A Gaussian process model for response time in conjoint surveys

Elea McDonnell Feit (joint work with Zhiya Zuo and Hongjun Ye)

21 July 2021

Agenda

What is a conjoint survey?

Response time in conjoint surveys

Application: electricity rate plan conjoint survey

What is a conjoint survey?

Designing new products

Should the minivan have a smaller cargo area so that we can give more leg room to the passengers?

Should we make the minivan larger, even though the fuel economy will go down?

Answers to these questions depend on what customers want

Just ask the customer?

Better designers spend time talking to potential customers about what they want and that is sort-of helpful

But customers typically want “everything” and if you listen to them you end up with “The Homer”

Estimating customer preferences with conjoint surveys

1. Ask customers to make 8–25 hypothetical product choices where product attributes are varied

Which of the following minivans would you buy?
Assume all three minivans are identical other than the features listed below.

	Option 1	**Option 2**	**Option 3**
	6 passengers	8 passengers	6 passengers
	2 ft. cargo area	3 ft. cargo area	3 ft. cargo area
	gas engine	hybrid engine	gas engine
	$35,000	$30,000	$30,000
I prefer (check one):	☐	☐	☑

2. Use a model to infer customer preferences for features from the choices

See Orme and Chrzan [2017] for a practical introduction to conjoint

Model for conjoint survey data

We assume the probability of choosing alternative j in task i is:

$$p(y_i = j) = \frac{\exp(\beta_r' x_{ij})}{\sum_{j'=1}^{J} \exp(\beta_r' x_{ij'})}$$

$$\beta_r \sim N_K(\theta, \Sigma)$$

x_{ij} is a vector of factor codings for the attributes of alternative j
β_r is a vector of estimated parameters for the respondent

Model for conjoint survey data

We assume the probability of choosing alternative j in task i is:

$$p(y_i = j) = \frac{\exp(\beta_r' x_{ij})}{\sum_{j'=1}^{J} \exp(\beta_r' x_{ij'})}$$

$$\beta_r \sim N_K(\theta, \Sigma)$$

x_{ij} is a vector of factor codings for the attributes of alternative j
β_r is a vector of estimated parameters for the respondent

The hierarchical prior regularizes the estimates across respondents via Bayesian shrinkage [Lenk et al., 1996]

Model for conjoint survey data

We assume the probability of choosing alternative j in task i is:

$$p(y_i = j) = \frac{\exp(\beta_r' x_{ij})}{\sum_{j'=1}^{J} \exp(\beta_r' x_{ij'})}$$

$$\beta_r \sim N_K(\theta, \Sigma)$$

x_{ij} is a vector of factor codings for the attributes of alternative j
β_r is a vector of estimated parameters for the respondent

The hierarchical prior regularizes the estimates across respondents via Bayesian shrinkage [Lenk et al., 1996]

$\beta_r' x_{ij}$ can be interpreted as the utility in a random utility model [McFadden and Train, 2000]

Example estimated preferences (not real data!)

We might find estimates for θ like this:

```
Coefficients :
          Estimate Std. Error  t-value  Pr(>|t|)
seat7    -0.535280   0.062360  -8.5837 < 2.2e-16 ***
seat8    -0.305840   0.061129  -5.0032 5.638e-07 ***
cargo3ft  0.477449   0.050888   9.3824 < 2.2e-16 ***
enghyb   -0.811282   0.060130 -13.4921 < 2.2e-16 ***
engelec  -1.530762   0.067456 -22.6926 < 2.2e-16 ***
price35  -0.913656   0.060601 -15.0765 < 2.2e-16 ***
price40  -1.725851   0.069631 -24.7856 < 2.2e-16 ***
```

6 seats (the base level) is preferred to 7 or 8 seat
More cargo space is preferred to less
Conventional engines are preferred to electric (by a lot!)
Lower prices are preferred (by a lot!)

Software for conjoint analysis

Commercial software (including designing and fielding the survey) is available from Sawtooth Software, Conjoint.ly and Qualtrics

mlogit or logitr packages provide estimation of HMNL by maximum likelihood

bayesm package provides an efficient Gibbs sampler for producing posterior samples

For an implementation in Stan, see my tutorial on choice modeling in Stan with Kevin Van Horn

The model is closely related to Savage, Betancourt and Vasserman's aggregate random coefficients logit in Stan

Response time in conjoint surveys

Motivation

Data from a conjoint survey does not always provide precise information about the parameters [Lenk et al., 1996, Sándor and Wedel, 2002]

We can observe the response time for a choice at essentially zero cost

If we can relate the response time to features of the choice task, then we might be able to extract additional information about preferences from the response time

Which choice will take longer to make?

Task A

job offer 1	job offer 2
high salary	low salary
8 working hours per day	8 working hours per day
5 working days	5 working days

Task B

job offer 1	job offer 2
high salary	low salary
9 working hours per day	6 working hours per day
5 working days	4 working days

Intuitively, Task A should be faster than Task B

Task order
Respondents adapt to repeated choices and answer subsequent questions faster [Haaijer et al., 2000, Otter et al., 2008]

$$z_{i1} = t_i$$

Utility difference between alternatives
Decision field theory [Busemeyer and Townsend, 1993] predicts that when one alternative is much better, choice will be faster and this has been confirmed in conjoint survey data [Diederich, 2003, Otter et al., 2008]

$$z_{i2} = |\beta_r x_{i1} - \beta_r x_{i2}|$$

More features related to response time

Average utility

People choose faster when faced with two attractive alternatives (approach-approach) [Diederich, 2003]

$$z_{i3} = \frac{1}{2}(\beta_r x_{i1} + \beta_r x_{i2})$$

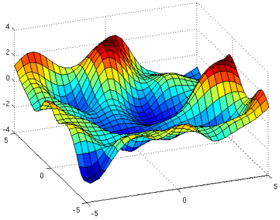
Attribute differences

When answering, respondents engage in an information search [Meißner and Decker, 2010, Shi et al., 2013], which is more time-consuming when many of the attributes are different

$$z_{i4} = \sum_k \beta_{rk} |x_{i1k} - x_{i2k}|$$

Relating features to response time

Example estimated 2-input Gaussian process



Source: Kernel Cookbook

A **Gaussian process (GP)** is a Bayesian approach for modeling a function that allows us to flexibly relate features of the choice task to the response time

The GP can capture ceiling and floor effects, “inverse U” shapes, “S” shapes, etc. in the relationship between the features and response time

Gaussian process for response time

The vector of response times for each task is modeled as a multivariate normal with a covariance matrix K

$$(RT_1, ..., RT_N)' \sim \mathcal{N}_N(0, K((z_1, ..., z_N)|\alpha, \rho_d, \sigma))$$

The covariance K is a function of the features of each choice task z_i. We use the popular squared exponential kernel:

$$K(z_i, z_{i'}|\alpha, \rho_d, \sigma) = \alpha^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{4}\frac{1}{\rho_d^2}(z_{id} - z_{i'd})\right) + I(i = i')\sigma^2$$

where α determines the average distance of the predicted response time from the mean response time, ρ_d determines how much the function changes along the dimension d, and σ is the noisiness of the response

Integrated model for choice and response time

Choice

$$p(y_i = j) = \frac{\exp(\beta_r x_{ij})}{\sum_{j'=1}^{J} \exp(\beta_r x_{ij'})} \qquad \beta_r \sim N_K(\theta, \Sigma)$$

Response time

$$RT \sim \mathcal{N}_N(0, K(z)) \quad K(z_i, z_{i'}) = \alpha^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{4}\frac{1}{\rho_d^2}(z_{id} - z_{i'd})\right) + I(i = i')\sigma^2$$

$$z_i = \begin{cases} t_i & \text{task order} \\ \frac{1}{2}(\beta_r x_{i1} + \beta_r x_{i2}) & \text{utility difference} \\ |\beta_r x_{i1} + \beta_r x_{i2}| & \text{average utility} \\ \sum_k \beta_{rk}|x_{i1k} - x_{i2k}| & \text{attribute difference} \end{cases}$$

Latent utility $\beta_r X_{ij}$ links together choice and response time

Stan code 1

```
transformed parameters {
  // for the multilogit
  cov_matrix[K] Sigma = quad_form_diag(Omega, tau);
  // util_matrix is a 4 dimenional matrix where
  // the 1st column is alter-based
  // the 2nd column is attr-based
  // the 3rd column is avg attractivess
  // the 4th column is Q
  matrix[N_seen, 4] util_matrix_seen;
  matrix[N, 4] util_matrix;
  matrix[N, N] L_K;
  vector[N] f_rt;
  for (i in 1:N_seen) {
    // identify respondent
    int r;
    r = RESPONDENT[i];
    util_matrix_seen[i, 1] = fabs(X[i,2]*Beta[,r]-X[i,1]*Beta[,r]);
    util_matrix_seen[i, 2] = sum(fabs(X[i,2].*Beta'[r,]-X[i,1].*Beta'[r,]));
    util_matrix_seen[i, 3] = (X[i,2]*Beta[,r]+X[i,1]*Beta[,r])/2;
    util_matrix_seen[i, 4] = Q[i];
  }
  // from 1 -> N_seen all observations;
  // from N_seen+1 -> N all grid values
  util_matrix = append_row(util_matrix_seen, util_matrix_pred);
  L_K = L_cov_exp_quad_ARD(util_matrix, alpha_rt, rho_rt, delta_rt);
  f_rt = L_K * eta_rt;
}
```

Stan code 2

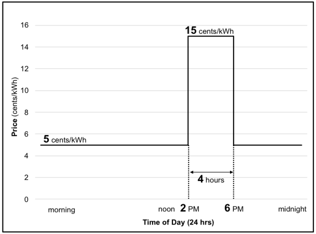
```
model {
  // for RT using GP
  // a flag for identifying which subject has been done before
  int flag[R];
  flag = rep_array(-1, R);

  // MULTILOGIT priors
  to_vector(Theta) ~ normal(0, 1);
  tau ~ normal(0, 0.3);
  Omega ~ lkj_corr(2);
  // RT priors
  rho_rt ~ inv_gamma(5, 5);
  alpha_rt ~ std_normal();
  sigma_rt ~ std_normal();
  eta_rt ~ std_normal();

  // drawing samples
  for (i in 1:N_train) {
    // identify respondent
    int r;
    r = RESPONDENT[i];
    // Theta: K by G
    // Z: G by R
    // Beta: K by R - individualized part worths
    if (flag[r]<0) {
        Beta[,r] ~ multi_normal(Theta*Z[,r], Sigma);
        flag[r] = 1;
    }
    // sample Y
    Y[i] ~ categorical_logit(X[i]*Beta[,r]);
  }
  RT ~ normal(f_rt[1:N_train], sigma_rt);
}
```

Application: electricity rate plan conjoint survey

Electricity rate plan choices

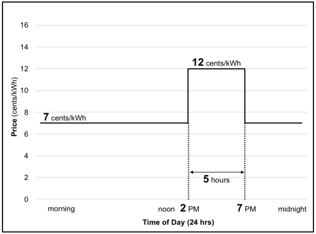
Local utility collaborated with Drexel Solutions Institute to understand how customers react to a new rate plan with a peak period

Plan A
Peak Hour: 2 pm – 6 pm
Peak Duration: 4 hours
Peak Rate: 15 cents/kWh



A

Plan B
Peak Hour: 2 pm – 7 pm
Peak Duration: 5 hours
Peak Rate: 12 cents/kWh



B

45 respondents each answered 14 binary choice tasks

Estimated choice parameters

Rates (prices) are more important than the duration of the peak rate

		Estimate	Post SD
Peak Rate	θ_1	1.806	0.021
Off-Peak Rate	θ_2	2.211	0.023
Peak Duration	θ_3	1.045	0.024
var(Peak Rate)	Σ_{11}	0.853	0.042
var(Off-Peak Rate)	Σ_{22}	1.120	0.031
var(Peak Duration)	Σ_{33}	0.663	0.021
cor(Peak, Off-Peak)	Ω_{12}	0.128	0.008
cor(Peak, Duration)	Ω_{13}	0.162	0.034
cor(Off-Peak, Duration)	Ω_{23}	0.102	0.026

Preferences vary substantially between respondents

Estimated response time parameters

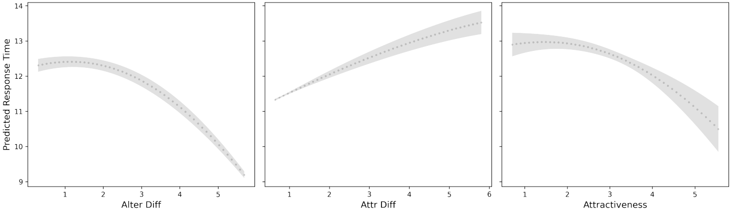
Average response time varies with the features

Individual response times are quite noisy around the average

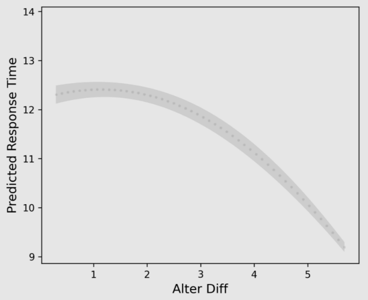
		Estimate	Post SD
Amplitude	α	2.519	0.024
Noise	σ	5.685	0.025
Utility Difference	ρ_1	7.070	0.263
Attrib. Difference	ρ_2	6.335	0.716
Average Utility	ρ_3	6.355	0.275
Task Order	ρ_4	3.766	0.081

GP response time

To understand the relationship between the four features and response time, we plot conditional response time predictions (slices)



Utility difference and response time

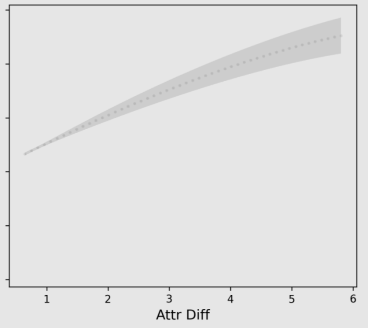


Conditional on $z_{i1} = 3$, $z_{i3} = 1.5$, $z_{i4} = 1.5$

$\uparrow$ utility difference $\Longrightarrow$
$\downarrow$ response time

ceiling effect

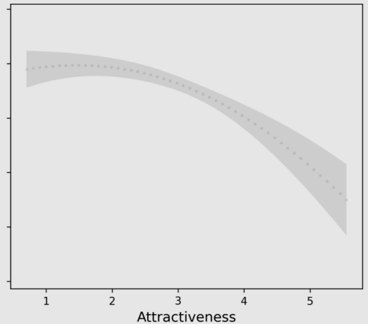
Attribute differences and response time



Conditional on $z_{i1} = 3$, $z_{i2} = 1.5$, $z_{i4} = 1.5$

$\uparrow$ attribute differences $\Longrightarrow$ $\uparrow$ response time

Average utility and response time



Conditional on $z_{i1} = 3$, $z_{i2} = 1.5$, $z_{i3} = 1.5$

$\uparrow$ average utility $\implies$

$\downarrow$ response time

Choice model parameters

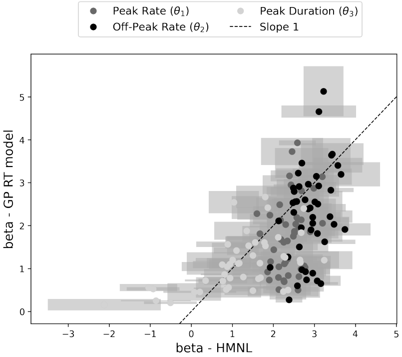
Adding response time does not seem to change our understanding of which attributes are important

But our estimates are more precise and we find there is more heterogeneity

		HMNL+ GP RT		Standard HMNL	
		Estimate	Post SD	Estimate	Post SD
Peak Rate	θ_1	1.806	0.021	2.329	0.068
Off-Peak Rate	θ_2	2.211	0.023	2.883	0.130
Peak Duration	θ_3	1.045	0.024	1.078	0.171
var(Peak Rate)	Σ_{11}	0.853	0.042	0.406	0.106
var(Off-Peak Rate)	Σ_{22}	1.120	0.031	0.425	0.091
var(Peak Duration)	Σ_{33}	0.663	0.021	1.359	0.205
cor(Peak, Off-Peak)	Ω_{12}	0.128	0.008	-0.041	0.139
cor(Peak, Duration)	Ω_{13}	0.162	0.034	-0.005	0.143
cor(Off-Peak, Duration)	Ω_{23}	0.102	0.026	0.094	0.139

Individual preferences (β_r) HMNL versus HMNL + RT GP

Individual preferences are more precisely estimated when response time is included in the model



What if we only observe response time and not the choice?

Estimating the model from response time RT_i without the observed choices y_i, we can still recover attribute preferences

		HMNL + GP RT		HMNL + GP RT without y_i	
		Estimate	Post SD	Estimate	Post SD
Attribute Preferences					
Peak Rate	θ_1	1.806	0.021	1.145	0.015
Off-Peak Rate	θ_2	2.211	0.023	1.270	0.020
Peak Duration	θ_3	1.045	0.024	1.641	0.001
var(Peak Rate)	Σ_{11}	0.853	0.042	1.029	0.025
var(Off-Peak Rate)	Σ_{22}	1.120	0.031	1.417	0.012
var(Peak Duration)	Σ_{33}	0.663	0.021	1.203	0.019
cor(Peak, Off-Peak)	Ω_{12}	0.128	0.008	-0.084	0.010
cor(Peak, Duration)	Ω_{13}	0.162	0.034	0.188	0.013
cor(Off-Peak, Duration)	Ω_{23}	0.102	0.026	-0.047	0.001

Out-of-sample predictive performance

The HMNL + GP RT model does a slighly worse job at predicting choices than model fitted to choice data alone

	Mean Squared Error	
	y_i	RT_i
Standard HMNL	0.078	-
HMNL + RT GP	0.130	35.550
HMNL + RT GP without observed y_i	0.264	45.166

But we can predict choice pretty well from response time alone

Summary

We learned about conjoint surveys, a tool for understanding consumer preferences for attributes

We developed an integrated model of choice and response time for conjoint surveys

- Better understanding of response time and decision making mechanism

We applied this to data from a conjoint survey on electric rate plans

Conjoint practitioners should be collecting and using response time

- Better choice predictions

Future work

Fit this model with other conjoint data sets (Do you have one?)

Extend the model to choices from sets of three or more alternatives

Figure out better ways to visualize the multi-input GP

Thanks!

Elea McDonnell Feit
with Zhiya Zuo & Hongjun Ye

Papers and tutorials at eleafeit.com

Reach me at eleafeit@gmail.com or @eleafeit on Twitter

References i

J. R. Busemeyer and J. T. Townsend. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, 100(3):432, 1993.

A. Diederich. Decision making under conflict: Decision time as a measure of conflict strength. *Psychonomic bulletin & review*, 10(1): 167–176, 2003.

R. Haaijer, W. Kamakura, and M. Wedel. Response latencies in the analysis of conjoint choice experiments. *Journal of Marketing Research*, 37(3):376–382, 2000.

P. J. Lenk, W. S. DeSarbo, P. E. Green, and M. R. Young. Hierarchical bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science*, 15(2):173–191, 1996.

D. McFadden and K. Train. Mixed mnl models for discrete response. *Journal of applied Econometrics*, 15(5):447–470, 2000.

M. Meißner and R. Decker. Eye-tracking information processing in choice-based conjoint analysis. *International Journal of Market Research*, 52(5):593–612, 2010.

B. Orme and K. Chrzan. *Becoming an Expert in Conjoint Analysis: Choice Modeling for Pros*. Sawtooth Software, 2017. ISBN 9780999367704. URL https://books.google.com/books?id=3MteswEACAAJ.

T. Otter, G. M. Allenby, and T. Van Zandt. An integrated model of discrete choice and response time. *Journal of Marketing Research*, 45(5):593–607, 2008.

Z. Sándor and M. Wedel. Profile construction in experimental choice designs for mixed logit models. *Marketing Science*, 21(4):455–475, 2002.

S. W. Shi, M. Wedel, and F. Pieters. Information acquisition during online decision making: A model-based exploration using eye-tracking data. *Management Science*, 59(5):1009–1026, 2013.