A Calibration Methodology for CHTS to FES Transition

Jay Breidt
Colorado State University

Joint work with Teng Liu and Jean Opsomer
Colorado State University

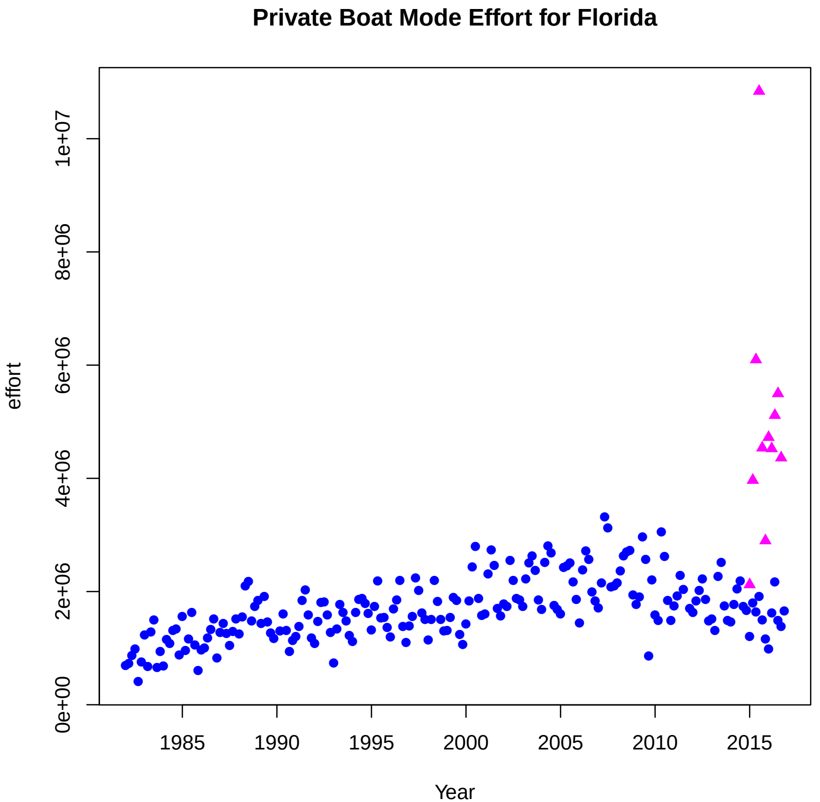
The problem

- Segal's Law: *A person with a watch knows what time it is. A person with two watches is never sure.*

- What about two surveys?

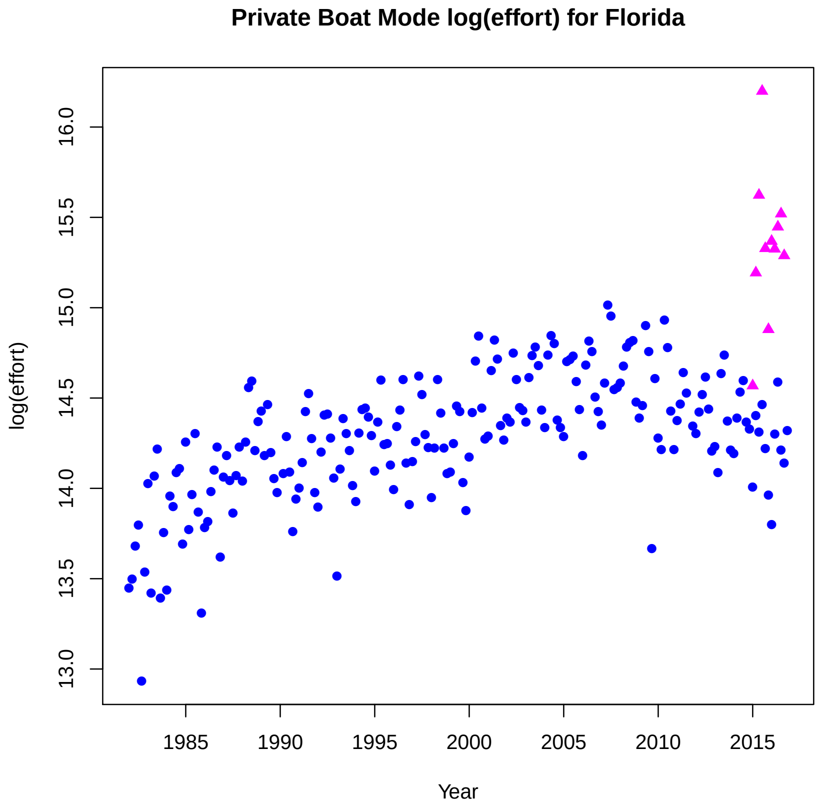
Florida private boats, original scale

- Available FES (Mail) effort estimates are consistently much higher than CHTS (Telephone)
- Limited number of overlapping waves (currently 2015:W1–2016:W5)



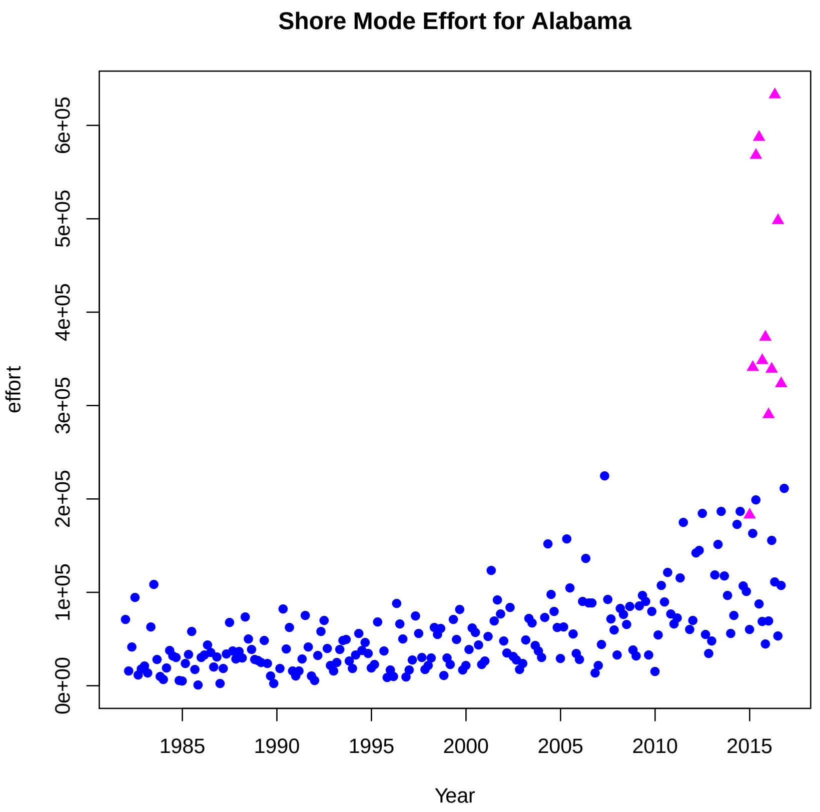
Florida private boats, log scale

- Available FES (Mail) effort estimates are consistently much higher than CHTS (Telephone)
- Limited number of overlapping waves (currently 2015:W1–2016:W5)



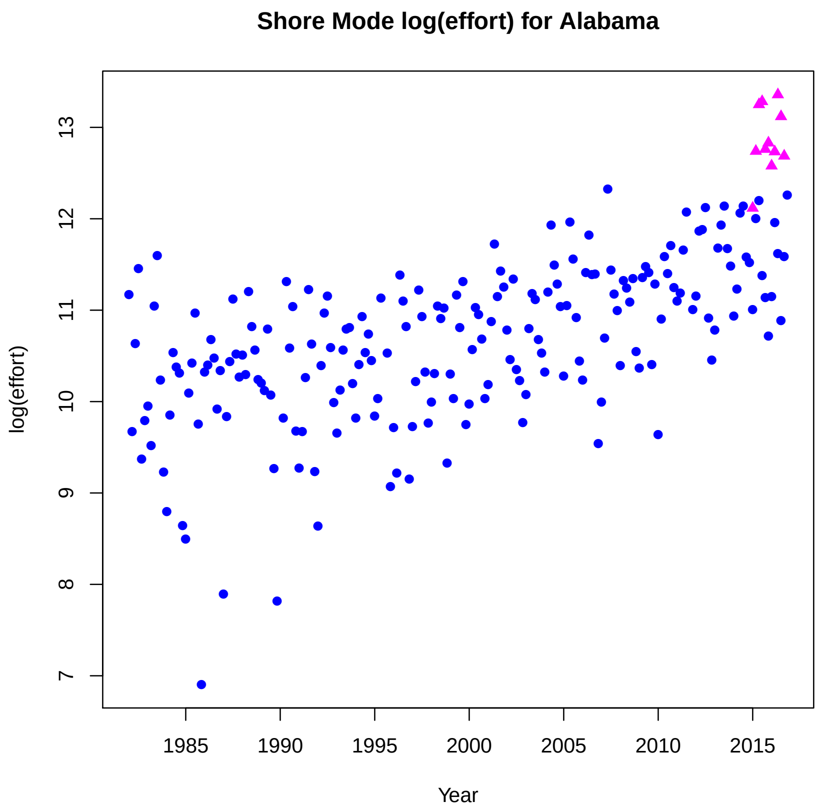
Alabama shore fishing, original scale

- Available FES (Mail) effort estimates are consistently much higher than CHTS (Telephone)
- Limited number of overlapping waves (currently 2015:W1–2016:W5)



Alabama shore fishing, log scale

- Available FES (Mail) effort estimates are consistently much higher than CHTS (Telephone)
- Limited number of overlapping waves (currently 2015:W1–2016:W5)

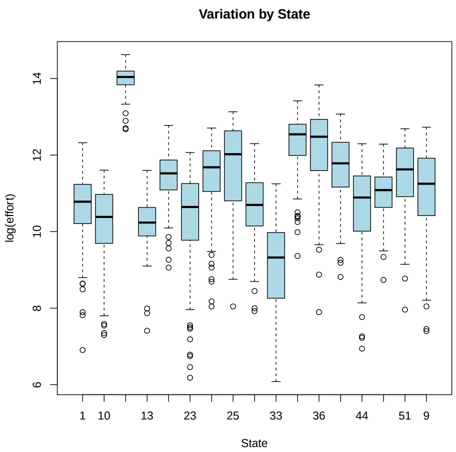


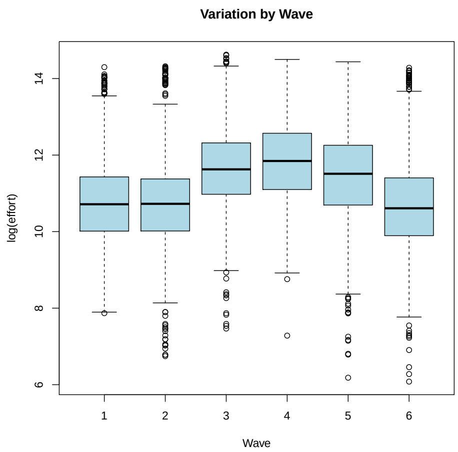
Calibration problem

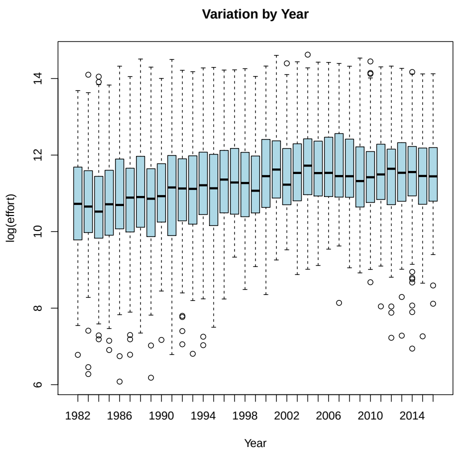
- Is there a way to convert from `Telephone` "units" to `Mail` "units" and vice versa?
- No judgement that one method is correct or even better: they are just different
- Want a defensible statistical approach, realizing that it will have to rely on some modeling assumptions

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 variation depends on 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"

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

for each state's effort series

- We do not observe `Effort` directly, but with `Sampling Error` and with `Method Effect`

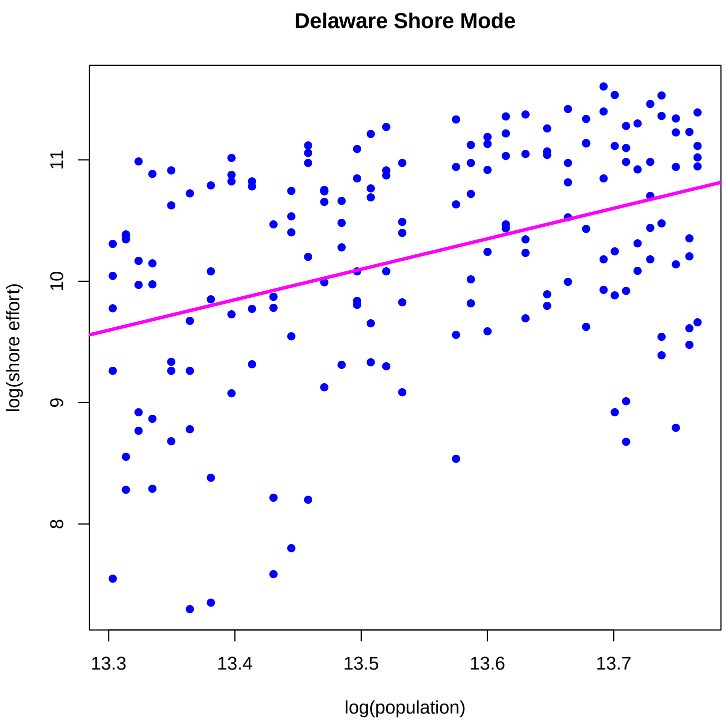
- Log-scale estimates can be written

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

- 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

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

- 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

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

- Further assume that telephone sampling error is independent of

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

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

Aside: Sampling Error Variances

- Derive theoretical expectation of design variance under mean model
- 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); similar model for mail (98.01% adjusted R^2 value)

  - empirical model is potentially useful for stable variance estimates, outside of calibration

- Set theoretical = empirical 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 solutions for σ^2_{Tst}, σ^2_{Mst}
  - treat these as fixed, known design variances in remainder

`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 responses $\neq$ measured responses
- Good sampling and measurement protocols minimize `Method Effects`
- `Method Effects` may change over time and cannot be entirely disentangled from `Effort=Trend+Seasonal+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 explain difference in Method Effects, not other sources of variation
- Estimating and extrapolating ($\texttt{Mail Method} - \texttt{Telephone Method}$) 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 2007–2014
- Estimates are proportion of wireless-only households
- Transform via empirical logits:

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

- 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
- Given the lack of other covariates to explain changes in `Method Effects`, all else is just a level shift:

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

  - e.g., any difference in measurement error between the two methods is assumed constant over time
  - another "extreme" is `Mail Method`−`Telephone Method`$=0$ in 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`, $\nu_{st} \sim \mathcal{N}(0, \psi)$

Complete model specification

- Putting it all together:

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

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

- In survey statistics, this is called the *Fay-Herriot model*

Fay-Herriot model

- 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 1000+ times in Google Scholar
  - built on powerful estimation and prediction techniques
  - supported by software: `sae` package in `R`
- With our formulation, calibration methodology is *exactly* an application of Fay-Herriot

Fay-Herriot approach

- Estimate ψ via REstricted Maximum Likelihood estimation (REML)
- Estimate $\boldsymbol{\beta}$ via Maximum Likelihood Estimation (MLE)
- Predict various unknown quantities via Empirical Best Linear Unbiased Prediction (EBLUP)
  - what would be the `Mail Target` equivalent for state s and past year-wave t, when no `Mail` estimate is available?

    $$\text{EBLUP}(M_{st}) \text{ combines all available information}$$

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

    $$\text{EBLUP}(T_{st}) \text{ combines all available information}$$

- Estimate mean squared error of resulting EBLUP's

EBLUP alternatives

- No adjustment at all: $\widehat{T}_{st}$
- Simple calibration from `Telephone Target` to `Mail Target`:

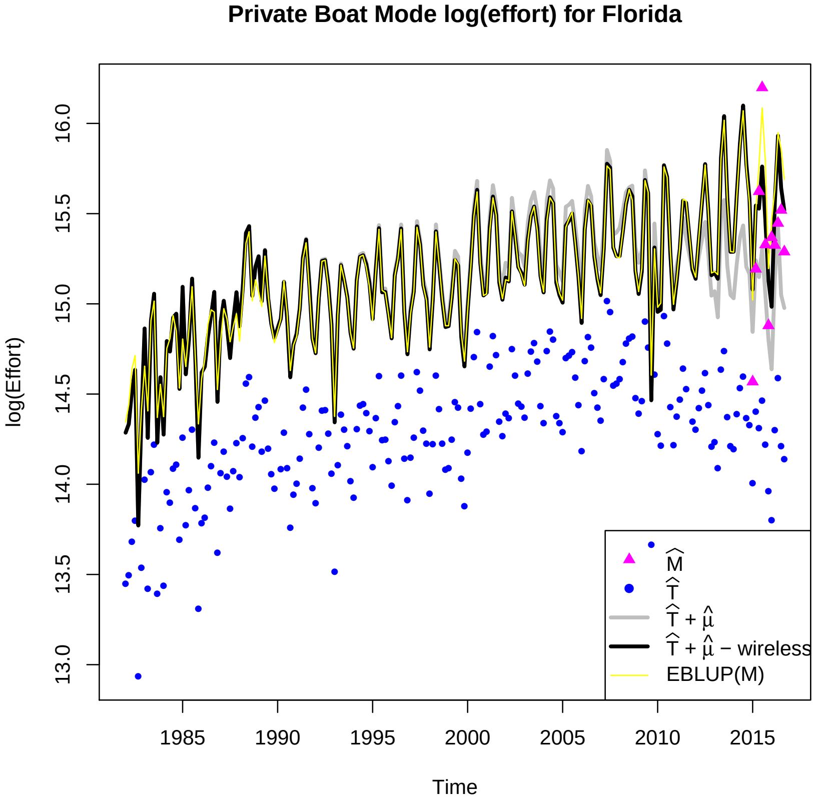
$$\texttt{Telephone} + \left(\texttt{Mail Method} - \texttt{Telephone Method}\right)$$

  – with level shift only: $\widehat{T}_{st} + \widehat{\mu}$
  – with level shift and wireless, $\widehat{T}_{st} + \widehat{\mu} - w_{st}\boldsymbol{b}'_{st}\widehat{\boldsymbol{\gamma}}$
  – sub-optimal, but simple and useful for comparison
- Empirical Best Linear Unbiased Predictor:

$$\mathsf{EBLUP}(M_{st}) = \widehat{\mu} + \boldsymbol{a}'_{st}\widehat{\boldsymbol{\alpha}} + \widehat{\nu}_{st}$$

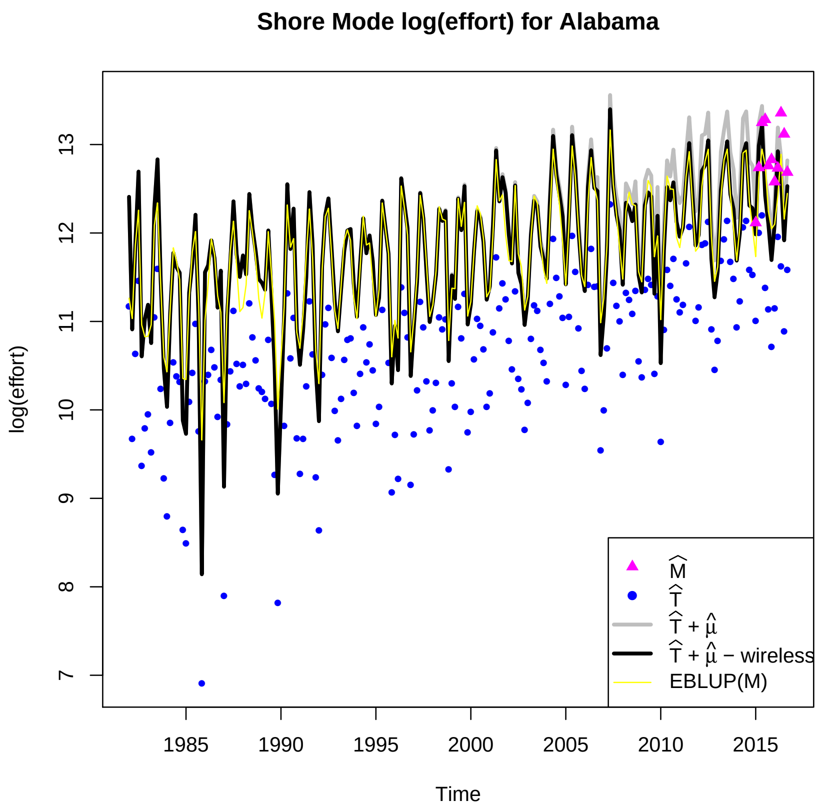
Florida private boats, log scale

- `Telephone` estimates calibrated to `Mail Target`



Alabama shore fishing, log scale

- `Telephone` estimates calibrated to `Mail Target`



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

Thank you!

Questions?