Estimating Response Propensity Models During Data Collection: Challenges and New Approaches

James Wagner

Survey Research Center
University of Michigan
Ann Arbor, MI

jameswag@isr.umich.edu
http://jameswagnersurv.blogspot.com/

January 26, 2016

Overview

Background

- Response propensity models fit during data collection can be useful
- Model estimates can be biased based on early data
- Use of data from previous surveys?
- Use of Bayesian models with informative priors
 - **Can we specify priors such that this bias is eliminated?**

Uses for Response Propensity Modeling

1. Measure predictors of response
   - *Example*: R-Indicators
2. Rank the cases with respect to estimated propensities
   - *Example*: Focus effort on low-probability cases
3. Prediction of expected output
   - *Example*: NSFG monitoring output

James Wagner Response Propensity Models

Problem

- Model may be mis-specified
- Interviewers may select cases based on unobserved characteristics
- Estimated coefficients may change over time
- Result: Predictions (especially early) may be biased

Comparison of Two Model Estimates
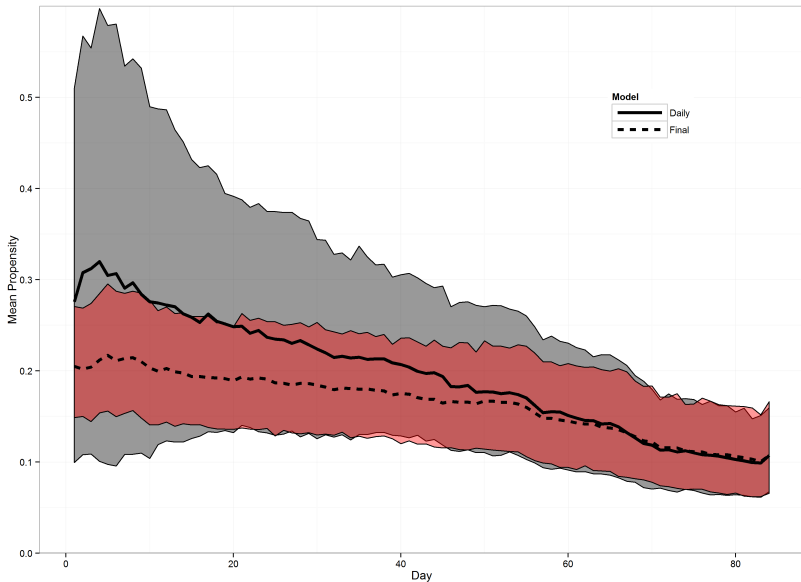
Two Examples

1. Call Scheduling Problem
   - Use Data from Previous Survey
2. Monitoring/Predicting Response
   - Bayesian Model with Informative Priors

Example 1: SCA

- Survey of Consumer Attitudes (SCA)
- Monthly survey
- RDD Sample
- Computerized sample management system
  - How to determine which case to call next?

James Wagner   Response Propensity Models

SCA: Scheduling Calls Based on Estimated Contact Propensities

- Wagner (2013) uses estimated contact propensities to schedule the next call
  - Estimates based on call records
  - Data includes records from two prior months and current month
  - Compare estimated contact probabilities across windows, rank the windows
  - Prioritize cases in the window for which they have the highest ranking
- What if these estimates depend upon the time at which the model is estimated?
- One estimate likely to be more efficient than another

James Wagner Response Propensity Models

SCA: Changes in Estimated Coefficients
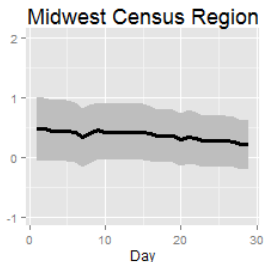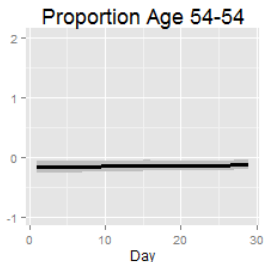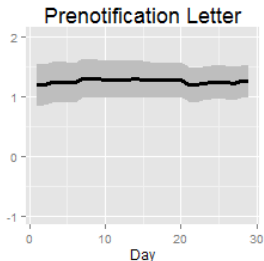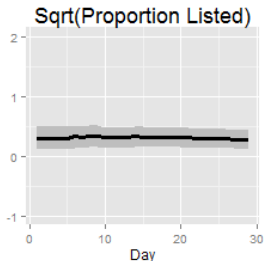
SCA: Changes in Actions Resulting from Different Estimates

Possible to compare two models:

1. Model using data available that day
2. Model using data available at the end

SCA: Changes in Actions Resulting from Different Estimates

| Change In Ranking	Percent
0	84.5
1	14.1
2	1.4
3	0.1

SCA: Conclusions

1. Previous data helpful in stabilizing estimates
2. Possibly some loss of efficiency due to poor prediction
3. Difficult to match the appropriate previous survey

Example 2: NSFG

- National Survey of Family Growth (NSFG)
- Quarterly data collection: 4 new samples each year
- Estimate "next call" response propensity models daily
    - Time-varying covariates included
   - For example: Number of calls, Ever refusal, Comments from informant

James Wagner   Response Propensity Models

Average Response Propensity for Active Cases by Day

James Wagner   Response Propensity Models
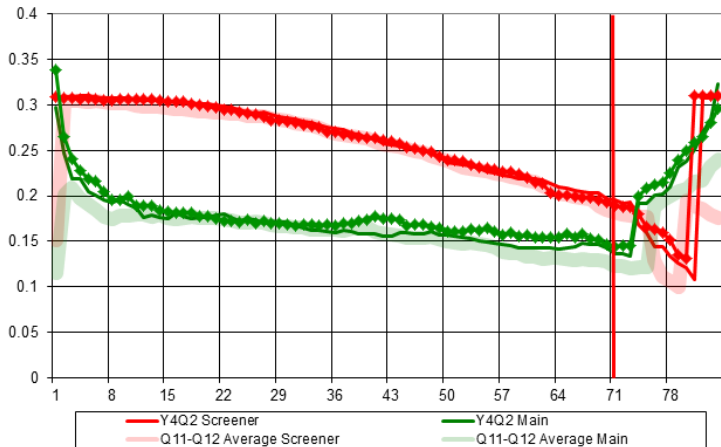
NSFG: Monitoring Mean Propensity

- Does timing of the estimate make a difference?
- Compare two models
  - Model using data available that day
  - Model using data available at the end

James Wagner   Response Propensity Models

NSFG: Comparison of Two Model Estimates
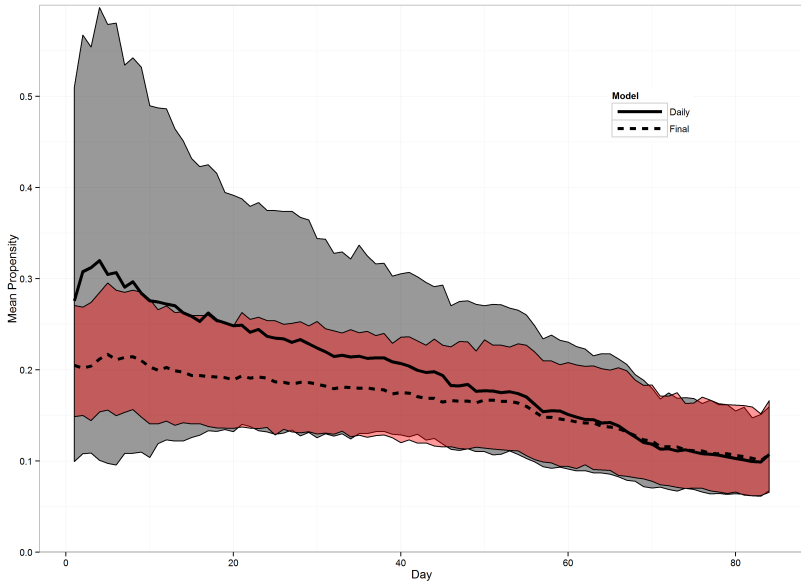
NSFG: Estimated Coefficients by Day

James Wagner   Response Propensity Models
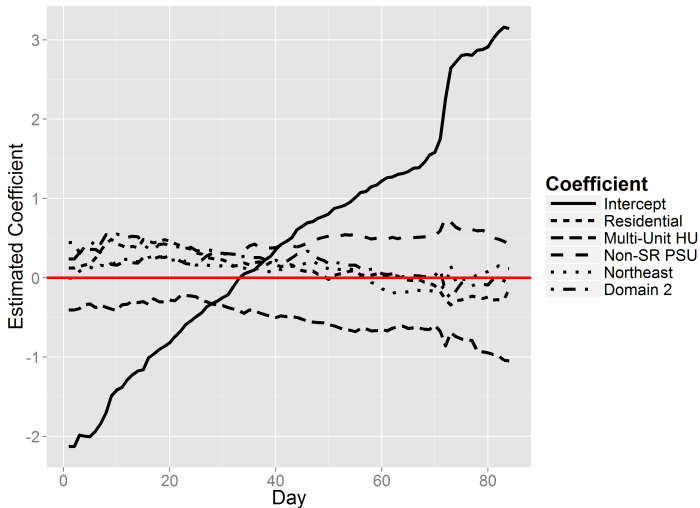
Final Estimates Vary across Quarters

Table: Selected Coefficient Estimates across Quarters

| Parameter	Q1	Q2	Q3	Q4
Intercept	-2.64	-2.44	-3.10	-2.54
urban	-0.04	0.03	0.00	0.01
numprevcalls	-0.08	-0.07	-0.10	-0.08
prevcontact	1.04	0.96	0.98	0.88
SCR_TEEN	0.05	0.08	0.10	0.05

NSFG: Can we specify a prior that attenuates bias?

Evaluate Three Different Priors

1. Use **all** the data from a previous quarter.
2. Use data from the **first half** of a previous quarter.
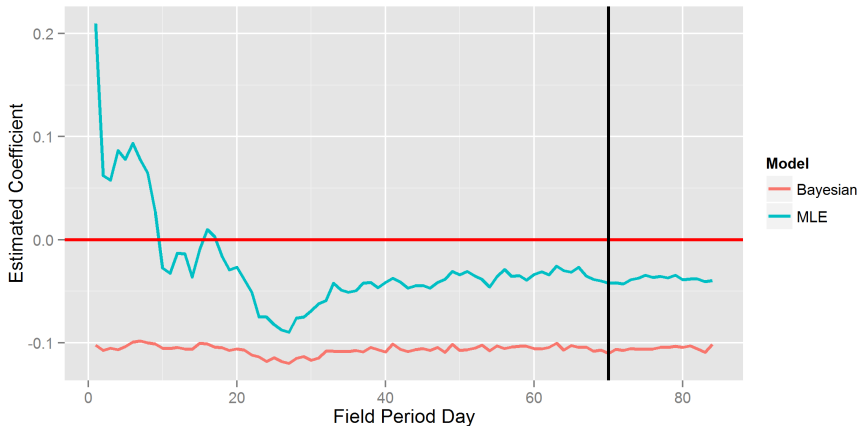3. Use data from the **last half** of a previous quarter.

In each case, placed a weak prior on the intercept.

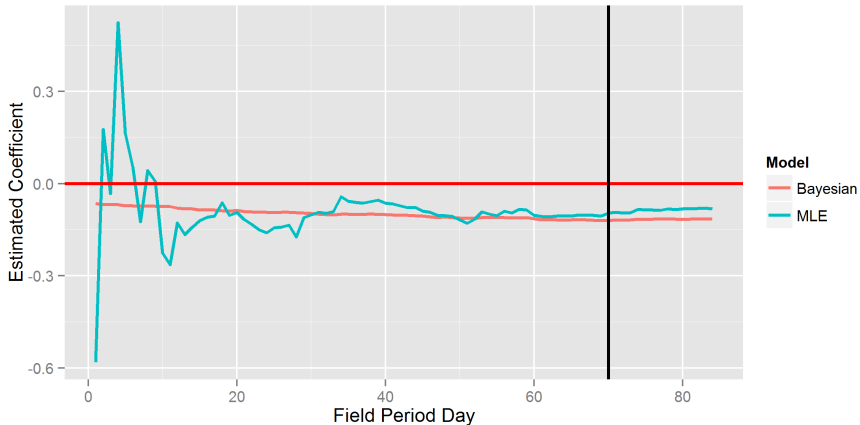# None of the priors did better than the current model
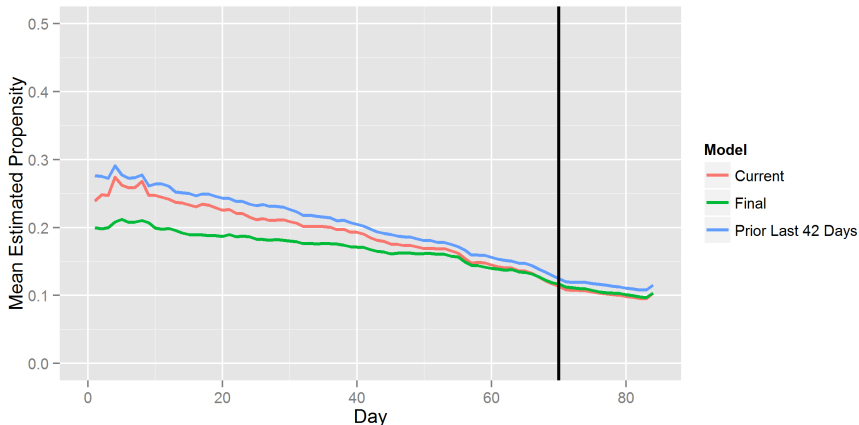
Estimated Coefficients Stabilized using Priors

Urban

Estimated Coefficients Stabilized using Priors

James Wagner Response Propensity Models

NSFG Conclusion

Lessons Learned

1. It is difficult to specify a prior in this setting.
2. May need to specify an informative prior for the intercept.
3. Setting prior for other purposes – ordering of cases – may be easier to do.
4. Informative priors may be helpful in the first few days or weeks.
5. Bayesian approach very useful for estimating models with new parameters (no previous data)

Conclusion

Conclusions

1. Using data from previous survey can be useful
2. Difficult to identify appropriate previous survey
3. Bayesian model provides a flexible solution
4. Specifying appropriate priors needs work

Conclusion

Next Steps

1. Prior setting for other purposes (ranking cases)
2. Use of priors early, but not late
3. Add parameters with no previous data (incentive experiment)

James Wagner   Response Propensity Models

Thank You!

Email: `jameswag@isr.umich.edu`

Blog: `http://jameswagnersurv.blogspot.com/`

James Wagner   Response Propensity Models