# A Small Area Estimation Approach for Reconciling Mode Differences in Two Surveys of Recreational Fishing Effort DRAFT: PLEASE DO NOT CITE OR DISTRIBUTE

F. Jay Breidt Teng Liu Jean D. Opsomer Colorado State University

June 10, 2017

#### Abstract

For decades, the National Marine Fisheries Service has conducted a telephone survey of United States coastal households to estimate recreational effort (the number of fishing trips) in saltwater. The effort estimates are computed for each of 17 US states along the coast of the Gulf of Mexico and the Atlantic Ocean, during six two-month waves (January-February through November-December). Recently, concerns about coverage errors in the telephone survey have led to implementation of a mail survey of the same population. Results from the mail survey are quite different from those of the telephone survey, due to coverage differences and mode effects, and a means of "calibrating" or reconciling the two sets of estimates is needed by fisheries managers and stock assessment scientists. We develop a log-normal model for the estimates from the two surveys, accounting for temporal dynamics through regression on population size and state-by-wave seasonal factors, and accounting in part for changing coverage properties through regression on wireless telephone penetration. Using the estimated design variances, we develop a regression model that is analytically consistent with the log-normal mean model. Finally, we use the modeled design variances in a Fay-Herriot small area estimation procedure to obtain empirical best linear unbiased predictors of the reconciled effort estimates for all states and waves.

## 1 Introduction

For decades, the National Marine Fisheries Service (NMFS) has conducted the Coastal Household Telephone Survey (CHTS) to collect recreational saltwater fishing effort (the number of fishing trips) from shore and private boat anglers in 17 US states along the coasts of the Atlantic Ocean and the Gulf of Mexico: Alabama, Connecticut, Delaware, Florida, Georgia, Louisiana, Maine, Maryland, Massachusetts, Mississippi, New Hampshire, New Jersey, New York, North Carolina, Rhode Island, South Carolina, and Virginia. Data collection occurs during a two-week period at the end of each twomonth sample period (or "wave"), yielding six waves for each year. However, samples are not obtained for every wave in every state; for example, many states have no wave 1 sample, reflecting minimal fishing effort during January and February in those states.

The CHTS uses random digit dialing (RDD) for landlines of households in coastal counties. RDD suffers from several shortcomings in this context, such as the inefficiency at identifying anglers (National Research Council, 2006), the declining response rate for telephone surveys (Curtin et al., 2005), and the undercoverage of anglers due to the increase in wireless-only households (Blumberg and Luke, 2013). Thus, after some experimentation, NMFS implemented the new Fishing Effort Survey (FES) that involves mailing questionnaires to a probability sample of postal addresses (Andrews et al., 2014).

The telephone-based CHTS and the mail-based FES have obvious methodological differences. The two surveys have different coverage properties, because they use very different frames: RDD of landlines for CHTS versus address-based sampling, with oversampling of addresses matched to licensed anglers, for FES. They have different nonresponse patterns, with overall FES response rates nearly three times higher than CHTS response rates (Andrews et al., 2014). Finally, the measurement processes are fundamentally different, due to the differences in asking about angling activity over the phone versus a paper form. Due at least in part to these methodological differences, there is a large discrepancy between the effort estimates from the CHTS and the FES estimates. Whatever the reasons for the discrepancy, it is of interest to fisheries managers and stock assessment scientists to be able to convert from the "units" of the telephone survey estimates to those of the mail survey estimates, and vice versa. This conversion is known as "calibration" in this context, and is not to be confused with the calibration method common in complex surveys. The calibration allows construction of a series of comparable estimates across time.

The data used for the calibration exercise come from the CHTS for most states and waves from 1982 to 2016, and from the FES for states and waves from 2015 to 2016. For each survey, the data consist of estimated total effort for shore fishing and for private boat fishing, along with estimated design variances and sample sizes, for each available state and wave.

The methodology described here uses effort estimates transformed via natural logarithms, for either shore or private boat fishing. Let  $M_{st}$  denote the estimated log-effort based on the mail survey in state s and year-wave tand let  $\hat{T}_{st}$  denote the estimated log-effort based on the telephone survey. We build a model that assumes that both mail and telephone estimates target a common underlying time series of true effort, but that each survey estimate is distorted both by sampling error and non-sampling error. The true effort series is further described with a classical time series model consisting of trend, seasonal, and irregular components. The sampling error series have properties that are well-understood based on features of the corresponding sampling designs, including well-estimated design variances. The non-sampling error cannot be completely disentangled from the true effort series. But given the overlap of mail and telephone estimates for some states and waves, the difference in the non-sampling errors can be estimated, and can be modeled with available covariates to allow extrapolation forward or backward in time. This extrapolation is a key part of the calibration procedure.

The combined model for the two sets of estimates and the underlying true effort series is a linear mixed model of a type that commonly appears in the context of area-level small area estimation, where it is known as the Fay-Herriot model (Fay and Herriot, 1979). In Fay-Herriot, it is standard to treat design variances as known. Our design variances are based on moderate to large sample sizes (minimum size n = 39) in each state and wave and so are well-estimated by the standards of small area estimation. A complication is that our design variances are on the original effort scale rather than the log scale. As an alternative to standard Taylor linearization, we develop a novel approach to transforming the estimated design variances that ensures analytic consistency between our mean model and our variance model.

The Fay-Herriot methodology leads to empirical best linear unbiased predictors (EBLUP's) of the mail target or the telephone target, and these constitute our calibrated effort series. Unlike the standard Fay-Herriot context, the EBLUP's require prediction at new sets of covariates. We adapt standard mean square error (MSE) approximations and estimates to this non-standard situation, and evaluate their performance via simulation. Finally, we apply the methods to the problem of calibrating past telephone survey estimates to the mail survey.

### 2 Model

#### 2.1 Mean model

We fix attention on one type of fishing behavior, either shore or private boat: the model development is identical in both cases. We assume that the telephone effort estimate  $\hat{T}_{st}$  is a design-unbiased estimator of the "telephone target"  $T_{st}$ , which includes both the true effort and survey mode effects due to the telephone methodology, while the mail effort estimate  $\hat{M}_{st}$  is a designunbiased estimator of the "mail target"  $M_{st}$ , which includes both the true effort and survey mode effects due to the mail methodology. That is,

$$\widehat{T}_{st} = T_{st} + e_{st}^T$$
 and  $\widehat{M}_{st} = M_{st} + e_{st}^M$ 

where the sampling errors  $\{e_{st}^T\}$  and  $\{e_{st}^M\}$  have zero mean under repeated sampling.

We assume that both the telephone target and the mail target contain the true effort series, which is further assumed to contain state-specific trends, due in part to changing state population sizes, state-specific seasonal effects that vary wave to wave, and irregular terms that are idiosyncratic effects not explained by regular trend or seasonal patterns. We model state-specific trends by using annual state-level population estimates from the US Census Bureau US Census Bureau (2016) on a log scale. We model a general seasonal pattern via indicators for the two-month waves, and allow the seasonal pattern to vary from state to state. The remaining irregular terms, denoted  $\{\nu_{st}\}$  below, represent real variation not explained by the regular trend plus

seasonal pattern, and are modeled as independent and identically distributed (iid) random variables with mean zero and unknown variance,  $\psi$ .

The survey mode effects present in the telephone and mail targets are non-sampling errors, including potential biases due to coverage error (population  $\neq$  sampling frame), nonresponse error (sample  $\neq$  respondents), and measurement error (true responses  $\neq$  measured responses). These effects may have their own trend and seasonality: for example, due to changes in the quality of the frame over time, changes in response rates over years or waves, changes in implementation of measurement protocols over time, etc. These non-sampling errors thus cannot be completely disentangled from the true effort series (a problem in every survey).

Because of the availability of overlapping effort estimates, however, the difference in the effort estimates is an unbiased estimator of the difference in the survey mode effects. These differences can then be modeled and extrapolated to other time points that do not have overlapping data, allowing calibration from the telephone target to the mail target, and vice versa. The extrapolation requires a model and suitable covariates, which in this setting means covariates that explain the change in measurement error, nonresponse error, or coverage error over time. The calibration thus relies critically on extrapolation, with the usual caveat that the calibrated values may be badly wrong if the model does not hold over the full range of time.

The changing proportion of wireless-only households is a potential covariate for explaining changes in coverage error over time for the landline-only telephone survey. Accordingly, we obtained June and/or December wirelessonly proportion estimates for each state from 2007–2014 from the National Health Interview Survey, conducted by the National Center for Health Statistics (Blumberg and Luke, 2013). We transformed these proportions via empirical logits and fitted the transformed values as state-specific lines with a slope change in 2010. The fitted model has an adjusted  $R^2$  value of 0.9948. Transforming back to proportions and extrapolating backward in time yields a series { $w_{st}$ } that is approximately zero prior to the year 2000.

Either trend or seasonal could contain survey mode effects. Accordingly, we allow for the possibility that trend and seasonal are different for mail versus telephone, and in particular we allow for the possibility that either trend or seasonal can change with the level of wireless. Our combined model then assumes

$$\widehat{T}_{st} = T_{st} + e_{st}^{T}$$

$$T_{st} = \mathbf{a}'_{st} \mathbf{\alpha} + 0 \cdot \mathbf{b}'_{st} \mathbf{\mu} + w_{st} \mathbf{c}'_{st} \mathbf{\gamma} + \nu_{st}$$

$$= [\mathbf{a}'_{st}, \mathbf{0}', w_{st} \mathbf{c}'_{st}] \mathbf{\beta} + \nu_{st}$$

$$= \mathbf{x}'_{Tst} \mathbf{\beta} + \nu_{st}$$

$$\widehat{M}_{st} = M_{st} + e_{st}^{M}$$

$$M_{st} = \mathbf{a}'_{st} \mathbf{\alpha} + 1 \cdot \mathbf{b}'_{st} \mathbf{\mu} + 0 \cdot \mathbf{c}'_{st} \mathbf{\gamma} + \nu_{st}$$

$$= [\mathbf{a}'_{st}, \mathbf{b}'_{st}, \mathbf{0}'] \mathbf{\beta} + \nu_{st}$$

$$= \mathbf{x}'_{Mst} \mathbf{\beta} + \nu_{st}, \qquad (1)$$

where

- $a_{st}$  is a vector of known covariates, including intercept, log(population), state indicators, wave indicators, and state by log(population) and state by wave interactions;
- $\boldsymbol{b}_{st}$  and  $\boldsymbol{c}_{st}$  are subvectors from  $\boldsymbol{a}_{st}$ ;
- $\beta' = [\alpha', \mu', \gamma']$  is a vector of unknown regression coefficients;
- the sampling errors  $\{e_{st}^T\}$  are independent  $\mathcal{N}(0, \sigma_{Tst}^2)$  random variables, with known design variances  $\sigma_{Tst}^2$ ;
- the sampling errors  $\{e_{st}^M\}$  are independent  $\mathcal{N}(0, \sigma_{Mst}^2)$  random variables, with known design variances  $\sigma_{Mst}^2$ ;
- the irregular terms  $\{\nu_{st}\}$ , representing real variation not explained by the regular trend plus seasonal pattern, are independent and identically distributed (iid)  $\mathcal{N}(0, \psi)$  random variables, with unknown variance  $\psi$ ;
- $\{e_{st}^T\}$ ,  $\{e_{st}^M\}$  and  $\{\nu_{st}\}$  are mutually independent.

The assumed independence of the sampling errors is justified by independent samples drawn state-to-state and wave-to-wave, and the assumed normality is justified by central limiting effects of moderate to large-size stratified samples in each state and wave. Further, we assume that because the mail and telephone surveys are selected and conducted independently, the sampling errors  $\{e_{st}^T\}$  and  $\{e_{st}^M\}$  are independent of one another. We use simulation to assess the sensitivity of some of our methods to the normality assumption on the random effects in §4.1 below. The design variances  $\{\sigma_{Tst}^2\}$  and  $\{\sigma_{Mst}^2\}$ are on the log scale, while the available design variance estimates  $\{\hat{V}_{Tst}\}$  and  $\{\hat{V}_{Mst}\}$  are on the original scale; we address this discrepancy in §2.2 below.

#### 2.2 Design variance model

Under the log-normal effort models (1), the variances of the sampling errors are given by

$$V_{Tst} = \operatorname{Var}\left(\exp(\widehat{T}_{st}) \mid T_{st}\right)$$
$$= \left\{\exp(\sigma_{Tst}^2) - 1\right\} \exp\left\{2T_{st} + \sigma_{Tst}^2\right\}$$
(2)

and

$$V_{Mst} = \operatorname{Var}\left(\exp(\widehat{M}_{st}) \mid M_{st}\right)$$
  
= {exp(\sigma\_{Mst}^2) - 1} exp {2M\_{st} + \sigma\_{Mst}^2}. (3)

We need to estimate  $\sigma_{Tst}^2$  and  $\sigma_{Mst}^2$ , incorporating the approximately designunbiased estimates  $\hat{V}_{Tst}$  and  $\hat{V}_{Mst}$  of  $V_{Tst}$  and  $V_{Mst}$ , respectively.

We follow an approach related closely to generalized variance function estimation (e.g., Ch. 7 of Wolter (2007)). Assume that given  $T_{st}$  and  $M_{st}$ , the empirical coefficients of variation (CV's) are log-normally distributed, independent of the effort estimates  $\widehat{T}_{st}$  and  $\widehat{M}_{st}$ :

$$\ln\left(\frac{\widehat{V}_{Tst}}{\exp(2\widehat{T}_{st})}\right) = \boldsymbol{d}'_{Tst}\boldsymbol{\delta}_0^T + \delta_1^T \ln(n_{Tst}) + \eta_{st}^T, \quad \eta_{st}^T \sim \mathcal{N}(0, \tau_T^2)$$
(4)

where  $d_{Tst}$  is a vector of known covariates (including state, wave, and state by wave interaction), and

$$\ln\left(\frac{\widehat{V}_{Mst}}{\exp(2\widehat{M}_{st})}\right) = \boldsymbol{d}'_{Mst}\boldsymbol{\delta}_0^M + \delta_1^M \ln(n_{Mst}) + \eta_{st}^M, \quad \eta_{st}^M \sim \mathcal{N}(0, \tau_M^2), \quad (5)$$

where  $d_{Mst}$  is a vector of known covariates. These models can be rewritten as regression models for the design variance estimates, with known offsets:

$$\ln\left(\widehat{V}_{Tst}\right) = 2\widehat{T}_{st} + \boldsymbol{d}'_{Tst}\boldsymbol{\delta}_0^T + \boldsymbol{\delta}_1^T\ln(n_{Tst}) + \eta_{st}^T, \quad \eta_{st}^T \sim \mathcal{N}(0, \tau_T^2)$$

and

$$\ln\left(\widehat{V}_{Mst}\right) = 2\widehat{M}_{st} + \boldsymbol{d}'_{Mst}\boldsymbol{\delta}_0^M + \delta_1^M \ln(n_{Mst}) + \eta_{st}^M, \quad \eta_{st}^M \sim \mathcal{N}(0, \tau_M^2).$$

Empirically, each of these models fits very well: 94.54% adjusted  $R^2$  value for telephone, and 98.01% adjusted  $R^2$  value for mail.

These empirical models may be of independent interest as generalized variance functions for variance estimation on the original scale: by plugging the point estimate, state, wave, and sample size into the fitted versions of (4) or (5), one obtains excellent point estimates of the coefficient of variation.

Assuming that  $\hat{V}_{Tst}$  is exactly unbiased for  $V_{Tst}$ , we then have from the log-normal CV model (4) and the assumed conditional independence of  $\hat{V}_{Tst}$  and  $\hat{T}_{st}$  given  $T_{st}$  that

$$\exp\left\{ \boldsymbol{d}_{Tst}^{\prime} \boldsymbol{\delta}_{0}^{T} + \boldsymbol{\delta}_{1}^{T} \ln(n_{Tst}) + \frac{\tau_{T}^{2}}{2} \right\} = \mathbb{E}\left[ \left. \frac{\widehat{V}_{Tst}}{\exp\left(2\widehat{T}_{st}\right)} \right| T_{st} \right]$$
$$= \mathbb{E}\left[ \widehat{V}_{Tst} \mid T_{st} \right] \mathbb{E}\left[ \exp\left(-2\widehat{T}_{st}\right) \mid T_{st} \right]$$
$$= V_{Tst} \exp\left(-2T_{st} + 2\sigma_{Tst}^{2}\right), \qquad (6)$$

and similarly

$$\exp\left\{ \boldsymbol{d}_{Mst}^{\prime}\boldsymbol{\delta}_{0}^{M} + \boldsymbol{\delta}_{1}^{M}\ln(n_{Mst}) + \frac{\tau_{M}^{2}}{2} \right\} = \mathbb{E}\left[ \left. \frac{\widehat{V}_{Mst}}{\exp\left(2\widehat{M}_{st}\right)} \right| M_{st} \right]$$
$$= \mathbb{E}\left[ \widehat{V}_{Mst} \mid M_{st} \right] \mathbb{E}\left[ \exp\left(-2\widehat{M}_{st}\right) \mid M_{st} \right]$$
$$= V_{Mst} \exp\left(-2M_{st} + 2\sigma_{Mst}^{2}\right). \tag{7}$$

Thus, we have from (2) and (6) that

$$\exp\left\{\boldsymbol{d}_{Tst}^{\prime}\boldsymbol{\delta}_{0}^{T}+\boldsymbol{\delta}_{1}^{T}\ln(n_{Tst})+\frac{\tau_{T}^{2}}{2}\right\}$$
$$=\left\{\exp(\sigma_{Tst}^{2})-1\right\}\exp\left\{2T_{st}+\sigma_{Tst}^{2}\right\}\exp\left(-2T_{st}+2\sigma_{Tst}^{2}\right)$$
$$=\exp(4\sigma_{Tst}^{2})-\exp\left(3\sigma_{Tst}^{2}\right)$$
(8)

and from (3) and (7) that

$$\exp\left\{ \boldsymbol{d}_{Mst}^{\prime}\boldsymbol{\delta}_{0}^{M} + \boldsymbol{\delta}_{1}^{M}\ln(n_{Mst}) + \frac{\tau_{M}^{2}}{2} \right\}$$
  
= {exp( $\sigma_{Mst}^{2}$ ) - 1} exp {2 $M_{st} + \sigma_{Mst}^{2}$ } exp (-2 $M_{st} + 2\sigma_{Mst}^{2}$ )  
= exp(4 $\sigma_{Mst}^{2}$ ) - exp (3 $\sigma_{Mst}^{2}$ ). (9)

The left-hand-side parameters of (8) can be estimated from (4) and the lefthand-side parameters of (9) can be estimated from (5). The resulting estimates of  $\sigma_{Tst}^2$  and  $\sigma_{Mst}^2$  can then be obtained by solving the equations (8) and (9), which are quartic polynomials in  $\exp(\sigma_{Tst}^2)$  and  $\exp(\sigma_{Mst}^2)$ . Using Descartes' rule of signs, it can be shown that each of these quartic equations has one negative real root, two complex conjugate roots, and one positive real root. The solutions for  $\sigma_{Tst}^2$  and  $\sigma_{Mst}^2$  are then the logarithms of the unique, positive real roots, which can be obtained via standard numerical procedures. While these solutions are in fact estimates, we will treat them as fixed and known in what follows, as is standard in the small area estimation techniques which we will apply in subsequent sections.

The resulting design variances on the log scale,  $\sigma_{Tst}^2$  and  $\sigma_{Mst}^2$ , are strongly correlated with the estimated variance approximations from Taylor linearization,  $\widehat{V}_{Tst} \exp\left(-2\widehat{T}_{st}\right)$  and  $\widehat{V}_{Mst} \exp\left(-2\widehat{M}_{st}\right)$ : 0.798 and 0.803, respectively. But they are not identical (see Figure 1), and the method described forces analytical consistency between the mean model and the variance model.

#### 2.3 Fay-Herriot small area estimation model

Define

$$m{x}'_{st} = \begin{cases} m{x}'_{Tst}, & ext{if no mail estimate is available;} \\ m{x}'_{Mst}, & ext{if no telephone estimate is available;} \\ (m{x}_{Tst} + m{x}_{Mst})'/2, & ext{otherwise.} \end{cases}$$



Figure 1: Estimated design variances for log-effort via Taylor linearization versus solution of the quartic polynomial equations (8) for telephone (left panel) and (9) for mail (right panel).

Then it is convenient to write

$$Y_{st} = \begin{cases} \widehat{T}_{st}, & \text{if no mail estimate is available;} \\ \widehat{M}_{st}, & \text{if no telephone estimate is available;} \\ \left(\widehat{T}_{st} + \widehat{M}_{st}\right)/2, & \text{otherwise;} \end{cases}$$

$$= \begin{cases} \boldsymbol{x}'_{Tst}\boldsymbol{\beta} + \nu_{st} + e^T_{st}, & \text{if no mail estimate is available;} \\ \boldsymbol{x}'_{Mst}\boldsymbol{\beta} + \nu_{st} + e^M_{st}, & \text{if no telephone estimate is available;} \\ (\boldsymbol{x}_{Tst} + \boldsymbol{x}_{Mst})'\boldsymbol{\beta}/2 + \nu_{st} + (e^T_{st} + e^M_{st})/2, & \text{otherwise;} \end{cases}$$

$$= \boldsymbol{x}'_{st}\boldsymbol{\beta} + \nu_{st} + e_{st}. \qquad (10)$$

This model then follows exactly the linear mixed model structure of Fay and Herriot (1979), with direct estimates  $Y_{st}$  equal to regression model plus random effect  $\nu_{st}$  plus sampling error with "known" design variance, given by

 $D_{st} = \begin{cases} \sigma_{Tst}^2, & \text{if no mail estimate is available;} \\ \sigma_{Mst}^2, & \text{if no telephone estimate is available;} \\ \frac{1}{4} \left( \sigma_{Tst}^2 + \sigma_{Mst}^2 \right), & \text{otherwise.} \end{cases}$ 

Averaging the telephone and mail estimates results in a small loss of information, since we are replacing two correlated observations with one observation, but allows the use of standard software for estimation.

### 3 Methods

#### 3.1 Estimation for the Fay-Herriot model

1

Define  $\mathcal{A} = \{(s,t) : Y_{st} \text{ is not missing}\}$  to be the set of all state by yearwave combinations for which we have an estimate from either survey. Let mdenote the size of the set  $\mathcal{A}$ . Define  $\mathbf{X} = [\mathbf{x}'_{st}]_{(s,t)\in\mathcal{A}}, \mathbf{Y} = [Y_{st}]_{(s,t)\in\mathcal{A}}$ , and

$$\boldsymbol{\Sigma}(\psi) = \operatorname{Var}(\boldsymbol{Y}) = \operatorname{diag}\{\psi + D_{st}\}_{(s,t)\in\mathcal{A}}.$$

Then

$$oldsymbol{Y} = oldsymbol{X}oldsymbol{eta} + [
u_{st}]_{(s,t)\in\mathcal{A}} + [e_{st}]_{(s,t)\in\mathcal{A}}.$$

If  $\psi$  were known, the best linear unbiased estimator (BLUE) of  $\beta$  would be

$$\widetilde{\boldsymbol{\beta}}_{\psi} = \left\{ \boldsymbol{X}' \boldsymbol{\Sigma}^{-1}(\psi) \boldsymbol{X} \right\}^{-1} \boldsymbol{X}' \boldsymbol{\Sigma}^{-1}(\psi) \boldsymbol{Y}.$$
(11)

Since  $\psi$  is not known, we replace it by a consistent estimator to obtain

$$\widehat{\boldsymbol{\beta}} = \left\{ \boldsymbol{X}' \boldsymbol{\Sigma}^{-1}(\widehat{\psi}) \boldsymbol{X} \right\}^{-1} \boldsymbol{X}' \boldsymbol{\Sigma}^{-1}(\widehat{\psi}) \boldsymbol{Y}.$$
(12)

We will use the Restricted Maximum Likelihood (REML) estimate  $\hat{\psi}$  unless otherwise indicated.

#### 3.2 Prediction

In the classical Fay-Herriot context, it is of interest to predict

$$oldsymbol{x}_{st}^{\prime}oldsymbol{eta}+
u_{st}$$

from (10). In our setting, however, we seek to predict

$$\phi_{st} = \boldsymbol{z}_{st}^{\prime} \boldsymbol{\beta} + \nu_{st}, \qquad (13)$$

where  $\boldsymbol{z}_{st}$  may not equal  $\boldsymbol{x}_{st}$ . For example, for a past time point with a telephone survey estimate but no mail survey estimate, we may want to use

$$oldsymbol{z}_{st}'=oldsymbol{x}_{Mst}'=[oldsymbol{a}_{st}',oldsymbol{b}_{st}',oldsymbol{0}']$$

to predict the mail target  $M_{st}$ , while for a future time point with a mail survey estimate but no telephone, we may want to use

$$oldsymbol{z}_{st} = [oldsymbol{a}_{st}',oldsymbol{0}',oldsymbol{0}']$$

to predict the telephone target, corrected for the wireless effect:  $T_{st} - w_{st} c'_{st} \gamma = a'_{st} \alpha + \nu_{st}$ .

Let  $\lambda_{st}$  denote a  $m \times 1$  vector with a one in the (s, t)th position and zero elsewhere. Under normality, it is well-known that the best mean square predictor of  $\phi_{st}$  in (13) is

$$\phi_{st}\left(\boldsymbol{\beta},\psi\right) = \boldsymbol{z}_{st}^{\prime}\boldsymbol{\beta} + \psi\boldsymbol{\lambda}_{st}^{\prime}\boldsymbol{\Sigma}^{-1}(\psi)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), \qquad (14)$$

which is feasible only if both  $\beta$  and  $\psi$  are both known. If only  $\psi$  is known, the best linear unbiased predictor (BLUP)

$$\phi_{st}\left(\widetilde{\boldsymbol{\beta}}_{\psi},\psi\right) = \boldsymbol{z}_{st}^{\prime}\widetilde{\boldsymbol{\beta}}(\psi) + \psi\boldsymbol{\lambda}_{st}^{\prime}\boldsymbol{\Sigma}^{-1}(\psi)(\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\psi))$$
(15)

is obtained by plugging the BLUE from (11) into (14). Finally, if neither  $\beta$  nor  $\psi$  is known, then the empirical best linear unbiased predictor (EBLUP) can be obtained by substituting a consistent estimator of  $\psi$  into (15):

$$\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}\right) = \boldsymbol{z}_{st}' \widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\psi}} \boldsymbol{\lambda}_{st}' \boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\psi}}) (\boldsymbol{Y} - \boldsymbol{X} \widehat{\boldsymbol{\beta}}), \tag{16}$$

where  $\widehat{\boldsymbol{\beta}}$  is given by (12). These EBLUP's are the proposed calibrated values on the log scale.

#### 3.3 Mean square error approximation

To assess the uncertainty of the calibrated values, we adapt the approach of Datta and Lahiri (2000) in approximating the mean square error (MSE) of the  $\phi_{st}\left(\hat{\boldsymbol{\beta}},\hat{\psi}\right)$  values. It can be shown that

$$MSE\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}\right)\right\} = E\left[\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}\right) - \phi_{st}\right\}^{2}\right]$$
$$= E\left[\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}_{\psi}, \psi\right) - \phi_{st}\right\}^{2}\right] + E\left[\left\{\phi_{st}\left(\boldsymbol{\beta}, \psi\right) - \phi_{st}\left(\widehat{\boldsymbol{\beta}}_{\psi}, \psi\right)\right\}^{2}\right]$$
$$+ E\left[\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}\right) - \phi_{st}\left(\boldsymbol{\beta}, \psi\right)\right\}^{2}\right]$$
$$= \dot{g}_{1st}(\psi) + \dot{g}_{2st}(\psi) + \dot{g}_{3st}(\psi) + o\left(m^{-1}\right), \qquad (17)$$

where

$$\begin{split} \dot{g}_{1st}(\psi) &= \frac{\psi D_{st}}{\psi + D_{st}}, \\ \dot{g}_{2st}(\psi) &= \left(\frac{\psi(\boldsymbol{z}_{st} - \boldsymbol{x}_{st})' + D_{st}\boldsymbol{z}_{st}'}{\psi + D_{st}}\right) \left[\sum_{u \in \mathcal{A}} (\psi + D_u)^{-1} \boldsymbol{x}_u \boldsymbol{x}_u'\right]^{-1} \\ &\times \left(\frac{\psi(\boldsymbol{z}_{st} - \boldsymbol{x}_{st})' + D_{st}\boldsymbol{z}_{st}'}{\psi + D_{st}}\right)', \end{split}$$

and

$$\dot{g}_{3st}(\psi) = \frac{2D_{st}^2}{(\psi + D_{st})^3} \frac{1}{\sum_{u \in \mathcal{A}} (\psi + D_u)^{-2}}.$$

The terms  $\dot{g}_{1st}(\psi)$  and  $\dot{g}_{3st}(\psi)$  are identical to the terms  $g_{1st}(\psi)$  and  $g_{3st}(\psi)$ in §4 of Datta and Lahiri (2000), while  $\dot{g}_{2st}(\psi)$  simplifies to  $g_{2st}(\psi)$  of that paper in the special case of  $\boldsymbol{z}_{st} = \boldsymbol{x}_{st}$ . We omit the proofs.

### 3.4 Mean square error estimation

We now propose an estimator of the MSE approximation in (17). Using arguments like those in §5 of Datta and Lahiri (2000), it can be shown that

and hence an approximately unbiased estimator of the MSE approximation in (17) is given by

mse 
$$\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\} = \dot{g}_{1st}(\hat{\psi}) + \dot{g}_{2st}(\hat{\psi}) + 2\dot{g}_{3st}(\hat{\psi}).$$
 (18)

We assess the quality of the asymptotic approximation (17) and its estimator (18) via simulation in §4.1.

#### 3.5 Prediction on the original scale

To compute predictors on the original scale, we back-transform by exponentiating the EBLUP from (16) and adjust for the nonlinearity of the backtransformation using the estimated MSE from (18):

$$\widehat{\exp(\phi_{st})} = \exp\left[\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}\right) + \frac{1}{2}\operatorname{mse}\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}\right)\right\}\right],\tag{19}$$

which is an estimator of the best mean square predictor under the normal model, and a standard adjustment even without the normality assumption.

### 4 Empirical results

#### 4.1 Simulation

In this section, we investigate the performance of our second-order approximation of MSE and the estimated MSE under a setting that mimics the calibration problem of this paper, but with a smaller number of observed time points: 17 states and six years (1985, 1995, 2005, 2010, 2015, and 2016) of six waves each, with telephone effort estimates for all waves, and with mail effort estimates for only the final two years. In this setting, m = (17 states)(6 waves)(6 years)=612. We took the wireless values and US Census population counts from the actual data.

We used as true regression coefficient values the estimates from model (10) fitted to shore data, with intercept, log(population), state indicators, wave indicators, state by log(population) interaction, and state by wave; plus wireless and its interactions with log(population), state indicators, and wave indicators; plus an indicator for presence of a mail survey estimate and the mail indicator's interactions with log(population), state indicators, and wave indicators. We also used  $\psi = 0.11$ , again from the fit of the model. The simulation model is similar to the final model selected in §4.2 below.

We considered three different patterns for the design variances  $\{D_{st}\}$ . First, we sampled six actual design variances for each simulated state, arranged the six into a "peaked" seasonal pattern, and replicated this seasonal pattern across all six years to create pattern (b). We considered two additional settings, by multiplying pattern (b) by 0.5 to yield pattern (a), and multiplying pattern (b) by 2.0 to yield pattern (c). The simulated sampling errors  $\{e_{st}\}$  in (10) were then generated independently as  $\mathcal{N}(0, D_{st})$  under each pattern.

Following Datta et al. (2005), we considered three distributions to simulate the normalized random effects:

- $\{\psi^{-1/2}\nu_{st}\}$  iid  $\mathcal{N}(0,1);$
- $\{\psi^{-1/2}\nu_{st}\}$  iid Laplace $(0, 1/\sqrt{2});$
- $\{\psi^{-1/2}\nu_{st}\}$  iid centered Exponential(1) (that is, exponential random variables centered to mean zero).

Under each distribution,  $E[\nu_{st}] = 0$  and  $Var(\nu_{st}) = \psi$ .

For each combination of sampling variance pattern and random effect distribution, we generated 1000 data sets from model (10). For each simulated data set, we used the R package sae (Molina and Marhuenda, 2015) to compute  $\hat{\psi}$  via REML and  $\hat{\beta}$ . We computed the EBLUP's in (16) for the mail targets  $\{M_{st}\}$ , approximated their MSE's using (17), and estimated their MSE's using (18). We then compared the approximations and the estimates to the true (Monte Carlo) MSE's over the 1000 simulated realizations.

Figure 2 shows plots of the MSE approximation and the estimated MSE versus the true MSE for each of the nine simulation scenarios. Here the gray dots are the MSE approximations and the black circles are the estimated MSE's. The approximations and estimates are nearly overlapping in all cases, indicating that the MSE estimates are essentially unbiased for the MSE approximations. Further, the points are all very close to the (0,1) reference line, indicating that the proposed methodology yields acceptable MSE estimates across a range of settings.

#### 4.2 Calibration of the CHTS and FES estimates

For the data described in §1, we used the **R** package **sae** (Molina and Marhuenda, 2015) to fit a number of models via maximum likelihood for both shore fishing and private boat fishing, and compared the models via their AIC values. The smallest model considered included intercept, log(population), state indicators, wave indicators, state by log(population) interaction, and state by wave interaction. That is, the smallest model includes no differences due to survey methodology and instead drops the terms  $b'_{st}\mu$  and



Figure 2: MSE approximation (solid gray dots) and estimated MSE's (open black circles) versus true MSE from Monte Carlo, for random effect distributions normal, Laplace, and centered exponential across the rows, and sampling error patterns (a), (b), and (c) across the columns.

 $w_{st}c'_{st}\gamma$  from (1). The largest model considered added wireless and its interactions with log(population), state indicators, wave indicators, and state by log(population), together with an indicator for presence of a mail survey estimate and the mail indicator's interactions with log(population), state indicators, and wave indicators. The omission of the higher order interactions between wireless and the mail indicator is due to parsimony: for the mail indicator in particular, there are only 17 states and 11 waves from which to estimate the parameters  $\mu$  in model (1).

Numerous submodels between the smallest and largest were considered; the best four models and additional reference models are given in Table 1 for shore fishing and Table 2 for private boat fishing. The tables are ordered by AIC values, with the best models at the top. The models that ignore some (largest minus all mail, largest minus all wireless) or all (smallest) of the survey mode differences are not competitive with the models that include these factors. The largest model considered is quite competitive, with the best models dropping a small number of interactions from that largest model.

While not the best model for either shore or private boat, the largest model minus the mail by log(population) interaction is third best in both cases. It is operationally convenient to use a common model for both calibrations, and this particular model is further convenient because, when extrapolating back in time, it involves only state by wave level shifts once the effect of wireless has died out. We therefore chose this model as the final model for both modes of fishing, and refitted it using REML to estimate the unknown variance  $\psi$ . We then computed EBLUP's of the mail target  $\{M_{st}\}$ for all states and waves.

An example for Alabama shore fishing is shown in Figure 3 and an example for Florida private boat fishing is shown in Figure 4. In each figure, we show the effects of successive adjustment, from the telephone log-effort estimates  $\{\hat{T}_{st}\}$ , to the estimates  $\{\hat{T}_{st} + \mathbf{b}'_{st}\hat{\boldsymbol{\mu}}\}$  that adjust only for mail methodology effects, to the estimates  $\{\hat{T}_{st} + \mathbf{b}'_{st}\hat{\boldsymbol{\mu}} - w_{st}\mathbf{c}'_{st}\hat{\boldsymbol{\gamma}}\}$  that adjust for both mail and wireless, and finally the EBLUP's themselves. As expected, the effect of wireless is only present in the later years since 2000, and is a relatively modest effect. The EBLUP can be seen as a smoothed version of the estimates adjusted for mail methodology and wireless effects.

Model is largest minus terms below:	$\log(\text{likelihood})$	AIC	df
mail:log(pop) and wireless:wave	-1803.53	3947.06	2798
mail:log(pop), mail:wave, wireless:wave	-1810.49	3950.99	2803
$\operatorname{mail:log(pop)}$	-1801.57	3953.14	2793
nothing (largest)	-1801.23	3954.47	2792
mail:log(pop) and mail:wave	-1808.48	3956.96	2798
mail:log(pop) and mail:state	-1821.50	3961.01	2809
mail interactions	-1828.03	3964.07	2814
wireless interactions	-1942.98	4161.97	2830
all interactions	-1969.05	4170.10	2852
all mail	-1935.15	4176.30	2815
all wireless	-1977.54	4229.09	2831
all mail and all wireless (smallest)	-2109.83	4447.66	2854

Table 1: Maximized log(likelihood), AIC and residual degrees of freedom (df) for various models fitted to effort estimates for shore fishing. See text for description of largest model.

Model is largest minus terms below:	$\log(\text{likelihood})$	AIC	df
mail interactions	-1336.00	2981.99	2816
mail:log(pop) and mail:wave	-1320.07	2982.13	2800
mail:log(pop)	-1315.48	2982.97	2795
mail:log(pop) and mail:state	-1331.70	2983.40	2811
nothing (largest)	-1314.83	2983.66	2794
mail:log(pop) and wireless:wave	-1323.26	2988.52	2800
mail:log(pop), mail:wave, wireless:wave	-1332.19	2996.37	2805
all mail	-1417.45	3142.90	2817
wireless interactions	-1463.00	3204.01	2832
all interactions	-1495.69	3225.37	2854
all wireless	-1548.81	3373.62	2833
all mail and all wireless (smallest)	-1611.74	3453.48	2856

Table 2: Maximized log(likelihood), AIC and residual degrees of freedom (df) for various models fitted to effort estimates for private boat fishing. See text for description of largest model.



Shore Mode log(effort) for Alabama

Figure 3: EBLUP's  $\left\{\phi_{st}\left(\hat{\boldsymbol{\beta}},\hat{\psi}\right)\right\}$  (gold curve) of mail targets  $\{M_{st}\}$  for shore fishing log-effort in Alabama. Blue dots are telephone log-effort estimates  $\{\hat{T}_{st}\}$  and pink triangles are mail log-effort estimates  $\{\widehat{M}_{st}\}$ . For comparison to EBLUP's, gray curve is the estimator  $\{\widehat{T}_{st} + \boldsymbol{b}'_{st}\hat{\boldsymbol{\mu}}\}$  that adjusts only for mail methodology effects, and black curve is  $\{\widehat{T}_{st} + \boldsymbol{b}'_{st}\hat{\boldsymbol{\mu}} - w_{st}\boldsymbol{c}'_{st}\hat{\boldsymbol{\gamma}}\}$  that adjusts for mail and wireless.



#### Private Boat Mode log(effort) for Florida

Figure 4: EBLUP's  $\left\{\phi_{st}\left(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\psi}}\right)\right\}$  (gold curve) of mail targets  $\{M_{st}\}$  for private boat fishing in Florida. Blue dots are telephone log-effort estimates  $\{\hat{T}_{st}\}$  and pink triangles are mail log-effort estimates  $\{\widehat{M}_{st}\}$ . For comparison to EBLUP's, gray curve is the estimator  $\{\hat{T}_{st} + \boldsymbol{b}'_{st}\hat{\boldsymbol{\mu}}\}$  that adjusts only for mail methodology effects, and black curve is  $\{\hat{T}_{st} + \boldsymbol{b}'_{st}\hat{\boldsymbol{\mu}} - w_{st}\boldsymbol{c}'_{st}\hat{\boldsymbol{\gamma}}\}$  that adjusts for mail and wireless.

### 5 Discussion

The proposed methodology accounts for various sources of variation in the effort series from each survey, including trend, seasonality and irregular terms in the true effort series, together with survey mode effects in the two series. The model assumes that differences in measurement and nonresponse errors between the two surveys would be stable over time, while the changes in coverage error over time due to growth in wireless-only households is explicitly modeled. Further, the methodology accounts for uncertainty due to sampling error, using a novel approach to ensure analytical consistency in mapping design variances estimated on the original scale to design variances estimated on the log scale.

As formulated in this paper, the calibration methodology turns out to follow a standard, well-established procedure: Fay-Herriot small area estimation. This means that the calibrated values turn out to empirical best linear unbiased predictors under a linear mixed model fitted using likelihood-based techniques. The method is flexible enough to provide optimal calibrated values for different problems: predicting mail targets using telephone-only data, or predicting telephone targets using mail-only data, for example.

Uncertainty is quantified via a mean square error approximation that adapts existing methods from the literature. Simulation results show that the mean square error approximation and its estimator are highly accurate for the kinds of sample sizes and sampling errors present in the calibration data. The methodology is readily implemented with standard software.

#### Acknowledgements.

We thank NOAA Fisheries scientists Rob Andrews and John Foster for assistance with the problem formulation and all aspects of the data compilation. NOAA affiliate Ryan Kitts-Jensen and Westat statistician Mike Brick provided useful discussion.

### References

Andrews, R., J. M. Brick, and N. A. Mathiowetz (2014). Development and testing of recreational fishing effort surveys. Technical report.

Blumberg, S. and J. Luke (2013). Wireless substitution: Early release of

estimates from the National Health Interview Survey, July–December 2012. Technical report, National Center for Health Statistics.

- Curtin, R., S. Presser, and E. Singer (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly* 69(1), 87.
- Datta, G. S. and P. Lahiri (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica* 10, 613–627.
- Datta, G. S., J. N. K. Rao, and D. D. Smith (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika* 92(1), 183–196.
- Fay, R. E. and R. A. Herriot (1979). Estimation of income from small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74, 269–277.
- Molina, I. and Y. Marhuenda (2015). sae: An R package for small area estimation. The R Journal 7/1, 81–98.
- National Research Council (2006). *Review of Recreational Fisheries Survey Methods.* Washington, DC: The National Academies Press.
- US Census Bureau (2016). State Population Totals Datasets: 2010-2016. https://www.census.gov/data/datasets/2016/demo/popest/ state-total.html.
- Wolter, K. M. (2007). Introduction to Variance Estimation (2 ed.). New York: Springer-Verlag Inc.