Using Bayesian Methods to Estimate Response Propensity Models During Data Collection

James Wagner

May 15, 2016

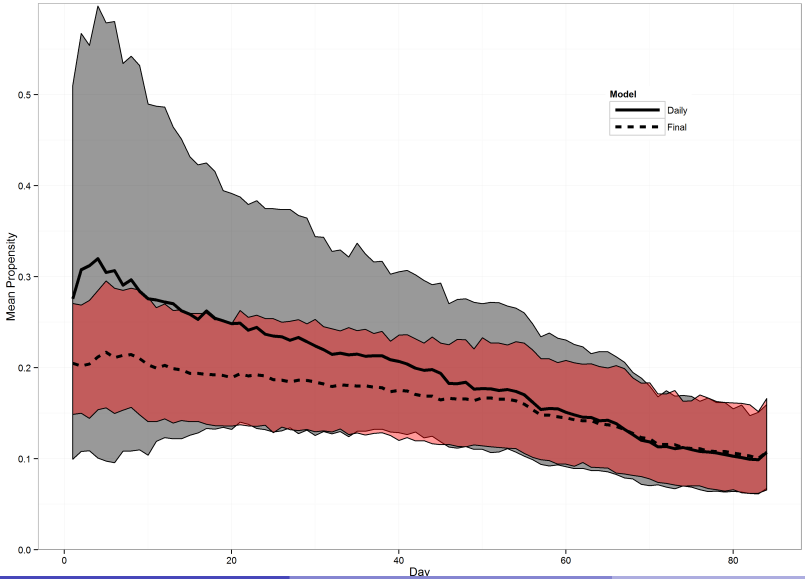
Uses for Response Propensity Modeling

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

Background

- Response propensity models fit during data collection can be useful
- Model estimates can be biased based on early data
- Bayesian models allow us to add information to the model as a prior
- **Can we specify priors such that this bias is eliminated?**

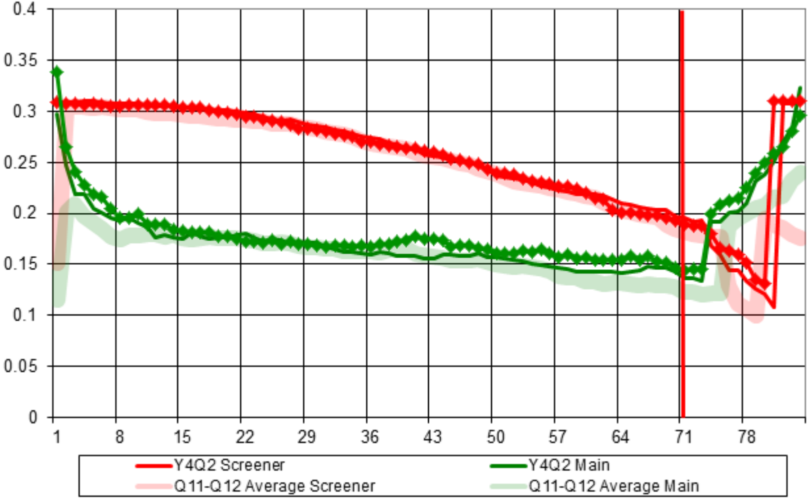
Comparison of Two Model Estimates



Example: National Survey of Family Growth

- NSFG has quarterly design, 12-week field period
- Survey on feritility and family formation
- Two stages of data collection
  1. **Screen** households to identify eligible persons
  2. **Interview** eligible persons
- Data include paradata, sampling frame data, commercial data, and (for stage 2) data from screening interview

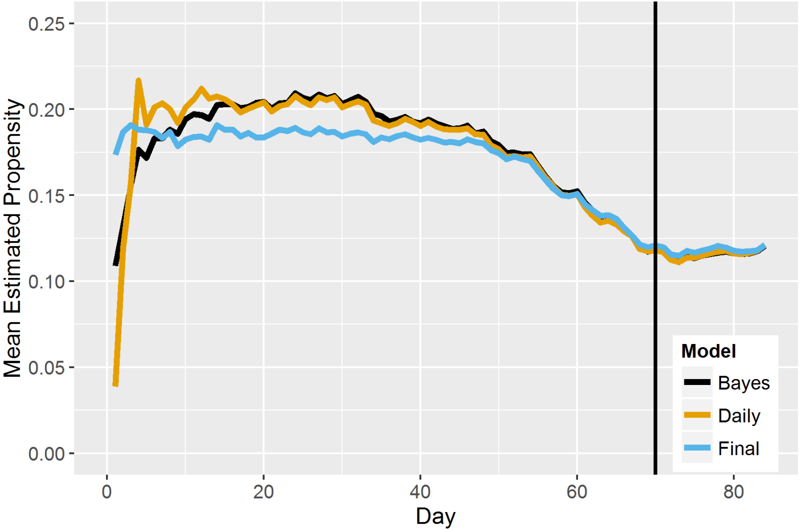
NSFG Monitors Mean Propensity of Active Cases



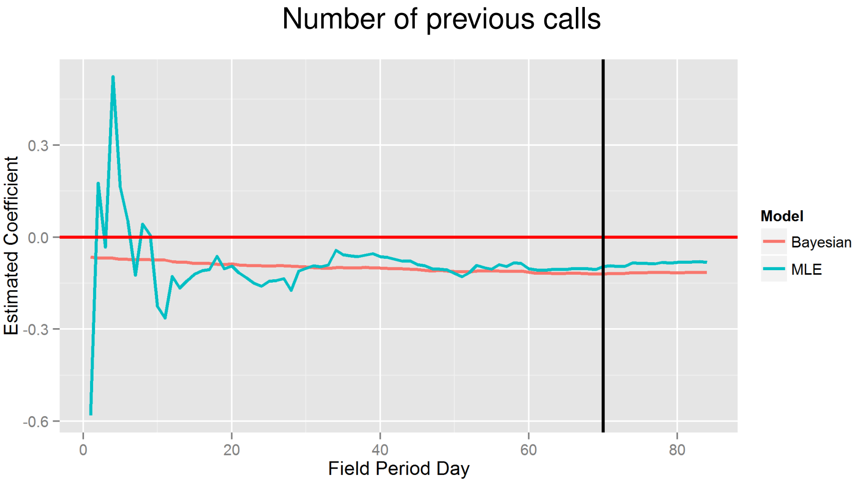
Can we specify a prior that attenuates bias?

- Set a prior for the intercept using a model
  - $0.209 \times ln(DayNumber) - 2.387$
- Priors for all other coefficients set using data from last 21 days of previous quarter
  - These are the "missing days" from the current quarter

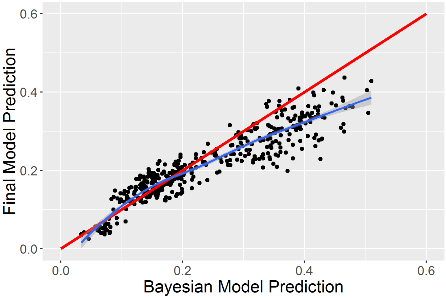
Overall mean: Bayesian model slightly better

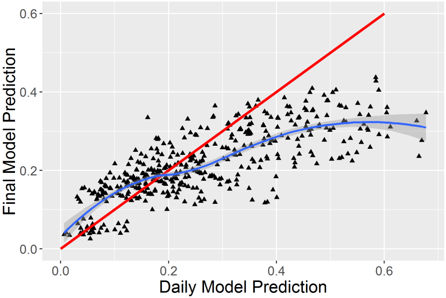


Estimated Coefficients Stabilized using Priors



Quintile assignment: Bayesian model big improvement





Quintile assignment: Bayesian model big improvement

Table: Bayes

	Bayes				
Final	0-20	20-40	40-60	60-80	80-100
0-20	76	25	11	0	0
20-40	31	55	22	4	0
40-60	5	31	66	7	2
60-80	0	3	10	75	24
80-100	0	0	0	26	86

Table: Daily

	Daily				
Final	0-20	20-40	40-60	60-80	80-100
0-20	67	18	18	6	3
20-40	29	31	33	19	0
40-60	12	46	30	21	2
60-80	4	16	26	29	37
80-100	0	1	4	38	69

	κ	ASE	z	Pr(>|z|)
Unweighted	0.551	0.025	21.7	2.18E-104
Weighted	0.748	0.013	46.13	0.00E+00

	κ	ASE	z	Pr(>|z|)
Unweighted	0.255	0.026	9.841	7.49E-23
Weighted	0.514	0.023	21.835	1.07E-105

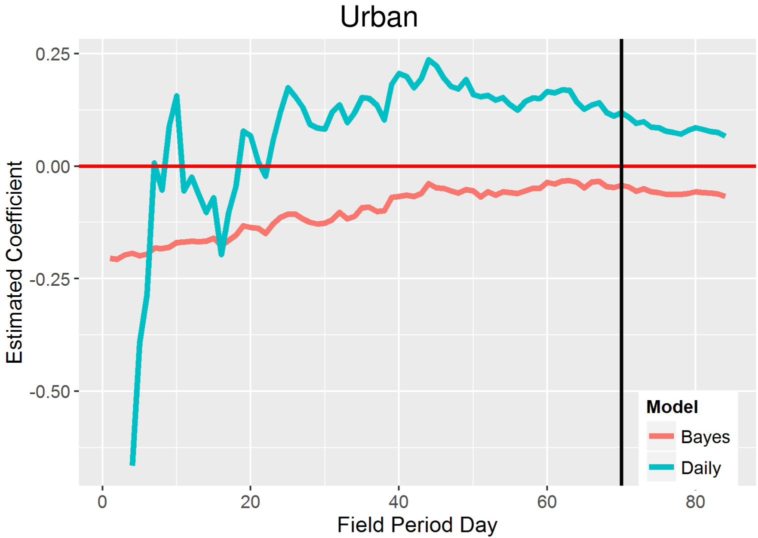
The agreement rate is is 64.0% for the Bayesian model predictions and 40.4% for the Daily model.

Identifying fourth percentile: Bayesian model big improvement

	Bayes	
Final	1-4	5-100
1-4	20	3
5-100	3	533

	Daily	
Final	1-4	5-100
1-4	13	10
5-100	10	526

Some estimated coefficients have poor priors



Conclusion

Lessons Learned

1. It is difficult to specify a prior in this setting.
2. Need to model the intercept over time.
3. Difficult to obtain accurate estimates of overall mean propensity.
4. Setting prior for other purposes – ordering of cases – is easier.

Conclusion

Next Steps

- Add parameters with no previous data (incentive experiment)
- Elicit priors from expert opinion and published research

Thank you!

email: jameswag@umich.edu
blog: http://jameswagnersurv.blogspot.com/