A Bayesian Model for Inference on Multiple Panel Public Opinion Surveys

Brittany Alexander *

Abstract

Multi-wave surveys, which track individual opinions over time using multiple administrations of the same survey, are common. Since they involve repeated measurements, the results from one survey date, or wave, are correlated with the next wave. However, combining information from two panel surveys with different numbers of waves or dates administered is non-trivial. We present a case study using Bayesian inference to combine two panels about terrorism policy from 2016. The first panel was a large 1730 individual two-wave probability-based panel with dropouts taken six months apart in May 2016 and November 2016. The second panel was a non-probability panel that had six waves taken every month from June 2016 to November 2016, had 779 respondents, also with dropouts, and includes a 108 person replenishment sample. We present an extension of multilevel regression with poststratification to model this data set, with an additional level of partial pooling across time and a multivariate likelihood for the repeated measures. We find this model produces more precise population estimates at individual time points without sacrificing the accuracy of predictions of individuals. In addition, the model finds that for most variables, there is no statistically significant change over time.

1. Introduction

This study involves analyzing two separate multi-wave surveys about Terrorism perceptions and policy preferences from 2016. The first survey is an extensive two-wave survey conducted in May and November of 2016, and is generally nationally representative and was taken using an address-based sample. The second survey is a six-wave survey administered monthly from June to November, and is not nationally representative. It comes from a panel of respondents recruited from web advertisements. The surveys present a sort of natural experiment on the effects of terrorist attacks on public perceptions and policy preferences relating to terrorism. Between waves of the surveys, terrorist attacks on U.S. soil by domestic and foreign actors and terrorist attacks in other western countries occurred. The two-wave survey with more respondents had too large a gap to monitor reactions to attacks, but the six-wave survey had fewer respondents and was not representative enough to detect changes over time at the population level.

We present a Bayesian model that uses both surveys to estimate population-level support and change over time in support of perceived likelihood, support for federal spending, and support for local spending. This new model can detect more minor changes over time because the uncertainty in the model is reduced by the additional information provided from the second survey.

Two separate survey companies conducted the two panels used in this study: Decision Research, which conducted the six-wave non-probability survey, and GFK, which conducted the two-wave probability survey. The GFK sample was collected using a subset of their KnowledgePanel, a probability-based panel designed to represent American adults. The GFK respondents were selected with an equal probability selection method, but the number of surveys given for weeks is restricted. The Decision Research survey used Decision Research's panel recruited from internet advertisements. The questionnaires were nearly identical, except new questions were added to later waves. Item non-response existed in

*Ph.D Candidate : Texas A&M University Dept. of Statistics. Pre-Doctoral Research Associate: Institute for Science, Technology & Public Policy

both panels but was much more prevalent in the GFK panel. The second wave of the GFK survey was restricted to those who completed the first survey. However, if a survey was missed in the Decision Research sample, that respondent was not invited to future waves except for the last wave.
Below is a table of the survey dates (all conducted in 2016) and the sample sizes. * Includes

	GFK W1	DR W1	DR W2	DR W3	DR W4	DR W5	DR W6	GFK W2
Date	May 5-17	June 6-12	July 11-15	Aug 10-17	Sep 11-15	Oct 12-17	Nov 10-12	Nov 10-15
Size	1730	671	658	640	618	600	779*	1210

108 person replenishment sample. There are four unique response patterns in the survey: the individuals in the GFK study who only completed the May 2016 wave, the individuals who did both waves in the GFK study, the individuals that did all six waves in the Decision Research Study, and the replenishment sample in the Decision research study that only completed the November 2016 wave. Each response pattern needs its regression because of the different structures of the data. In the case of those respondents who completed more than one wave, the regression will be multivariate and model all the individuals' responses at each time point together.

Between several waves, terrorist attacks occurred in both the U.S. and other western countries. Additionally, these surveys were taken across a Presidential campaign where Donald Trump made fighting Islamic terrorism a part of his platform. Thus, this presents a natural experiment on whether the presence of terrorist attacks and the discussion of the attacks impact public perceptions and views of terrorism. Below is a timeline detailing different events during the period of the surveys.

Figure 1: Timeline of Events

Based on the timeline, there are four primary points of interest: the change from time 1 to time 3 to measure the effects of the Orlando and Dallas attacks, the change from time 3 to time 4 to measure the effects of the convention, the change from time 5 to time 6 to measure the effects of the New Jersey and New York bombings, and the change from time 6 to time 7 to measure the effects of Donald Trump's election.
In previous research (Liu 2019) individual change over time in these variables was observed, but there was on average little change in most of the variables. These articles did not address population-level change over time.
There are five response variables: concern for certain types of terrorist attacks (Concern Type), the perceived likelihood of an attack in the next six months (Likelihood), support

for federal spending to prevent terrorism (Fed Spend), support for local spending to prevent terrorism (Local Spend), and support for a list of counterterrorism policies (Policy). These questions are Likert scales that are approximately continuous. When items on the Likert scale are averaged, the average is approximately normally distributed. The primary goal is to perform regressions at each time and estimate the support for policies and perceptions of risk and other attitudinal characteristics in the population while simultaneously developing a deeper understanding of the stability of beliefs on terrorism across six months at both the individual and population level.

The goal is to fit regressions under a Bayesian framework at each time point using three distinct but related models. Bayesian modeling is a natural fit for this problem as Bayesian models can be easily adapted to allow the sharing of information across various groups and handle data with different structures. Additionally, a Bayesian framework allows the Multilevel regression with poststratification (Gelman and Little 1997) to handle survey non-response with both surveys even though one survey is a non-probability sample and the other is a probability sample.

The first model takes advantage of the repeated measurements of the respondents and allows the regression coefficients to be partially pooled towards each other to provide better stability for the estimates at time points in the smaller Decision research study. The pooling is generated from a hierarchical prior for the regression coefficients. Thus, each regression coefficient has the same prior at every time.

The second model selects one set of responses for each respondent and allows the coefficients to again be partially pooled. Essentially this structure allows the regression coefficients to learn from the data at every time point. The second model is to create a dataset without repeated measurements by selecting one time point per respondent as done in the variable selection process to use in the analysis and then keep a hierarchical prior on the coefficients. The second model is less computationally intensive than the first, but it does not incorporate how individuals' responses change over time.

The tertiary model is to fit individual unrelated Bayesian regressions at each time point. The third model is the fastest method, but it doesn't take advantage of the repeated measures or allow for pooling across time. This is a model used primarily to compare the results from the primary and secondary models and validate that the new models do not sacrifice accuracy.

2. Methodology

2.1 Missing Data

For most of the items, the respondents completed almost every response. But since the goal was to build a regression model, missing items present problems. The survey combined "don't know" and "refused." "Refused" is when a respondent chooses not to disclose their answer, perhaps due to privacy concerns. In an online survey, sometimes people submit low-quality data and skip questions to finish quicker. The missingness patterns were examined, and it seems reasonable that when there is a small number of missing items, the missingness was random. Typically, only a single item in a question was missing. For example, a respondent answered every item about if they remember a terrorist attack except one or might not answer their concern for an armed attack, but answered every other question about armed attacks. This missingness pattern seems more characteristic of random mistakes by respondents or a respondent not having an opinion. Some respondents had a significant number of missing values, which may indicate missingness, not at random. If a respondent had more than 5 missing items per wave, they were removed from the sample.

Since all the multi-item questions would later be averaged, and the items within a scale are highly correlated, this multiple imputation strategy seems reasonable.
The multiple imputation was done once using the algorithm in the R package mi. This imputation method is not model-based and aims to iteratively impute the data from a regression given the rest of the data. The algorithm can also handle categorical data. It is a Bayesian approach that uses Gibbs sampling. The imputation was not weighted, but the demographic covariates had no missing data and were covariates in the multiple imputation. Four multiply imputed data sets were generated using four chains of MCMC iterations used by the R function. The mi package has extensive diagnostics that were examined to show that the algorithm converged and the imputed responses looked reasonable. Using the Rubin rules of multiple imputation, the final estimates are done by fitting the model to each imputed data set and then setting the final estimate of the coefficient to be the average of the coefficients, and the variance of the estimates as the average of the variance of each parameter across the datasets plus the covariance of the parameter between the datasets.
The multiple imputation method used was not weighted or model-based. The method in Quartagno, Carpenter, and Goldstein (2020) is designed for surveys and is weighted and model-based, but this was not ideal for this particular problem. A method that was not dependent on the model was ideal because multiple regressions were run, and the data set had many parameters and observations. In addition, demographic variables were included as covariates in the multiple imputations, which provided a structure to impute the data based on similar respondents.
Since multiple imputation was used, the model was fit on each of the four imputed data sets. Then the results were combined using the Rubin rules (Rubin 1996).
Let Q be the parameter of interest, and $\hat{Q}_1$,$\hat{Q}_2$,$\hat{Q}_3$,$\hat{Q}_4$ be the estimates of Q from each estimate. Let $\hat{\mu_Q}$ be the posterior mean of Q, and $\hat{\sigma^2_Q}$ be the posterior variance of Q. Then:

$$\hat{\mu_Q} = E(Q|Y_{obs}) = E[E(Q|Y_{obs}, Y_{mis})|Y_{obs}] \tag{1}$$

$$= \frac{1}{4}(\hat{Q}_1 + \hat{Q}_2 + \hat{Q}_3 + \hat{Q}_4) \tag{2}$$

$$\hat{\sigma^2_Q} = E[Var(Q|Y_{obs}, Y_{mis})|Y_{obs}] + Var(E(Q|Y_{obs}, Y_{mis})|Y_{obs}) \tag{3}$$

$$= \frac{1}{4}(Var(\hat{Q}_1) + Var(\hat{Q}_2) + Var(\hat{Q}_3) + Var(\hat{Q}_4)) + Var(\frac{1}{4}(\hat{Q}_1 + \hat{Q}_2 + \hat{Q}_3 + \hat{Q}_4)) \tag{4}$$

Where Var is the sample variance operator.

2.2 Multilevel Regression With Poststratification

The GFK sample was designed to be nationally representative, but it had some non-response and did not represent U.S. adults. The Decision Research sample was from a panel recruited online non-randomly and was not nationally representative. One of the goals of this study is to understand the population level perceptions and preferences about terrorism, and to accomplish this, the model must adjust for the non-representativeness of the data. In a Bayesian framework, a standard tool to adjust for non-response is Multilevel Regression with Poststratification (MRP), first detailed in Gelman & Little (1997). MRP involves including various demographic variables (age, gender, education, income, etc.) inside a regression. Often a few interactions of these variables are used in the regression.

Then an estimate can be predicted for each combination of the variables, often called post-stratification cells. Then a weighted average is used to estimate the population mean θ_t from J post-stratification cells, each with size N_j and total population size N, as seen in Formula 1:

$$\hat{\theta} = \sum_{j=1}^{J} \frac{N_j \hat{\theta_j}}{N} \tag{5}$$

Since the weighted average accounts for the population prevalence of each cell, the MRP estimator is representative of the populations even if the sample was not representative. Thus, the primary difference between classical survey weighting and MRP is that MRP is a weighted average of the regression results, but classical survey weighting is a weighted average of respondents.

MRP is especially beneficial in cases where the number of respondents per post-stratification cell is slight, such as Wang et al. (2015), which used an opt-in panel of Xbox users to forecast the 2012 election at the state level. There were 345858 individuals in the survey but 176256 poststratification cells. MRP uses hierarchical priors as part of the model to allow for partial pooling across time. The prior in the Bayesian standpoint provides a starting point for the model on what the parameters should look like. Hierarchical priors allow for the parameters to learn from similar parameters.

In standard MRP, this pooling usually occurs within variables so that the distribution of the coefficient for college-educated individuals learns somewhat from the coefficient for individuals with only a high school education. This "learning" smooths the effects and increases the precision in estimating support at both the population and individual levels. Typically, there are no significant variations across levels of demographic variables. MRP starts by assuming that the effects of the different levels of a demographic variable come from a common distribution of effects similar to a random effect model. But if the data suggest this assumption does not hold, the model can adapt to allow for differences between variable levels.

MRP has been shown to account for differential non-response for both probability and non-probability samples. For example, Wang et al. (2015) used an opt-in panel of Xbox users to forecast the 2012 election, and the forecast was comparable to a standard probability pre-election poll. This example shows that MRP is a natural choice to handle differential non-response in the polls. // Gelman et al. (2016) provided an extension of MRP using multiple independent national polls over 21 years to provide estimates of support for same-sex marriage over time. But since these polls were independent, the model in Gelman et al. (2016) could not be directly applied to this data.

MRP can be extended to incorporate multiple polls over time. The theory behind Formula 1 is not dependent on the model structure, provided that the underlying model is Bayesian and fits with some form of Markov chain Monte Carlo method. We use the following variables in our poststratification: age, gender, age and gender interaction, race, education, and Census geographic region. This creates 512 poststratification cells at each time.

2.3 Primary Model

Let μ_{it} define the mean of the normal distribution representing the response at time t for respondent i. Let s_i denote the response pattern for respondent i: $s_i = 1$ if the respondent was measured only at time 1 in the GFK panel, $s_i = 2$ if the respondent was measured at time 1 and time 7 in the GFK panel, $s_i = 3$ if the respondent answered waves 2-7 in the Decision Research panel, $s_i = 4$ if the respondent answered only wave 7 in the Decision

Research panel.

$\mu_{itk} = \alpha_0 + \alpha^R_{j[i]} + \alpha^A_{j[i]} + \alpha^E_{j[i]} + \alpha^G_{j[i]} + \alpha^L_{j[i]} + \alpha^{A*G}_{j[i]} + X^T_{it}\beta_t$ for time t, and respondent i.

$$Y_{i1} \mid s_i = 1 \sim N\left(\mu_{it}, \sigma^2_{t=1}\right) \tag{6}$$

$$\begin{bmatrix} Y_{i1} \\ Y_{i7} \end{bmatrix} \mid s_i = 2 \sim N\left(\begin{bmatrix} \mu_{i1} \\ \mu_{i7} \end{bmatrix}, \Sigma^2_{s=2}\right) \tag{7}$$

$$\begin{bmatrix} Y_{i2} \\ Y_{i3} \\ Y_{i4} \\ Y_{i5} \\ Y_{i6} \\ Y_{i7} \end{bmatrix} \mid s_i = 3N\left(\begin{bmatrix} \mu_{i2} \\ \mu_{i3} \\ \mu_{i4} \\ \mu_{i5} \\ \mu_{i6} \\ \mu_{i7} \end{bmatrix}, \Sigma^2_{s=3}\right) \tag{8}$$

$$Y_{i7} \mid s_i = 4 \sim N\left(\mu_{i7}, \sigma^2_{t=7}\right) \tag{9}$$

Where $\alpha^R_{j[i]}$ represents the appropriate intercept for the respondents race, $\alpha^A_{j[i]}$ represents the appropriate intercept for the respondents age, $\alpha^E_{j[i]}$ represents the appropriate intercept for the respondents education level, $\alpha^G_{j[i]}$ represents the appropriate intercept for the respondents gender, and $\alpha^{A*G}_{j[i]}$ represents the appropriate intercept for the interaction of the respondents age and gender.
We assign the following prior distributions to the data.

$$\sigma_{\alpha_{pop}} \overset{iid}{\sim} N_+(0,1) \tag{10}$$

$$\sigma_{\alpha_t} \overset{iid}{\sim} lognormal(log(\sigma_{\alpha_{pop}}), 0.1) \tag{11}$$

$$\alpha_{k_t} | \sigma \overset{iid}{\sim} Normal(0, \sigma_{\alpha_t}) \tag{12}$$

$$\tag{13}$$

$$\Sigma_s \; LKJ(2) \tag{14}$$

Where LKJ is a LKJ prior with $\eta = 2$ (Lewandowski 2009).
Variance terms from the normal distribution have the following prior:

$$\sigma_s \sim N_+(0,1) \tag{15}$$

$$\tag{16}$$

We assign the following prior distributions to the regression coefficients where K is a demographic variable such as education, and k is a level of that demographic variable.

$$\sigma_{\alpha_K} \overset{iid}{\sim} N_+(0,1) \tag{17}$$

$$\sigma_{\alpha_t} | \sigma_{\alpha_K} \overset{iid}{\sim} lognormal(log(\sigma_{\alpha_K}), 0.1) \tag{18}$$

$$\alpha_{k_t} | \sigma_{\alpha_K}, \sigma_{\alpha_t} \overset{iid}{\sim} Normal(0, \sigma_{\alpha_t}) \tag{19}$$

$$\tag{20}$$

Where $N_+(0,1)$ is a standard normal distribution truncated to positive numbers.

This framework allows for partial pooling across time and demographic groups even though there are no repeated measurements in the survey. This model fits in the programming language Stan and its interface to R, another statistical language.

2.4 Secondary Model

In the secondary model, we create a new data set that only includes one wave of data per respondent. This new data set reduces the complexity of the model because all the data is statistically independent. The regression model for respondent i at time t is as follows: $Y_i t\ Normal(\alpha_0 + \alpha_R + \alpha_E + \alpha_R + \alpha_A + \alpha_G + \alpha_{A*G}\sigma_t^2)$.
Where $\alpha_{j[i]}^R$ represents the appropriate intercept for the respondents race, $\alpha_{j[i]}^A$ represents the appropriate intercept for the respondents age, $\alpha_{j[i]}^E$ represents the appropriate intercept for the respondents education level, $\alpha_{j[i]}^G$ represents the appropriate intercept for the respondents gender, and $\alpha_{j[i]}^{A*G}$ represents the appropriate intercept for the interaction of the respondents age and gender.
We assign the following prior distributions to the regression coefficients where K is a demographic variable such as education, and k is a level of that demographic variable.

$$\sigma_{\alpha_K} \overset{iid}{\sim} N_+(0,1) \tag{21}$$

$$\sigma_{\alpha_t}|\sigma_{\alpha_K} \overset{iid}{\sim} lognormal(log(\sigma_{\alpha_K}), 0.1) \tag{22}$$

$$\alpha_{k_t}|\sigma_{\alpha_K}, \sigma_{\alpha_t} \overset{iid}{\sim} Normal(0, \sigma_{\alpha_t}) \tag{23}$$

$$\tag{24}$$

Variance terms from the likelihood have the following prior:

$$\sigma_{s_{pop}} \overset{iid}{\sim} N_+(0,1) \tag{25}$$

$$\sigma_{s_t}|\sigma_{s_{pop}} \overset{iid}{\sim} lognormal(log(\sigma_{s_{pop}}), 0.1) \tag{26}$$

$$\tag{27}$$

Where $\sigma_{s_{pop}}$ is a hyperparameter for variance at each time. Where $Normal_+(0,1)$ is a standard normal distribution truncated to positive numbers. This model is fit in the programming language Stan and its interface to R, another statistical language.

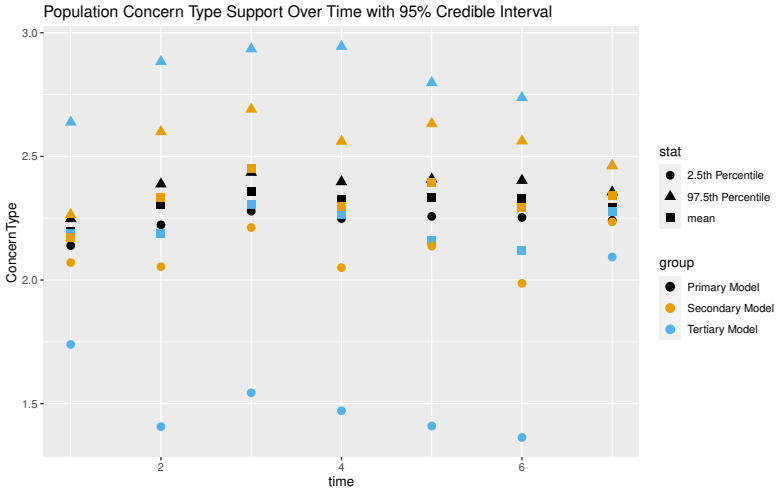
2.5 Tertiary Model

In the third model, we ignore the structure of the repeated measures of the data and use standard MRP fit in the Rstanarm R package. We use the stan_lmer function with its defaults, apply it to each imputed data set, and then apply Rubin's rules to combine the models fit on each data set. The purpose of the tertiary model is to compare the primary and secondary models too.

3. Results

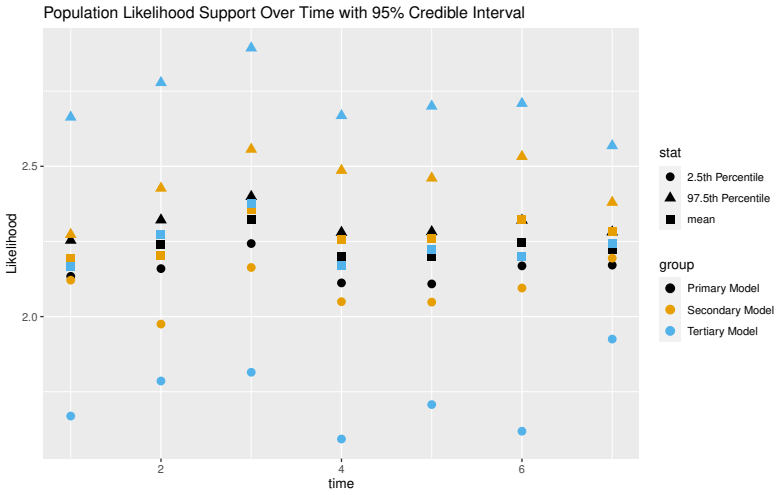
3.1 Population Means Across Time

Now a series of plots of the population-level support is constructed for each variable. The predicted mean and the 2.5th and 97.5th percentile are shown for each model. Since the models are Bayesian, the 2.5th and 97.5th percentile of the posterior distribution comprise a 95% credible interval similar to a confidence interval. Later, the change over time will be discussed in a different series of plots.
The first response variable deals with concern about of eight different types of terrorist attacks. In the below plot, the average concern across the attacks is displayed over time.



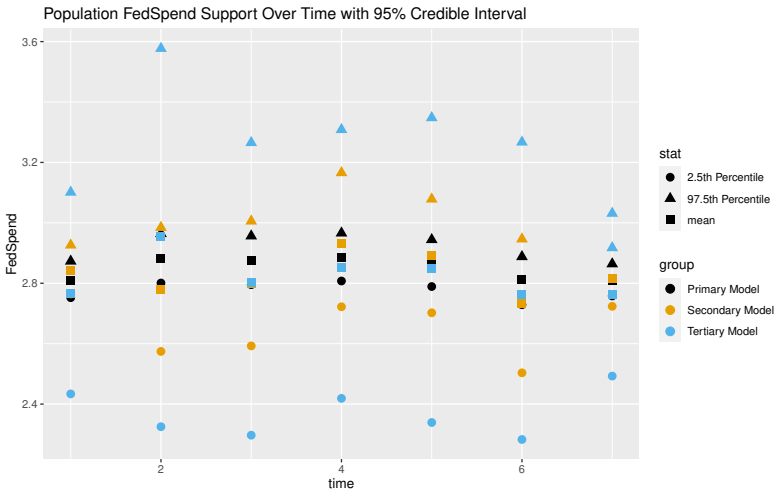
The scale for this question is: not concerned (1), somewhat concerned (2), very concerned (3), extremely concerned (4). We can see that typical values are approximately 2.125-2.375, suggesting that people are somewhat concerned about terrorism on average. Population means were relatively stable, and the difference is not practically significant.

The next response variable deals with perceived likelihood of eight different types of terrorist attacks. In the below plot, the average likelihood across the attacks is displayed over time.



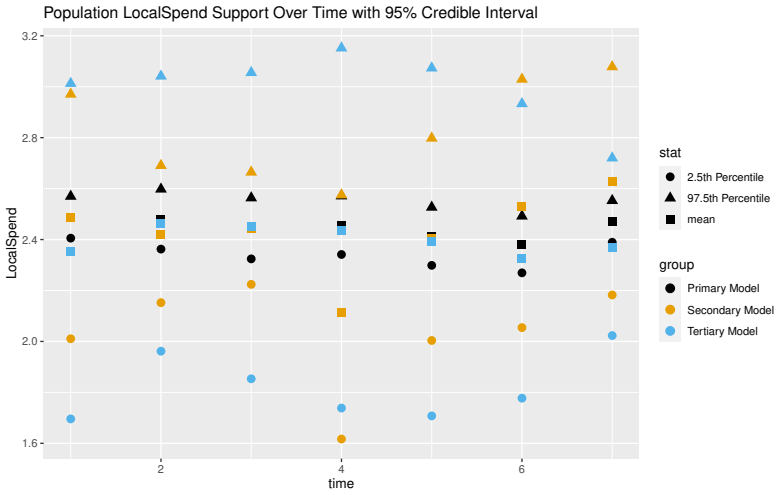
The scale for likelihood was 1: not at all likely, 2 slightly likely, 3 somewhat likely, and 4 very likely. The estimate of population-level support hovers around approximately 2.25, which suggests most of the types of attacks were viewed either as slightly likely or somewhat likely.

The next response variable deals with support for federal governmental spending to prevent a range of eight different types of terrorist attacks. In the below plot, the average support for federal spending across the attacks is displayed over time.



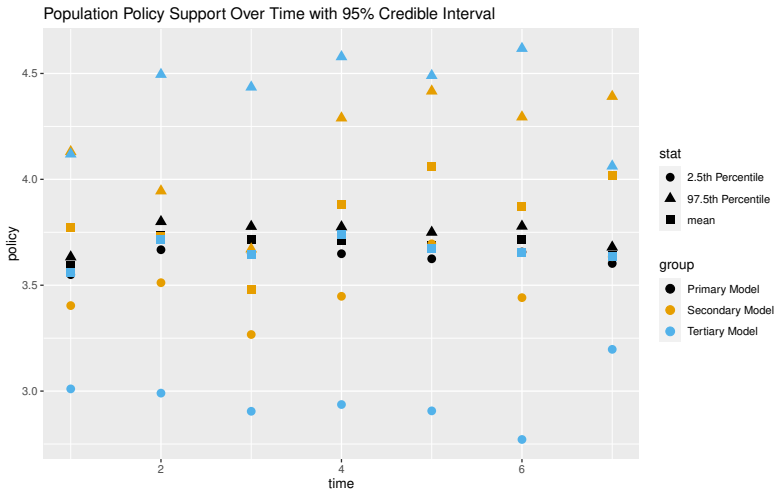
The scale of this question is very low (1), somewhat low (2), pretty high (3), very high (4). We can see that the mean is about 2.8 to 3, corresponding to most items being 3 with a few items as 2. This means that support for federal spending is relatively strong. We also see stability in the estimates across time.

The next response variable deals with support for local governmental spending to prevent a range of eight different types of terrorist attacks. In the below plot, the average support for local spending across the attacks is displayed over time.



The scale of this question is very low (1), somewhat low (2), pretty high (3), very high (4). The average value is approximately 2.5, slightly but not significantly lower than support for federal spending.

The next response variable deals with support for a broad range of counterterrorism policies. In the below plot, the average support for the policies across the policies is displayed over time.

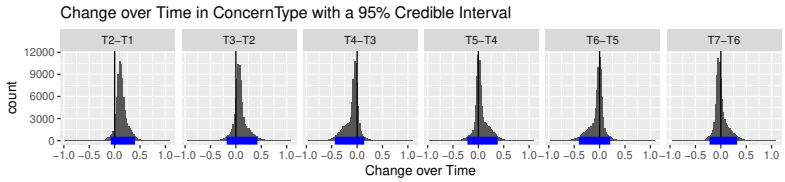


The scale for this question is strongly oppose (1), somewhat oppose (2), neutral (3), somewhat support (4), strongly support (5). We generally see that the policies are more supported than opposed since the average is closer to 4 than 3.

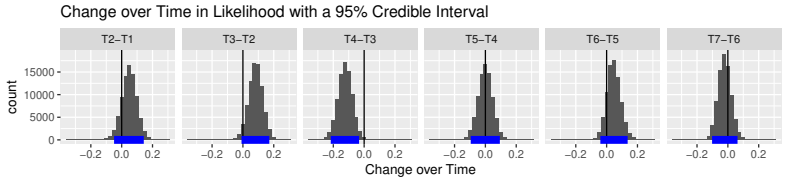
3.2 Change over Time

Next we look at the posterior distributions of change over time from one time point to the next time point. We use the primary model because it uses all information and has the highest precision. We show histograms of the difference between the two waves using samples taken from the model. The histograms have a 95% credible interval in blue below, and have a vertical line at 0. If the credible interval contains 0, it is plausible there was not statistically significant change over time. If the interval does not contain 0 there is evidence for change over time.

First we look at change over time in ConcernType. Here we see that for every difference



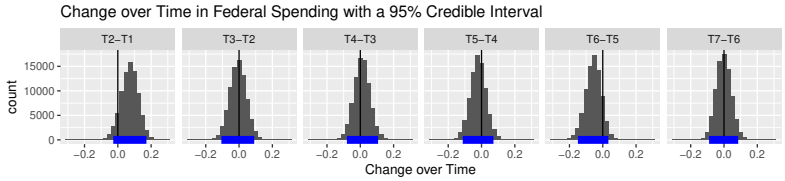
in time points, zero is in all credible intervals, and the median appears to be approximately zero.
Next we look at change over time in Likelihood. We do see here that zero was contained in



the intervals for T2-T1, T5-T4, T6-T5, T7-T6. Between time 3 (July) and time 2 (June) the credible interval is very close to not containing zero. And between time 4 (August) and time 3 we see zero not contained in the interval and a observed decrease. The Orlando terrorist attack occurred after most people in the decision research study took their time two survey.

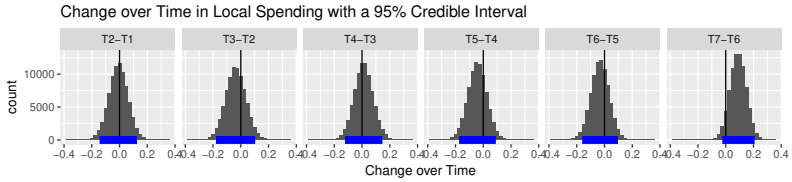
The Orlando Pulse shooting was the deadliest terrorist attack on U.S. soil since 9/11, and was at the time, the deadliest mass shooting in American history (Straub et al. 2017). The June survey was included responses from before and after the Orlando shooting, with 36% occurring after the Orlando attack. And between time 2 and time 3 for all respondents, the Dallas police shooting occurred. There was not another major US terrorist attack until September, so the decrease between time 4 and time 3 could be a decrease in perceived as time from the last attack increased. The September attack was the NY and NJ bombings which resulted in 31 causalities but no deaths. The lesser severity of that attack might explain the lack of change in perceived likelihood.

Next we look at change over time in Federal Spending. Here we see that for every difference



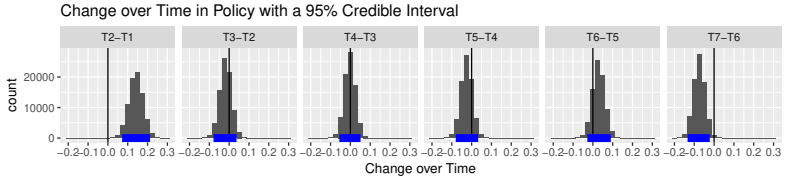
in time points, zero is in all credible intervals, and the median appears to be approximately zero.

Next we look at change over time in Local Spending. Here we see that for every difference



in time points, zero is in all credible intervals, and the median appears to be approximately zero.

Next we look at change over time in Policy Support. Here we see an increase in policy

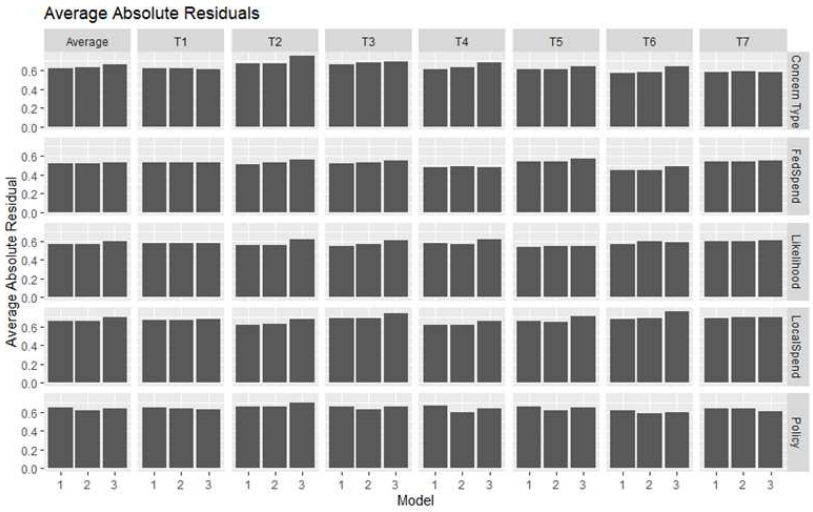


support from time 2 (June) to time 1 (May), and a decrease from time 7 (November) to time 6 (October). About a third of the respondents who took the survey at time 2 did so after Orlando which might explain the change over time. The decrease from November to October might be related to the 2016 Election because the survey at time 7 happened after the election. There wasn't a terrorist attack between time 6 and time 7.

Overall, for the survey questions of interest, we see that for the most part there is no change between wave to wave. These models would have detected a change of just a single item moving a single unit across the population, so if no change is observed then it is very likely no practically significant change over time occurred. Individuals did move, but the movement was in both directions. Since the mean estimate are not practically different in the three models, the partial pooling over time in the primary and secondary model probably did not have practically significant effects on the mean prediction, indicating that the partial pooling has minimal effect on the mean estimates.

3.3 Validating the Methods

The first and second models have a novel structure that has not been tested before on public opinion data. Since it is unclear how good the models are, the data will be split into a testing and training data set to validate the models. Regression models are optimized on the data used to fit them and sometimes do not perform well when applied to new data. If the model performs well on testing data not used to fit the model, this model could likely be applied to similar datasets with similar performance. All models fit 80% of the data so that the remaining 20% can be used to validate the model, which is a standard procedure to evaluate the model better. Berk and Picard (1990) detail the importance of validating regression models with data not used to fit the model. We compute the average absolute residuals of the predictions in the testing data set to measure the model's accuracy. We define this as $\epsilon_{it} = |\hat{Y}_{it} - Y_{it}|$, where $\hat{Y}_{it}$ is the model's prediction for the ith respondent at time t, and $Y_{it}$is the actual value of respondent i. Below is a bar chart of the average absolute residual of each of the three models at each of the seven-time points for all the fit models and an absolute average residual across time.



This bar chart shows that the three models have similar accuracy in predicting individual respondents, with the primary model tending to have slightly lower residuals. The difference in residuals is slightly more noticeable in time points 2-6, where only the decision research respondents were measured. Overall, this plot suggests this model can roughly predict a respondent's views within .5 a point on the Likert scale and that the primary and secondary models appear not to sacrifice the performance of prediction at the individual level.

Combined, the estimates of the population means and average residuals suggest that the primary and secondary model use of partial pooling across time is not harmful for prediction at the population and individual level. The primary and secondary models present an improvement relative to previous work,

4. Discussion

4.1 Reduction in Uncertainty

The 95% credible intervals show a remarkable decrease in uncertainty from the primary model in the population level variables relative to standard MRP. For example, in the likelihood regression at time one, the width of the 95% credible interval for the population

mean is 0.11 in the primary model compared to 1.01 in the tertiary model. It appears there are multiple possible explanations for the decrease in uncertainty. The first is relatively obvious: the primary model considers a significantly larger amount of data, and the partial pooling allows this data to provide more precise estimates at the individual level across time. Evidence for this is observed in the data. For the first MRP cell, the standard deviation in the primary model is 0.058, but in the tertiary model, it is 0.78, which suggests that much of the reduction in uncertainty for the primary model relative to the tertiary model is explained by the reduction in uncertainty at the individual level. This explanation of the uncertainty reductions validates the reduction of uncertainty in the model.

4.2 Future Work

The primary and secondary models are promising methods to track changes over time. The secondary model could be modified to model categorical data. These models also provide a framework to combine non-probability and probability samples. A benefit of these models applied to this data is that the inference from the probability sample helps to improve the inference in the non-probability sample, but the larger sample size of the probability sample carries more weight in the analysis. It would probably be wise in future work aimed at studying changes of time in this method to use a larger probability sample in the beginning and end with non-probability samples in the middle.

4.3 Conclusion

: We see that there was no change over time for almost every variable of interest, and the models could detect relatively minor changes over time. Concern for attacks increased from May to July, and the increase remained steady until November. This finding implies short-term stability and does confirm previous work. This stability was also observed in a context. The perceived likelihood for a terrorist attack was, on average, slightly likely. Most Americans support new counterterrorism policies and federal and local counterterrorism spending. Using the primary and secondary models to track changes over time is a promising approach and can be applied to many other public opinion questions. The success of the second model shows it is not necessary to have repeated measures to benefit from the increased certainty of the partial pooling from the multilevel regression. Further research can explore the use of the primary and secondary models for different lengths of time. In addition, further research should explore adapting this approach to categorical data.

5. References

Gelman, A., Lax, J., Phillips, J., Gabry, J., & Trangucci, R. (2016). Using multilevel regression and poststratification to estimate dynamic public opinion. Unpublished manuscript, Columbia University, 2.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. Journal of multivariate analysis, 100(9), 1989-2001.

Liu, X., J. Mumpower, K. Portney and A. Vedlitz. 2019. "Perceived Risk of Terrorism and Policy Preferences for Government Counterterrorism Spending: Evidence from a U.S. National Panel Survey," Risk Hazards & Crisis in Public Policy, Vol. 10(1):102-135.

Stan Development Team. 2021. Stan Modeling Language Users Guide and Reference Manual, 2.27. https://mc-stan.org

Picard, R. R., & Berk, K. N. (1990). Data splitting. The American Statistician, 44(2), 140-147.

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. International Journal of Forecasting, 31(3), 980-991.

M Quartagno, J R Carpenter, H Goldstein, Multiple Imputation with Survey Weights: A Multilevel Approach, Journal of Survey Statistics and Methodology, Volume 8, Issue 5, November 2020, Pages 965–989, https://doi.org/10.1093/jssam/smz036

Straub, Frank, Jack Cambria, Jane Castor, Ben Gorban, Brett Meade, David Waltemeyer, and Jennifer Zeunik. 2017. Rescue, Response, and Resilience: A Critical Incident Review of the Orlando Public Safety Response to the Attack on the Pulse Nightclub. Critical Response Initiative. Washington, DC: Office of Community Oriented Policing Services.

6. Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Number 1624296. Additional support for this project was provided by the Institute for Science, Technology, and Public Policy, The Bush School of Government and Public Service, Texas A&M University.

Portions of this research were conducted with high performance research computing resources provided by Texas A&M University (https://hprc.tamu.edu).