NCRM National Centre for Research Methods

An assessment of the utility of a Bayesian framework to improve response propensity modelling

Gabriele Durrant, Eliud Kibuchi, Patrick Sturgis, Olga Maslovskaya
University of Southampton

Joint Statistical Meetings 29th July -3rd Aug 2017

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
UNIVERSITY OF Southampton
MANCHESTER 1824 The University of Manchester
THE UNIVERSITY OF EDINBURGH

Outline

- Introduction
- Research question
- Data
- Models
- Results
- Summary
- Discussion

Introduction

- Surveys are faced with decreasing response rates and increasing survey costs
- How do we effectively estimate and understand response behaviour?
- Use of response propensity (RP) models
- How predictive are standard RP models?
  - Tend to have low predictive power (2% to 8% in terms of pseudo R^2)

Introduction

- Improving predictive power of RP models
  - Paradata and new auxiliary variables
- Other proposed method
  - Use of informative priors via Bayesian framework
- Sources of informative priors
  - Expert judgements
  - Previous/historical data

Main Research Question

- Does the specification of informative priors derived from previous wave's data improve the predictive and discriminative power of RP models?

- Aim:
  - To explore the effect of conditioning previous wave data on predictive and discrimination ability of RP models

Data

UK Understanding Society Survey (wave 1 to 5)

- Large-scale household longitudinal survey which aims to capture and explain individual and households health, social and economic stability.

- Consists of:
  - General Population (GP) sample, Ethnic Minority Boost (EMB) sample, British Household Panel Survey (BHPS)

- Initial GP sample: 24,784 households (Great Britain and NI)

- Methodological challenge:
  - Linking households across waves: Individuals are followed using individual ID's (stable across waves) instead of households ID's

Data

- To further explore the effect of borrowing previous wave information on small sample sizes:
  - Subsamples consisting of 10%, 5% and 2% of main sample are obtained

Data

Explanatory and Response Variables

- Paradata and auxiliary data:
  - Call records, interviewer's observation and survey variables

- Response variables

Variable	**Categorisation**
Outcome	Successful (at least one interview) vs Unsuccessful (no interview)
Length of call sequence	Short (up to 6 calls) vs Long (7 + calls)

Models

- Model form: Logistic regression for modelling binary outcome

- Main modelling approach: Bayesian
- Frequentist approach: For evaluation of explanatory capacity of models in terms of pseudo-R^2

- Why Bayesian approach?
  - Allows the coherent and optimal incorporation of prior information as new data become available.
  - This may lead to improved predictions due to better and stable parameter estimates.

Models

- Implementation: Integrated Nested Laplace Approximation (INLA) a fast and accurate approximation method for Bayesian inference

- Prior distributions:
  - Uninformative normal priors for the initial wave models
  - Informative priors derived from previous wave data posterior results
  - Misspecified informative priors for global sensitivity analysis – allows comparison of posterior results over variations in prior variance

- Model selection:
  - Watanabe-Akaike information criterion for Bayesian

Models

Different specifications prior distributions used for each wave

Prior type	Specification of prior distribution for regression parameters in wave n	Model name
Vague	$\beta_k^n \sim N(0,0.001)$	M1
Informative	$\beta_k^n \sim \mathrm{N}\left(\beta_k^{n-1}, \left(\sigma_k^{n-1}\right)^2\right)$	M2
	$\beta_k^n \sim \mathrm{N}\left(\beta_k^{n-1}, \left(\sigma_k^{n-1} \times 0.1\right)^2\right)$	M3
	$\beta_k^n \sim \mathrm{N}\left(\beta_k^{n-1}, \left(\sigma_k^{n-1} \times 2\right)^2\right)$	M4
	$\beta_k^n \sim \mathrm{N}\left(\beta_k^{n-1}, \left(\sigma_k^{n-1} \times 5\right)^2\right)$	M5
	$\beta_k^n \sim \mathrm{N}\left(\beta_k^{n-1}, \left(\sigma_k^{n-1} \times 10\right)^2\right)$	M6
	$\beta_k^n \sim N\left(\beta_k^{n-1}, \left(\sigma_k^{n-1} \times 100\right)^2\right)$	M7

Where n and k and represents wave and regression parameters respectively

Models

Evaluation and model performance

- Aim to explore the ability of models to predict final call outcome and length of call sequence.

- Model fit:
  - Watanabe Akaike Information Criteria (WAIC) for Bayesian
  - Akaike information Criteria (AIC) for frequentist

- Accuracy of models evaluated using out-of-sample testing
  - Discrimination (sensitivity and specificity)
  - Prediction (positive and negative predicted value)
  - Area under the Receiver Operating characteristic (ROC) curve

Results

Evaluation criteria for frequentist models using AIC, Nagelkerke's pseudo R^2 and WAIC for Bayesian models for waves 2 and 3

Wave	Model	AIC	Nagelkerke R^2 (%)	WAIC
2	Frequentist	12559.00	7.40	-
	M1	-	-	12561.34
3	frequentist	8700.50	4.91	-
	M1	-	-	8701.35
	M2	-	-	8704.27
	M3	-	-	8856.92
	M4	-	-	8692.62
	M5	-	-	8695.86
	M6	-	-	8699.08
	M7	-	-	8701.32

Results

Results of classification table and Area under the ROC curve, sensitivity, specifity, positive predictive values (PPV) and negative predictive values (NPV) for the final call outcome for waves 2 and 3

wave	Modelling approach	Classification (%)	AUC (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
2	frequentist	77.5	64.3	52.0	78.0	3.0	99.0
	M1	77.5	64.3	53.0	78.0	3.0	99.0
3	frequentist	81.7	62.4	25.0	82.0	0.0	100.0
	M1	81.7	62.4	27.0	82.0	0.0	100.0
	M2	81.7	62.0	50.0	82.0	0.0	100.0
	M3	81.7	56.5	Nan	82.0	0.0	100.0
	M4	81.7	62.4	40.0	82.0	0.0	100.0
	M5	81.7	62.4	30.0	82.0	0.0	100.0
	M6	81.7	62.4	27.0	82.0	0.0	100.0
	M7	81.7	62.3	27.0	82.0	0.0	100.0

Summary

- Use of informative priors based on previous wave data does not significantly improve RP models fit in terms of WAIC. This is because the likelihood (current wave data) dominates the posterior and variations in priors means have almost no impact on estimated posterior means.
- Misspecified informative priors with tight variance i.e. $(\beta_k^n \sim \mathrm{N}\left(\beta_k^{n-1}, \left(\sigma_k^{n-1} \times 0.1\right)^2\right)$ have an overall poor model fit. This is because a tight prior on the parameter β_k^{n-1} leads to a tight posterior for β_k^n. The prior is very peaked relative to the posterior (prior is much informative compared to data) which makes prior to dominate the posterior.
- Use of informative priors does not improve predictive and discrimination ability of RP models because posterior is dominated by the likelihood.

Summary

- The power of informative priors derived from previous waves may also have been affected by the changing data generating mechanisms due to responsive strategies introduced during survey. This implies that time would be an important consideration when choosing data to derive informative priors from (i.e. quarterly data may be more informative than yearly data)

- Effectiveness of the informative priors is also influenced by the explanatory capacity of the variables used in the model. When covariates variables in the model have weak explanatory strength on the outcome of interest , their informative priors also tend to be weak. This implies effective borrowing of previous wave information is dependent on ability of paradata and auxiliary to explain variability of outcome of interest.

Conclusions

- Results contribute to better understanding of the use of previous wave data as informative priors for subsequent waves analyses in longitudinal studies
- These findings set framework for exploration of other types of informative priors such as those elicited from expert judgements