Predicting Lag Between First Attempt and First Contact Using Partial Data vs a Bayesian Approach

Stephanie Coffey, Michael Elliott, James Dahlhamer

5th Workshop on Adaptive and Responsive Survey Design

Overview

- Monitoring Metrics During Data Collection
- Illustrative Metric
  - Lag between first attempt and first contact
- Data Sources
- Methods for Estimation
- Methods for Comparison and Evaluation
- Results
- Discussion

Monitoring Field Metrics

- Monitoring relies on expectations
  - Metrics related to progress, cost, or quality
  - Short term monitoring to ensure current data collection is "on track"
    - Is workload 50% resolved halfway through survey period?
    - Is the cooperation rate similar to last implementation?
    - Is cost per case roughly as expected?
  - Long term monitoring to understand trends
    - Is the refusal rate increasing?
    - Is the cost or time spent to obtain a complete increasing?

Monitoring Field Metrics

- Metrics are monitored against expectations
- If current metrics diverge from expectations, can intervene:
  - Operational updates
    - Feedback to interviewers
    - Reassignment of cases or workload
  - General change in procedures
    - Limit contact attempts to control costs
  - Start of new data collection operation
    - Subsampling
- Expectations might come from:
  - Past implementations of the same survey
  - Similar surveys
  - External literature

Monitoring Field Metrics

- Accurate expectations can lead to effective interventions
  - Balance detail with usability
  - Reflect what should occur during data collection
  - Account for what is currently happening
- Currently used methods often:
  - Rely heavily on historical information
  - Are very high level (top line estimates)
  - Accounting for current information is *ad hoc*
- Want to combine historical data and current information
  - A natural approach is to use Bayesian methods; prior incorporates historical information.

Example Metric: Lag Between 1st Attempt and Contact

- Interested in the expected time lag between the first attempt and first contact
- Without contact in a CAPI survey:
  - Interview cannot occur
  - Eligibility may remain unknown
  - Affects response rates, nonresponse adjustments
- Noncontact can be a significant part of nonresponse
- Goal:
  - Want to predict expected lag – what is a reasonable time frame for interviewers?
  - Can provide feedback to interviewers, or adjust workload

Example Data:
National Health Interview Survey

- Large, national survey on health outcomes in the US
  - Sponsored by the National Center for Health Statistics
  - Fielded by Census Bureau
  - Cross-sectional monthly samples, n ≈ 6,000
  - Interviewer administered, in-person only
- Progress metrics used for monitoring:
  - Resolution rate (may not be response)
  - Productivity rate (cost/hours per complete)
- Data available: 4 Months of NHIS Data (February – May, 2016)
  - Sampling frame is mostly basic geographic information
  - Enriched frame with the planning database (PDB)
  - Interviewer characteristics from employment data
  - Derived case reassignment indicator from operational data
  - Estimated the lag (in days) using contact history instrument (CHI)

Historical Mean Lag (3 Months)

- Benefit:
  - Intuitive
  - Simple number that can be easily compared to current data
- Drawbacks:
  - All cases are roughly “the same”
  - Expected lag is the same regardless of when case is first attempted during data collection

Month	All Contacted Cases		All Contacted Cases with Lag > 0	
	Cases	Mean Lag (SD)	Cases	Mean Lag (SD)
02/2016	6059	3.40 (5.89)	2376	8.68 (6.53)
03/2016	6627	3.43 (6.21)	2565	8.87 (7.16)
04/2016	6085	3.41 (5.92)	2396	8.66 (6.60)
3 Month	18771	3.42 (6.01)	7337	8.74 (6.78)

Historical Mean Lag (3 Months)

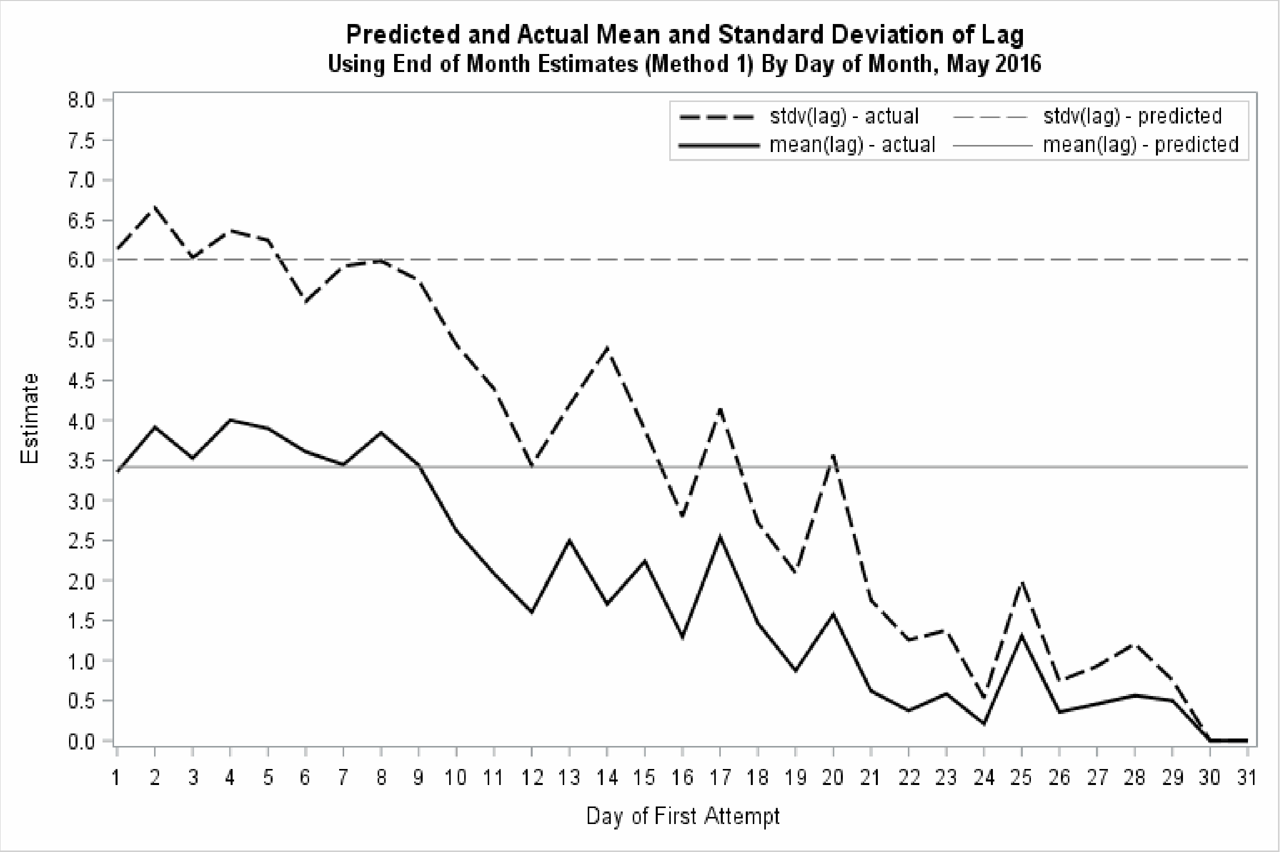
- Benefit:
  - Intuitive
  - Simple number that can be easily compared to current data
- Drawbacks:
  - All cases are roughly “the same”
  - Expected lag is the same regardless of when case is first attempted during data collection

Month	All Contacted Cases		All Contacted Cases with Lag > 0	
	Cases	Mean Lag (SD)	Cases	Mean Lag (SD)
02/2016	6059	3.40 (5.89)	2376	8.68 (6.53)
03/2016	6627	3.43 (6.21)	2565	8.87 (7.16)
04/2016	6085	3.41 (5.92)	2396	8.66 (6.60)
3 Month	18771	3.42 (6.01)	7337	8.74 (6.78)

Day of 1st Attempt	April 2016
	Mean Lag (SD)
1 – 7	3.83 (6.32)
8 – 14	2.73 (5.00)
15 – 21	1.42 (3.14)
22+	0.57 (1.39)

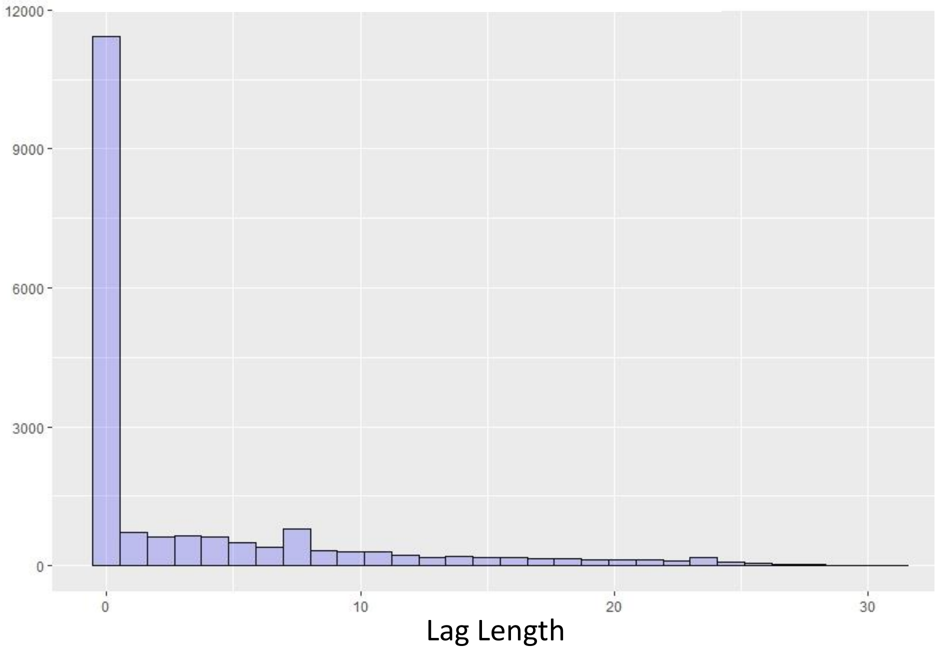
Historical Mean vs Current Lag

- Mean is a reasonable expectation early, but not after Day 8
- Later in field period, there is less time and the lag has to be shorter
- Standard deviation not very useful for constructing confidence intervals
  - 90% = +/- 10 days
  - 95% = +/- 12 days

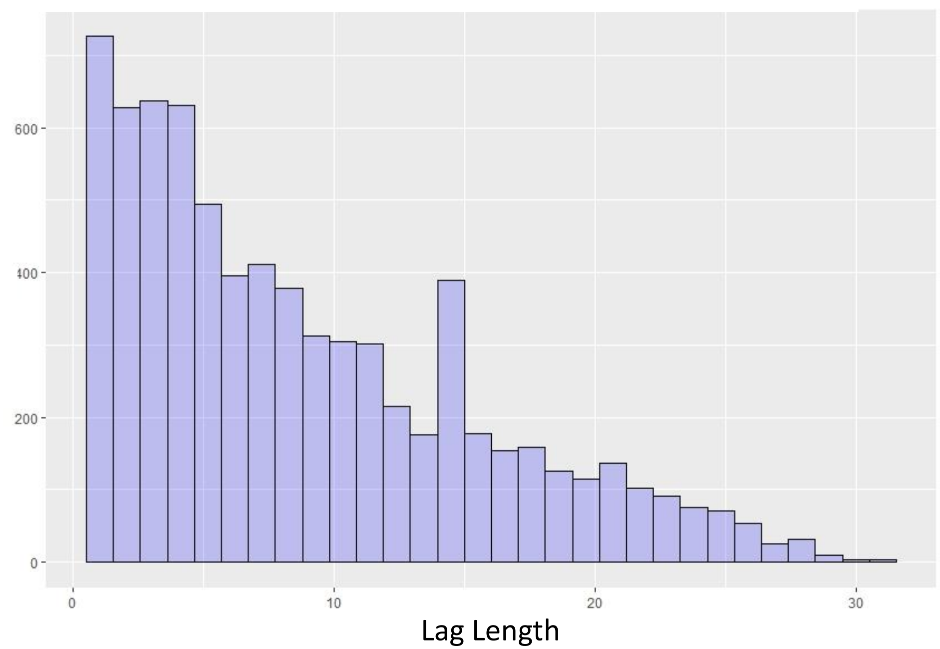


Distribution of Lags

Distribution of Lags (Historical Data)



Distribution of Nonzero Lags (Historical Data)



- Significant number of zero-lag cases (~60%)
- For lags > 0, nonlinear relationship between lag length and frequency
- The single mean value does not reflect the time-varying nature of lag length

Model-Based Metrics

- Benefit: Can create a more tailored prediction
  - Include time-varying covariates
  - Include covariates that provide geographic nuance
  - Include covariates about interviewers characteristics
- Drawbacks
  - More complex than a mean
  - Concern about misspecification
- Model needs to accommodate:
  - Zero-length lags (60% of cases)
  - Cases where the attempt happened but the contact has not (right censored cases)
- Selected: Weibull Hurdle Model
  - Hurdle allows prediction of zeroes
  - Weibull is a right-skewed distribution that easily accommodates right censoring

Weibull Hurdle Model

- Combination of two models
  - Prediction for likelihood of crossing the hurdle
    - Hurdle = Not making contact on 1st attempt
  - Prediction of survival time, given the hurdle is crossed

$p(y_i) = \begin{cases} (1-\pi_i) \\ (\pi_i) g(y_i) \end{cases}$, $g(y_i)$ is the Weibull distribution, where:

$\pi_i = \frac{\rho_i}{(1+\rho_i)}$ is the logistic link function

$\rho_i = \exp(\boldsymbol{\gamma}'\boldsymbol{z_i})$

$g(y_i) = \frac{\alpha}{\lambda_i^{\alpha}} y_i^{\alpha-1} e^{-\left(\frac{y_i}{\lambda_i}\right)^{\alpha}}$, $y_i \geq 0$ where:

α is the shape parameter of the Weibull distribution

$\lambda_i = \exp(\boldsymbol{\beta}'\boldsymbol{x_i})$ is the inverse link and is the scale parameter

$\hat{y}_i = (\pi_i) g(y_i)$ is the estimate of the outcome using the estimated parameters

Covariates Used for Models

Lag Between Initial Attempt and Initial Contact	Dependent Variable
Regional Office (RO)	Highest Level of Field Organization at Census: 6 Levels for the US
Interviewer Employment Tenure	Whether the FR was new or established at the Bureau: 0-1 year, 2+ years
Interviewer Education Level	The highest education completed by the interviewer
Reassignment Indicator	Indicator to identify cases reassigned between initial attempt and initial contact
Reassignment Date	Date of the reassignment of a case between interviewers
Calday	The numbered day of the month when the first attempt was made on the case
Block Group PDB Variables	
% Mobile Homes	Continuous Variable (0 to 100)
% Not HS Graduates	Continuous Variable (0 to 100)
% Without Health Insurance	Continuous Variable (0 to 100)
% Urbanized Population	Continuous Variable (0 to 100)
% Vacant Units	Continuous Variable (0 to 100)

Methods for Estimation

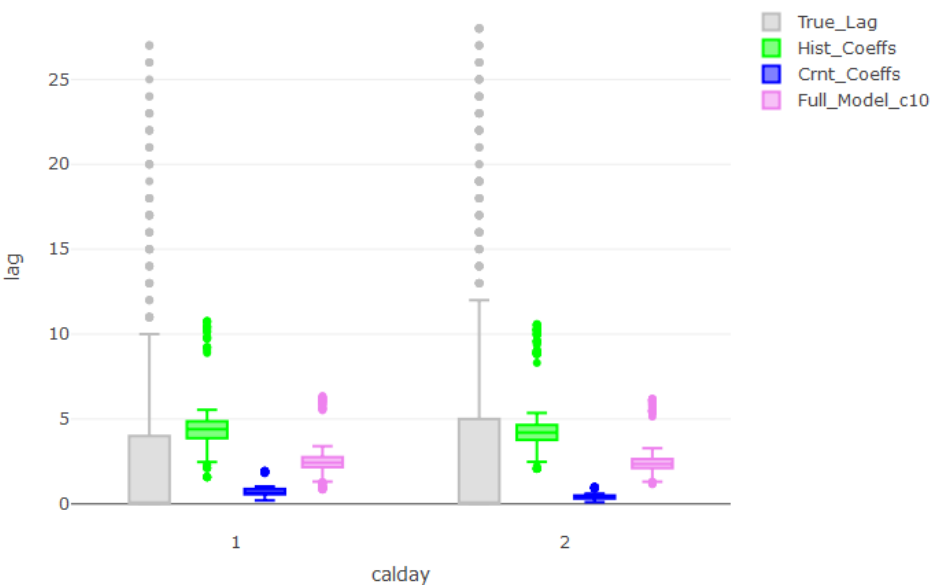
- Historical Mean
  - Use historical data to estimate average expected lag for cases
- Model-based prediction #1 – Historical Data Only
  - Use historical data to estimate coefficients
  - Predict lag on current sample dataset
- Model-based prediction #2 – Current Data Only
  - Use current round of data collection only to estimate coefficients and predict lag
- Model-based prediction #3 – Combination
  - Bayesian framework – full model
  - Use historical data to develop priors for model coefficients
  - Update predictions as current round data is aggregated to update expected lags

Methods for Evaluation

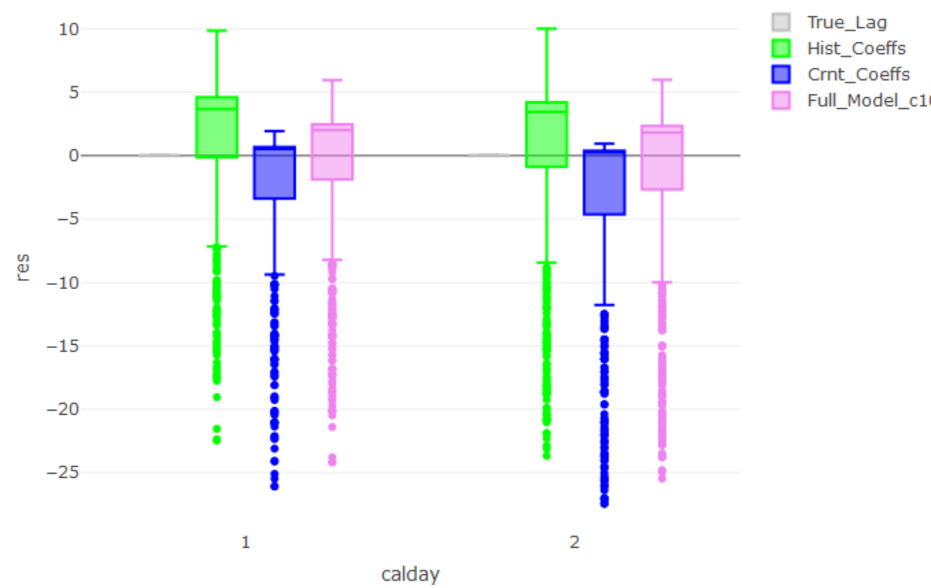
- For each of the four methods:
  - Boxplots of Actual Lags vs. Predicted Lags
  - Boxplots of Residuals for Four Methods
  - Variance, Bias, and Mean Squared Error
- Predictions Change Over Time
  - Different cases are in the model at different times
  - Estimates of Lag change over time
  - Look at results at three points in time:
    - Day 5, 15, 25
- Change Strength of Prior
  - c determines how strong the prior is
  - Use c = 2 (1/2 month); c = 1 (1 month); c = ½ (2 months)

Predicted Lags and Residuals by Day of 1st Attempt (Through Day 02)

Predicted Lags by Prediction Method



Predicted Lag Residuals by Prediction Method

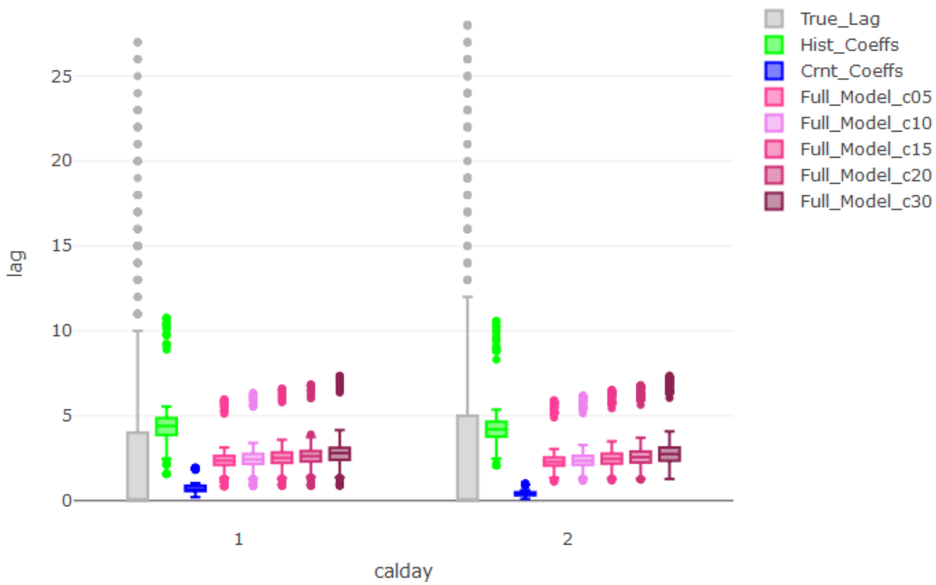


Initial Observations

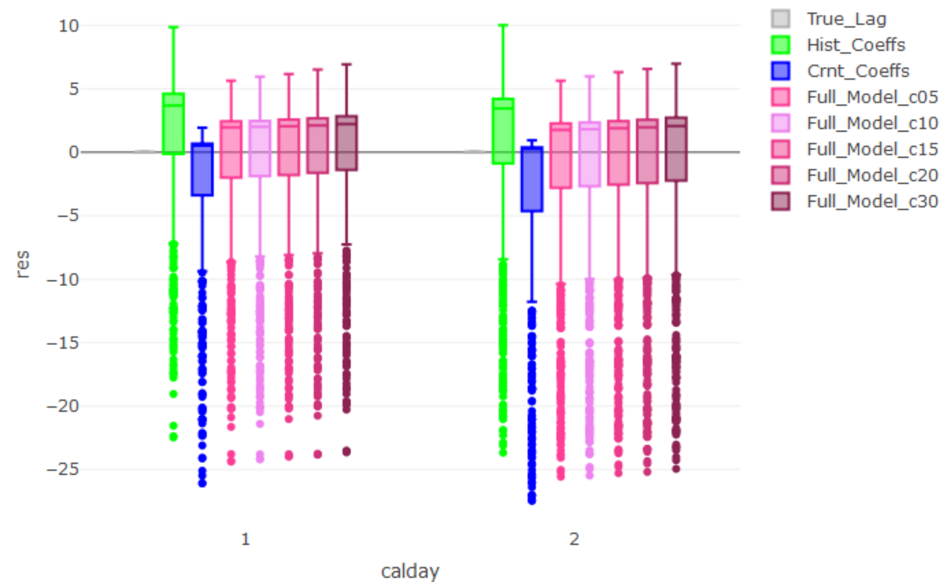
- Full model predictions fall between historic- and current-only models
- Residuals when the prior is equivalent to a month of data seem to be better distributed
- Current model has very little to work with (on day 2, lags cannot be larger than 1)

Predicted Lags and Residuals by Day of 1st Attempt (Through Day 02)

Predicted Lags by Method (varying c)



Predicted Lag Residuals by Method (varying c)

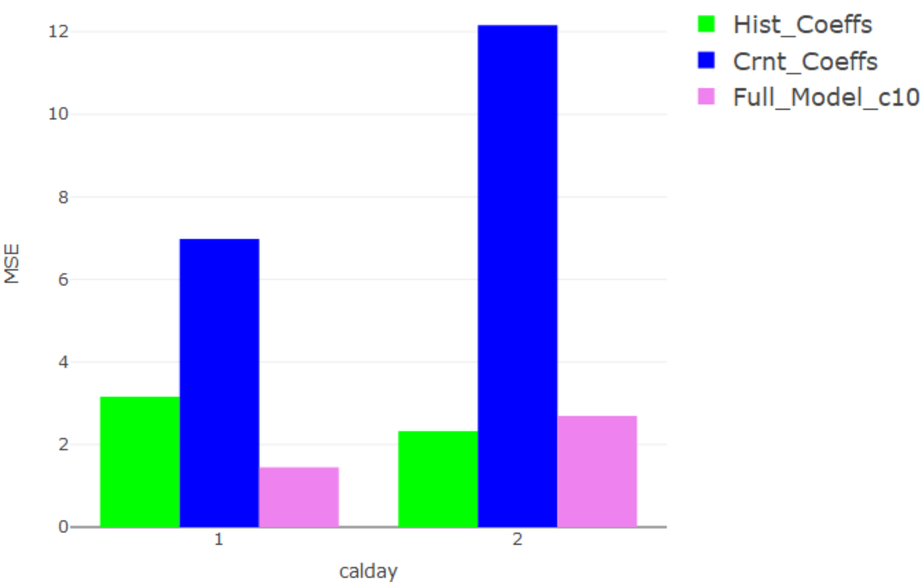


Strength of Prior

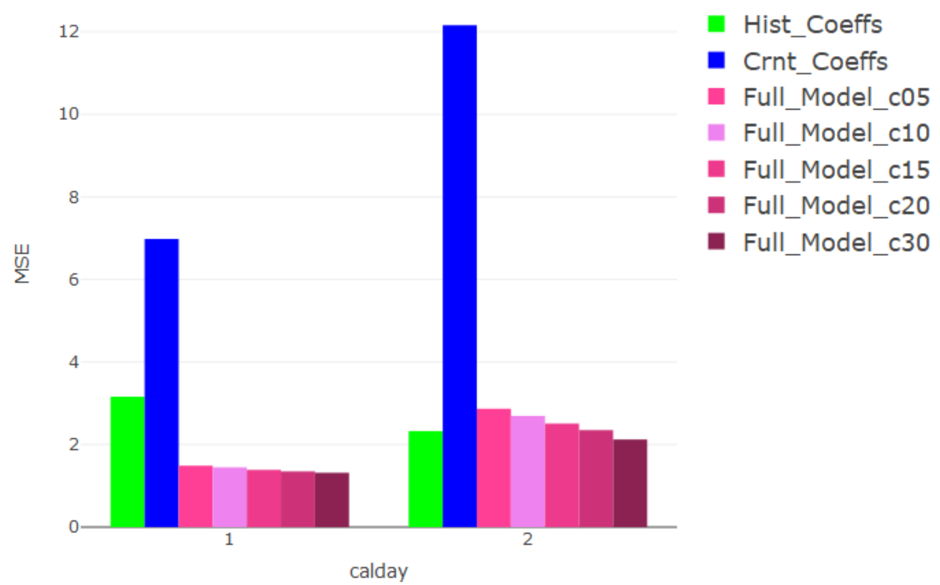
- Increase the variance to reduce the effect of the prior on posterior estimation
  - C = 0.5 → Prior = 2 months of data; C = 1.0 → Prior = 1 month of data
  - C = 1.5 → Prior = 2/3 month of data; C = 3.0 → Prior = 3 months of data
- Early in data collection, relaxing the prior does not have much of an effect

Lag Prediction MSE by Day of 1st Attempt (Through Day 02)

Prediction MSE by Prediction Method



Prediction MSE by Method (varying c)

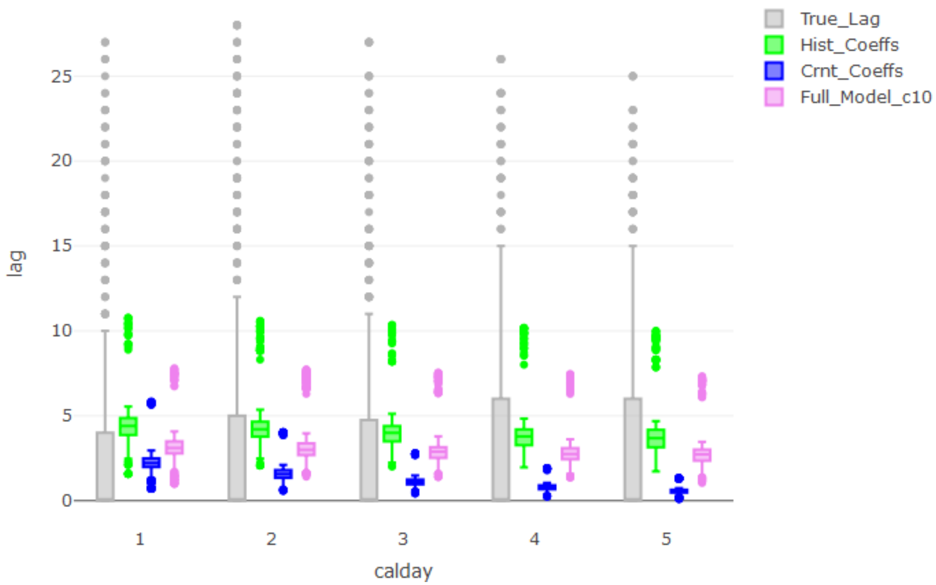


Initial Observations

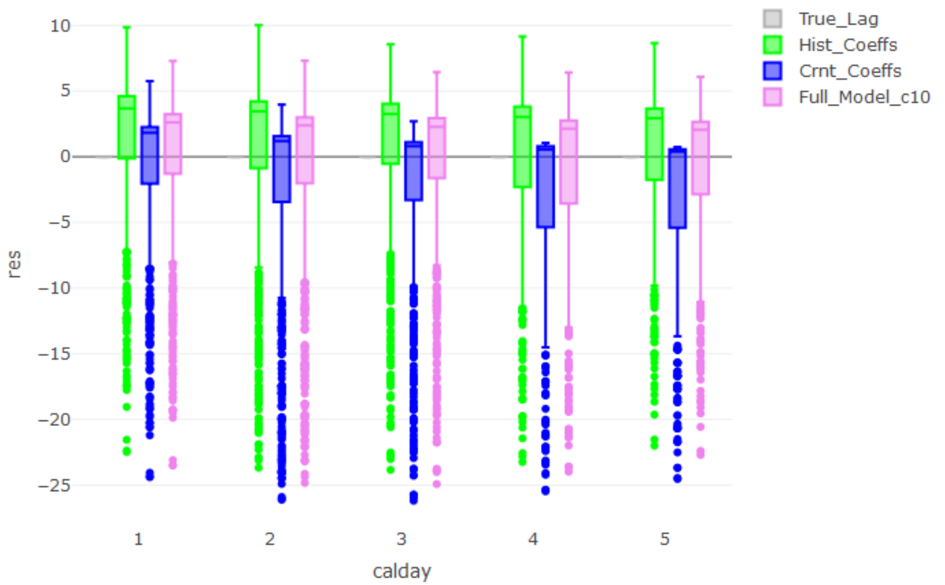
- This early in data collection, current data may diverge from historic data
- Could lead to slightly higher MSE on Day 2 than the historic-only model
- Very large reductions in MSE over the current-only model

Actual vs. Predicted Lag by Day of 1st Attempt (As of Day 5)

Predicted Lags by Prediction Method

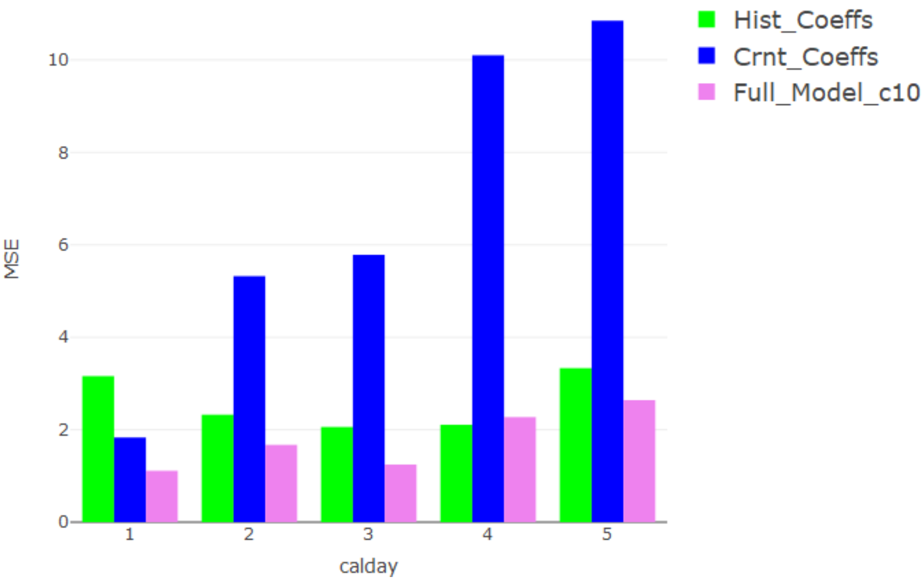


Predicted Lag Residuals by Prediction Method

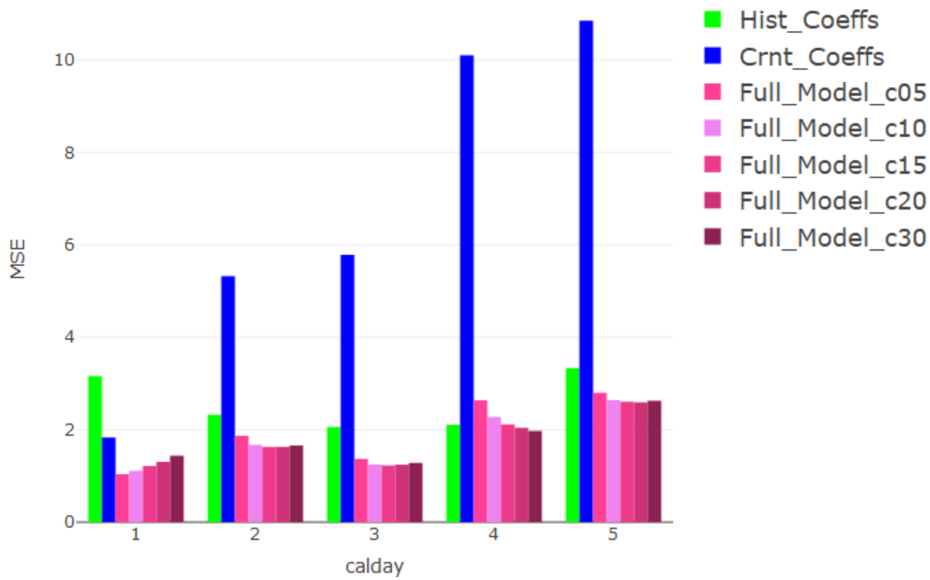


Actual vs. Predicted Lag by Day of 1st Attempt (As of Day 05)

Prediction MSE by Prediction Method



Prediction MSE by Method (varying c)

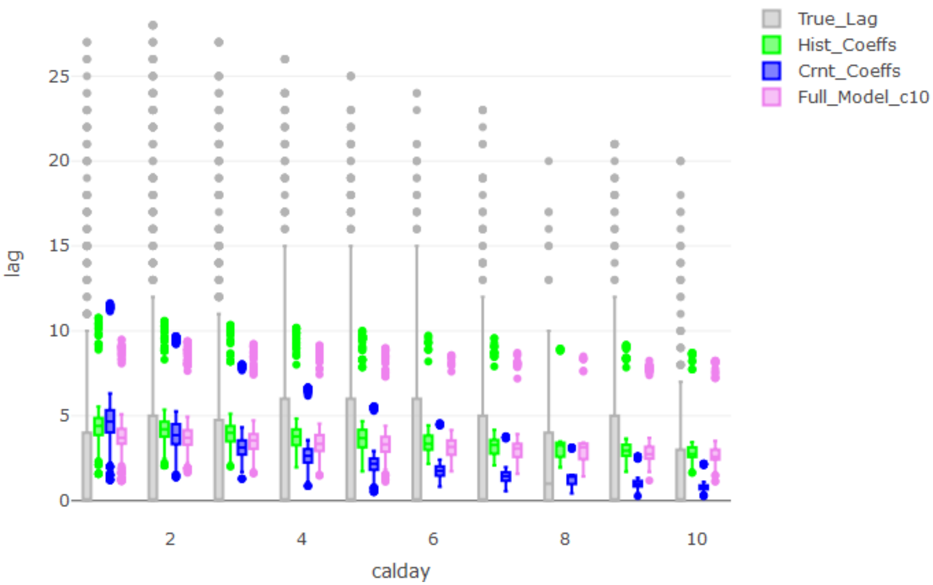


Observations:

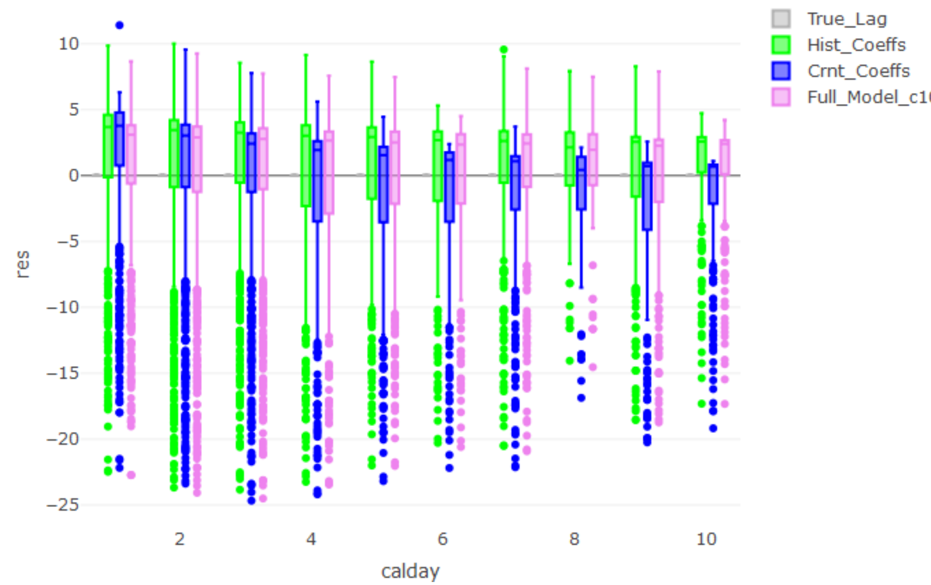
- Current data outcomes could be stabilizing, so full model is improvement over both the historic-only and current-only models
- Current data still only partially accumulated – relaxing the prior only has small effects
- Very large reductions in MSE over the current-only model

Actual vs. Predicted Lag
by Day of 1st Attempt (As of Day 10)

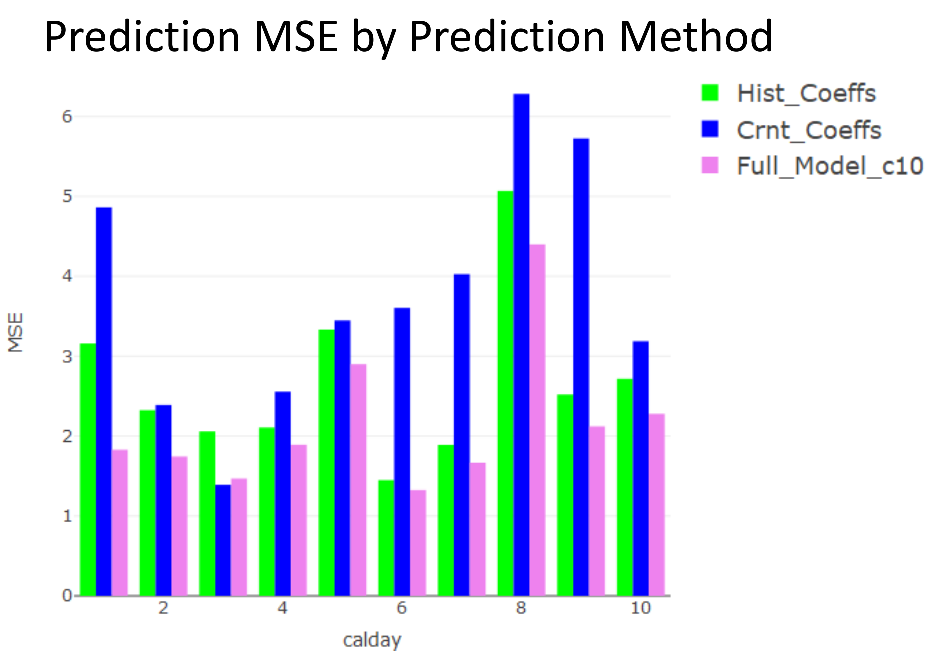
Predicted Lags by Prediction Method

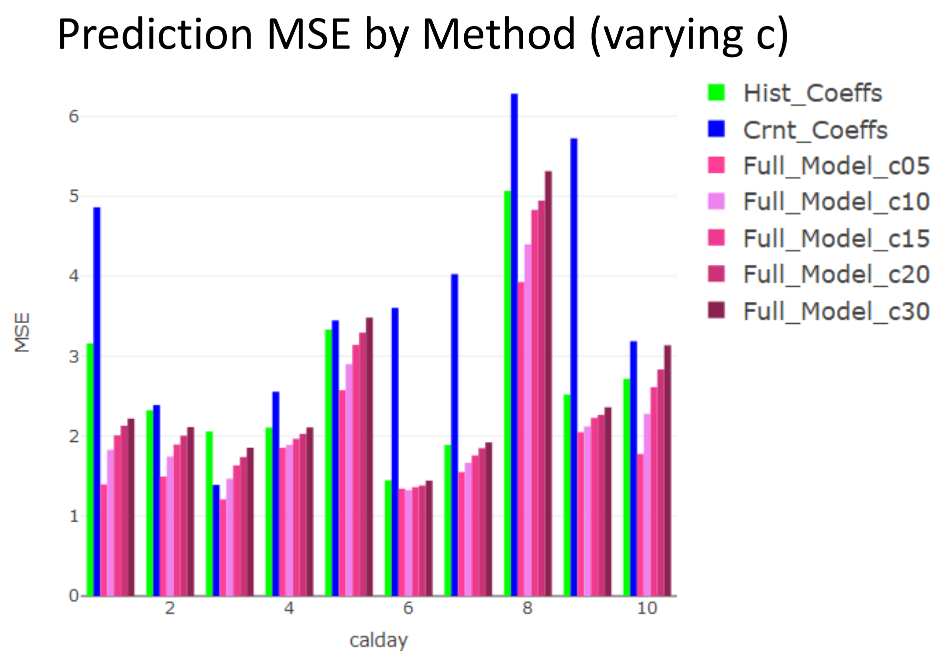


Predicted Lag Residuals by Prediction Method



Actual vs. Predicted Lag by Day of 1st Attempt (As of Day 10)





Observations:

- Relaxing the prior now has a visible effect, allowing the posterior predictions to approach the current-only model predictions
- When c = 1.0 (one month), posterior predictions still have lower MSE than either model individually
- Effect of prior decreases from this point, especially when relaxing the prior

Summary

- Full Bayesian model performed better than historical- or current-only model
  - Large gains in prediction MSE over current-only model
  - Modest gains over historical-only model
    - May see larger gains if the current data is “very different”
- Most of the gains occurred early in data collection
  - Day 10+ the effect of prior was reduced (esp. as c increased)
  - For interviewer feedback, can have more impact earlier
- It took very little information
  - Day of first attempt
  - Experience of interviewer
  - Aggregate geographic data
- A longitudinal survey could provide richer covariates to further improve upon this application
  - Models could be stratified to create expectations for different skill levels of interviewers or account for difficulty level of cases.

Contact Information

Stephanie Coffey: stephanie.coffey@census.gov

Michael Elliott: mrelliot@umich.edu

Thank you!