rows corresponding to the 50 states and the city of Washington, D.C., which is treated here like a state for the purposes of the model. Each column contains the state shares of that region's 5–17 population. For the states not in a given region, their population share is zero.

In comparing results from the different benchmarked predictors, we shall examine not the predictors themselves, but the adjustments they make to the best linear unbiased predictor, $\tilde{\theta}$, as are shown in equations (11), (13), and (22). These adjustments depend on the term $W^T(y - \tilde{\theta})$, which contains the discrepancies between the benchmark targets $W^T y$, which here are the direct Current Population Survey estimates of the regional 5–17 poverty rates, and the corresponding model-based best linear unbiased predictions, $W^T \tilde{\theta}$. For our data and model, these discrepancies, expressed as percentages as are the poverty rates, are $W^T(y - \tilde{\theta}) = (0{\cdot}74, -0{\cdot}73, -0{\cdot}55, 0{\cdot}17)^T$ for the Northeast, Midwest, South, and West regions. Given the form of $W$, it can be shown that the adjustments to $\tilde{\theta}$ from each $\hat{\theta}_{\mathrm{QL}}$ will allocate each regional discrepancy among the states of that region. The adjustments will thus be positive for all states in the Northeast and West, and negative for all states in the Midwest and South. It can further be shown that: ($i$) when $\Omega = I$, the adjustments vary in size proportional to $\text{Pop}_i$; ($ii$) when $\Omega = \text{diag}(\text{Pop}_i)$, the adjustments are all constant at their regional discrepancies; and ($iii$) when $\Omega = \text{diag}(\text{Pop}_i^2)$, the adjustments vary in size proportional to $1/\text{Pop}_i$. The first and third results can be seen in Fig. 1, which plots adjustments to $\tilde{\theta}$ from the various benchmarked predictors for the states of the South region. In Fig. 1 and later in Fig. 2, we use $\hat{\theta}_{\text{QL1}}$, $\hat{\theta}_{\text{QL2}}$, and $\hat{\theta}_{\text{QL3}}$ to identify the cases with $\Omega = I$, $\Omega = \text{diag}(\text{Pop}_i)$, and $\Omega = \text{diag}(\text{Pop}_i^2)$, respectively. Notice that the adjustments from $\hat{\theta}_{\text{QL3}}$ for the smallest states of the South region are quite large in magnitude, much larger than any of the adjustments from $\hat{\theta}_{\text{QL1}}$. Figure 1 omits the plot for $\hat{\theta}_{\text{QL2}}$ due to the constancy of its benchmark adjustments across the states of the South region.

The benchmark adjustments from $\hat{\theta}_{\text{int}}$ also depend on the regional discrepancies $W^T(y - \tilde{\theta})$, but are not simply allocations of the discrepancies among states within the regions since the matrix $V(\tilde{\theta})$ in (13) is not diagonal. A similar remark applies to $\hat{\theta}_{\Sigma_e W}$. As a result, Fig. 1 shows that the adjustments from $\hat{\theta}_{\text{int}}$ and $\hat{\theta}_{\Sigma_e W}$ for the states of the South are not all negative.

We also examined results from benchmarking to the direct Current Population Survey estimate of the national 5–17 poverty rate, for which $W$ is a vector containing the 51 state shares of the national 5–17 population. For $\hat{\theta}_{\text{int}}$ and the $\hat{\theta}_{\text{QL}}$, we found that the adjustments from regional benchmarking tended to be larger than those made from national benchmarking. This is because there are far fewer states within each region over which to accumulate the required regional adjustments which are, apart from the West, substantially larger in magnitude than the national discrepancy, which turns out to be $-0{\cdot}195$ percent. However, this was not generally the case for the adjustments from $\hat{\theta}_{\Sigma_e W}$, which in some cases were larger in magnitude for the national benchmarking.

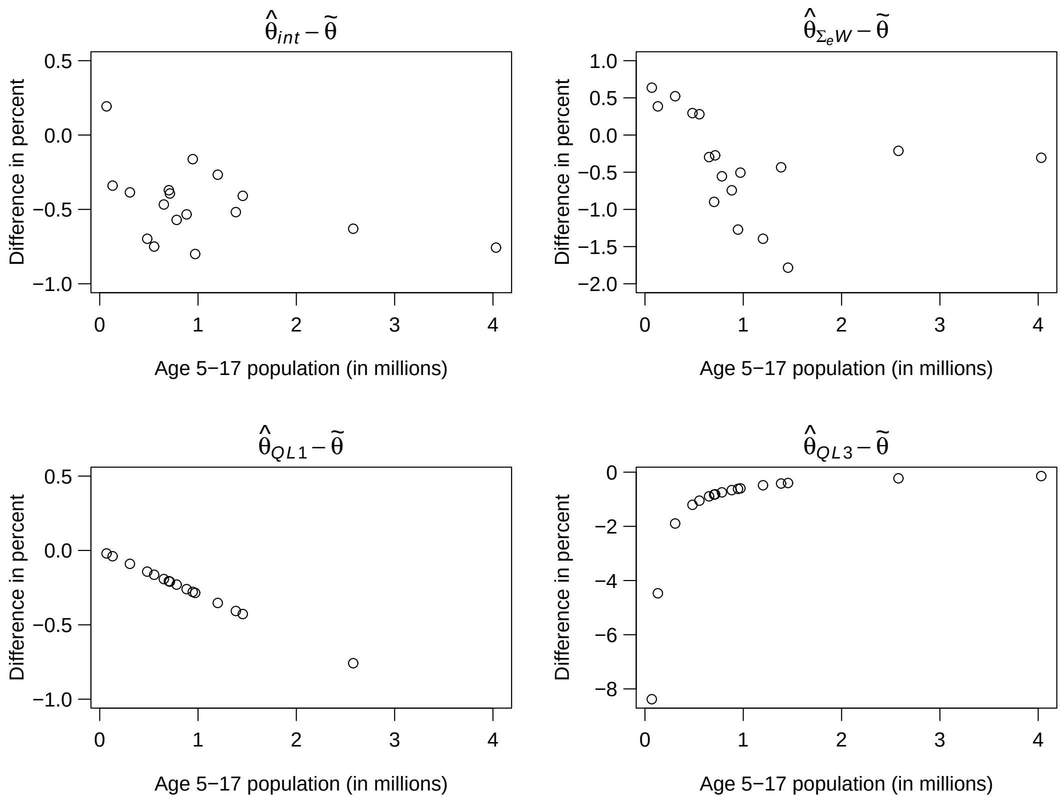


Figure 1: Estimated state child poverty rates in percent, 1998, South region of the U.S. The graphs plot the changes to the estimated poverty rates in going from the model-based best linear unbiased predictors, $\tilde{\theta}_i$, to the various small area predictors that are benchmarked in alternative ways to the direct Current Population Survey estimate for the South region. The changes in the estimated child poverty rates are plotted against the state age 5–17 populations.

We now compare mean squared errors for the various regional benchmarking predictors, examining the multiplicative percentage increases of these mean squared errors relative to those of the best linear unbiased predictor, which are the diagonal elements of $V(\tilde{\theta})$. For the $\hat{\theta}_{\text{QL}}$, the mean squared error increases are computed from (12) with the three versions of $\Omega$. The mean squared errors for $\hat{\theta}_{\text{int}}$ are also computed from (12) by setting $\Omega^{-1} = V(\tilde{\theta})$; see Remark 2. These computations all assume that model (1) is true and $\sigma_u^2$ is known. Mean squared errors for $\hat{\theta}_{\Sigma_e W}$ are computed from (19) assuming that $G = \Sigma_e W$ in the augmented model (15). From Remark 6, the prediction error, $\theta - \hat{\theta}_{\Sigma_e W}$, depends on just $u$ and $e$, not on $\delta$. Hence, the mean squared error of $\hat{\theta}_{\Sigma_e W}$ is the same whether computed under model (1) or model (15), and so is comparable to the mean squared errors of $\hat{\theta}_{\text{int}}$ and $\hat{\theta}_{\text{QL}}$ assuming model (1) is true.

The mean squared error percent increases vary widely across the different predictors, as can be seen from the plots for the states of the South region in Fig. 2. By far the largest increases occur for small states for $\hat{\theta}_{\text{QL3}}$. The single largest increase is 804 percent for Washington, D.C., though some small states in the other regions also show increases of several hundred percent. This shows the cost to mean squared errors for small states of emphasizing large states in the loss function to keep their mean squared error percent increases tiny. For some small states, the mean squared errors of $\hat{\theta}_{\text{QL3}}$ exceed the sampling variances of the direct survey estimates, $y_i$. Mean squared error percent increases for $\hat{\theta}_{\text{QL2}}$ are quite modest, with the largest two of these only somewhat exceeding 10 percent. Mean squared error percent increases for $\hat{\theta}_{\text{QL1}}$ are tiny for small states, but substantial for some large states, the largest exceeding 50 percent.

Mean squared error percent increases for $\hat{\theta}_{\text{int}}$ are variable, tending to be largest for large states. Those for $\hat{\theta}_{\Sigma_e W}$ are also variable, though not in any clearly size-dependent way, and while a few are somewhat large, they are not nearly as large as are the largest for $\hat{\theta}_{\text{QL3}}$.

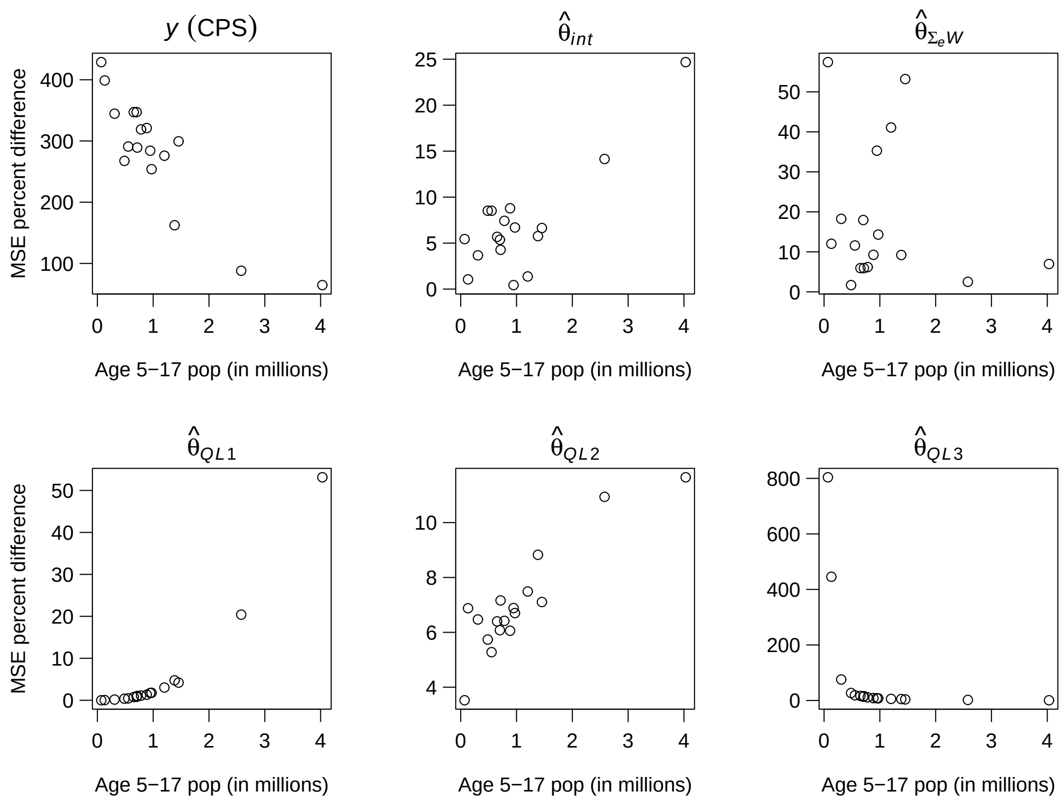


Figure 2: Mean squared errors of estimated state child poverty rates, 1998, South region of the U.S. The first graph plots the multiplicative percentage differences in mean squared errors between the direct Current Population Survey estimates, $y_i$, and the best linear unbiased predictors, $\tilde{\theta}_i$. The other graphs plot the multiplicative percentage differences in mean squared errors between the various benchmarked predictors and the $\tilde{\theta}_i$. The mean squared error percentage differences are plotted against the state age 5–17 populations.

## APPENDIX: PROOFS OF RESULTS AND FURTHER DISCUSSION

Proofs and derivations of the results of the paper are provided below, along with some additional discussion. To make this material more self-contained, we restate results before giving the proofs.

### Material for Section 2, External Benchmarking

**Lemma 1.** Let $x$, $y$, $z$ be zero mean random vectors with $(x^T, y^T, z^T)^T$ having a finite and positive definite covariance matrix. Let $P(\ \mid\ )$ be general notation for linear projection so that, e.g., $P(x \mid y)$ is the linear projection of $x$ on $y$. Let $r = z - P(z \mid y)$. Then

$$\begin{aligned} P(x \mid y, z) &= P(x \mid y, r) = P(x \mid y) + P(x \mid r), \\ \text{var}\{x - P(x \mid y, z)\} &= \text{var}\{x - P(x \mid y)\} - \text{var}\{P(x \mid r)\}. \end{aligned}$$

*Proof.* The first result is an immediate consequence of standard results on linear projections since $\mathcal{L}(y, z) = \mathcal{L}(y, r)$ and $r$ is orthogonal to $y$. (We let $\mathcal{L}(y, z)$ denote the vector space spanned by the elements of $y$ and $z$.) For the variance result, write

$$x - P(x \mid y) = \{x - P(x \mid y, z)\} + \{P(x \mid y, z) - P(x \mid y)\}.$$

The two terms on the right hand side are orthogonal, so that

$$\begin{aligned} &var\{x - P(x \mid y)\} = var\{x - P(x \mid y, z)\} + var\{P(x \mid y, z) - P(x \mid y)\} \\ &\quad \Rightarrow var\{x - P(x \mid y, z)\} = var\{x - P(x \mid y)\} - var\{P(x \mid r)\}. \end{aligned}$$ □

**Theorem 1.** Under the model given by (1) and (4),

$$\begin{aligned} \tilde{\theta}_{y,t} &= \tilde{\theta} + \text{cov}(\theta - \tilde{\theta}, t - \tilde{t})V(\tilde{t})^{-1}(t - \tilde{t}) \\ \text{MSE}(\tilde{\theta}_{y,t}) \equiv E(\theta - \tilde{\theta}_{y,t})^2 &= V(\tilde{\theta}) - \text{cov}(\theta - \tilde{\theta}, t - \tilde{t})V(\tilde{t})^{-1}\text{cov}(\theta - \tilde{\theta}, t - \tilde{t})^T, \end{aligned}$$

where

$$
\begin{aligned}
\tilde{t} &= W^T\tilde{\theta} + C^T Q^{-1}(I - P_X)y \\
\mathrm{cov}(\theta - \tilde{\theta}, t - \tilde{t}) &= V(\tilde{\theta})W - \{I - \Sigma_e Q^{-1}(I - P_X)\}C \\
V(\tilde{t}) &= W^T V(\tilde{\theta})W + V(\tilde{\eta}) - W^T\{I - \Sigma_e Q^{-1}(I - P_X)\}C \\
&\quad - C^T\{I - \Sigma_e Q^{-1}(I - P_X)\}^T W \\
V(\tilde{\eta}) &= \Sigma_\eta - C^T Q^{-1}(I - P_X)C.
\end{aligned}
$$

*Proof.* To apply Lemma 1, we identify $\theta$ with $x$, $y$ with $y$, and $t$ with $z$. Since linear projections are conditional expectations given that normality is temporarily assumed, this yields

$$
\begin{aligned}
E(\theta \mid y, t) &= E(\theta \mid y) + E\{\theta \mid t - E(t \mid y)\} \\
&= E(\theta \mid y) + \mathrm{cov}(\theta, t \mid y)\mathrm{var}(t \mid y)^{-1}\{t - E(t \mid y)\}, \qquad (24) \\
\mathrm{var}(\theta \mid y, t) &= \mathrm{var}(\theta \mid y) - \mathrm{var}\left[E\{\theta \mid t - E(t \mid y)\}\right] \\
&= \mathrm{var}(\theta \mid y) - \mathrm{cov}(\theta, t \mid y)\mathrm{var}(t \mid y)^{-1}\mathrm{cov}(\theta, t \mid y)^T. \qquad (25)
\end{aligned}
$$

In (24) and (25) we have used the fact that, for any $(x, y, z)$ that are jointly normal, $\mathrm{cov}(x, z \mid y) = \mathrm{cov}\{x - E(x \mid y), z - E(z \mid y)\} = \mathrm{cov}\{x, z - E(z \mid y)\}$. Letting $\beta \sim N(0, \sigma_\beta^2 I)$, we easily obtain the quantities needed for (24) and (25) as follows, taking their limits as $\sigma_\beta^2 \to \infty$. These results then hold without the normality assumption. Recall that, as $\sigma_\beta^2 \to \infty$, $\Sigma_y^{-1} \to Q^{-1}(I - P_X)$, $E(\theta \mid y) \to \tilde{\theta}$, and $\mathrm{var}(\theta \mid y) \to V(\tilde{\theta})$ . Thus,

$$
E(t \mid y) = W^T E(\theta \mid y) + E(\eta \mid y) \to W^T\tilde{\theta} + C^T Q^{-1}(I - P_X)y \equiv \tilde{t}, \qquad (26)
$$

$$
\begin{aligned}
\mathrm{cov}(\theta, t \mid y) &= \mathrm{var}(\theta \mid y)W + \mathrm{cov}(\theta, \eta \mid y) \\
&= \mathrm{var}(\theta \mid y)W + \{\mathrm{cov}(\theta, \eta) - \mathrm{cov}(\theta, y)\Sigma_y^{-1}\mathrm{cov}(y, \eta)\} \\
&= \mathrm{var}(\theta \mid y)W + 0 - (\Sigma_y - \Sigma_e)\Sigma_y^{-1}C \\
&\to V(\tilde{\theta})W - \{I - \Sigma_e Q^{-1}(I - P_X)\}C, \qquad (27)
\end{aligned}
$$

and (27) equals $\text{cov}(\theta - \tilde{\theta}, t - \tilde{t})$. Then,

$$\text{var}(\eta \mid y) = \Sigma_\eta - C^T \Sigma_y^{-1} C \to \Sigma_\eta - C^T Q^{-1}(I - P_X)C \equiv V(\tilde{\eta}),$$

and we have

$$\begin{aligned} \text{var}(t \mid y) &= W^T \text{var}(\theta \mid y) W + \text{var}(\eta \mid y) + W^T \text{cov}(\theta, \eta \mid y) + \text{cov}(\eta, \theta \mid y) W \\ &\to W^T V(\tilde{\theta}) W + V(\tilde{\eta}) - W^T \{I - \Sigma_e Q^{-1}(I - P_X)\} C \\ &\quad - C^T \{I - \Sigma_e Q^{-1}(I - P_X)\}^T W \qquad (28) \\ &\equiv V(\tilde{t}) = \text{var}(t - \tilde{t}). \end{aligned}$$

Substituting the expressions (26)–(28) into (24) and (25) proves the theorem. □

## Material for Section 3, Internal Benchmarking

Theorem 2 below provides the internally benchmarked predictor under the general quadratic loss, $L_\Omega(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^T \Omega (\hat{\theta} - \theta)$, where the weight matrix $\Omega$ is assumed to be known and positive definite. Before proving the theorem, we state and prove a needed lemma.

**Lemma 2.** *For any statistic $Ly$ with $E(Ly) = 0$ for all $\beta$, $E\{(\tilde{\theta} - \theta)(Ly)^T\} = 0$.*

*Proof.* $E(Ly) = 0$ for all $\beta$ is equivalent to $LX = 0$. Note that $P_X X = X$, $P_X Q L^T = X(X^T Q^{-1} X)^{-1} X^T L^T = 0$, and, from equation (2) of Section 2, $\tilde{\theta} - \theta = e - \Sigma_e Q^{-1}(I - P_X)y$. Hence,

$$\begin{aligned} E\{(\tilde{\theta} - \theta)(Ly)^T\} &= E\left[\{e - \Sigma_e Q^{-1}(I - P_X)y\}(y - X\beta)^T L^T\right] \\ &= \Sigma_e L^T - \Sigma_e Q^{-1} Q L^T + \Sigma_e Q^{-1} P_X Q L^T \\ &= 0. \end{aligned}$$

□

**Theorem 2.** Consider the class of linear unbiased predictors $\hat{\theta} = Ky$ of $\theta$ which satisfy, in addition, $W^T K y = W^T y$ with probability 1, i.e., $W^T K = W^T$. Then, under the quadratic loss $L_\Omega(\theta, Ky)$, $E\{L_\Omega(\theta, Ky)\} \geq E\{L_\Omega(\theta, \hat{\theta}_{\mathrm{QL}})\}$, with equality holding if and only if $Ky = \hat{\theta}_{\mathrm{QL}}$ with probability 1, where

$$\hat{\theta}_{\mathrm{QL}} \equiv \tilde{\theta} + \Omega^{-1}W(W^T\Omega^{-1}W)^{-1}W^T(y - \tilde{\theta}). \tag{29}$$

*Proof.* Note first that $W^T K_0 y = W^T y$. We now use the identity

$$\begin{aligned}
E\{(Ky-\theta)^T\Omega(Ky-\theta)\} &= E\left[\mathrm{tr}\{\Omega(Ky-\theta)(Ky-\theta)^T\}\right] \\
&= \mathrm{tr}\left[\Omega E\{(Ky-\theta)(Ky-\theta)^T\}\right] \\
&= \mathrm{tr}\left[\Omega E\{(Ky-\tilde{\theta}+\tilde{\theta}-\theta)(Ky-\tilde{\theta}+\tilde{\theta}-\theta)^T\}\right] \\
&= \mathrm{tr}\left[\Omega E\{(Ky-\tilde{\theta})(Ky-\tilde{\theta})^T + (\tilde{\theta}-\theta)(\tilde{\theta}-\theta)^T\}\right] \\
&= E\{(Ky-\tilde{\theta})^T\Omega(Ky-\tilde{\theta})\} + E\{(\tilde{\theta}-\theta)^T\Omega(\tilde{\theta}-\theta)\}, \quad (30)
\end{aligned}$$

where the fourth equality is a consequence of Lemma 2, since $E(Ky - \tilde{\theta}) = 0$. Next we use the algebraic identity

$$\begin{aligned}
(Ky-\tilde{\theta})^T\Omega(Ky-\tilde{\theta}) &= (Ky-K_0y+K_0y-\tilde{\theta})^T\Omega(Ky-K_0y+K_0y-\tilde{\theta}) \\
&= (K_0y-\tilde{\theta})^T\Omega(K_0y-\tilde{\theta}) + (Ky-K_0y)^T\Omega(Ky-K_0y),
\end{aligned}$$

since the cross product terms, $(Ky - K_0y)^T\Omega(K_0y - \tilde{\theta})$ and its transpose, are easily seen to be zero by substituting $\Omega^{-1}W(W^T\Omega^{-1}W)^{-1}W^T(y - \tilde{\theta})$ for $K_0y - \tilde{\theta}$ and noting that $(Ky - K_0y)^T W = 0$. Thus,

$$\begin{aligned}
E\{(Ky-\theta)^T\Omega(Ky-\theta)\} &= E\{(\tilde{\theta}-\theta)^T\Omega(\tilde{\theta}-\theta)\} + E\{(K_0y-\tilde{\theta})^T\Omega(K_0y-\tilde{\theta})\} \\
&\quad + E\{(Ky-K_0y)^T\Omega(Ky-K_0y)\} \\
&= E\{(K_0y-\theta)^T\Omega(K_0y-\theta)\} \\
&\quad + E\{(Ky-K_0y)^T\Omega(Ky-K_0y)\},
\end{aligned}$$

since (30) also holds when we set $K = K_0$. This proves the result. □

*On the similarity of $\hat{\theta}_{QL}$ to solutions to other, related problems*

As $\hat{\theta}_{\text{QL}}$ given by (29) above and $\hat{\theta}_{\text{int}}$ given by equation (13) of Section 3 are of similar form, so too are solutions to many other benchmarking and related problems that involve minimizing some sort of quadratic objective function, typically a sum-of-squares criterion or a variance, under linear constraints on the resulting estimators. Dagum and Cholette (2006) and Knottnerus (2003) provide several examples. Despite these similar solutions, differences in the problems being solved should be kept in mind. For example, many authors assume that $\theta$ is a vector of fixed parameters being estimated, not a vector of stochastic quantities being predicted.

To cite one specific example, H. J. Boonstra, in an unpublished 2004 Statistics Netherlands working paper, obtained an estimator of the same form as (29) for the problem of minimum variance linear adjustment, or calibration, of estimators to satisfy general linear constraints, such as when adjusting table entries to force agreement with specified margins. Despite the similar forms of the results, the problem Boonstra considered differs from ours in two respects. First, Boonstra minimized the variance of a linear function of his adjusted estimators, whereas we minimized a mean squared error criterion. Second, he considered only linear adjustments to the original estimators that, in the notation of our problem, would be of the form $\tilde{\theta} + FW^T(y - \tilde{\theta})$, with $F$ to be determined, whereas we considered general linear estimators, $Ky$, in Theorem 2.

## Material for Section 4, Internal Benchmarking via an Augmented Model/Diffuse Prior

For reference, we first repeat equations (15)–(18) of 4 which give the augmented model, the best linear unbiased predictor and its mean squared error under that model, and the expression for the projection matrix $P_{[X|G]}$. These are

$$
\begin{aligned}
y &= \theta + e, \qquad \theta = X\beta + G\delta + u, && (31)\\
\hat{\theta}_G &= y - \Sigma_e Q^{-1}(I - P_{[X|G]})y, && (32)\\
\text{MSE}(\hat{\theta}_G) &\equiv \text{var}(\theta - \hat{\theta}_G) = \Sigma_e - \Sigma_e Q^{-1}(I - P_{[X|G]})\Sigma_e, && (33)\\
P_{[X|G]} &= [X \mid G]\left([X \mid G]^T Q^{-1}[X \mid G]\right)^{-1}[X \mid G]^T Q^{-1}. && (34)
\end{aligned}
$$

We now restate and then prove Theorem 3.

**Theorem 3.** In model (31) assume that $[X \mid G]$ and $[X \mid \Sigma_e W]$ both have full rank. Then $\hat{\theta}_G$ given by (32) satisfies the internal benchmark constraints, $W^T\hat{\theta}_G = W^T y$, if and only if $G = \Sigma_e W R_1 + X R_2$ for some $q \times q$ nonsingular matrix $R_1$ and $p \times q$ matrix $R_2$. Furthermore, for such a $G$, $\hat{\theta}_G$ is invariant to alternative choices of $R_1$ and $R_2$.

*Proof.* First, suppose that $G = \Sigma_e W R_1 + X R_2$, in which case

$$
[X \mid G] = [X \mid \Sigma_e W]\begin{bmatrix} I & R_2 \\ 0 & R_1 \end{bmatrix}. \qquad (35)
$$

Since $[X \mid G]$ is assumed to be full rank, so must be both matrices on the right hand side of (35), which implies that $R_1$ must be nonsingular. Then, (35) also implies that $\mathcal{L}[X \mid \Sigma_e W] = \mathcal{L}[X \mid G]$, and so from (32)

$$
W^T\hat{\theta}_G - W^T y = -(\Sigma_e W)^T Q^{-1}(I - P_{[X|G]})y = -(\Sigma_e W)^T Q^{-1}(I - P_{[X|\Sigma_e W]})y \qquad (36)
$$

where $P_{[X|\Sigma_e W]}$ is defined as in (34), substituting $\Sigma_e W$ for $G$. But the columns of $I - P_{[X|\Sigma_e W]}$ are orthogonal to those of $\Sigma_e W$, so the right hand side of (36) is zero and the benchmark constraints are satisfied.

To prove the reverse implication, assume that the benchmark constraints hold so that, from (32), $(\Sigma_e W)^T Q^{-1}(I - P_{[X|G]})y = 0$ for any $y$. Since $y$ can be any vector in $R^m$, $(I - P_{[X|G]})y$ can be any vector in $\mathcal{L}[X \mid G]^{\perp}$, the orthogonal complement of $\mathcal{L}[X \mid G]$.

This shows that every column of $\Sigma_e W$ must be orthogonal to $\mathcal{L}[X \mid G]^{\perp}$. The columns of $\Sigma_e W$ are thus all in $\mathcal{L}[X \mid G]$, which implies that $\Sigma_e W = XH_1 + GH_2$ for some $p \times q$ matrix $H_1$ and $q \times q$ matrix $H_2$. Thus,

$$[X \mid \Sigma_e W] = [X \mid G] \begin{bmatrix} I & H_1 \\ 0 & H_2 \end{bmatrix}.$$

Since $[X \mid \Sigma_e W]$ is assumed to have full rank, $H_2$ must be nonsingular. Therefore, $G = \Sigma_e W H_2^{-1} - XH_1H_2^{-1}$, and setting $R_1 = H_2^{-1}$ and $R_2 = -H_1H_2^{-1}$ gives the desired result.

To prove the invariance of $\hat{\theta}_G$ to alternative choices of $R_1$ and $R_2$ for defining $G$, note that since $\mathcal{L}[X \mid \Sigma_e W] = \mathcal{L}[X \mid G]$ and $P_{[X|\Sigma_e W]} = P_{[X|G]}$, from (32) we have that, for any $G$ satisfying our assumptions, $\hat{\theta}_G = y - \Sigma_e Q^{-1}(I - P_{[X|G]})y = y - \Sigma_e Q^{-1}(I - P_{[X|\Sigma_e W]})y$, and the last expression does not depend on $R_1$ and $R_2$. □

*On the dependence of $\hat{\beta}$ and $\hat{\delta}$ on the choice of G*

While $\hat{\theta}_G$ is invariant to alternative choices of $G$, this is not true of the generalized least squares estimates of $\beta$ and $\delta$. To see this, we first define $\hat{X} = (G^TQ^{-1}G)^{-1}G^TQ^{-1}X$ and $\tilde{X} = G\hat{X} = P_GX$, where $P_G = G(G^TQ^{-1}G)^{-1}G^TQ^{-1}$ is the projection matrix that projects onto $\mathcal{L}(G)$, the vector space spanned by the columns of $G$, under the inner product $< a, b >= a^TQ^{-1}b$. Thus, $I - P_G$ is the projection matrix that projects onto $\mathcal{L}(G)^{\perp}$, the orthocomplement of $\mathcal{L}(G)$. The following results are easily verified:

$$Q^{-1}(I - P_G) = (I - P_G)^TQ^{-1} = (I - P_G)^TQ^{-1}(I - P_G). \tag{37}$$

We now reparameterize $X\beta + G\delta$ in (31) to $(X - \tilde{X})\beta + G\alpha$, where $\alpha = \hat{X}\beta + \delta$. As this orthogonalizes $[X \mid G]$ to $[(X - \tilde{X}) \mid G]$ with respect to the inner product $< a, b >= a^TQ^{-1}b$, it shows that, using (37),

$$\begin{aligned} \hat{\beta}_G &= \{(X - \tilde{X})^TQ^{-1}(X - \tilde{X})\}^{-1}(X - \tilde{X})^TQ^{-1}y \\ &= \{X^TQ^{-1}(I - P_G)X\}^{-1}X^T(I - P_G)^TQ^{-1}y, \end{aligned} \tag{38}$$

$$\text{var}(\hat{\beta}_G) = \{X^T Q^{-1}(I - P_G)X\}^{-1}.$$

So $\hat{\beta}_G$ and $\text{var}(\hat{\beta}_G)$ depend on $G$ through $P_G$. We could get analogous expressions for $\hat{\delta}_G$ and $\text{var}(\hat{\delta}_G)$.

*Proof that* $\Sigma_\tau^{-1}(I - P_{X,\tau}) \to Q^{-1}(I - P_{[X|G]})$ *as* $\tau^2 \to \infty$*, so that (20) and (21) of Section 4 converge to (32) and (33).*

Recall that $\Sigma_\tau = Q + \tau^2 GG^T$. Then, completely analogous to equation (5) of Section 2, as $\tau^2 \to \infty$

$$\Sigma_\tau^{-1} = Q^{-1} - Q^{-1}G(\tau^{-2}I + G^T Q^{-1} G)^{-1} G^T Q^{-1} \to \Sigma_\infty^{-1},$$

where

$$\Sigma_\infty^{-1} = Q^{-1}(I - P_G), \tag{39}$$

with alternative equivalent expressions for $\Sigma_\infty^{-1}$ given by equation (37). We now need to show that $\Sigma_\infty^{-1}(I - P_{X,\infty}) = Q^{-1}(I - P_{[X|G]})$, where $P_{X,\infty} = \lim_{\tau\to\infty} P_{X,\tau} = X(X^T\Sigma_\infty^{-1}X)^{-1}X^T\Sigma_\infty^{-1}$.

Since $[(X - \tilde{X}) \mid G]$ and $[X \mid G]$ are related by the nonsingular linear transformation,

$$[(X - \tilde{X}) \mid G] = [X \mid G] \begin{bmatrix} I & 0 \\ -\hat{X} & I \end{bmatrix}, \tag{40}$$

and $[X \mid G]$ has full rank, it follows that $[(X - \tilde{X}) \mid G]$ has full rank, as does $(X - \tilde{X})$. Thus, we define the matrix that projects onto $\mathcal{L}(X - \tilde{X})$, again under the inner product $< a, b >= a^T Q^{-1} b$, as

$$P_{(X-\tilde{X})} = (X - \tilde{X})\{(X - \tilde{X})^T Q^{-1}(X - \tilde{X})\}^{-1}(X - \tilde{X})^T Q^{-1}. \tag{41}$$

We analogously define $\tilde{G} = P_X G$ and

$$P_{(G-\tilde{G})} = (G - \tilde{G})\{(G - \tilde{G})^T Q^{-1}(G - \tilde{G})\}^{-1}(G - \tilde{G})^T Q^{-1}, \tag{42}$$

which projects onto $\mathcal{L}(G-\tilde{G})$. We also defined in Section 4

$$P_{[X|G]} = [X \mid G]\left([X \mid G]^T Q^{-1} [X \mid G]\right)^{-1} [X \mid G]^T Q^{-1},$$

which projects onto $\mathcal{L}[X \mid G]$. Since $X-\tilde{X}$ has full column rank, $X^T\Sigma_\infty^{-1}X = (X-\tilde{X})^T Q^{-1}(X-\tilde{X})$ is nonsingular. Also, $(X-\tilde{X})^T Q^{-1} G = 0$, implying that $P_{(X-\tilde{X})}P_G = 0$. Similarly, $P_{(G-\tilde{G})}P_X = 0$. The result $\Sigma_\infty^{-1}(I - P_{X,\infty}) = Q^{-1}(I - P_{[X|G]})$ is then established by the following lemma.

**Lemma 3.** *For model (31) with $\Sigma_\infty^{-1}$ given by (39), and assuming $[X \mid G]$ has full rank, we have*

$$\begin{aligned}
\Sigma_\infty^{-1}\{I - X(X^T\Sigma_\infty^{-1}X)^{-1}X^T\Sigma_\infty^{-1}\} &= Q^{-1}(I - P_{(X-\tilde{X})})(I - P_G) \\
&= Q^{-1}(I - P_{(X-\tilde{X})} - P_G) \\
&= Q^{-1}(I - P_{[X|G]}) \\
&= Q^{-1}(I - P_{(G-\tilde{G})})(I - P_X) \\
&= Q^{-1}(I - P_{(G-\tilde{G})} - P_X).
\end{aligned}$$

*Proof.* From (37), (39) (41), and (42), $\Sigma_\infty^{-1}\{I - X(X^T\Sigma_\infty^{-1}X)^{-1}X^T\Sigma_\infty^{-1}\}$ can be written as

$$Q^{-1}\left[(I-P_G) - (X-\tilde{X})\{(X-\tilde{X})^T Q^{-1}(X-\tilde{X})\}^{-1}(X-\tilde{X})^T Q^{-1}(I-P_G)\right]$$
$$= Q^{-1}(I - P_{(X-\tilde{X})})(I - P_G).$$

From (40) we see that $\mathcal{L}[X \mid G] = \mathcal{L}[(X-\tilde{X}) \mid G]$ so that $P_{[(X-\tilde{X})|G]} = P_{[X|G]}$. Since $X-\tilde{X}$ is the residual from projecting $X$ on $G$, $(X-\tilde{X})$ and $G$ are orthogonal, i.e., $(X-\tilde{X})^T Q^{-1} G = 0$, which implies that $P_{[(X-\tilde{X})|G]} = P_{(X-\tilde{X})} + P_G$. Hence,

$$I - P_{[X|G]} = I - P_{(X-\tilde{X})} - P_G = (I - P_{(X-\tilde{X})})(I - P_G)$$

since $P_{(X-\tilde{X})}P_G = 0$. The last two results of the lemma follow by symmetry since $P_{[X|G]} = P_{X-\tilde{X}} + P_G = P_{G-\tilde{G}} + P_X$. $\square$

*Relation between the diffuse prior and transformation approach predictors*

For the case where the regression parameters $\beta$ are known, the predictor $\hat{\theta}_\infty \equiv \lim_{\tau\to\infty} \hat{\theta}_\tau$ can be shown to be that of the transformation approach proposed by Ansley and Kohn (1985) to deal with initial conditions for nonstationary time series models. In the present context, the vector $\delta$ takes the role of the initial conditions, and the transformation approach finds the optimal predictor of $\theta$ among the class of linear predictors $\hat{\theta}$ for which the prediction error, $\theta - \hat{\theta}$, does not depend on $\delta$. This predictor can be obtained by making the linear transformation

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} W^T \\ J^T \end{bmatrix} y = \begin{bmatrix} W^T(X\beta + G\delta + u + e) \\ J^T(X\beta + u + e) \end{bmatrix} \tag{43}$$

obtained by requiring that $J$ be any $m \times (m-q)$ matrix of full rank satisfying $J^TG = 0$. To satisfy the benchmark constraints, we set $G = \Sigma_e W$. We then define $\hat{\theta} = y - \hat{e}$ where $\hat{e}$ is the best linear predictor of $e$ based on $z_2$, i.e., $\hat{e} = \text{cov}(e, z_2)[\text{var}(z_2)]^{-1}(z_2 - J^TX\beta)$. Ansley and Kohn showed that this predictor is the same as the predictor obtained by placing a diffuse prior on the nonstationary initial conditions, i.e., by letting $\text{var}(\delta) = \tau^2 I$ with $\tau^2 \to \infty$ in model (31).

To apply the transformation approach predictor when $\beta$ is estimated, we would substitute $\hat{\beta}$ for $\beta$ in $\hat{e}$, $\hat{\beta}$ being the estimator of $\beta$ from the model for $z_2$. This estimator can be shown to be the limit as $\tau^2 \to \infty$ of $\hat{\beta}_\tau = (X^T\Sigma_\tau^{-1}X)^{-1}X^T\Sigma_\tau^{-1}y$, which is $(X^T\Sigma_\infty^{-1}X)^{-1}X^T\Sigma_\infty^{-1}y = \{(X-\tilde{X})^TQ^{-1}(X-\tilde{X})\}^{-1}(X-\tilde{X})^TQ^{-1}y = \hat{\beta}_G$, given by (38).

*Derivation of equation (22) of Remark 7.*

To obtain equation (22), we let $G = \Sigma_e W$ in equation (16) of Section 3, and proceed as follows:

$$\begin{aligned}
\hat{\theta}_{\Sigma_e W} &= y - \Sigma_e Q^{-1}(I - P_{[X|G]})y \\
&= y - \Sigma_e Q^{-1}(I - P_X - P_{(G-\tilde{G})})y \\
&= \tilde{\theta} + \Sigma_e Q^{-1} P_{(G-\tilde{G})} y \\
&= \tilde{\theta} + \Sigma_e Q^{-1} M (M^T Q^{-1} M)^{-1} M^T Q^{-1} y \\
&= \tilde{\theta} \; + \Sigma_e Q^{-1}(I - P_X)\Sigma_e W\{W^T \Sigma_e (I - P_X)^T Q^{-1}(I - P_X)\Sigma_e W\}^{-1} \\
&\qquad\qquad \times W^T \Sigma_e (I - P_X)^T Q^{-1} y
\end{aligned}$$

The second line above uses Lemma 3, the third line uses equation (2) of Section 2 for $\tilde{\theta}$, the fourth line uses $P_{(G-\tilde{G})} = M(M^T Q^{-1} M)^{-1} M^T Q^{-1}$ where $M = G - \tilde{G} = (I - P_X)G = (I - P_X)\Sigma_e W$, and in the fifth line we simply substitute for $M$. To get equation (22), we use the easily verified results $(I - P_X)^T Q^{-1}(I - P_X) = Q^{-1}(I - P_X) = (I - P_X)^T Q^{-1}$, and then use the fact that $W^T \Sigma_e (I - P_X)^T Q^{-1} y = W^T \Sigma_e Q^{-1}(I - P_X) y = W^T (y - \tilde{\theta})$, using equation (2) of Section 2 again.

## References


Ansley, C. F. and Kohn, R. (1985). Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. *Annals of Statistics*, 13:1286–1316.

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83:28–36.

Bell, W. R. and Hillmer, S. C. (1983). Modeling time series with calendar variation. *Journal of the American Statistical Association*, 78:526–534.

Dagum, E. B. and Cholette, P. A. (2006). *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series*. Springer-Verlag, New York.

Datta, G. S., Ghosh, M., Steorts, R., and Maples, J. J. (2011). Bayesian benchmarking with applications to small area estimation. *Test*, 20:574–588.

Datta, G. S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small-area estimation problems. *Statistica Sinica*, 10:613–628.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.

Durbin, J. and Quenneville, B. (1997). Benchmarking by state space models. *International Statistical Review*, 65:23–48.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *Journal of the American Statistical Association*, 74:269–277.

Hillmer, S. C. and Trabelsi, A. (1987). Benchmarking of economic time series. *Journal of the American Statistical Association*, 82:1064–1071.

Knottnerus, P. (2003). *Sample Survey Theory: Some Pythagorean Perspectives*. Lecture Notes in Statistics. Springer-Verlag, New York.

Pfeffermann, D. (1984). On extensions of the Gauss–Markov theorem to the case of stochastic regression coefficients. *Journal of the Royal Statistical Society, Series B: Methodological*, 46:139–148.

Pfeffermann, D. and Barnard, C. H. (1991). Some new estimators for small-area means with application to the assessment of farmland values. *Journal of Business and Economic Statistics*, 9:73–84.

Pfeffermann, D. and Nathan, G. (1981). Regression analysis of data from a cluster sample. *Journal of the American Statistical Association*, 76:681–689.

Pfeffermann, D. and Tiller, R. (2006). Small-area estimation with state-space models subject to benchmark constraints. *Journal of the American Statistical Association*, 101(476):1387–1397.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association*, 78(476):47–59.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. John Wiley and Sons, New York, second edition.

Rao, J. N. K. (2003). *Small Area Estimation*. John Wiley and Sons, Hoboken, New Jersey.

Rivest, L.-P. and Vandal, N. (2003). Mean squared error estimation for small areas when the small area variances are estimated. In Roberts, G. and Bellhouse, D., editors, *Proceedings of the International Conference on Recent Advances in Survey Sampling*, Ottawa, Canada. Laboratory for Research in Statistics and Probability, Carleton University.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects (disc: p. 32-51). *Statistical Science*, 6:15–32.

Trabelsi, A. and Hillmer, S. C. (1990). Bench-marking time series with reliable benchmarks. *Applied Statistics*, 39:367–379.

Wang, J. and Fuller, W. A. (2003). The mean squared error of small area predictors constructed with estimated error variances. *Journal of the American Statistical Association*, 98:716–723.

Wang, J., Fuller, W. A., and Qu, Y. (2008). Small area estimation under a restriction. *Survey Methodology*, 34(1):29–36.

You, Y. and Rao, J. N. K. (2000). Hierarchical Bayes estimation of small area means using multi-level models. *Survey Methodology*, 26(2):173–181.

You, Y. and Rao, J. N. K. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *The Canadian Journal of Statistics*, 30(3):431–439.

You, Y., Rao, J. N. K., and Hidiroglou, M. (2012). On the performance of self benchmarked small area estimators under the Fay-Herriot area level model. *Survey Methodology*. to appear.