Modeling ABS Survey Response Rates

An AmeriSpeak Case Study

October 24, 2023
Stas Kolenikov, Patrick Coyle, and David Dutwin
2023 FCSM Research & Policy Conference

Introduction

Our role: Survey statisticians

- Sampling, weighting, nonresponse analysis

Goal: Create a sampling design calibration tool.

Method: Combine tract-level response rate data for surveys fielded by NORC to Census Planning Database (PDB) tract-level data on response propensity. Model survey response rate as a function of the ACS 2016-2020 Self-Response Rate.

- *Semi-parametric modeling*: generalized additive models (GAM) (Wood 2017)

- *Parametric modeling*

Census Low Response Score

- Smoothed / predicted nonresponse rate to the Census self-completion modes (web, mail) in multi-mode surveys.

- Typical range of 0 to 40.

- Predictions are based on the tract / census block group demographic characteristics.

- Released in some versions of the Planning Data Base (https://www.census.gov/topics/research/guidance/planning-databases.html)
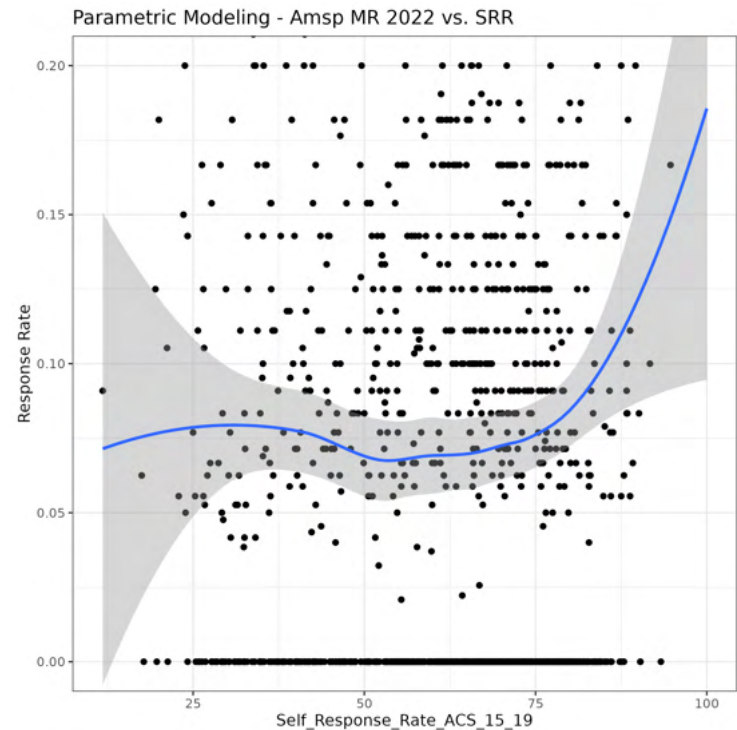
Semi-parametric modeling

Model Descriptions

- Generalized additive models (GAM).
    - Smoothing splines (cubic spline basis)
    - Flexible nonlinearity
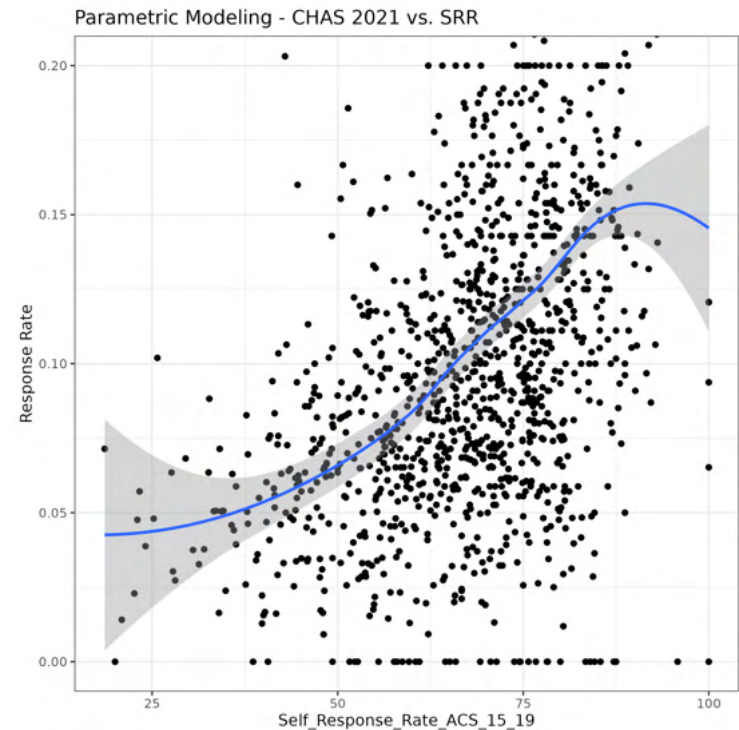
2022 AmeriSpeak Mail Recruitment

- Weak correlation…

Statistic	Value
Effective degrees of freedom (edf)	2.266
F-statistic	0.542
p-value	0.087

Parametric Modeling - Amsp MR 2022 vs. SRR

2021 Colorado Health Access Survey

- Strong correlation...

Statistic	Value
Effective degrees of freedom (edf)	6.580
F-statistic	31.895
p-value	0.000

Parametric Modeling - CHAS 2021 vs. SRR

2022 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation

Statistic	Value
Effective degrees of freedom (edf)	1.236
F-statistic	3.124
p-value	0.000
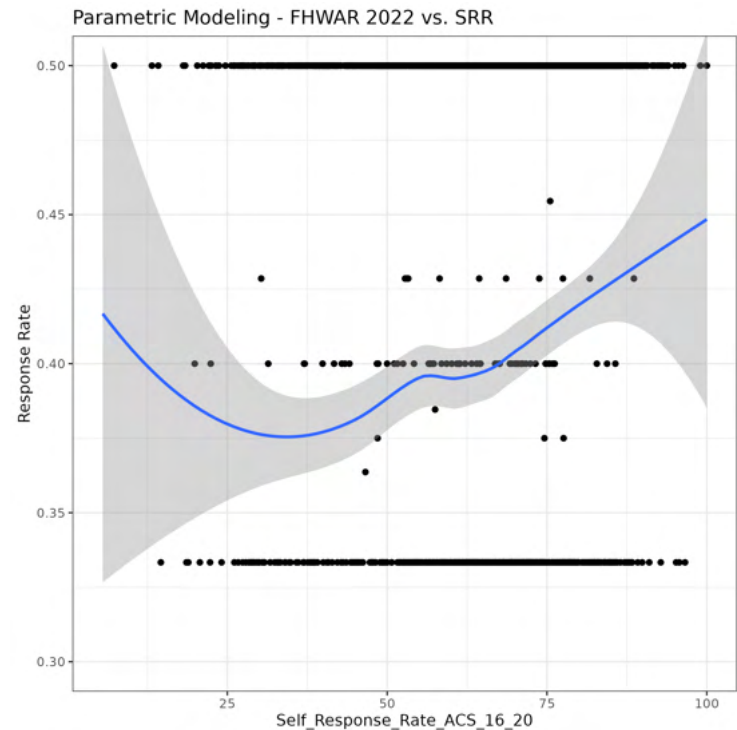
Parametric Modeling - FHWAR 2022 vs. SRR

Parametric modeling

Model Descriptions

- **Data input**:
  - x, the ACS Self-Response Rates for census tracts divided by 100.
  - x_L and x_H, the min/max ACS Self Response Rates.
- **User input**: r_L and r_H, the min/max expected tract-level response rates for a new survey.
- **Output**: $\hat{r}$, the predicted tract-level response rates for a new survey.

1. **Rescale**:

$$\hat{r} = r_L + (r_H - r_L) * \frac{x - x_L}{x_H - x_L}$$

2. **Response rate power transform**:

$$\hat{r} = r_L + (r_H - r_L) * x^\alpha, \qquad \alpha > 0$$

3. **Nonresponse rate power transform**:

$$\hat{r} = r_H - (r_H - r_L) * (1 - x)^\beta, \qquad \beta > 0$$

Model Descriptions

- **Data input**:
    - x, the ACS Self-Response Rates for census tracts divided by 100.
    - x_L and x_H, the min/max ACS Self Response Rates.
- **User input**: r_L and r_H, the min/max expected tract-level response rates for a new survey.
- **Output**: $\hat{r}$, the predicted tract-level response rates for a new survey.
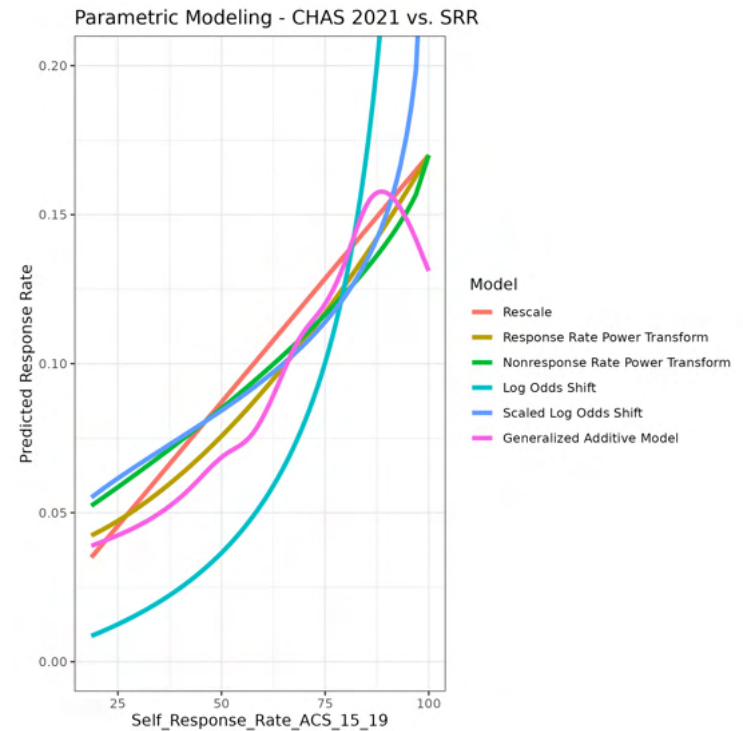
4. **Log odds shift**: `

$$\log\left(\frac{\hat{r}}{1-\hat{r}}\right) = \log\left(\frac{x}{1-x}\right) + \alpha, \qquad \alpha < 0$$

`

5. **Scaled log odds shift**: `

$$\log\left(\frac{\hat{r}}{1-\hat{r}}\right) = \beta * \log\left(\frac{x}{1-x}\right) + \alpha, \qquad \alpha < 0, \qquad \beta > 0$$

`

Parametric Modeling - CHAS 2021

| Model	Parameters	RMSE
Rescale	slope = 1.266	0.061
Response Rate Power Transform	alpha = 1.73	0.061
Nonresponse Rate Power Transform	beta = 0.667	0.062
Log Odds Shift	shift = -3.253	0.082
Scaled Log Odds Shift	shift = -2.38, scale = 0.311	0.063
Generalized Additive Model	edf = 6.58	

Parametric Modeling - CHAS 2021 vs. SRR

Parametric Modeling - FHWAR 2022

Model	Parameters	RMSE
Rescale	slope = 0.772	0.451
Response Rate Power Transform	alpha = 2.031	0.451
Nonresponse Rate Power Transform	beta = 0.52	0.451
Log Odds Shift	shift = -1.125	0.470
Scaled Log Odds Shift	shift = -0.454, scale = 0.088	0.451
Generalized Additive Model	edf = 1.236	
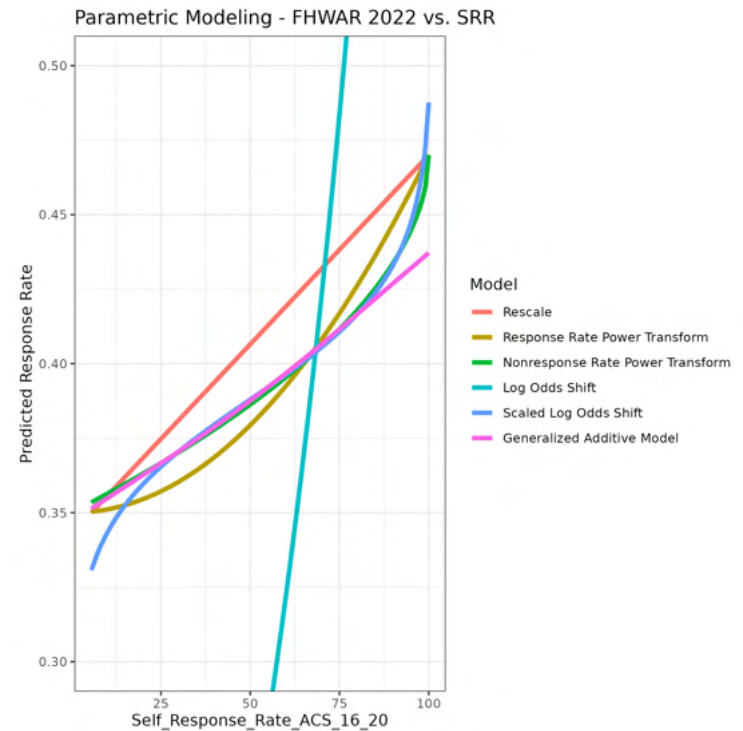
AmeriSpeak - Evaluating Nonresponse Follow-Up (NRFU)
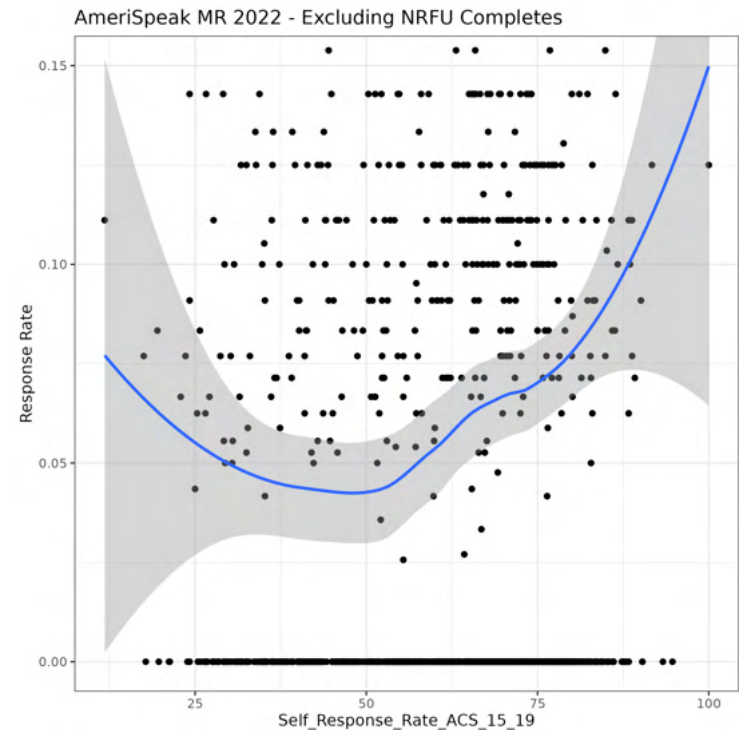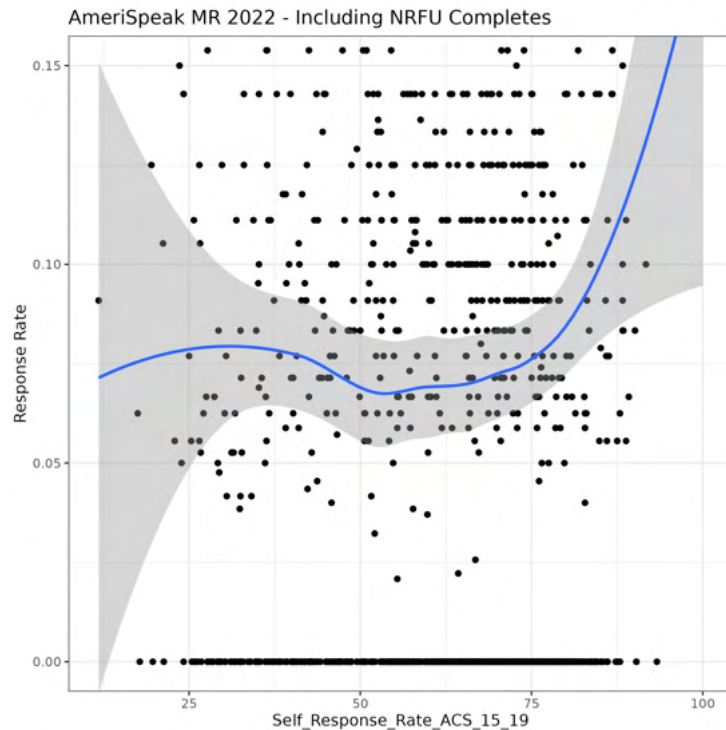
Conclusions

- AmeriSpeak boosts its response rate in particular areas with a non-response followup (NRFU) program and other interventions such as Spanish-language materials. This is effective at combating systematic non-response.

- Census did not release a Low Response Score with the 2022 PDB.

- Need for cross-validation in parametric modeling.

Thank you.

Stas Kolenikov
Principal Statistician
kolenikov-stas@norc.org

Patrick Coyle
Statistician
coyle-patrick@norc.org

David Dutwin
Senior Vice President, Business
Ventures and Innovation
dutwin-david@norc.org

Research You Can Trust

NORC at the University of Chicago