Calibration Methodology for CHTS to FES Transition

Jay Breidt
Colorado State University

Joint work with Teng Liu, Colorado State University
and Jean Opsomer, Westat

The need for survey redesign

- van den Brakel et al. (2020) *International Statistical Review*:

  A key requirement of repeated surveys conducted by national statistical institutes is the comparability of estimates over time, resulting in uninterrupted time series describing the evolution of finite population parameters. This is often an argument to keep survey processes unchanged as long as possible. It is nevertheless inevitable that a survey process will need to be redesigned from time to time, for example, to improve or update methods or implement more cost-effective data collection procedures.

Transitioning from telephone to mail

- Olson et al. (2020), Transitions From Telephone Surveys to Self-Administered and Mixed-Mode Surveys: AAPOR Task Force Report, *Journal of Survey Statistics and Methodology*:

  > Telephone surveys have been a ubiquitous method of collecting survey data, but the environment for telephone surveys is changing. Many surveys are transitioning from telephone to self-administration or combinations of modes for both recruitment and survey administration.

Discontinuities due to survey redesign

- van den Brakel et al. (2020) *International Statistical Review*:

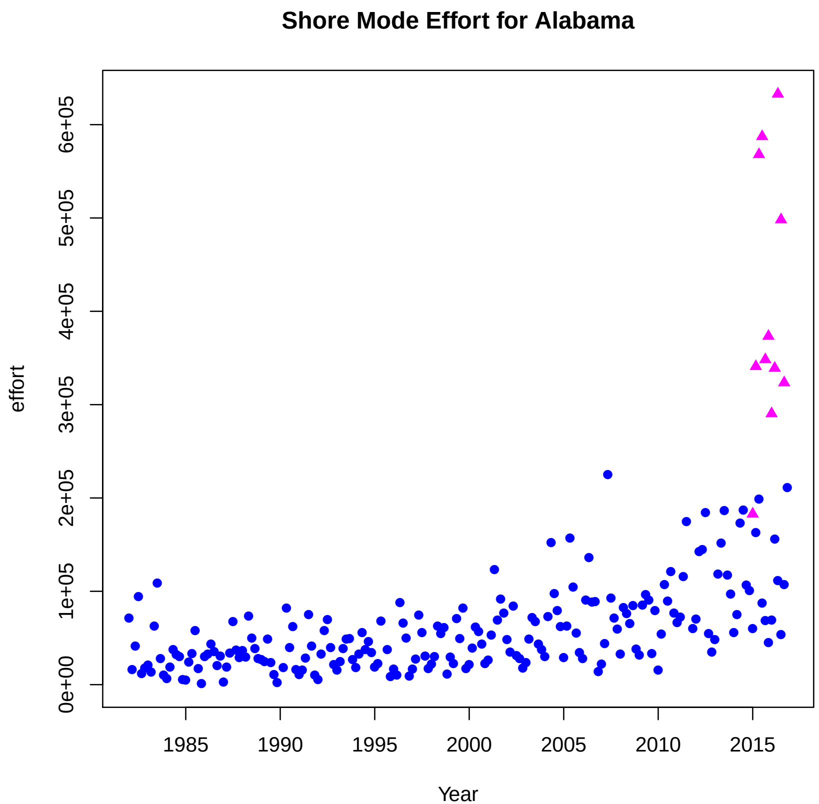
  Survey samples contain, besides sampling errors, different sources of non-sampling errors that have a systematic effect on the outcomes of a survey. As long as the survey process is kept constant, this bias component is not visible for the comparability over time. If, however, one or more components of the survey process are modified, the biases induced by these non-sampling errors are changed, likely to be visible and misinterpreted as finite population parameter changes. Major redesign of the underlying survey process, therefore, generally has systematic effects on the survey estimates, disturbing comparability with figures published in the past.

Discontinuity in fishing effort

- Major change in survey methodology, leading to major change in survey estimates
- MRIP committed to develop calibration method that enabled construction of a new, consistent time series
- How to approach this problem statistically?
  - identify sources of uncertainty
  - make any assumptions explicit
  - use best practices under given assumptions
  - assess sensitivity to failure of model assumptions

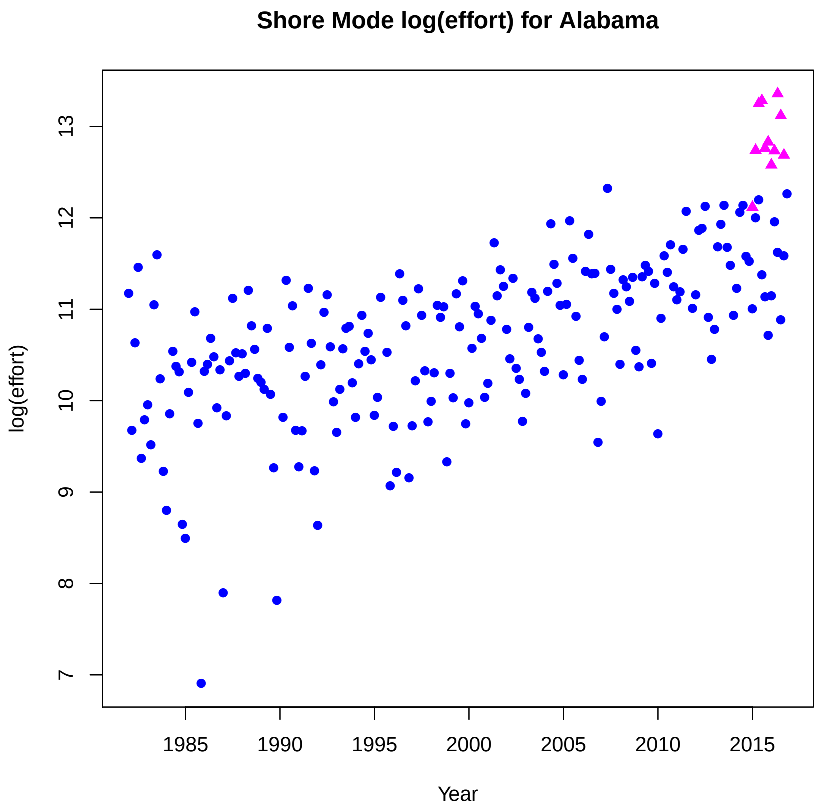
Alabama shore fishing, original scale

- Available FES (Mail) effort estimates are consistently much higher than CHTS (Telephone)
- Limited number of overlapping waves



Alabama shore fishing, log scale

- Available FES (Mail) effort estimates are consistently much higher than CHTS (Telephone)
- Limited number of overlapping waves

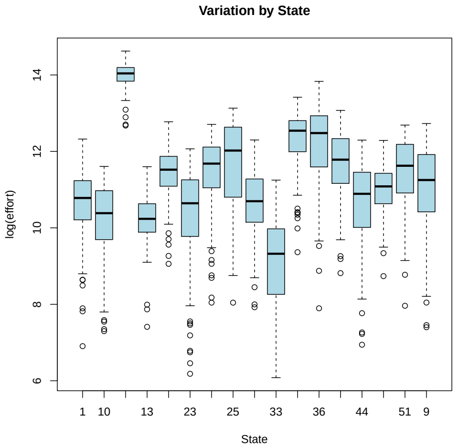


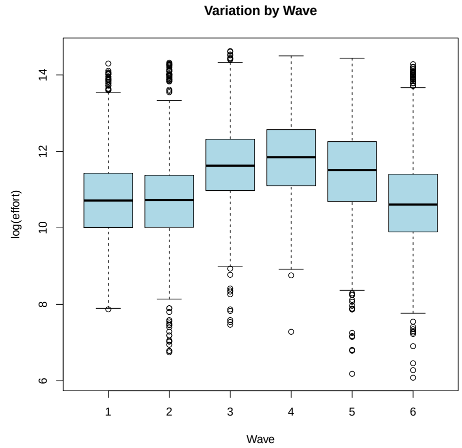
Calibration problem

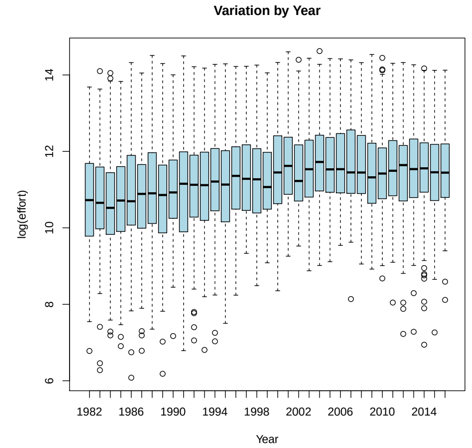
- MRIP goal: *"...estimates that best represent what would have been produced had the new FES design been used prior to 2017"*
- Is there a way to convert from `Telephone` "units" to `Mail` "units" and vice versa?
- Want a defensible statistical approach, realizing that it will have to rely on some modeling assumptions
- Our statistical methodology will make no judgment that one method is correct or even better: they are just different

Start by identifying sources of variation

- Mail and telephone estimates differ from each other and vary over time and space







Describing spatio-temporal variation

- Explain as much of the *shared* spatio-temporal variation as possible, then model the mail-telephone differences
- Both Mail and Telephone should "see" spatio-temporal variation:
  - `Trend`: effort varies over years in part due to population changes
  - `Seasonal`: effort varies wave-to-wave, and this pattern varies state-to-state
  - `Irregular`: true effort has additional, real variation not explained by regular `Trend` + `Seasonal` pattern
- Model is then `Effort` = `Trend` + `Seasonal` + `Irregular` for each state

Sources of variation, II

- Model is "classical decomposition" of time series analysis,

$$\texttt{Effort} = \texttt{Trend} + \texttt{Seasonal} + \texttt{Irregular}$$

for each state's effort series

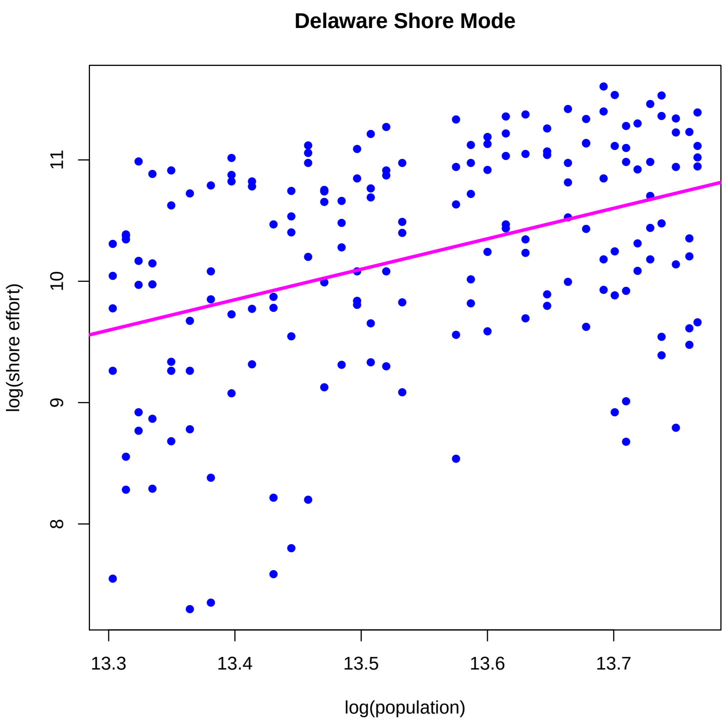
- We do not observe `Effort` directly, but with `Sampling Error` and with `Method Effect`
- Log-scale estimates can be written

$$\begin{aligned}\texttt{Telephone} &= \texttt{Telephone Method} + \texttt{Effort} \\ &\quad + \texttt{Telephone Sampling Error} \\ \texttt{Mail} &= \texttt{Mail Method} + \texttt{Effort} \\ &\quad + \texttt{Mail Sampling Error}\end{aligned}$$

- We'll discuss `Effort`, then `Sampling Error`, then `Method Effects`

Modeling Trend in Effort

- Use state-specific population sizes to describe trend
  - from the US Census Bureau, have state-specific population sizes
  - decennial census plus demographic methods in non-census years



Modeling Trend + Seasonal in Effort

- Construct dummy variables (indicator variables) for six two-month waves, one set for each state
- Trend + Seasonal model is then

$$\begin{aligned} \boldsymbol{a}'_{st}\boldsymbol{\alpha} &= \texttt{state} + \texttt{log(pop)} + \texttt{state*log(pop)} \\ &\quad + \texttt{wave} + \texttt{state*wave} \end{aligned}$$

- Simple model accounts for much of the variation in `Telephone`:

	R^2_{adj}	Residual SE	df
Shore all	0.841	0.544	2869
Shore prior to 2000	0.847	0.561	1431
Shore 2000 and later	0.849	0.475	1335
Boat all	0.878	0.493	2871
Boat prior to 2000	0.890	0.487	1436
Boat 2000 and later	0.893	0.424	1332

Modeling `Irregular` in `Effort`

- `Irregular`: true effort has additional, real variation not explained by regular `Trend` + `Seasonal` pattern
- By definition, we cannot explain it
- Instead, we model `Irregular` as a random quantity, with mean zero, and unknown variance to be estimated:

`Irregular`	independent and identically distributed as Normal with mean zero and variance ψ
$\{\nu_{st}\}$	iid $\mathcal{N}(0, \psi)$

Sampling Error

- Sampling Error properties for telephone and mail are well-understood from their respective designs
  - zero-mean, hence

$$\begin{aligned}\texttt{Telephone} &= \texttt{Telephone Method} + \texttt{Effort} \\ &\quad + \texttt{Telephone Sampling Error}\end{aligned}$$

    is an unbiased estimator of

$$\texttt{Telephone Target} = \texttt{Telephone Method} + \texttt{Effort}$$

  - *design variance* = variance of sampling error can be estimated from the sample (and converted from original scale to log scale)

Sampling Error Models

- Further, sampling error is from within-state stratified sample of moderate to large size
- Assume that

$$
\begin{aligned}
\texttt{Telephone Sampling Error} &\sim \text{independent Normals} \\
&\ \text{with mean zero and variance } \sigma^2_{Tst} \\
\{e^T_{st}\} &\sim \text{independent } \mathcal{N}\left(0, \sigma^2_{Tst}\right)
\end{aligned}
$$

- Further assume that telephone sampling error is independent of

$$
\begin{aligned}
\texttt{Mail Sampling Error} &\sim \text{independent Normals} \\
&\ \text{with mean zero and variance } \sigma^2_{Mst} \\
\{e^M_{st}\} &\sim \text{independent } \mathcal{N}\left(0, \sigma^2_{Mst}\right)
\end{aligned}
$$

Sampling Error Variances

- We have estimates $\widehat{V}_{Tst}$ and $\widehat{V}_{Mst}$ of the design variances V_{Tst} and V_{Mst} on the original scale
  - *not* estimates of σ^2_{Tst} and σ^2_{Mst} on the log scale
  - common approach in this setting is to apply "Taylor linearization" to approximate the variance on the transformed scale
- We have a novel approach for this problem that (unlike Taylor approximation) forces analytical consistency between mean model and variance model

Technical aside: `Sampling Error` Variances, (1)

- Derive theoretical expectation of design variance estimator under mean model:
- $\widehat{V}_{Tst}$ is design-unbiased for

$$
\begin{aligned}
V_{Tst} &= \mathsf{Var}\left(\exp(\widehat{T}_{st}) \mid T_{st}\right) \\
&= \left\{\exp(\sigma^2_{Tst}) - 1\right\} \exp\left\{2T_{st} + \sigma^2_{Tst}\right\}
\end{aligned}
$$

- $\widehat{V}_{Mst}$ is design-unbiased for

$$
\begin{aligned}
V_{Mst} &= \mathsf{Var}\left(\exp(\widehat{M}_{st}) \mid M_{st}\right) \\
&= \left\{\exp(\sigma^2_{Mst}) - 1\right\} \exp\left\{2M_{st} + \sigma^2_{Mst}\right\}
\end{aligned}
$$

Technical aside: Sampling Error Variances, (2)

- Build empirical model for the design variance estimates:

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

for telephone (94.54% adjusted R^2 value); and

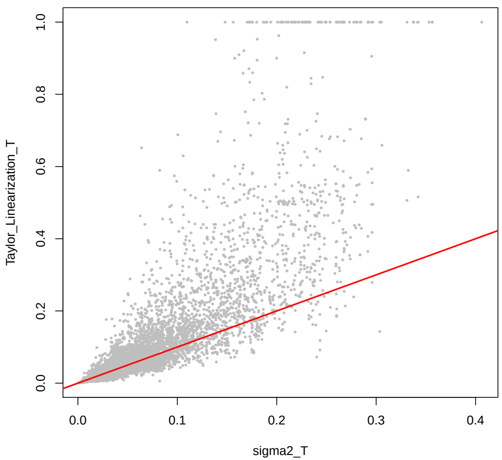
$$\ln\left(\widehat{V}_{Mst}\right) = 2\widehat{M}_{st} + \boldsymbol{d}'_{Mst}\boldsymbol{\delta}_0^M + \delta_1^M \ln(n_{Mst}) + \eta_{st}^M, \quad \eta_{st}^M \sim \mathcal{N}(0, \tau_M^2).$$

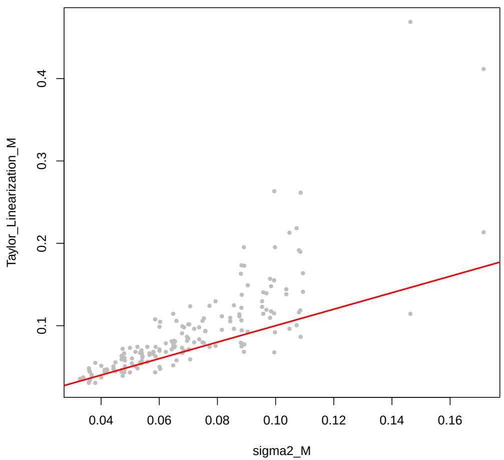
for mail (98.01% adjusted R^2 value)

  - empirical model is potentially useful for stable variance estimates, outside of calibration
  - derive theoretical expectation of design variance under empirical model

Technical aside: Sampling Error Variances, (3)

- Set E [mean model] = E [empirical model] and solve for σ^2_{Tst}, σ^2_{Mst}
  - two quartic equations, each with one real positive, one real negative, and two complex roots
  - result is unique positive solution for σ^2_{Tst}, σ^2_{Mst}
  - treat these as fixed, known design variances in remainder
- Not equivalent to Taylor linearization, but correlated:





Method Effects are Nonsampling Errors

- We have Sampling Error because sampling frame $\neq$ sample
- Method Effects include potential biases due to *Nonsampling Errors:*
  - coverage error: population $\neq$ sampling frame
  - nonresponse error: sample $\neq$ respondents
  - measurement error: true effort $\neq$ measured responses
- Good sampling and measurement protocols minimize Method Effects
- Method Effects may change over time and cannot be entirely disentangled from $\texttt{Effort} = \texttt{Trend} + \texttt{Seasonal} + \texttt{Irregular}$

Confounding of `Method Effects` with `Effort`

- `Telephone` is an unbiased estimator of `Telephone Method`+`Effort`
- But the nonsampling errors in `Telephone Method` could have . . .
  - `trend`: change in quality of frame over time, change in overall response rates over time, change in measurement protocols over time
  - `seasonal`: varying nonresponse by wave, . . .
  - `irregular`: idiosyncratic nonsampling errors from state to state and wave to wave
- Similarly, `Mail` is an unbiased estimator of `Mail Method`+`Effort`
  - `Mail Method` may have its own `trend`, `seasonal`, `irregular`

Confounding of `Method Effects` with `Effort`, II

- We cannot disentangle these `Method Effects` from true `Effort`
- This is a problem in *every survey*, and we try to mitigate it through
  - good frame development and maintenance
  - nonresponse followup and adjustment
  - testing of measurement protocols
  - training of field staff
  - ...
- We cannot estimate `Method Effects` from the sample itself (if we could, we would always estimate and remove it!)

Estimating the difference in method effects

- Model is

$$\begin{aligned}\texttt{Mail} &= \texttt{Mail Method} + \texttt{Effort} \\ &\quad + \texttt{Mail Sampling Error} \\ \texttt{Telephone} &= \texttt{Telephone Method} + \texttt{Effort} \\ &\quad + \texttt{Telephone Sampling Error}\end{aligned}$$

- We cannot disentangle `Mail Method` or `Telephone Method` from `Effort`, but with overlapping estimates,

$$\begin{aligned}\texttt{Mail} - \texttt{Telephone} &= \texttt{Mail Method} - \texttt{Telephone Method} \\ &\quad + \texttt{Mail Sampling Error} \\ &\quad - \texttt{Telephone Sampling Error}\end{aligned}$$

is an unbiased estimator of the *difference* in `Method Effects`

Modeling Mail Method—Telephone Method

- We can estimate the *difference* in Method Effects, given overlap in the surveys
  - limited overlapping data with which to explore the difference
- If we can model the difference, we can extrapolate to other time points that do not have overlapping data:
  - covariates need to be available forward and backward in time
  - covariates need to have explanatory power for difference in Method Effects
- Estimating and extrapolating (Mail Method − Telephone Method) forward and backward allows "calibration" for

$$\texttt{Telephone Target} \rightleftarrows \texttt{Mail Target}$$

Covariates for `Mail Method—Telephone Method`?

- Extrapolation has its usual dangers! Does the model hold over time?
  - if the model does not hold over the full range of time, our calibrated values can be badly wrong
  - assess sensitivity to failure of model stability over time
- Measurement error changing over time? Covariates that explain such a change?
- Nonresponse error changing over time? Covariates that explain such a change?
- Coverage error changing over time? Covariates that explain such a change?
  - *wireless-only households*

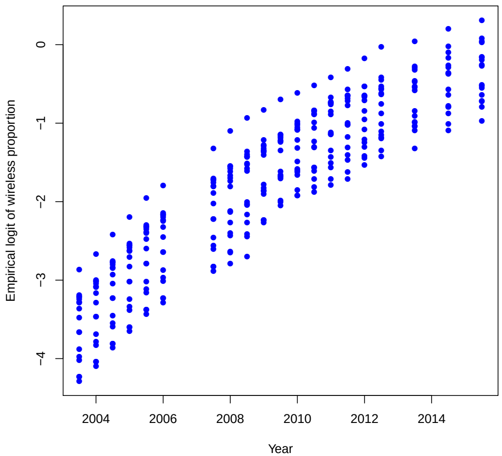
Available wireless-only household data

- From National Health Interview Survey (NCHS), we have June and/or December estimates for each state from 2007–2015
- Estimates are proportion of wireless-only households
- Transform via empirical logits:

$$\text{logit} = \log\left(\frac{\text{proportion wireless-only}}{1 - \text{proportion wireless-only}}\right)$$

Fitting and extrapolating wireless-only household data

- Fit logits as state-specific lines with slope change in 2010:
  `Adjusted R-squared:  0.9948`



- Transform back to proportions and extrapolate forward and backward in time: `wireless`$=\{w_{st}\}$
  - extrapolated proportion is approximately zero prior to 2000

Incorporating `wireless`

- `wireless` and its interactions with `state`, `wave`, `log(pop)` and interactions help to explain some variation
- `wireless` is highly significant statistically: strong evidence that it should not be dropped from model
- But practical effect is less pronounced:

	R^2_{adj}	Residual SE	df
Shore all without wireless	0.841	0.544	2869
Shore all with wireless	0.856	0.518	2761
Boat all without wireless	0.878	0.493	2871
Boat all with wireless	0.896	0.455	2763

Model for Mail Method−Telephone Method

- Includes wireless and its interactions with trend and seasonal factors
- Extrapolates sensibly in time
  - extrapolates back in time to zero in every state around year 2000
  - extrapolates (eventually) to one forward in time
- Other than the effect of wireless and its interactions, we allow for other differences in measurement error and write

$$\texttt{Mail Method}-\texttt{Telephone Method} = \boldsymbol{b}'_{st}\boldsymbol{\mu} - w_{st}\boldsymbol{c}'_{st}\boldsymbol{\gamma}$$

  - we approach this empirically but parsimoniously
  - note that if we choose $\boldsymbol{b}_{st} = \mathbf{0}$, then Mail Method−Telephone Method$= 0$ in the past, meaning don't calibrate

The combined model

- Recap on the model

$$
\begin{aligned}
\texttt{Telephone} &= \texttt{Telephone Method} + \texttt{Effort} \\
&\quad + \texttt{Telephone Sampling Error} \\
&= \texttt{Telephone Target} + \texttt{Telephone Sampling Error} \\
\texttt{Mail} &= \texttt{Mail Method} + \texttt{Effort} + \texttt{Mail Sampling Error} \\
&= \texttt{Mail Target} + \texttt{Mail Sampling Error}
\end{aligned}
$$

- We know a lot about both `Sampling Error` terms
- We can estimate and model `Telephone Target` and `Mail Target`
- Inside that model is ($\texttt{Mail Method} - \texttt{Telephone Method}$), where the biggest assumptions lie

Notation

- $\widehat{T}_{st}$ = natural log of telephone effort estimate in state s, year-wave t
- T_{st} = `Telephone Target`
- e^T_{st} = `Telephone Sampling Error`, $e^T_{st} \sim \mathcal{N}(0, \sigma^2_{Tst})$
- $\widehat{M}_{st}$ = natural log of mail effort estimate in state s, year-wave t
- M_{st} = `Mail Target`
- e^M_{st} = `Mail Sampling Error`, $e^M_{st} \sim \mathcal{N}(0, \sigma^2_{Mst})$
- ν_{st} = `Irregular`, $\nu_{st} \sim \mathcal{N}(0, \psi)$

- Putting it all together:

$$
\begin{aligned}
\widehat{T}_{st} &= T_{st} + e^T_{st}, \quad e^T_{st} \sim \mathcal{N}(0, \sigma^2_{Tst}) \\
T_{st} &= \boldsymbol{a}'_{st}\boldsymbol{\alpha} + 0 \cdot \boldsymbol{b}'_{st}\boldsymbol{\mu} + w_{st}\boldsymbol{c}'_{st}\boldsymbol{\gamma} + \nu_{st} \\
&= [\boldsymbol{a}'_{st}, \mathbf{0}', w_{st}\boldsymbol{c}'_{st}]\,\boldsymbol{\beta} + \nu_{st} \\
&= \boldsymbol{x}'_{Tst}\boldsymbol{\beta} + \nu_{st} \\
\widehat{M}_{st} &= M_{st} + e^M_{st}, \quad e^M_{st} \sim \mathcal{N}(0, \sigma^2_{Mst}) \\
M_{st} &= \boldsymbol{a}'_{st}\boldsymbol{\alpha} + 1 \cdot \boldsymbol{b}'_{st}\boldsymbol{\mu} + 0 \cdot \boldsymbol{c}'_{st}\boldsymbol{\gamma} + \nu_{st} \\
&= = \left[\boldsymbol{a}'_{st}, \boldsymbol{b}'_{st}, \mathbf{0}'\right]\boldsymbol{\beta} + \nu_{st} \\
&= \boldsymbol{x}'_{Mst}\boldsymbol{\beta} + \nu_{st},
\end{aligned}
$$

where $\boldsymbol{\beta}' = [\boldsymbol{\alpha}', \boldsymbol{\mu}', \boldsymbol{\gamma}']$

- This is closely related to the *Fay-Herriot model* of survey statistics

Using the Fay-Herriot approach

- It is convenient to write

$$
\begin{aligned}
Y_{st} &= \begin{cases} \widehat{T}_{st}, & \text{if no mail estimate is available;} \\ \widehat{M}_{st}, & \text{if no telephone estimate is available;} \\ \left(\widehat{T}_{st} + \widehat{M}_{st}\right)/2, & \text{otherwise;} \end{cases} \\
&= \begin{cases} \boldsymbol{x}'_{Tst}\boldsymbol{\beta} + \nu_{st} + e^{T}_{st}, & \text{if no mail;} \\ \boldsymbol{x}'_{Mst}\boldsymbol{\beta} + \nu_{st} + e^{M}_{st}, & \text{if no telephone;} \\ (\boldsymbol{x}_{Tst} + \boldsymbol{x}_{Mst})'\boldsymbol{\beta}/2 + \nu_{st} + (e^{T}_{st} + e^{M}_{st})/2, & \text{otherwise;} \end{cases} \\
&= \boldsymbol{x}'_{st}\boldsymbol{\beta} + \nu_{st} + e_{st}.
\end{aligned}
$$

- Disadvantage: small loss of information due to averaging
- Advantage: calibration methodology is *exactly* an application of Fay-Herriot

Fay-Herriot small area estimation approach

- R.E. Fay III and R.A. Herriot. "Estimates of income for small places: an application of James-Stein procedures to census data." *Journal of the American Statistical Association* (1979): 269–277.
  - standard and well-studied methodology for *small area estimation*
  - cited 1400+ times in Google Scholar
  - built on powerful estimation and prediction techniques
  - supported by theory, including for mean square error estimation
  - supported by software: `sae` package in `R`

Estimation for Fay-Herriot

- If ψ were known, estimate $\boldsymbol{\beta}$ via best linear unbiased estimator (BLUE)

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

  where

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

  and $D_{st} =$ variance of sampling error

- Since ψ is not known, replace it by a consistent estimator to obtain

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

- We use REstricted Maximum Likelihood (REML) estimator $\widehat{\psi}$

Prediction examples

- What would be the `Mail Target` equivalent for state s and past year-wave t, when no `Mail` estimate is available?

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

- What would be the `Telephone Target` equivalent for state s and future year-wave t, with no `Telephone` estimate?

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

Prediction examples, continued

- What would be the `Telephone Target` equivalent for state s and past or future year-wave t, with or without `Telephone` estimate, adjusted for `wireless`?

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

- In each case, prediction involves a new set of covariates

$$\phi_{st} = \boldsymbol{z}'_{st}\boldsymbol{\beta} + \nu_{st}$$

instead of the original covariates

$$\boldsymbol{x}'_{st}\boldsymbol{\beta} + \nu_{st}$$

Prediction for Fay-Herriot

- If $\boldsymbol{\beta}$ and ψ were known, best mean square predictor under normality would be

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

- If only ψ was known, use BLUE of $\boldsymbol{\beta}$ to construct best linear unbiased predictor (BLUP)

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

- If neither is known, then use empirical best linear unbiased predictor (EBLUP):

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

Summary of Fay-Herriot methodology

- Estimate ψ and $\boldsymbol{\beta}$
- Predict various unknown quantities via Empirical Best Linear Unbiased Prediction (EBLUP)
- Also need a measure of uncertainty:
  - approximate mean squared error (MSE) of resulting EBLUP's
  - estimate mean squared error of resulting EBLUP's
- Transform EBLUP's back to original scale

- Adapts Datta and Lahiri (2000):

$$
\begin{aligned}
\mathsf{MSE}\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}},\hat{\psi}\right)\right\} &= \mathsf{E}\left[\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}},\hat{\psi}\right)-\phi_{st}\right\}^2\right] \\
&= \mathsf{E}\left[\left\{\phi_{st}\left(\widetilde{\boldsymbol{\beta}}_{\psi},\psi\right)-\phi_{st}\right\}^2\right]+\mathsf{E}\left[\left\{\phi_{st}\left(\boldsymbol{\beta},\psi\right)-\phi_{st}\left(\widetilde{\boldsymbol{\beta}}_{\psi},\psi\right)\right\}^2\right] \\
&\quad +\mathsf{E}\left[\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}},\hat{\psi}\right)-\phi_{st}\left(\boldsymbol{\beta},\psi\right)\right\}^2\right] \\
&= \dot{g}_{1st}(\psi)+\dot{g}_{2st}(\psi)+\dot{g}_{3st}(\psi)+o\left(m^{-1}\right),
\end{aligned}
$$

where

$$
\dot{g}_{1st}(\psi)=\frac{\psi D_{st}}{\psi+D_{st}},
$$

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

and

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

Technical aside: MSE estimation

- Adapting Datta and Lahiri (2000), it can be shown that

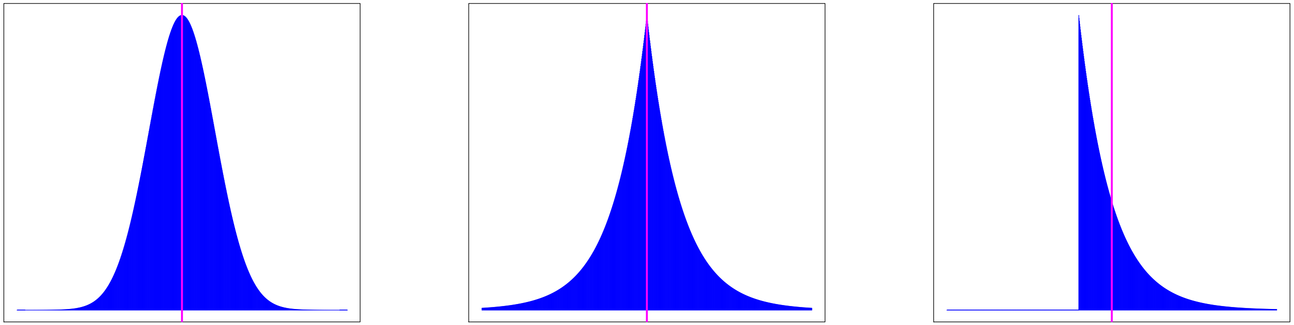
$$
\begin{aligned}
\mathsf{E}\left[\dot{g}_{1st}(\hat{\psi})\right] &\simeq \dot{g}_{1st}(\psi) - \dot{g}_{3st}(\psi) \\
\mathsf{E}\left[\dot{g}_{2st}(\hat{\psi})\right] &\simeq \dot{g}_{2st}(\psi) \\
\mathsf{E}\left[\dot{g}_{3st}(\hat{\psi})\right] &\simeq \dot{g}_{3st}(\psi)
\end{aligned}
$$

- Hence an approximately unbiased estimator of the MSE approximation is

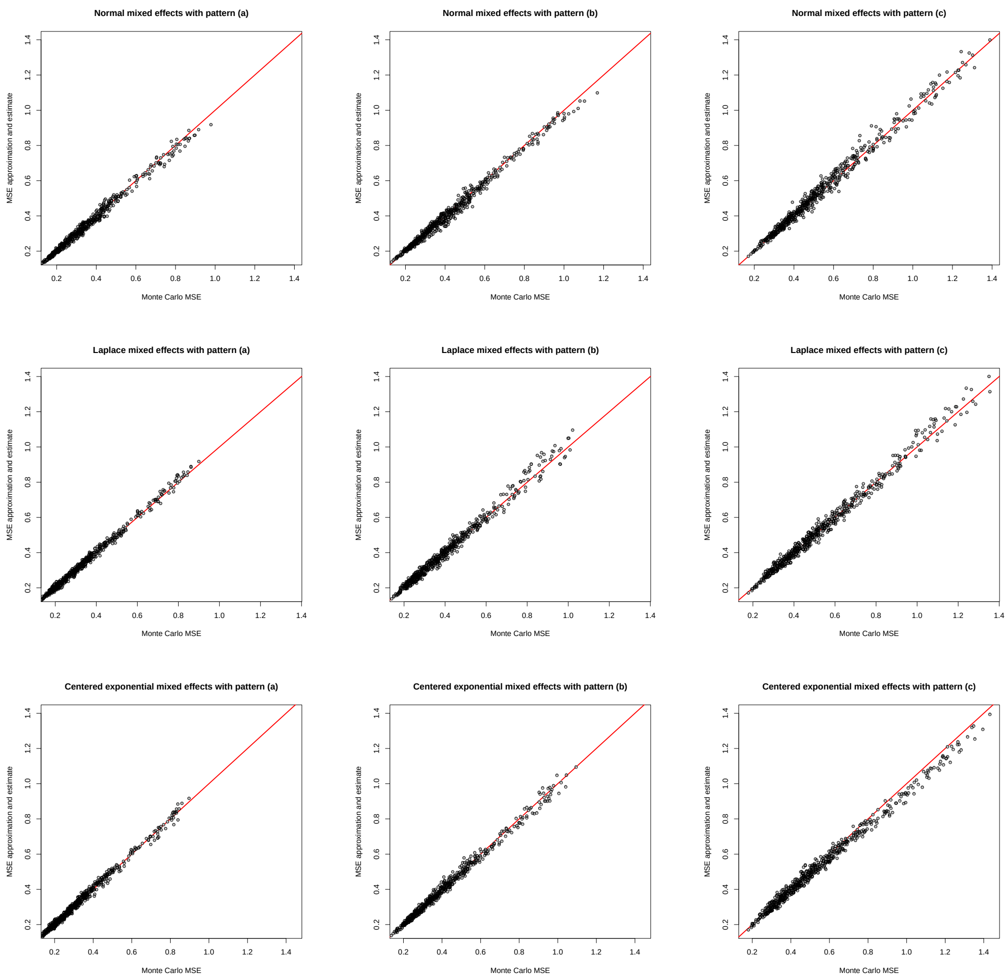
$$
\mathsf{mse}\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\} = \dot{g}_{1st}(\hat{\psi}) + \dot{g}_{2st}(\hat{\psi}) + 2\dot{g}_{3st}(\hat{\psi})
$$

Assess MSE approximation/estimation via simulation

- 17 states and six years, `Telephone` in all years, `Mail` in final two
- Model similar to final fitted model
- Three design variance patterns:
  - pattern (b): sample actual design variances, arrange in seasonal cycle, and replicate across years
  - pattern (a) is $(1/2)\times$(b) and pattern (c) is $2\times$(b)
- Three densities for ν_{st}: normal, Laplace, centered exponential



Simulation MSE (1000 replications) vs. estimates



Prediction on the original scale

- MSE approximation and estimator work well in simulation
  - robust to non-normality of $\{\nu_{st}\}$
- Estimator of MSE is used in the back transformation to the original effort scale:

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

Choosing a model

- Smallest model: drops $\boldsymbol{b}'_{st}\boldsymbol{\mu}$ and $w_{st}\boldsymbol{c}'_{st}\boldsymbol{\gamma}$, keeps only

$$\texttt{log(pop)} + \texttt{wave} + \texttt{state} * (1 + \texttt{log(pop)} + \texttt{wave})$$

- Largest model: adds $\boldsymbol{b}'_{st}\boldsymbol{\mu}$

$$\texttt{mail} * (1 + \texttt{state} + \texttt{log(pop)} + \texttt{wave})$$

  and adds $w_{st}\boldsymbol{c}'_{st}\boldsymbol{\gamma}$

$$\texttt{wireless}*(1 + \texttt{state} + \texttt{log(pop)} + \texttt{wave} + \texttt{state*log(pop)})$$

  to smallest model

- 80 models in the suite: all $2^7 = 128$ models between (smallest+`mail`+`wireless`) and largest, but dropping models like

$$\texttt{wireless}*(1 + \cancel{\texttt{state}} + \texttt{log(pop)} + \texttt{wave} + \texttt{state*log(pop)})$$

Comparing models in the suite

- AIC = Akaike's Information Criterion
  - smaller AIC means better expected out-of-sample prediction
  - rewards models that fit observed data well
  - penalizes models with too many parameters
- Out-of-sample prediction MSE
  - predict 2018:W1–W2 for all available states (not used in fit)
  - smaller MSE means better predictions
- df = degrees of freedom
  - larger df means fewer estimated parameters

Models ordered on AIC: shore fishing

Model is largest minus terms below:	MSE	AIC	df
mail:log(pop), mail:state, wireless:wave	0.0837	4564.28	3022
mail:state, wireless:wave	0.0899	4564.69	3021
mail:log(pop) and wireless:wave	0.1350	4564.86	3006
wireless:wave	0.1354	4566.85	3005
mail:log(pop) and mail:state	0.0840	4570.45	3017
nothing (largest)	0.1343	4573.28	3000
mail interactions	0.2104	4580.51	3022
wireless interactions	0.3694	4719.05	3038
all interactions	0.3341	4742.84	3050
all wireless	0.4745	4758.73	3029
all mail	1.9466	4838.73	3023
all mail and all wireless (smallest)	2.7443	5106.70	3052

Models ordered on AIC: private boat fishing

Model is largest minus terms below:	MSE	AIC	df
nothing (largest)	0.2068	3314.55	2990
mail:log(pop)	0.2124	3314.56	2991
mail:log(pop) and wireless:wave	0.2163	3316.42	2996
wireless:wave	0.2241	3316.47	2995
mail:log(pop) and mail:state	0.2050	3322.73	3007
mail:state	0.1910	3323.00	3006
mail interactions	0.2272	3362.27	3012
all mail	0.7046	3501.23	3013
wireless interactions	0.4004	3520.33	3028
all interactions	0.4615	3646.78	3050
all wireless	0.5421	3750.03	3029
all mail and all wireless (smallest)	1.2677	3901.82	3052

Model selection

- Same model for both fishing modes is desirable
- Highly parsimonious model is desirable
- Among the three models that are AIC-best in *both* fishing modes, we chose the model dropping mail:log(pop) and mail:state
  - competitive AIC
  - smaller out-of-sample prediction mean square error
  - more parsimonious (greater df due to fewer parameters)
- Selected model implies past adjustments (pre-wireless) are just level shifts for each wave (on log scale):

$$\texttt{mail} * (1 + \cancel{\texttt{state}} + \cancel{\texttt{log(pop)}} + \texttt{wave})$$

Sensitivity to assumptions

- For calibration to the distant past, the methodology requires extrapolation over a long time window
- Model assumes stable differences between `Telephone` and `Mail` (outside of `wireless` effects) over that window
- If the differences are not stable, anything is possible!
  - one fairly extreme case is that `Mail`$\rightarrow$`Telephone` in the distant past

Summary

- Models account for various sources of variation, including `Trend`, `Seasonality`, `Irregular`, `Sampling Error`, and non-sampling `Method Effects`
  - model assumes measurement and nonresponse differences between the surveys are stable over time
  - model assumes coverage error has changed over time due to growth in wireless-only households
- As formulated, calibration methodology turns out to follow a standard, well-established procedure: *Fay-Herriot small area estimation*
- Yields optimal predictions = calibrated values, under the assumptions of the model

Acknowledgements

- Thanks to NOAA Fisheries scientists Rob Andrews and John Foster for assistance with problem formulation and all aspects of data compilation
- Also thanks to NOAA affiliate Ryan Kitts-Jensen and Westat statistician Mike Brick for useful discussion

Thank you!

Questions?