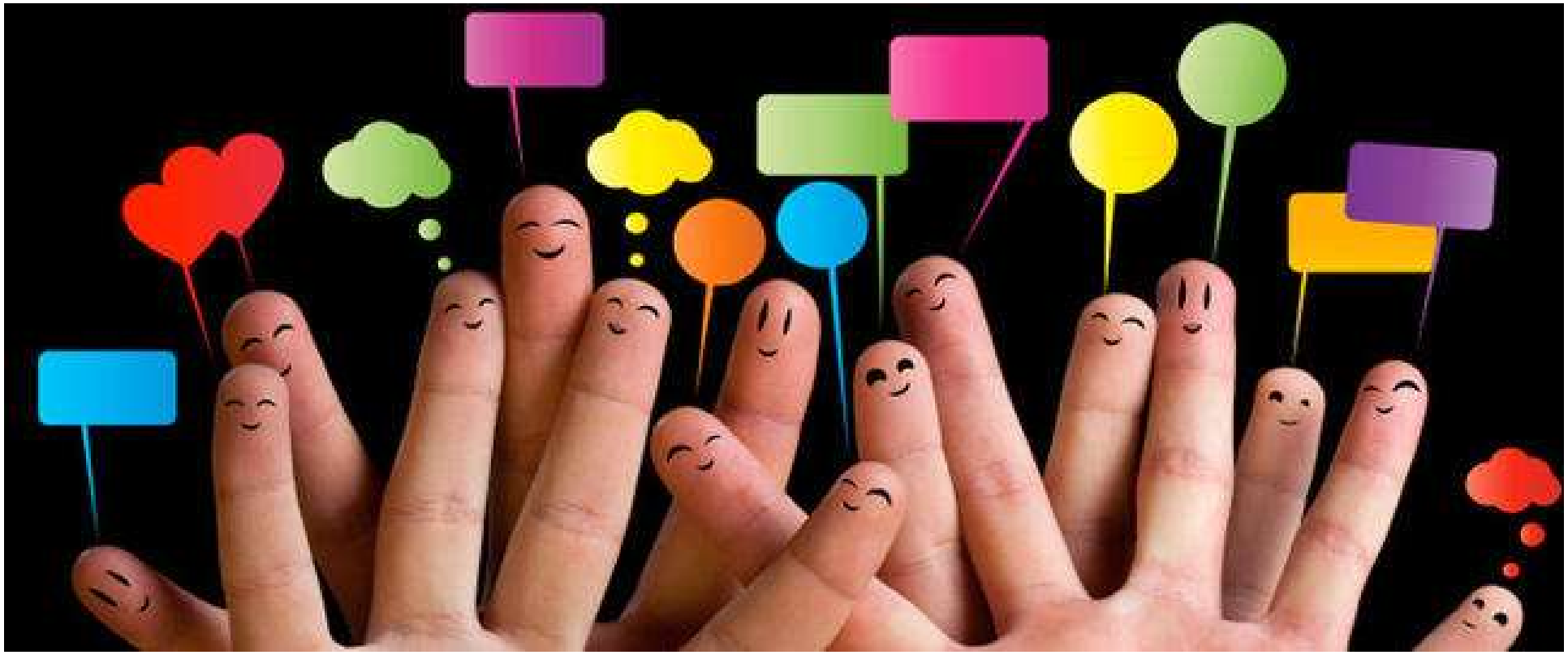


Revisiting Ensembles - A Straightforward Approach to Increase Model Accuracy?

bms marketing research + strategy

Kevin‘s Ensembles

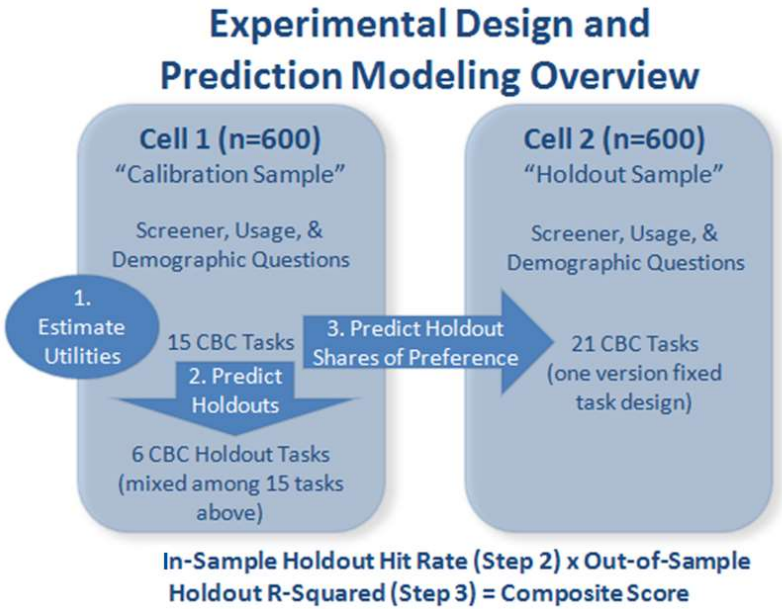
- In his 2015 Sawtooth Conference Paper Kevin Lattery presented a comprehensive work on ensemble techniques in conjoint analysis
  (A Machine Learning Approach to Conjoint Analysis: Boosting and Blending Ensembles, 2015)
- Lattery demonstrated that merging several models to create ensemble solutions could increase prediction accuracy for Latent-Class
- Blending and Boosting Techniques were used to find out how ensembles improve choice model predictions
- Lattery concluded that blending with mean values works well and further that machine learning approaches (Neural Networks) as well as boosting slightly further increase the model fit
- However, as it is very time and computational intensive to achieve such slight improvements, the question remained if this effort really pays off

Bryan‘s Ensembles

- Orme did the same for HB models and again for a diverse set of choice models – in the second case merging the predictions from models that were in part ensemble solutions themselves (Comment on Lattery’s Conjoint Analysis Ensembles, 2015 and Findings of the 2016 Sawtooth Software CBC Modeling Prize Competition, 2016)
- Whereas these papers have shown the general potential of this approach, it also became obvious that it might require some time and effort to create ensembles that outperform standard solutions by a relevant margin

The 2016 Sawtooth Software Predictive Modeling Competition

- Sawtooth Software provided the dataset and hosted this modeling competition:



Source: Orme 2017

Predictive Modeling Competition: Models & Winners

- 5 of the Top 10 Teams used Ensemble Solutions:

Rank	Team	Methods & Model Specification
1	MIReS	HB (GAUSS implementation), interaction between Stateroom and Ship Amenities, budget constraints, fuzzy consideration set model
2	Nutcracker	HB (Sawtooth Software) 16 separate models in an ensemble, varying priors, covariates, utility constraints, and attribute codings
3	SKIM 2	HB (Sawtooth Software) single model for hit rates with ASC for position effect; ensemble of HB solutions for share predictions
4	Landsberger Strasse	HB (R – bayesm with a Dirichlet Process Prior)
5	Scooter-QX	Latent Class Analysis (Q Software) ensemble of 20 solutions using a mixture modeling approach, distributional assumption: Multivariate Normal - Full Covariance
6	Sawtooth 3	HB and Latent Class (Sawtooth Software) ensemble of 40 utility runs with varying starting points for LC and varying covariates for HB, default part-worth coding
7	Sawtooth 2	HB (Sawtooth Software) single priors-optimized run, default part-worth coding
8	Illuminas	HB (Sawtooth Software) single run, default priors, with alt-specific position effects
9	Landsberger Strasse 2	HB (Sawtooth Software) 4 separate models in an ensemble with different covariates for hit rates; shares of preference based on HB models run separately within 5 LC segments
10	Yoda	HB (Sawtooth Software), survey questions about preferences for levels of unordered attributes coded as augmented tasks, some utility constraints

Source: Orme 2017

Findings of the Competition

- Tuning HB to have lower prior variance than the CBC/HB software defaults improved the predictions
- With HB utilities (tuned for optimal priors), logit rule (Share of Preference) on the draws and RFC had essentially the same predictive performance
- Ensembles help to improve Within-Sample Hitrate
- Ensembles help to improve Out-of-Sample Hitrate
- Ensembles across all team solutions beats the single best team
- Overall finding is, that the improvements are not really very large (0.612/0.624)
- ‘All this extra ensembling work is interesting and gives the tiny edge needed to win a competition like this, but the big question is whether it is practical to do it for actual client work where speed and efficiency in project execution is valued’ (Orme 2017)

Motivation for this Paper

- As we were somewhat successful in the Sawtooth Modeling Prize Competition by achieving Best Within-Sample Hit Rate with a rather quick and straightforward approach - blending a small set of models that only differed with regard to the covariates applied during estimation - we wondered if we could replicate these results
- We also want to evaluate the potential this solution has on Out Of Sample Prediction Accuracy and want to revisit the question ‘Is it all worth it?’, i.e. discuss the ratio of time/effort put into the ensemble solution compared to the increase in prediction accuracy we can achieve
- Based on three comparable price conjoint studies (FMCG studies) with comparable sample sizes and almost equal sets of covariates (purchase behavior and demographic information), we replicated our approach and explored the effects on Within and Out-Of-Sample prediction accuracy

Objectives of this Paper

- The sample of each study was split for Within-Sample and Out-of-Sample Validation. With this setup we

1. Evaluate if creating ensembles leads to an increase in prediction accuracy compared to a standard Sawtooth HB Model with default settings, increased # of iterations and a prior-optimized Model (Optimization via Model Explorer).

2. Explore if the accuracy of predicted choices indeed increases and if the increase can be considered relevant in comparison to the standard solution or another low-effort solution (prior-optimized HB).

Datasets for the Exercise

- All Datasets based on Price Only, FMCG studies with comparable number of respondents, SKUs and tasks

Project	N=	SKUs	Price Points per SKU	Tasks/Concept per Task	Model Specifics	Covariates
FMCG 1	1055	32	5-6	15/4-9+None	2 Sets of Tasks (5+10)	Socio-demographic, Purchase Behavior
FMCG 2	1098	30	6	15/4-11+None	2 Sets of Tasks (5+10)	Socio-demographic Purchase Behavior
FMCG 3	1000	15	3-6	15/5-8+None	3 Sets of Tasks (3x5), Relative Pricing	Socio-demographic, Purchase Behavior

- FMCG 1 = 1 SKU + 32 Price Attributes + NONE = 169 parameter
- FMCG 2 = 1 SKU + 30 Price Attributes + NONE = 180 parameter
- FMCG 3 = 1 SKU + 15 Price Attributes + NONE = 78 parameter
- All three studies complex in number of parameters, but easy to answer choice task with only 2 attributes in display (SKU, Price)

Validation Benchmarks

1. **HB standard settings**
   1. 10.000 / 10.000
   2. Prior Variance 1.0; 5 Degrees of Freedom
   3. Acceptance Rate 30%

2. **HB increased # of iteration**
   1. 190.000 / 10.000
   2. Prior Variance 1.0; 5 Degrees of Freedom
   3. Acceptance Rate 30%

3. **Prior optimized HB**
   1. Sample Size = original; Sampling replication count 10
   2. 2.000 / 5.000
   3. Perform search for optimal prior variance and degrees of freedom
   4. 0.8,1.0,1.2,1.4,1.6 / 10,20,30,40,50

Benchmark 3 Prior optimized HB: Prior Optimal Search

- Optimal prior search only indicates potential for small improvements

FMCG 1

```
Hit Rate Summary across Replicates:

                Prior D.F.
                10      20      30      40      50
Prior    0.80   0.56110 0.56202 0.56404 0.56193 0.56532
Var      1.00   0.56138 0.56156 0.56239 0.56376 0.56642
         1.20   0.56936 0.56560 0.56376 0.56312 0.56202
         1.40   0.56211 0.56037 0.56789 0.56376 0.56532
         1.60   0.56495 0.56468 0.57128 0.56220 0.56404

Optimal Priors from Grid Search, Maximum Hit Rate (0.571284403669725) Found at:
Prior Variance: 1.6
Prior degrees of freedom: 30
```

FMCG 2

```
Hit Rate Summary across Replicates:

                Prior D.F.
                10      20      30      40      50
Prior    0.80   0.45963 0.46852 0.45528 0.46583 0.46639
Var      1.00   0.46630 0.46731 0.46824 0.47176 0.46880
         1.20   0.46250 0.46731 0.46102 0.46657 0.47019
         1.40   0.46435 0.46620 0.46731 0.46981 0.47176
         1.60   0.46509 0.47083 0.46620 0.47324 0.46685

Optimal Priors from Grid Search, Maximum Hit Rate (0.473240740740741) Found at:
Prior Variance: 1.6
Prior degrees of freedom: 40
```

FMCG 3

```
Hit Rate Summary across Replicates:

                Prior D.F.
                10      20      30      40      50
Prior    0.80   0.62297 0.62277 0.62178 0.62297 0.61990
Var      1.00   0.62010 0.61980 0.62089 0.62238 0.62139
         1.20   0.62376 0.61960 0.61743 0.62168 0.62277
         1.40   0.61802 0.61931 0.62129 0.62248 0.62158
         1.60   0.61901 0.62069 0.62000 0.61980 0.61950

Optimal Priors from Grid Search, Maximum Hit Rate (0.623762376237624) Found at:
Prior Variance: 1.2
Prior degrees of freedom: 10
```

Validation Benchmark Results - Hitrates

- As expected our three benchmark estimations don‘t differ that much
- All three work well and the differences are mainly caused due to different MCMC-Chains

➔ Proper HB settings and well defined models result in stable estimates!

Hitrate in %	Within Sample			Out Of Sample		
	FMCG 1	FMCG 2	FMCG 3	FMCG 1	FMCG 2	FMCG 3
Default Prior - 10k/10k	56,24%	**52,87%**	**67,46%**	45,06%	42,76%	55,75%
Default Prior - 190k/10k	57,34%	52,78%	66,86%	45,21%	44,60%	**55,91%**
Optimized Prior - 190k/10k	**57,34%**	51,76%	67,39%	**45,32%**	**44,64%**	55,65%

Validation Benchmark Results - RMSE

- In addition to the hitrates we calculated the root mean squared error (RMSE) for a comparison of two pricing scenarios.
- We compared a base case scenario with all SKUs at market price with a second scenario with price increase of 1 price point for all SKUs.
- Out-of-sample we have no part-worth estimates, therefore we used logCounts (Johnson 2006).

RMSE	Out Of Sample		
	FMCG 1	FMCG 2	FMCG 3
Default Prior - 10k/10k	1,84	**1,75**	1,51
Default Prior - 190k/10k	**1,79**	**1,75**	**1,49**
Optimized Prior - 190k/10k	1,84	**1,75**	1,51

- Differences between our benchmark estimates are tiny.
- This result confirm our finding: All benchmark estimates deliver stable results

Benchmark Results – Difference in Which Choices are Predicted Correct

- While the percentage of correctly predicted choices differs only slightly between benchmark solutions, there are rather large differences looking at individual predicted choices

➔ This indicates the potential for ensemble solutions!

% of Differing Choice Predictions FMCG1 - Within Sample	Within Sample		
	Default Prior - 10k/10k	Default Prior - 190k/10k	Optimized Prior - 190k/10k
Default Prior - 10k/10k	0,00%		
Default Prior - 190k/10k	18,07%	0,00%	
Optimized Prior - 190k/10k	16,24%	16,51%	0,00%

% of Differing Choice Predictions FMCG2 - Within Sample	Within Sample		
	Default Prior - 10k/10k	Default Prior - 190k/10k	Optimized Prior - 190k/10k
Default Prior - 10k/10k	0,00%		
Default Prior - 190k/10k	24,81%	0,00%	
Optimized Prior - 190k/10k	23,70%	24,26%	0,00%

% of Differing Choice Predictions FMCG3 - Within Sample	Within Sample		
	Default Prior - 10k/10k	Default Prior - 190k/10k	Optimized Prior - 190k/10k
Default Prior - 10k/10k	0,00%		
Default Prior - 190k/10k	7,52%	0,00%	
Optimized Prior - 190k/10k	7,26%	6,34%	0,00%

The Covariates

- For further improvements we selected a set of covariates which are available in most of our studies:
  - 5 socio-demographic variables + 9 binary variable on shopping behavior
  - The 9 behavioral variables can be grouped in three clusters (brand, price, NPD)

Age	Gender	HH Size (Adult)	Employment Status	Income	Shopping Behavior
18-29	Male	1	Part Time	Low (<2.5/50k)	Role of Brand
30-39	Female	2	Full Time	Medium (<5/100k)	Role of Price
40-49		3	Not Employed	High (>5/100k)	Role of NPD
50-59		3+			
60+					

Estimations with Covariates

- Increased differentiation between within- and out-of-sample predictions through this additional information
- Within sample: As HB fits the data perfectly only small differences detected
- Out of sample: Benefit from the additional information of the covariates in the estimation

Hitrate in %	Within Sample			Out Of Sample		
	FMCG 1	FMCG 2	FMCG 3	FMCG 1	FMCG 2	FMCG 3
Default Prior - 10k/10k	56,24%	52,87%	67,46%	45,06%	42,76%	55,75%
Age	55,14%	51,94%	64,75%	44,16%	**43,68%**	53,76%
Empstat	53,94%	52,13%	67,13%	44,04%	**43,93%**	**56,79%**
Gender	**56,33%**	**53,70%**	67,19%	**47,37%**	**45,47%**	55,49%
HHSize	55,23%	51,11%	67,13%	**45,73%**	**44,20%**	55,21%
Income	55,69%	**53,89%**	66,07%	**45,72%**	**46,12%**	55,25%
Role of Brand	55,14%	51,20%	67,06%	**47,87%**	**44,01%**	**56,02%**
Role of Price	54,68%	51,67%	**67,52%**	**45,84%**	**44,07%**	**57,18%**
Role of NPD	54,86%	50,74%	66,93%	44,91%	**43,90%**	**56,91%**
Brands differ	55,14%	51,39%	67,26%	44,86%	**42,94%**	**56,01%**
Know which brand I buy	55,23%	**53,61%**	**67,85%**	44,62%	**45,23%**	**57,36%**
Buy same as last time	55,87%	51,02%	66,86%	**46,33%**	**44,58%**	54,62%
Compare prices	55,41%	51,57%	67,06%	**45,16%**	**44,74%**	55,74%
Looking for offers	56,06%	52,59%	**67,59%**	**45,73%**	**45,15%**	**56,76%**
Price knowledge	**56,51%**	50,56%	67,13%	**47,94%**	**44,07%**	55,02%
Interest in new products	55,60%	52,69%	66,67%	**46,15%**	**44,79%**	54,73%
Quality should be improved	55,78%	52,41%	66,73%	**45,62%**	**44,43%**	55,31%
Difficult to find the right product	**56,61%**	52,50%	67,26%	**48,11%**	**45,73%**	**55,86%**
Behavior	54,40%	51,57%	66,53%	**48,10%**	**45,82%**	**57,39%**

Ensemble Solution 1

- The best single variable (covariate) of each role-cluster was used to create Ensemble 1
- For each study the (best) variable Ensembles were different
- These ensembles outperform the default solution with marginal increases in prediction accuracy Within Sample and with larger increases Out of Sample

Hitrate in %	Within Sample			Out Of Sample		
	FMCG 1	FMCG 2	FMCG 3	FMCG 1	FMCG 2	FMCG 3
Default Prior - 10k/10k	56,24%	52,87%	67,46%	45,06%	42,76%	55,75%
10k/10k – Buy same as last time	55,87%	/	/	46,33%	/	/
10k/10k – Price knowledge	56,51%	/	/	47,94%	/	/
10k/10k – Difficult find right product	56,61%	/	/	48,18%	/	/
Ensemble	**57,16%**	/	/	**50,41%**	/	/
10k/10k – Know which brand I buy	/	53,61%	/	/	45,23%	/
10k/10k – Looking for offers	/	52,59%	/	/	45,15%	/
10k/10k – Interest in new products	/	52,69%	/	/	44,79%	/
Ensemble	/	**52,96%**	/	/	**46,93%**	/
10k/10k - Know which brand I buy	/	/	67,85%	/	/	57,36%
10k/10k - Looking for offers	/	/	67,59%	/	/	56,76%
10k/10k - Difficult find right product	/	/	67,26%	/	/	55,86%
Ensemble	/	/	**67,72%**	/	/	**57,38%**

Ensemble Solution 2

- Combining the best socio-demographic covariates with the single best behavior covariates to a ensemble also results (with one exception) in better predictions, but again different ensembles used for different studies

Hitrate in %	Within Sample			Out Of Sample		
	FMCG 1	FMCG 2	FMCG 3	FMCG 1	FMCG 2	FMCG 3
Default Prior - 10k/10k	56,24%	52,87%	67,46%	45,06%	42,76%	55,75%
10k/10k - Gender	56,33%	/	/	47,37%	/	/
10k/10k – Price knowledge	56,51%	/	/	47,92%	/	/
10k/10k – Difficult find the right product	56,61%	/	/	48,11%	/	/
Ensemble	**55,87%**	/	/	**49,86%**	/	/
10k/10k – Gender	/	53,70%	/	/	45,47%	/
10k/10k – Income	/	53,61%	/	/	46,12%	/
10k/10k – Know which brand I buy	/	53,89%	/	/	45,23%	/
Ensemble	/	**54,81%**	/	/	**46,88%**	/
10k/10k – Know which brand I buy	/	/	67,52%	/	/	56,76%
10k/10k – Looking for offers	/	/	67,85%	/	/	57,36%
10k/10k – Price knowledge	/	/	67,59%	/	/	56,76%
Ensemble	/	/	**67,72%**	/	/	**57,79%**

Ensemble Solution 3

- Ensemble with all 9 covariates performed best, especially for Out-of-Sample predictions, but only slight improvements measured
- However, using all 9 binary variables as covariates also results in good predictions

Hitrate in %	Within Sample			Out Of Sample		
	FMCG 1	FMCG 2	FMCG 3	FMCG 1	FMCG 2	FMCG 3
Default Prior - 10k/10k	56,24%	52,87%	67,46%	45,06%	42,76%	55,75%
Covariates (a-i)	54,40%	51,57%	66,53%	48,10%	45,82%	57,39%
a) Brands differ	55,14%	51,39%	67,26%	44,86%	42,94%	56,01%
b) Know which brand I buy	55,23%	53,61%	67,85%	44,62%	45,23%	57,36%
c) Buy same as last time	55,87%	51,02%	66,86%	46,33%	44,58%	54,62%
d) Compare prices	55,41%	51,57%	67,06%	45,16%	44,74%	55,74%
e) Looking for offers	56,06%	52,59%	67,59%	45,73%	45,15%	56,76%
f) Price knowledge	56,51%	50,56%	67,13%	47,94%	44,07%	55,02%
g) Interest in new products	55,60%	52,69%	66,67%	46,15%	44,79%	54,73%
h) Quality should be improved	55,78%	52,41%	66,73%	45,62%	44,43%	55,31%
i) Difficult to find the right product	56,61%	52,50%	67,26%	48,11%	45,73%	55,86%
Ensemble (a-i)	**56,33%**	**54,26%**	**67,19%**	**48,13%**	**45,98%**	**57,62%**

Ensemble Solutions - Overview

- Best out-of-sample predictions are derived by ensembles of selected behavioral covariates resp. socio-demographic and selected behavioral covariates.
- Simply running the 9 behavioral binary covariates in one estimation also seems to be a practical solution leading to close results

Hitrate in %	Within Sample			Out Of Sample		
	FMCG 1	FMCG 2	FMCG 3	FMCG 1	FMCG 2	FMCG 3
Default Prior - 10k/10k	56,24%	**52,87%**	**67,46%**	45,06%	42,76%	55,75%
Default Prior - 190k/10k	57,34%	52,78%	66,86%	45,21%	44,60%	**55,91%**
Optimized Prior - 190k/10k	**57,34%**	51,76%	67,39%	**45,32%**	**44,64%**	55,65%
Ensemble Role-Clusters (1)	57,16%	**52,96%**	**67,72%**	**50,41%**	**46,93%**	**57,38%**
Ensemble socio-demographic+behavior (2)	55,87%	**54,81%**	**67,72%**	**49,86%**	**46,88%**	**57,79%**
Ensemble of binary behavior questions (3)	56,33%	**54,26%**	67,19%	**48,13%**	**45,98%**	**57,62%**
Behaviour Covariates (a-i)	54,40%	51,57%	66,53%	**48,10%**	**45,82%**	**57,39%**

Ensemble Solutions - RMSE

- RMSE based on comparing the two pricing scenarios confirms, that best out-of-sample predictions are derived by ensembles of socio-demographic and selected behavioral covariates.
- RMSE values confirm the findings from the hitrates

RMSE	Out Of Sample		
	FMCG 1	FMCG 2	FMCG 3
Default Prior - 10k/10k	1,84	**1,75**	1,51
Default Prior - 190k/10k	**1,79**	**1,75**	**1,49**
Optimized Prior - 190k/10k	1,84	**1,75**	1,51
Ensemble Role-Clusters (1)	1,79	1,74	**1,47**
Ensemble socio-demographic+behavior (2)	1,72	1,72	1,46
Ensemble of binary behavior questions (3)	**1,73**	1,73	1,48
Behaviour Covariates (a-i)	1,84	**1,71**	**1,47**

- Out-of-sample pricing scenarios show same influence from covariates as results based on hitrates.

Findings

- Based on three comparable price conjoint studies, we replicated our competition approach and explored the effects on prediction accuracy
- Ensembles, as theoretically plausible, are superior compared to the individual estimations they consist of
- In-line with the predictive modeling competition we see only small improvements compared to the benchmark estimates
- Within Sample we could show that all estimations converge nearly to the same solutions. Furthermore the Covariates seem not to improve the estimates in a noticeable amount and HB fits as good as possible to the dataset
- Out-of-sample, the covariates pay-off the effort of longer computational time
- Study specific ensembles out of behavioral and socio-demographic variables seem to be most efficient

Conclusions

- Overall even with standard settings CBC/HB does a good job and all ensemble solutions only slightly improve the outcome
- The blending of several HB models differentiated by different covariates leads to an increased accuracy in predicting respondents' choices within and especially out of sample
- Blending behavioral and socio-demographic covariates could be recommended as a cost-efficient & low-effort standard procedure to slightly increase prediction accuracy compared to standard solutions
- If one would not invest the additional work and computational time, covariates could be straight forward implemented
- More important still is to have an appropriate ratio between parameters and amount of collected data

Future Work

- Test our finding on non-FMCG studies
- Could we find a better set of covariates, that could be used for a wide variety of CBC models?
- The predictive modeling competition shows, that other estimation models (mixture of normal, Dirichlet process prior) could result in further improvements
- How could we improve predictions with “really” sparse data – could ensembles help?

Thank you!

References

- **Allenby, G.M.; Rossi, P.E. (2006):** Hierarchical Bayes Models, in: Grover, R.; Vriens, M. (Hrsg.): The Handbook of Marketing Research: Uses, Misuses, and Future Advances, S. 418-440, SAGE Publications Inc., Thousand Oaks.
- **Hein, M.; Kurz, P.; Steiner, W.(2013):** Limits for Parameter Estimation in Choice-Based Conjoint Analysis: A Simulation Study, European Conference on Data Analysis 2013.
- **Hein, M., Kurz, P., Steiner, W.(2019):**On the effect of HB covariance matrix prior settings: A simulation study. Journal of Choice Modelling (forthcoming).
- **Hein, M., Kurz, P., Steiner, W.(2019):** Analyzing the Capabilities of the HB Logit Model for Choice-Based Conjoint Analysis: A Simulation Study. In: Journal of Business Economics (forthcoming).
- **Kurz, P; Binner,S. (2011):** Added Value through Covariates in HB Modeling?, Proceedings of the 2011 Sawtooth Software Conference.
- **Lattery, K. (2015):** A Machine Learning Approach to Conjoint Analysis: Boosting and Blending Ensembles, Proceedings of the 2015 Sawtooth Software Conference.
- **Liakhovitski, D.; Shmulyian, F. (2011):** Covariates in Discrete Choice Models: Are They Worth the Trouble?" ART Form Presentation.
- **Nassar, N. (2016):** The Winning Choice Model; A Semi-Compensatory One, Proceedings of the 2016 Sawtooth Software Conference.
- **Orme, B. (2016):** Comment On Lattery's Conjoint Analysis Ensembles, Proceedings of the 2015 Sawtooth Software Conference.
- **Orme, B. (2016):** Findings of the 2016 Sawtooth Software CBC Modeling Prize Competition, Proceedings of the 2016 Sawtooth Software Conference.
- **Orme, B. (2016):** What Are the Optimal HB Priors Settings for CBC and MaxDiff Studies?, Sawtooth Software Resarch Paper Series
- **Johnson, R., Orme, B., Pinnell, J. (2006):** Simulating Market Preference with Build Your Own Data Proceedings of the 2006 Sawtooth Software Conference.
- **Sentis, K. and Li, L. (2001):** One Size Fits All or Custom Tailored: Which HB Fits Better?" Proceedings of the 2001 Sawtooth Software Conference.
- **Sentis, K.; Geller, V. (2011):** The Impact of Covariates on HB Estimates, Proceedings of the 2011 Sawtooth Software Conference.

Appendix

Appendix 1: Benchmark 3 Prior optimized HB: Prior Optimal Search

- Optimal prior search shows only small improvements
- Most changes are not stable between replications
- This result is expected, because there is enough information in the data and therefore the uninformative “Inverted Wishart Prior” cancels out

Replication	FC 0,8+10	FC 0,8+20	FC 0,8+30	FC 0,8+40	FC 0,8+50	FC 1+10	FC 1+20	FC 1+30	FC 1+40	FC 1+50	FC 1,2+10	FC 1,2+20	FC 1,2+30	FC 1,2+40	FC 1,2+50	FC 1,4+10	FC 1,4+20	FC 1,4+30	FC 1,4+40	FC 1,4+50	FC 1,6+10	FC 1,6+20	FC 1,6+30	FC 1,6+40	FC 1,6+50	Δ prior	Δ prior - repl.
1	0,536	0,565	0,565	0,555	0,565	0,569	0,544	0,567	0,553	0,565	0,557	0,567	0,553	0,572	0,549	0,555	0,557	0,584	0,567	0,541	0,558	0,565	0,569	0,564	0,562	0,049	0,017
2	0,588	0,549	0,569	0,567	0,568	0,572	0,557	0,572	0,559	0,582	0,573	0,565	0,572	0,548	0,563	0,565	0,567	0,574	0,564	0,567	0,582	0,552	0,581	0,545	0,558	0,043	0,011
3	0,581	0,564	0,554	0,575	0,575	0,559	0,550	0,565	0,572	0,567	0,587	0,555	0,546	0,571	0,582	0,574	0,564	0,561	0,561	0,566	0,567	0,572	0,562	0,576	0,577	0,041	0,009
4	0,570	0,533	0,569	0,550	0,554	0,552	0,561	0,561	0,540	0,560	0,573	0,564	0,567	0,540	0,552	0,576	0,542	0,558	0,560	0,569	0,552	0,551	0,568	0,561	0,569	0,043	0,011
5	0,562	0,570	0,576	0,577	0,560	0,561	0,562	0,562	0,585	0,579	0,571	0,564	0,573	0,567	0,562	0,564	0,554	0,579	0,590	0,569	0,570	0,572	0,574	0,576	0,572	0,036	0,004
6	0,550	0,547	0,546	0,535	0,555	0,551	0,534	0,550	0,544	0,553	0,539	0,541	0,549	0,549	0,554	0,532	0,545	0,550	0,549	0,551	0,534	0,543	0,560	0,550	0,547	0,028	-0,005
7	0,580	0,575	0,567	0,560	0,568	0,561	0,583	0,576	0,588	0,571	0,577	0,586	0,566	0,578	0,561	0,570	0,583	0,564	0,569	0,584	0,573	0,591	0,578	0,567	0,577	0,031	-0,001
8	0,549	0,583	0,561	0,581	0,572	0,558	0,583	0,556	0,574	0,563	0,572	0,572	0,577	0,573	0,565	0,566	0,564	0,563	0,583	0,558	0,560	0,551	0,591	0,579	0,558	0,042	0,010
9	0,534	0,562	0,564	0,565	0,557	0,582	0,562	0,541	0,567	0,556	0,583	0,542	0,552	0,557	0,553	0,557	0,550	0,568	0,541	0,576	0,571	0,573	0,563	0,552	0,551	0,049	0,017
10	0,562	0,572	0,569	0,555	0,579	0,550	0,580	0,572	0,554	0,569	0,562	0,598	0,583	0,577	0,578	0,561	0,576	0,577	0,554	0,572	0,583	0,575	0,567	0,552	0,570	0,049	0,017
Δ replication	0,054	0,050	0,030	0,046	0,025	0,032	0,050	0,035	0,048	0,028	0,049	0,057	0,037	0,038	0,033	0,044	0,041	0,035	0,049	0,043	0,050	0,048	0,031	0,034	0,030	0,032	

Optimal prior
1,6 / 30

Appendix 2: Benchmark 3 Prior optimized HB: Prior Optimal Search

- Comparing the more with the less complex study, we can show, that the changes with different settings are smaller when more information in the data is available
- The small differences cancel out if we increase the number of iterations and allow the model to converge
- MNL/HB shows again it‘s advantage and stability

Replication	FC 0,8+10	FC 0,8+20	FC 0,8+30	FC 0,8+40	FC 0,8+50	FC 1+10	FC 1+20	FC 1+30	FC 1+40	FC 1+50	FC 1,2+10	FC 1,2+20	FC 1,2+30	FC 1,2+40	FC 1,2+50	FC 1,4+10	FC 1,4+20	FC 1,4+30	FC 1,4+40	FC 1,4+50	FC 1,6+10	FC 1,6+20	FC 1,6+30	FC 1,6+40	FC 1,6+50	Δ prior	Δ prior - repl.
1	0,475	0,477	0,452	0,475	0,480	0,474	0,472	0,484	0,481	0,472	0,466	0,476	0,469	0,486	0,467	0,460	0,468	0,463	0,481	0,496	0,481	0,486	0,465	0,479	0,481	0,044	0,021
2	0,459	0,458	0,453	0,455	0,469	0,475	0,468	0,476	0,466	0,468	0,479	0,454	0,474	0,462	0,472	0,469	0,449	0,462	0,474	0,464	0,471	0,453	0,469	0,482	0,463	0,033	0,010
3	0,448	0,451	0,441	0,463	0,463	0,467	0,457	0,470	0,464	0,460	0,449	0,465	0,456	0,471	0,469	0,460	0,452	0,468	0,465	0,463	0,452	0,472	0,475	0,478	0,453	0,037	0,014
4	0,444	0,431	0,448	0,447	0,458	0,453	0,444	0,471	0,445	0,445	0,451	0,433	0,434	0,449	0,462	0,467	0,446	0,440	0,451	0,456	0,451	0,450	0,442	0,455	0,456	0,041	0,018
5	0,471	0,490	0,472	0,467	0,461	0,477	0,464	0,449	0,479	0,480	0,487	0,464	0,434	0,456	0,481	0,479	0,469	0,471	0,466	0,487	0,478	0,470	0,450	0,468	0,478	0,056	0,032
6	0,460	0,478	0,456	0,451	0,469	0,457	0,467	0,459	0,470	0,468	0,471	0,478	0,461	0,457	0,460	0,458	0,476	0,466	0,464	0,466	0,477	0,467	0,486	0,462	0,471	0,035	0,012
7	0,444	0,462	0,456	0,464	0,450	0,453	0,466	0,467	0,490	0,469	0,444	0,469	0,483	0,475	0,472	0,473	0,479	0,475	0,476	0,464	0,457	0,480	0,476	0,481	0,454	0,045	0,022
8	0,462	0,470	0,441	0,477	0,471	0,469	0,471	0,462	0,470	0,462	0,461	0,465	0,436	0,465	0,456	0,458	0,480	0,469	0,463	0,471	0,472	0,463	0,454	0,466	0,456	0,044	0,020
9	0,481	0,500	0,485	0,484	0,470	0,485	0,487	0,471	0,498	0,490	0,476	0,488	0,497	0,483	0,470	0,481	0,494	0,490	0,493	0,474	0,468	0,478	0,486	0,490	0,479	0,032	0,009
10	0,450	0,469	0,448	0,476	0,472	0,454	0,477	0,472	0,455	0,474	0,441	0,482	0,464	0,461	0,492	0,439	0,449	0,469	0,467	0,476	0,444	0,490	0,460	0,472	0,478	0,053	0,030
Δ replication	0,037	0,069	0,044	0,037	0,030	0,032	0,043	0,035	0,053	0,044	0,046	0,055	0,063	0,037	0,035	0,042	0,048	0,050	0,042	0,040	0,038	0,040	0,044	0,035	0,029	0,023	

Optimal prior 1,4 / 30

Replication	FC 0,8+10	FC 0,8+20	FC 0,8+30	FC 0,8+40	FC 0,8+50	FC 1+10	FC 1+20	FC 1+30	FC 1+40	FC 1+50	FC 1,2+10	FC 1,2+20	FC 1,2+30	FC 1,2+40	FC 1,2+50	FC 1,4+10	FC 1,4+20	FC 1,4+30	FC 1,4+40	FC 1,4+50	FC 1,6+10	FC 1,6+20	FC 1,6+30	FC 1,6+40	FC 1,6+50	Δ prior	Δ prior - repl.
1	0,615	0,615	0,613	0,605	0,581	0,613	0,615	0,62	0,615	0,617	0,616	0,607	0,619	0,62	0,608	0,612	0,609	0,613	0,618	0,616	0,612	0,61	0,612	0,609	0,608	0,039	0,004
2	0,631	0,647	0,636	0,633	0,623	0,633	0,632	0,635	0,642	0,639	0,637	0,631	0,641	0,636	0,642	0,634	0,623	0,629	0,621	0,633	0,625	0,619	0,63	0,622	0,629	0,028	-0,007
3	0,626	0,635	0,618	0,621	0,601	0,623	0,637	0,626	0,62	0,63	0,622	0,631	0,619	0,627	0,622	0,61	0,627	0,619	0,611	0,615	0,617	0,617	0,61	0,613	0,612	0,036	0,001
4	0,599	0,619	0,608	0,618	0,595	0,605	0,605	0,616	0,624	0,601	0,606	0,608	0,618	0,613	0,598	0,588	0,601	0,624	0,605	0,594	0,592	0,602	0,609	0,603	0,598	0,036	0,001
5	0,655	0,628	0,632	0,643	0,626	0,646	0,639	0,64	0,631	0,638	0,646	0,636	0,631	0,634	0,635	0,629	0,629	0,633	0,624	0,633	0,632	0,628	0,614	0,628	0,624	0,042	0,007
6	0,627	0,64	0,628	0,613	0,604	0,619	0,618	0,625	0,637	0,624	0,627	0,613	0,622	0,634	0,638	0,624	0,624	0,629	0,626	0,622	0,616	0,616	0,631	0,631	0,634	0,036	0,001
7	0,623	0,615	0,618	0,591	0,592	0,615	0,61	0,619	0,607	0,62	0,612	0,612	0,62	0,613	0,621	0,609	0,615	0,627	0,607	0,619	0,607	0,621	0,624	0,619	0,624	0,036	0,001
8	0,612	0,601	0,603	0,617	0,598	0,611	0,607	0,606	0,615	0,601	0,6	0,615	0,601	0,605	0,611	0,603	0,611	0,604	0,604	0,602	0,604	0,611	0,598	0,607	0,611	0,019	-0,016
9	0,604	0,618	0,609	0,621	0,599	0,611	0,615	0,619	0,621	0,599	0,625	0,624	0,619	0,62	0,617	0,619	0,616	0,616	0,618	0,614	0,62	0,621	0,618	0,616	0,62	0,026	-0,009
10	0,629	0,623	0,624	0,627	0,6	0,624	0,633	0,633	0,621	0,623	0,62	0,623	0,626	0,621	0,622	0,63	0,627	0,631	0,622	0,618	0,614	0,619	0,625	0,616	0,622	0,033	-0,002
Δ replication	0,056	0,046	0,033	0,051	0,045	0,041	0,034	0,034	0,035	0,04	0,046	0,029	0,04	0,031	0,044	0,046	0,028	0,029	0,022	0,039	0,04	0,026	0,033	0,028	0,036	0,035	

Optimal prior 1,0 / 10