Predicting Future Participation in AmeriSpeak Panel

Bayesian mixed model analysis

FCSM / October 25, 2022

Stas Kolenikov, David Dutwin, Ipek Bilgen, Poulami Maitra

AmeriSpeak

Problem statement / background

In both online panels and longitudinal surveys, panel attrition and low participation rates impact survey estimates, total survey error, as well as panel operation costs.

Panel retention and high survey cooperation rates are also crucial to support studies of lower incidence populations and to potentially minimize nonresponse bias.

A potential indicator of low participation and panel attrition/turnover is panelist dissatisfaction.

By the Numbers

49K

Participating Households

(50 States + DC)

1.2K+

Client Surveys Completed

(Since June 2015)

22%

Recruitment Response Rate

(AAPOR RR3)

34%

Recruitment Response Rate

(For recruitment years with NRFU)

AmeriSpeak.norc.org/research

Panel Attrition

AmeriSpeak

WHY?

Can we predict attrition and decline in participation over time? If so, how?

- Predicting panel attrition (i.e., drop out) and nonresponse to subsequent surveys using panelist's self-reported data, as well as paradata
- Identification of attrition and nonresponse patterns
- Preventing or forestalling the decline of interest and dropout
  - cheaper than recruiting new panelists
  - helps maintaining the panel representativeness

WHAT?

- Is it feasible to predict panel attrition and nonresponse to subsequent surveys using overall and over time panelist self-reported future participation scores?
- Who has lower panelist self-reported future participation scores?
  - If (and how) their participation scores change over time?
- Who is expected to drop out faster based on our models?

HOW?

- Predicting future drop out and nonresponse to subsequent surveys using *change* in survey-level panelist participation propensities and self-reported future participation scores over time
- Bayesian cross-classified latent trajectory model combining paradata (unit response; survey features) and survey data (FUTURE)

FUTURE question:

How likely are you to participate in subsequent survey invitations from AmeriSpeak?

- (1) Not at all,
- (2) somewhat,
- (3) very, or
- (4) extremely likely?

Introduced in June 2021, asked no more frequently than once every two weeks

Analysis

AmeriSpeak

AmeriSpeak data

- Year 2021 recruits
- 98 surveys with at least 2,500 invited participants
  - small surveys are usually various pilots
- 3,366 panelists
- 65,277 invitations
  - 13,911 completes and 51,366 incompletes/refusals.

Model in equations

A two-way/cross classified latent trajector model

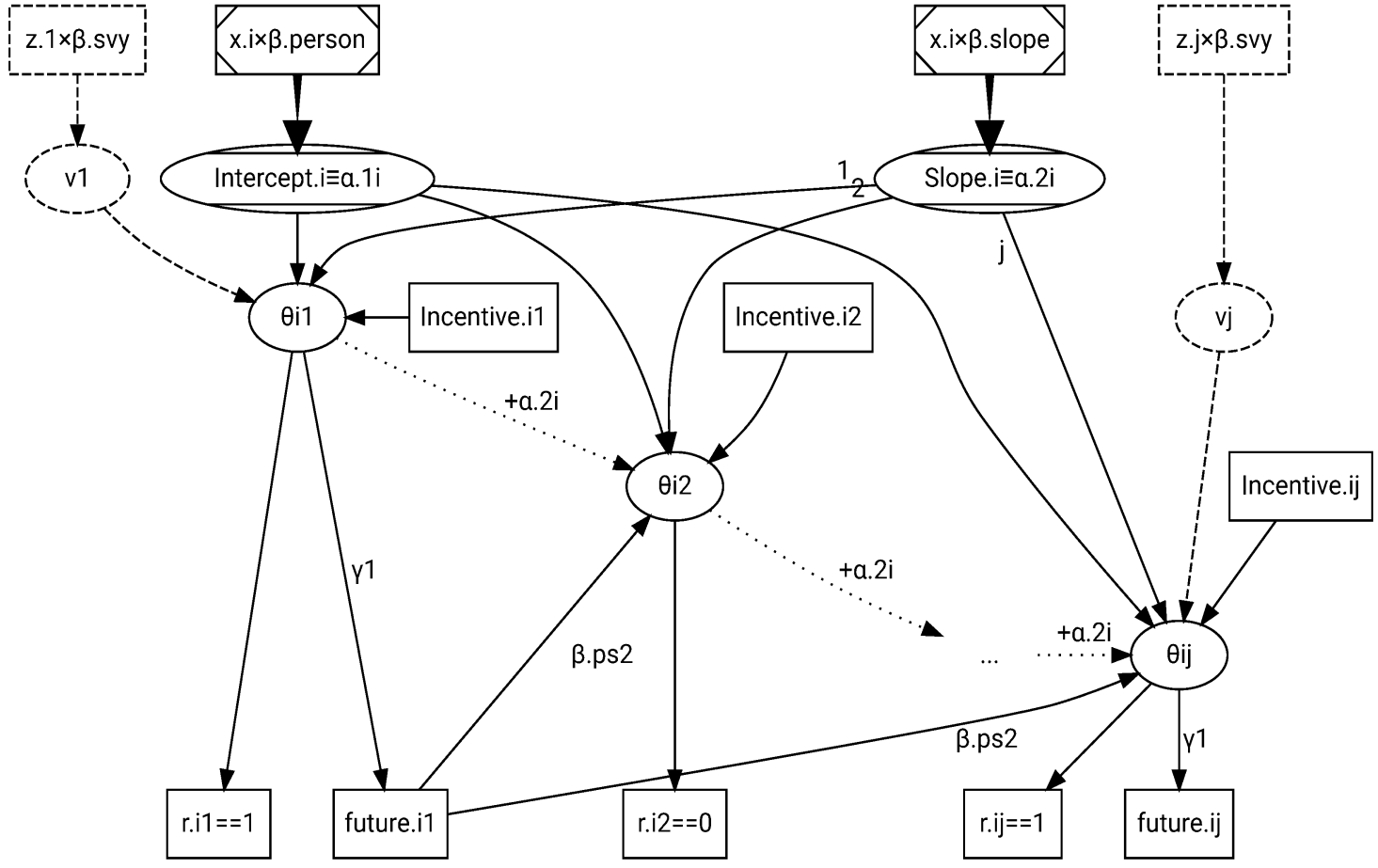
$$
\begin{aligned}
\theta_{ij} &= \text{logit}\,\mathbb{P}[r_{ij}=1], \quad r_{ij} \equiv \text{panelist } i \text{ completes survey } j \\
\theta_{ij} &= \beta'_{\text{p·s}} x_{ij} + \text{Intercept}_i + \text{Slope}_i\, m_i(j) + \nu_j \\
\text{Intercept}_i &= \beta'_{\text{person}} x_i + e_{1i}, \quad \text{Slope}_i = \beta'_{\text{slope}} x_i + e_{2i} \\
\nu_j &= \beta'_{\text{svy}} z_j + \varepsilon_j,
\end{aligned}
$$

$$
\mathbb{P}[\text{future}_{ij} = m | r_{ij}] = \mathbb{P}_\Lambda(\kappa_{m-1} < \gamma_1 + \gamma_2 \theta_{ij} \leq \kappa_m)
$$

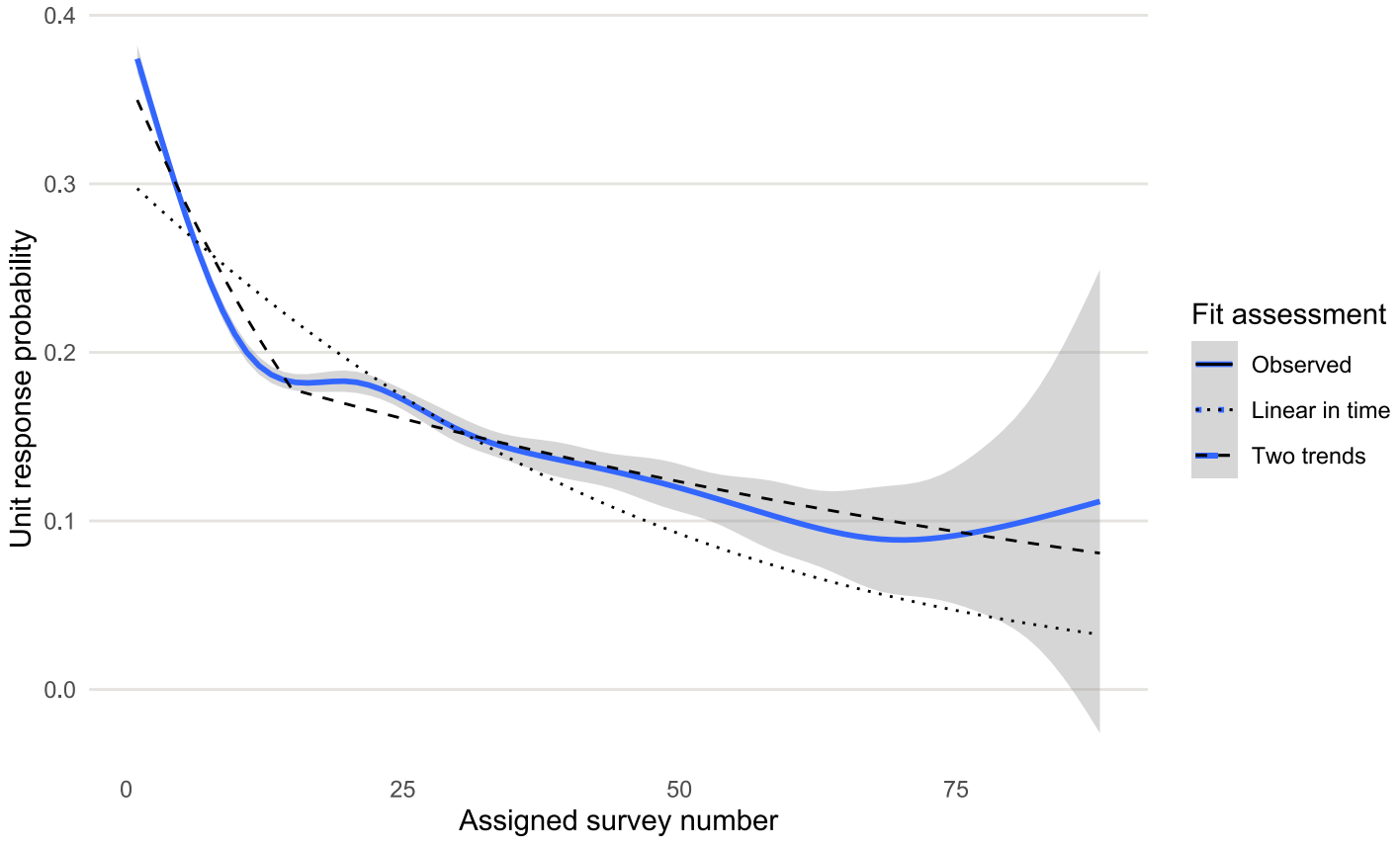
$$
\mathbb{P}_\Lambda(\kappa_{m-1} < x \leq \kappa_m) = \Lambda(\kappa_m) - \Lambda(\kappa_{m-1}), \quad \Lambda(z) = \left[1 + \exp(-Z)\right]^{-1}
$$

$m_i(j) \equiv$ sequential number of survey i among the j-th panelist roster of invites.

Model as a diagram



Need more slopes



Significant variables

parameter	mean	2.5%	97.5%
beta_person_intercept[Const.]	-3.193	-3.534	-2.862
beta_person_intercept[race4]	-0.056	-0.104	-0.006

No significant variables is... good?

- Initial recruitment is effective in bringing in a variety of respondents who are equally motivated to start

Significant variables

parameter	mean	2.5%	97.5%
beta_person_slope1[Const.]	-0.108	-0.186	-0.023
beta_person_slope1[ln_weights]	-0.038	-0.055	-0.024
beta_person_slope1[age]	0.004	0.002	0.005
beta_person_slope1[educ_5]	0.022	0.002	0.042
beta_person_slope1[race4]	-0.069	-0.113	-0.023
beta_person_slope2[ln_weights]	0.044	0.022	0.067
beta_person_slope2[age]	-0.004	-0.006	-0.002

Faster attrition:

- younger
- less educated
- Hispanic (`race`==4)

Significant variables

The FUTURE question appears twice in the model:

- (1) as the contemporaneous outcome of the latent propensity to respond...

parameter	mean	2.5%	97.5%
gamma_future[slope]	3.797	3.597	4.005
gamma_future[Const.]	4.632	4.333	4.951

The relevant part is the slope coefficient -- which is the slope of the regression of the FUTURE question on the latent response propensity.

Significant variables

The FUTURE question appears twice in the model:

- ... (2) as the (time-varying) predictor of the response propensity in the future
  - have to jump through some data management hoops to define "last known value of FUTURE"

parameter	mean	2.5%	97.5%
beta_person_by_survey[future12]	1.022	0.869	1.176
beta_person_by_survey[future3]	1.616	1.471	1.771
beta_person_by_survey[future4]	2.832	2.689	2.979

The "desirable" monotonicity is observed: higher values of FUTURE move the propensity up quite substantially:

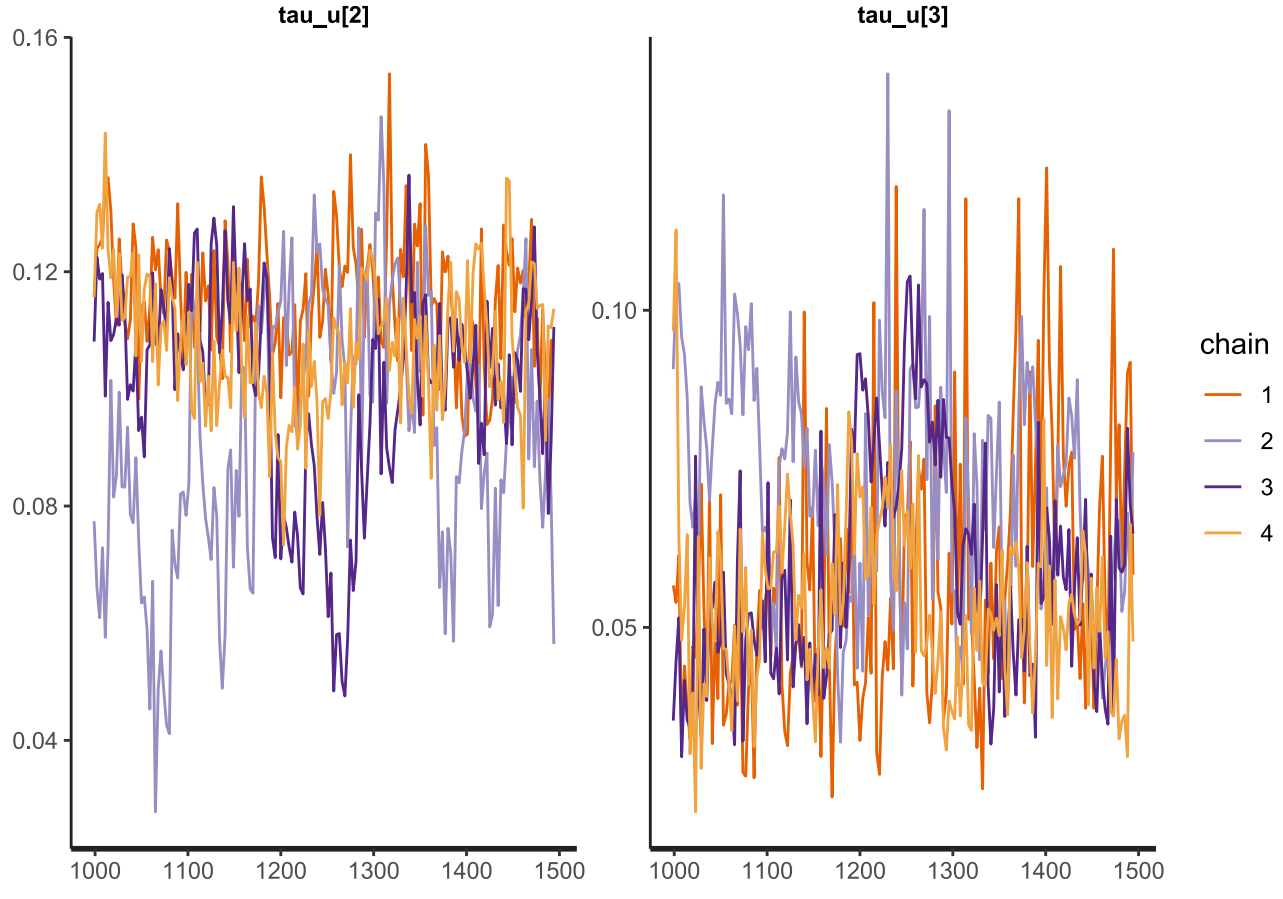
- hypothetical unit response rate = 20% with FUTURE == 4
- counterfactual unit response rate = 4% with FUTURE == 1, 2

Bayesian diagnostics

4 chains; 1000 warmup draws; 500 simulation draws; adaptive delta parameter of 0.92; tree depth of 10; thinning by 2

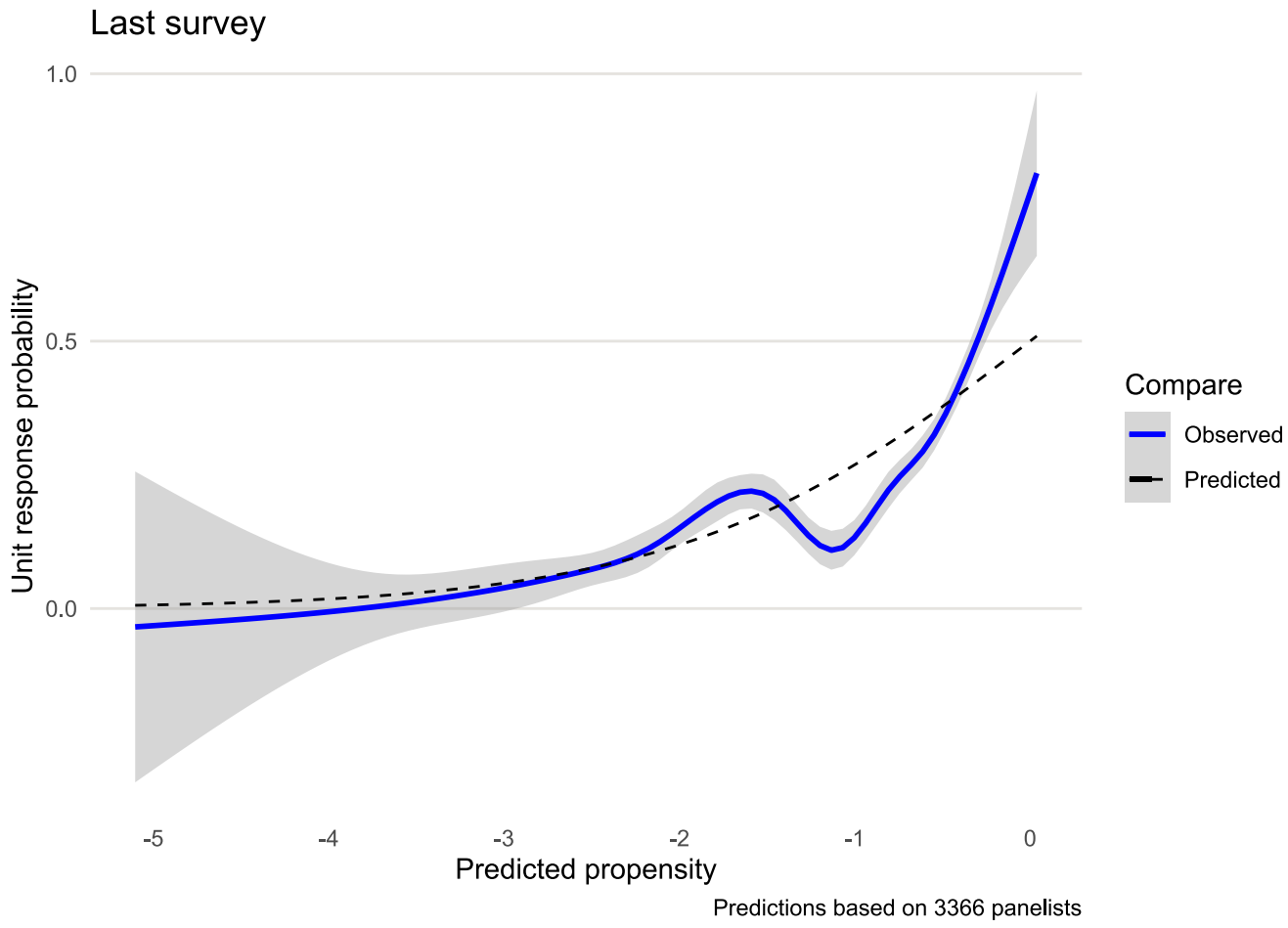
- For regression slopes, all values of the variance reduction factor (Rhat statistic) were below 1.05.
- For the model estimates, the values ranged from 0.997 to 1.040.
- For the person-level random effects, the values ranged from 0.996 to 1.020.
- The Rhat statistics for the variance components indicated lack of chain convergence. The Rhat for τ_1 was 1.157, and for τ_2, 1.130.
- There were no divergent transitions in the simulation part of the chain.
- The treedepth never approached the maximum value in the simulation part of the chain.

Bayesian diagnostics



Goodness of fit

Nowcast of the response propensity: iron out the kinks and wiggles



Parting words

AmeriSpeak

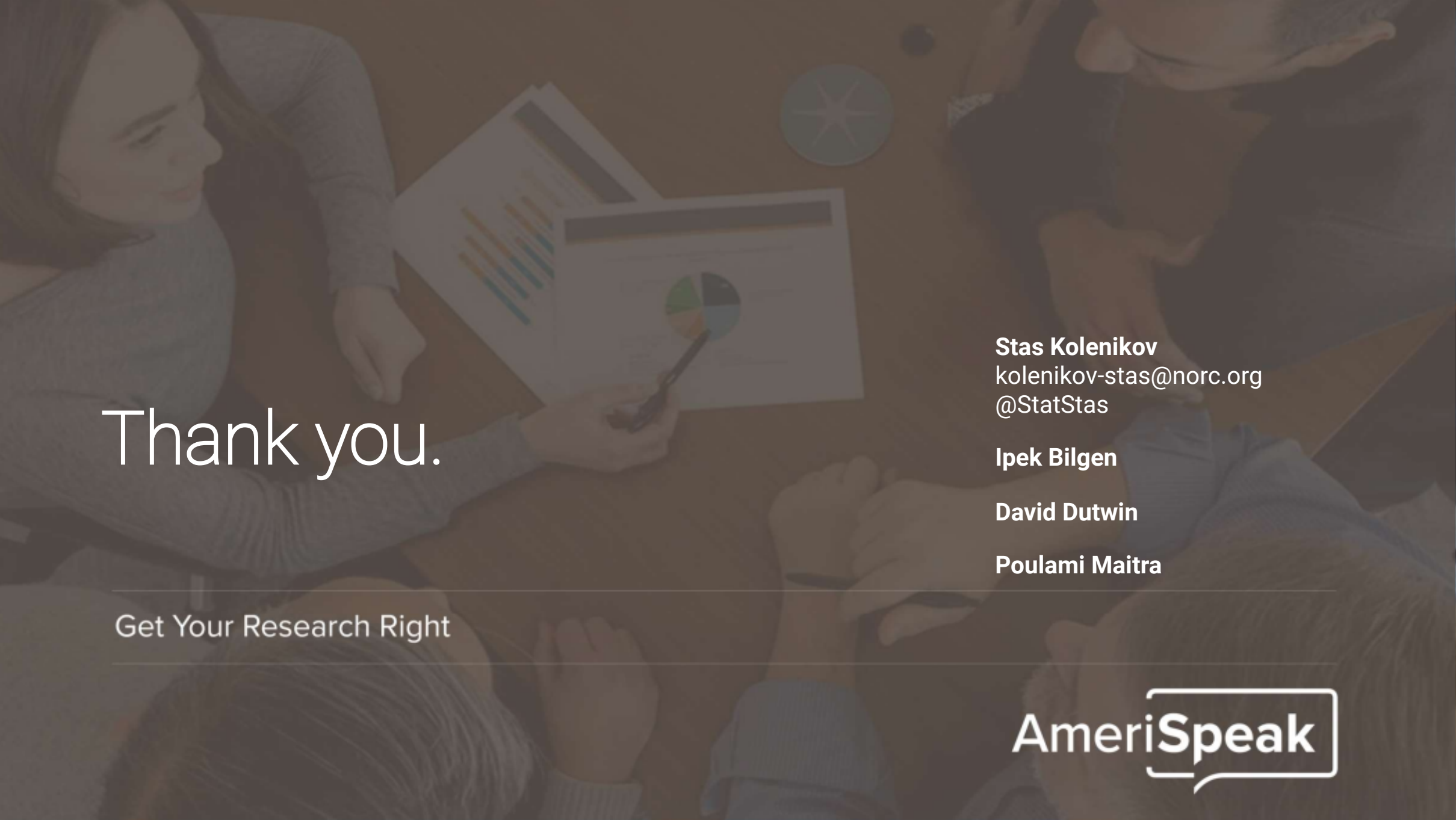
Conclusions and discussion

- Moderately sophisticated Bayesian modeling exercise
- Explainable patterns of decline in participation
  - lower education, younger, ethnic minorities
- Self-reported intent to continue panel participation is valid
- Identify high-risk panelists
  - greatest (most negative) slopes
  - decline in self-reports

- Next steps: interventions for the highest risk panelists

- Still needed: social theory of attrition

  - accelerated time?
  - mixture model?
- Still needed: social theory of panel time
  - current model: number of invitations
  - calendar time?



Thank you.

Stas Kolenikov
kolenikov-stas@norc.org
@StatStas

Ipek Bilgen

David Dutwin

Poulami Maitra

Get Your Research Right

AmeriSpeak