ART2015
Advanced Research Techniques Forum
June 14 – 17, 2015 • San Diego, CA

A new modeling tool for identifying meaningful segments and their willingness to pay:

Improving Validity by Reducing the Confound between Scale and Preference Heterogeneity

Jay Magidson

Jeff Dumont

Jeroen K. Vermunt

Statistical Innovations
Thinking outside the brackets!™

RSG
the science of insight

Tilburg University

Overview

- Heterogeneity in discrete choice data may relate to:
  - Preferences or willingness to pay (WTP)
  - Differences in the amount of error variance (scale heterogeneity).

- We use simulated and real data to investigate the effectiveness of a new Scale-Adjusted latent class (SALC) modeling tool to separate preference from scale heterogeneity.

- We compare Latent Class and (standard) HB regarding their ability to deal with scale heterogeneity.

What is a Scale Factor?

- The scale factor for respondent *i*, is a term by which **all** part-worth parameters β in the MNL choice model are multiplied. In this respect it differs from Preference heterogeneity.
- It relates to the amount of **consistency** or **certainty** in that person's *expected choices*.
- Traditional HB and Latent Class (LC) methods ignore scale heterogeneity or assume that it is equal for all respondents.
- The preference vs. scale confound may result in spurious segments that differ only in scale and not differ in their preferences or willingness to pay.

"All choice models confound scale and [preference part-worth] parameter estimates. The confound is particularly problematic in complex models like random coefficients [HB-like] models and latent class models if one cannot separate scale and [preference] parameters."

Louviere and Eagle (2006)

"So, the bottom line is that one cannot estimate individual-level [preference] parameters from choice models unless one can separate scale and [preference] parameter estimates." …

Simulation Design

- 3 Segments differing in preference
- 4 attributes (Brand, Feature, Price, and None)
- 9 choice sets per respondent
- Each task (set) has 3 alternatives

Attributes	True Part-worth Utilities		
Brand	Class 1	Class 2	Class 3
A	0	0	0
B	1	-1	0
C	-1	-2	0
Feature	-0.5	0.5	0
Price	-0.5	-1	-1.5
None	-3.5	-3.5	-3.5

Example Choice set:

Brand	Feature	Price
A	1	$1
B	2	$2
None		

LC vs. HB: Very Different Approaches

Segment and WTP Parameter Recovery:

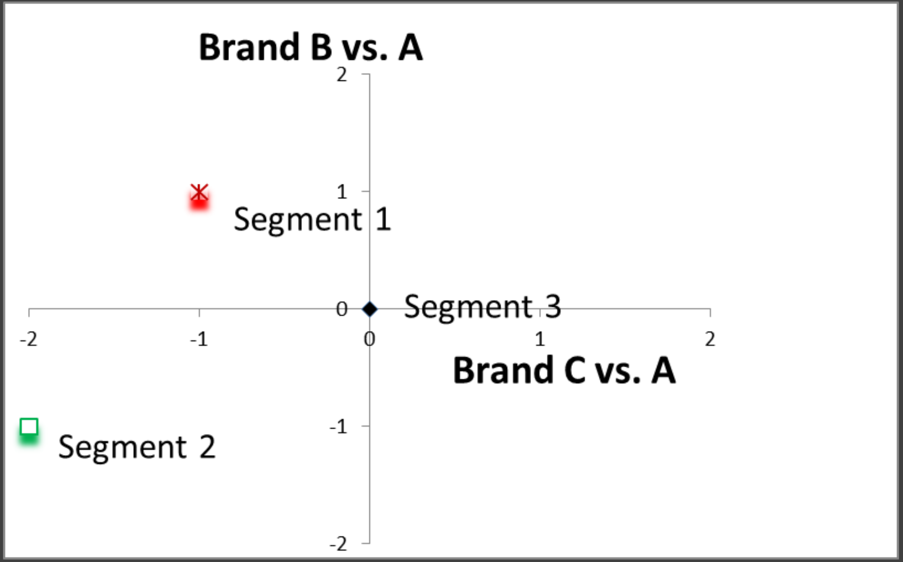
LC: 1) First get segments and class-specific parameters
-- traditional LC assumes no scale heterogeneity
2) Use posterior membership probabilities as weights to get individual-level WTP coefficients

HB: 1) First get individual-level WTP coefficients
-- Upper level HB model assumes multivariate normality
-- The lower (Individual) level parameters confound preference and scale heterogeneity
2) Then cluster these coefficients to get segments

HB: This is like initially assuming that no segments exist (Step 1) and then trying to find them (Step 2).

Traditional Latent Class Analysis (LCA) assumes no within-segment Variation

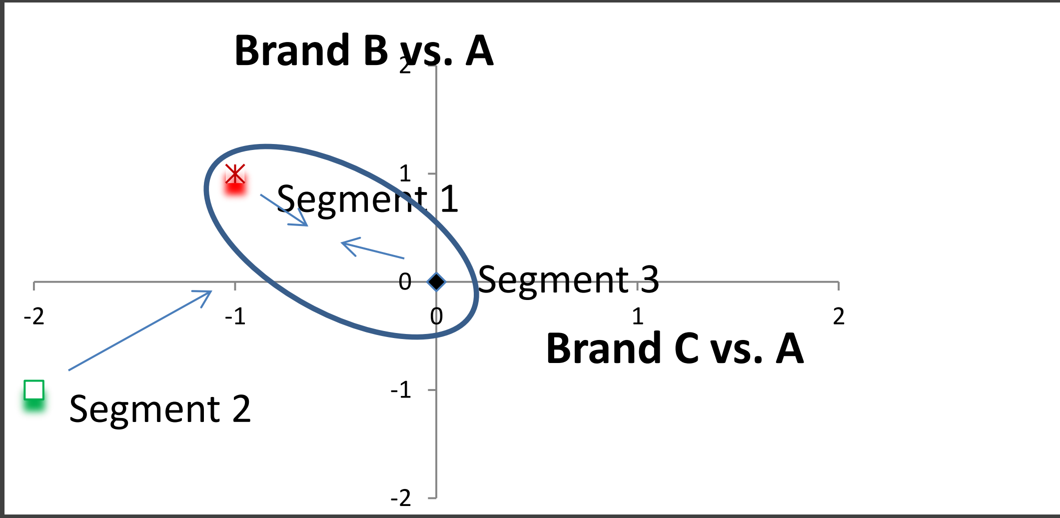
Attributes	True Part-worth Utilities		
Brand	Class 1	Class 2	Class 3
A	0	0	0
B	1	-1	0
C	-1	-2	0
Feature	-0.5	0.5	0
Price	-0.5	-1	-1.5
None	-3.5	-3.5	-3.5



- Preference heterogeneity due to 3 latent classes only
- With respect to Brand Preference:
  - Class 1 prefers Brand B to A and A over C
  - Class 2 prefers Brand A to both B and C
  - Class 3 is indifferent between Brands A, B, and C.
- No scale heterogeneity

HB: MVN Prior Inconsistent with LC Segments

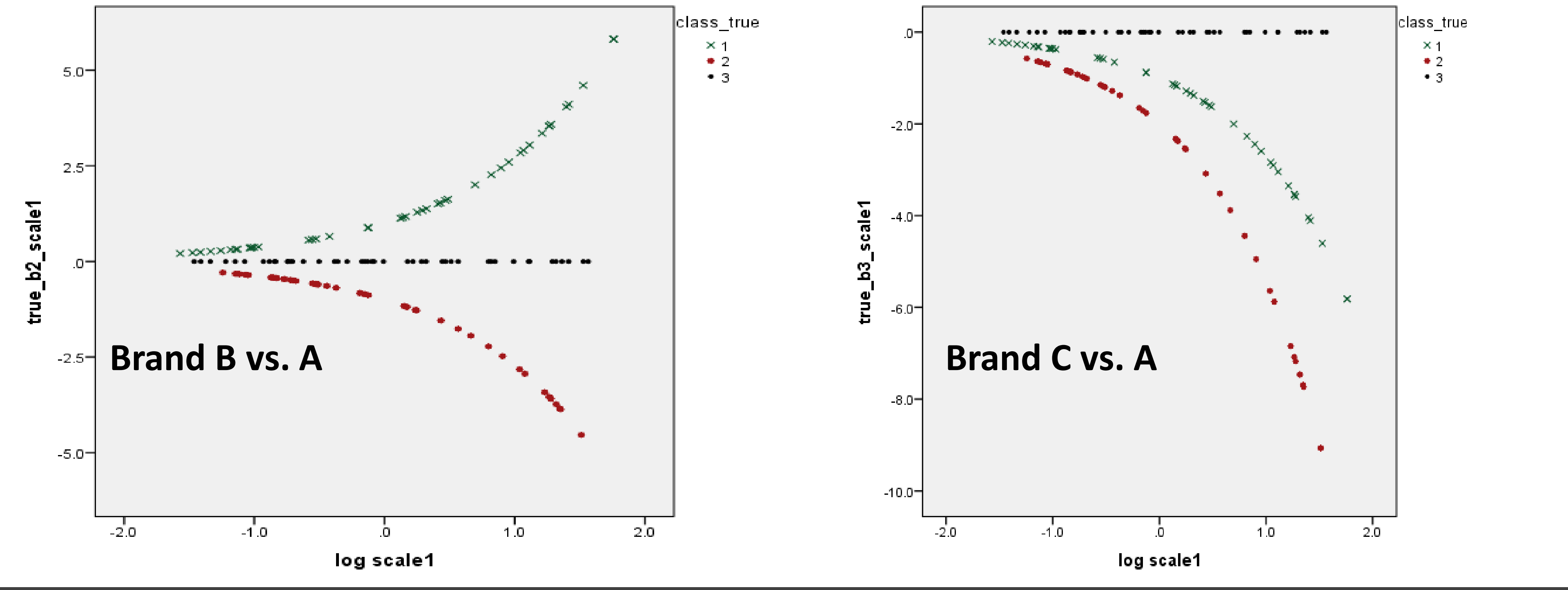
Attributes	True Part-worth Utilities		
Brand	Class 1	Class 2	Class 3
A	0	0	0
B	1	-1	0
C	-1	-2	0
Feature	-0.5	0.5	0
Price	-0.5	-1	-1.5
None	-3.5	-3.5	-3.5



Regarding individual Brand coefficients: C vs. A and B vs. A BVN ellipse above:

– May capture segments 1 and 3 but 'center mass' assumption is inconsistent with bimodality – yields 'regression to mean'
– Segment 2 cases would also be regressed towards overall mean

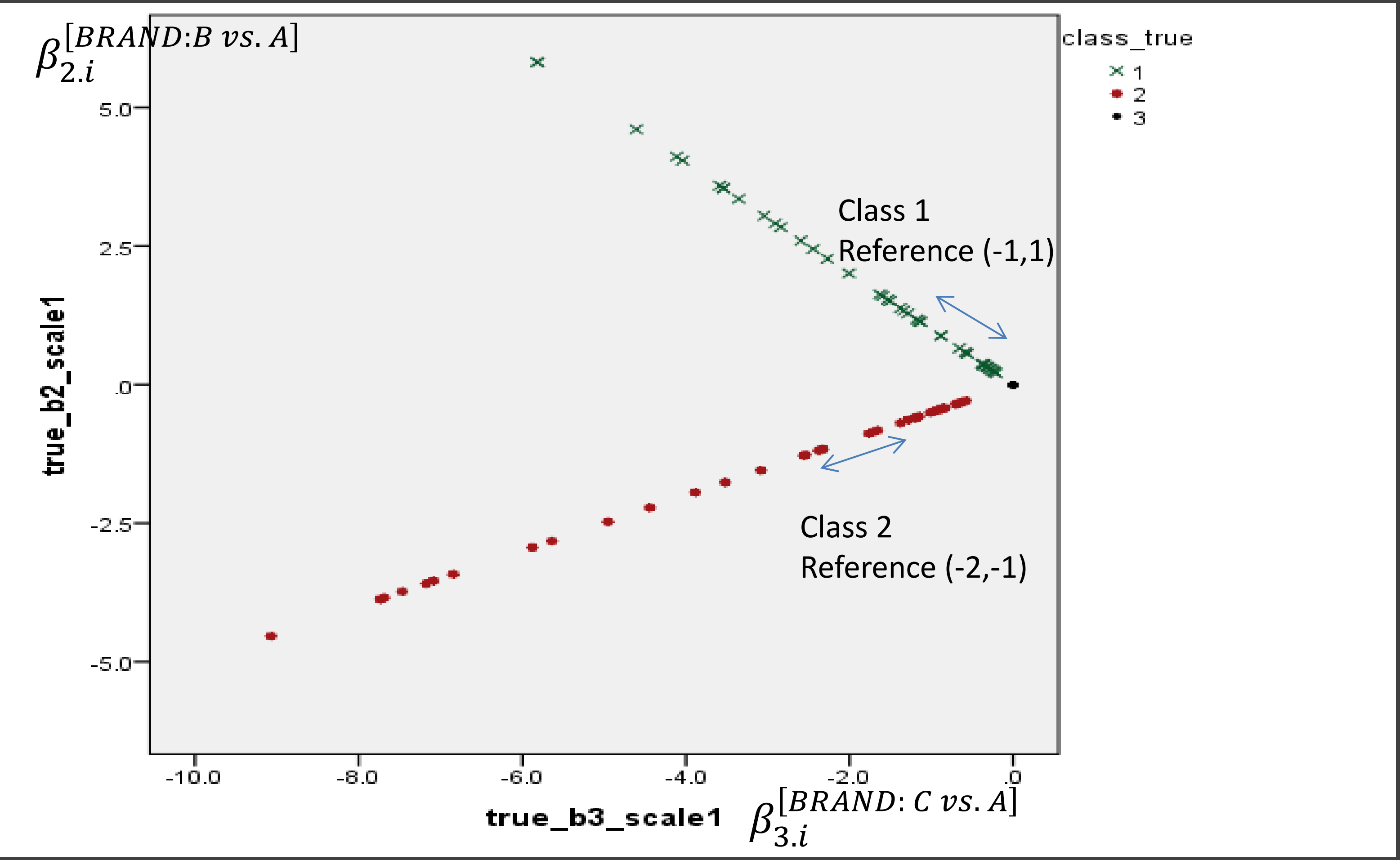
Scale Adjusted Latent Class (SALC) Models assume LC Segments + Scale Heterogeneity



- Log-scale ~N(0,1)
- Respondents with larger scale are easier to classify correctly.

Attributes	True Part-worth Utilities		
Brand	Class 1	Class 2	Class 3
A	0	0	0
B	1	-1	0
C	-1	-2	0

Plot of Individual coefficients $\beta^{[BRAND]}_{i,j}$ in Brand Space: (C vs. A, B vs. A)



- Preference Heterogeneity
  - Class 1 prefers B over A and A over C
  - Class 2 prefers A over both B and C
  - Class 3 is indifferent between brands A, B, and C.

Attributes	True Part-worth Utilities		
Brand	Class 1	Class 2	Class 3
A	0	0	0
B	1	-1	0
C	-1	-2	0

New Tool for Separating Scale and Preference Heterogeneity

$$\beta_{j.i}^{[BRAND]} = \exp(\lambda_i - \lambda_0)\beta *_{j.i}^{[BRAND]}$$

- Log-scale parameters λ_i are estimated simultaneously with preference parameters β*
- The more attributes, the greater the power to separate scale from preference
- For purposes of identification, the log-s**cale parameter** λ_i is determined relative to a fixed reference point λ_0, and is modeled using individual or group-level observed or latent variables (Vermunt, 2013):
  - Latent *continuous* scale factor: log-scale factor follows normal distribution (with mean λ_0 **= 0**)
  - Latent *categorical* scale classes (sClasses): with the log-scale factors for one of the sClasses being set to λ_0 **= 0.**
- Implemented in Latent GOLD and LG Choice (version 5.0)
- Improves over earlier approach Magidson and Vermunt (2007)

Part-worths (β) and Scale-adjusted Part-worths (β*)

- Each part-worth is expressed as a product of scale and (scale-adjusted) preference. For Brand j and Price k we have:

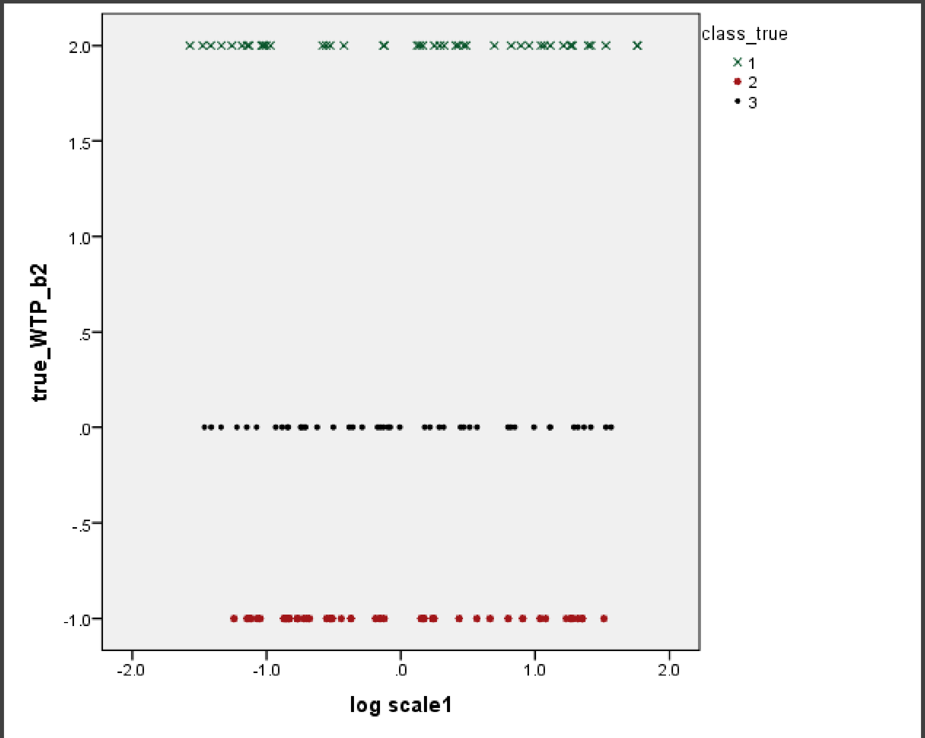
$$\beta_{j.i}^{[BRAND]} = \exp(\lambda_i - \lambda_0)\beta*_{j.i}^{[BRAND]}$$

$$\beta_{k.i}^{PRICE} = \exp(\lambda_i - \lambda_0)\beta*_{k.i}^{[\Pr ice]}$$

- Compare LC with HB in recovering preference heterogeneity:
  - Focus on Willingness-To-Pay (WTP) coefficients since they are scale-free and do not depend upon any particular reference λ_0:

$$WTP_{j.i}^{[BRAND]} = \frac{\exp(\lambda_i - \lambda_0)\beta*_{j.i}^{[BRAND]}}{\exp(\lambda_i - \lambda_0)(-\beta*_{k.i}^{[\Pr ice]})} = -\frac{\beta*_{j.i}^{[BRAND]}}{\beta*_{k.i}^{[\Pr ice]}}$$

WTP Coefficients are Scale-Free



Attributes	True Part-worth Utilities		
Brand	Class 1	Class 2	Class 3
A	0	0	0
B	1	-1	0
C	-1	-2	0
Feature	-0.5	0.5	0
Price	-0.5	-1	-1.5
None	-3.5	-3.5	-3.5

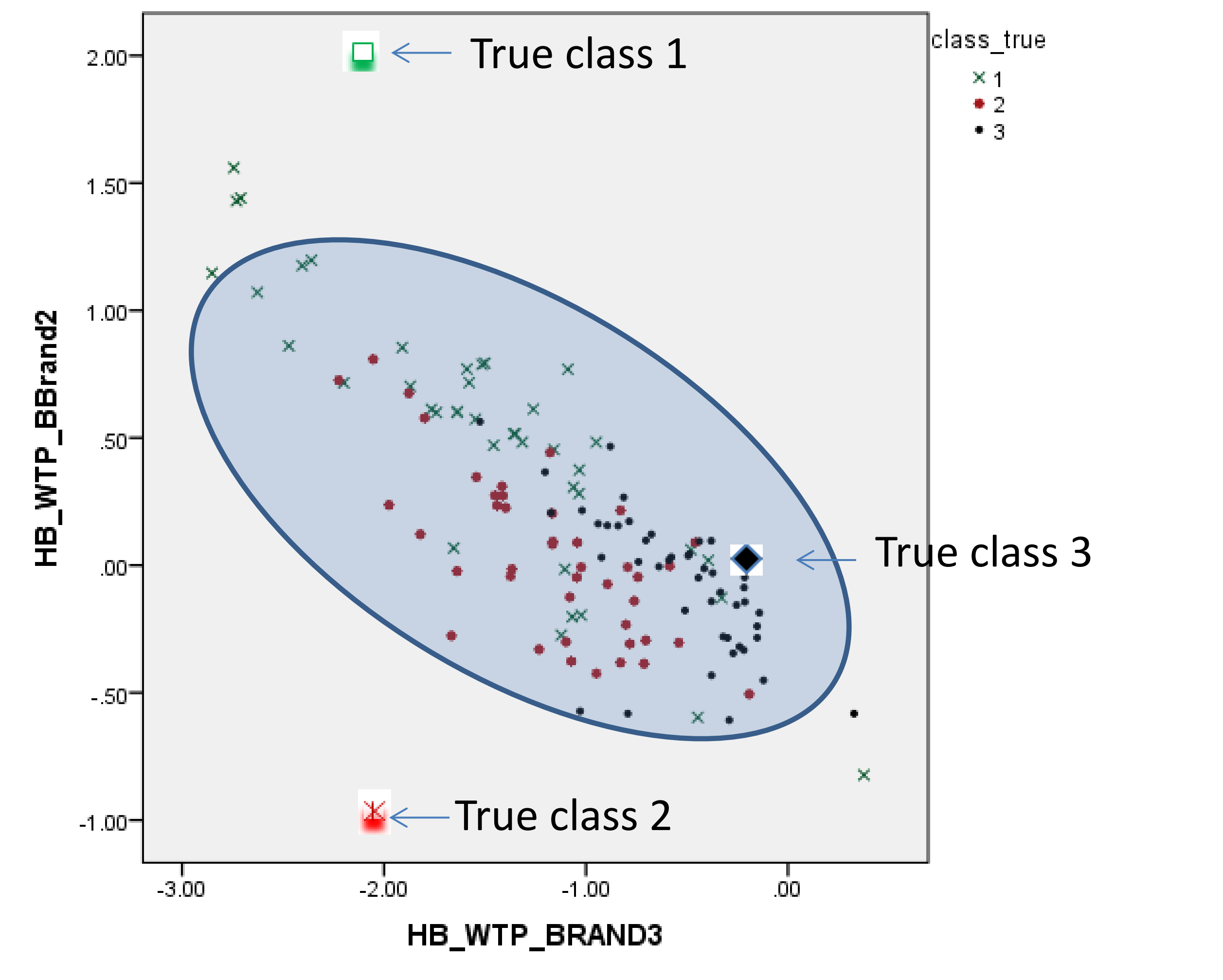
Attributes	True WTP		
Brand	Class 1	Class 2	Class 3
A	0	0	0
B	2	-1	0
C	-2	-2	0
Feature	-1	0.5	0
Price	-1	-1	-1
None	-7	-3.5	-2.3

- Preference and WTP heterogeneity due to 3 latent classes
- With respect to Brand A vs. B:
  - Class 1 prefers B to A and would pay a premium of $2 for B
  - Class 2 prefers A to B and would pay a premium of $1 for A
  - Class 3 is indifferent between brands A and B

Simulation #1

- Data (N=135) simulated according to SALC
- Compare SALC results to standard LC and HB
  - Segment recovery: correct classification rates
  - WTP Parameter recovery – Median Absolute Error
  - Predictions (brand preference share)
  - Hit rate and overfitting

Results: MVN Yields Poor Segment Recovery



Attributes	True WTP		
Brand	Class 1	Class 2	Class 3
A	0	0	0
B	2	-1	0
C	-2	-2	0

Neither segments nor WTP parameters recovered well – 'Bayesian shrinkage' due to MVN assumption.

Regardless which clustering method used, segment recovery < 55% correct.

44% who prefer Brand B (Class 1) are mistakenly predicted to pay a premium for brand A over B!

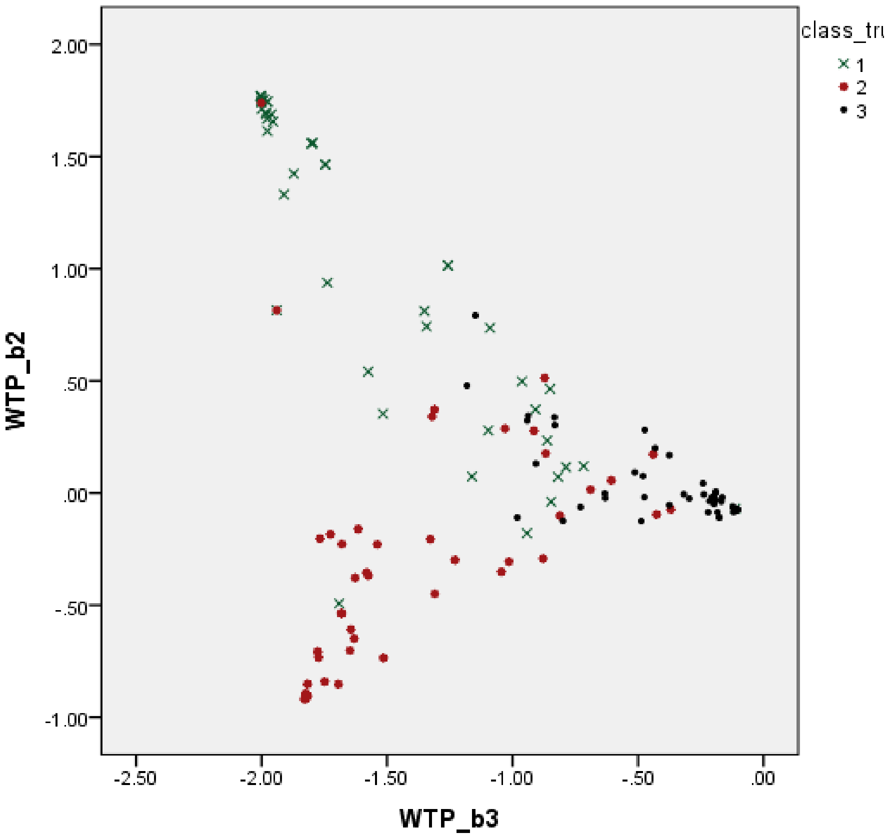
HB predicts 58% prefer brand B while true population rate is 50%.

Hit rate (holdouts 58% vs. 62% SALC and xx% LC)
(in-sample) = 70% vs. 68% SALC and 67% LC(overfitting)

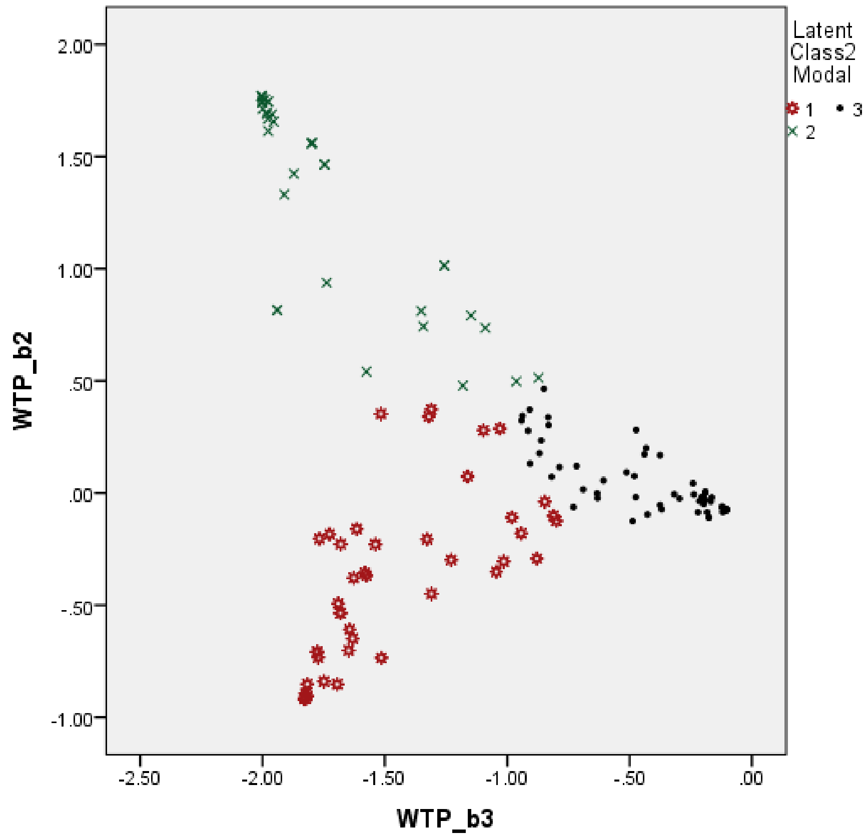
MAE= 0.88 (more than twice LC)
Segment recovery (LC) = 54.6%

Results from SALC N=135, low scale simulation

Symbols denote True Classes



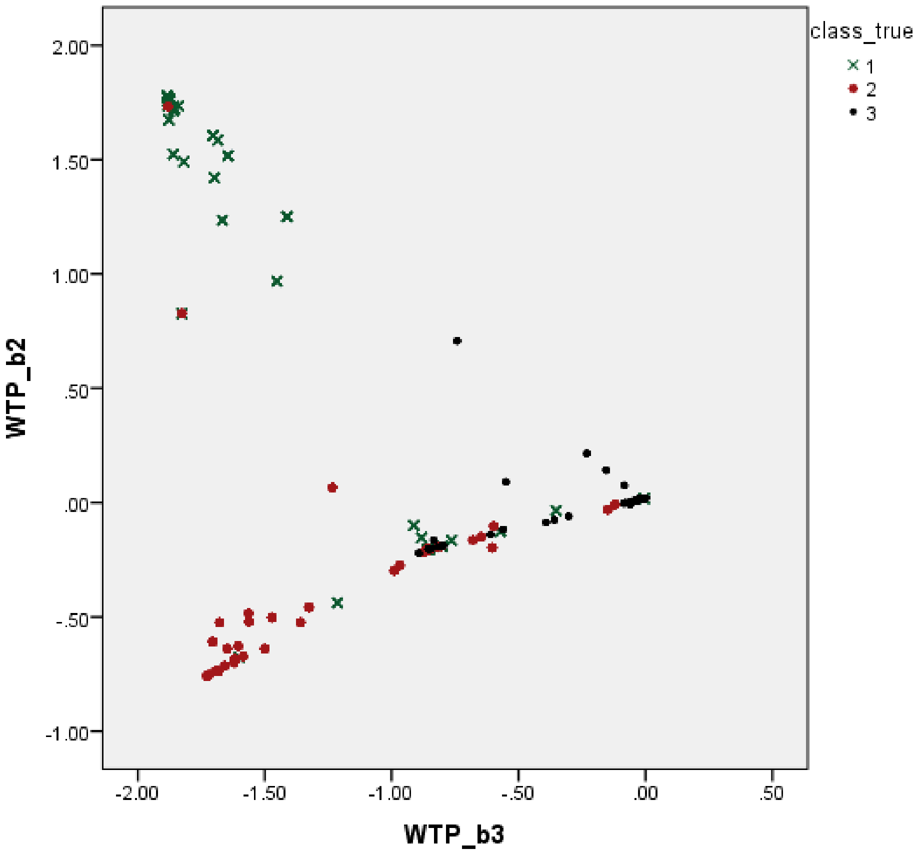
Symbols denote SALC Class Assignments

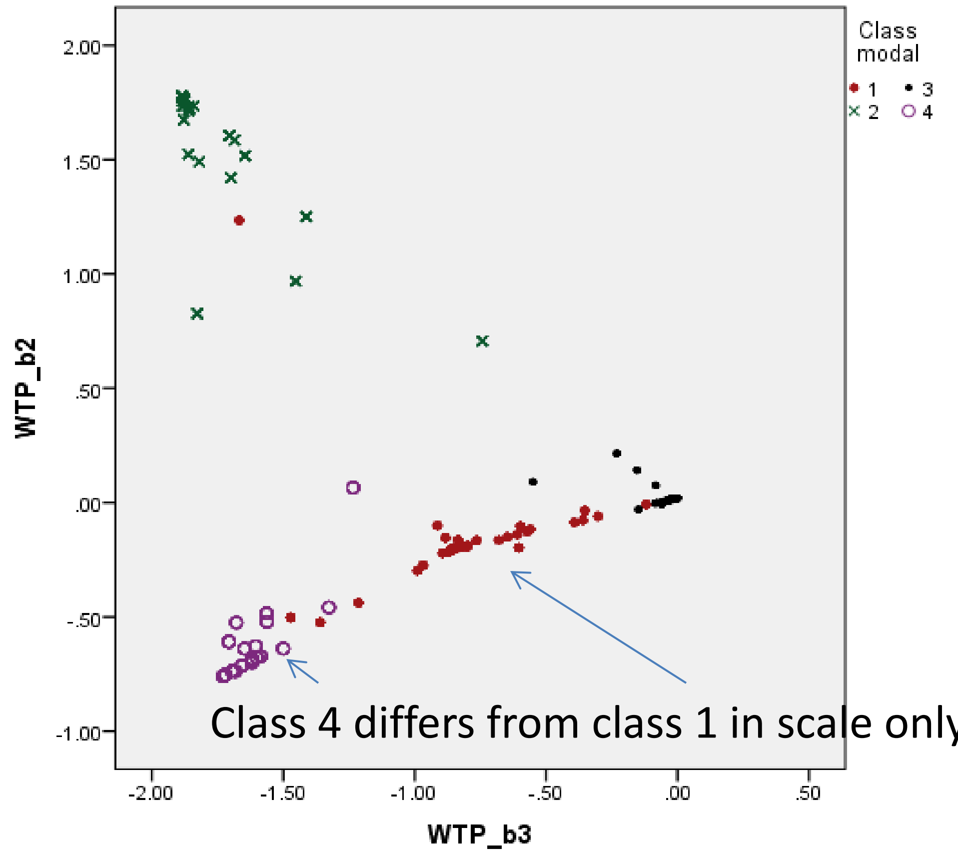


Overall 50% prefer brand B
Hit rate (in-sample) = 68%

Median Absolute Error (MAE) = 0.33
Segment Recovery: 69.2%

Standard LC Yields 4 Segments

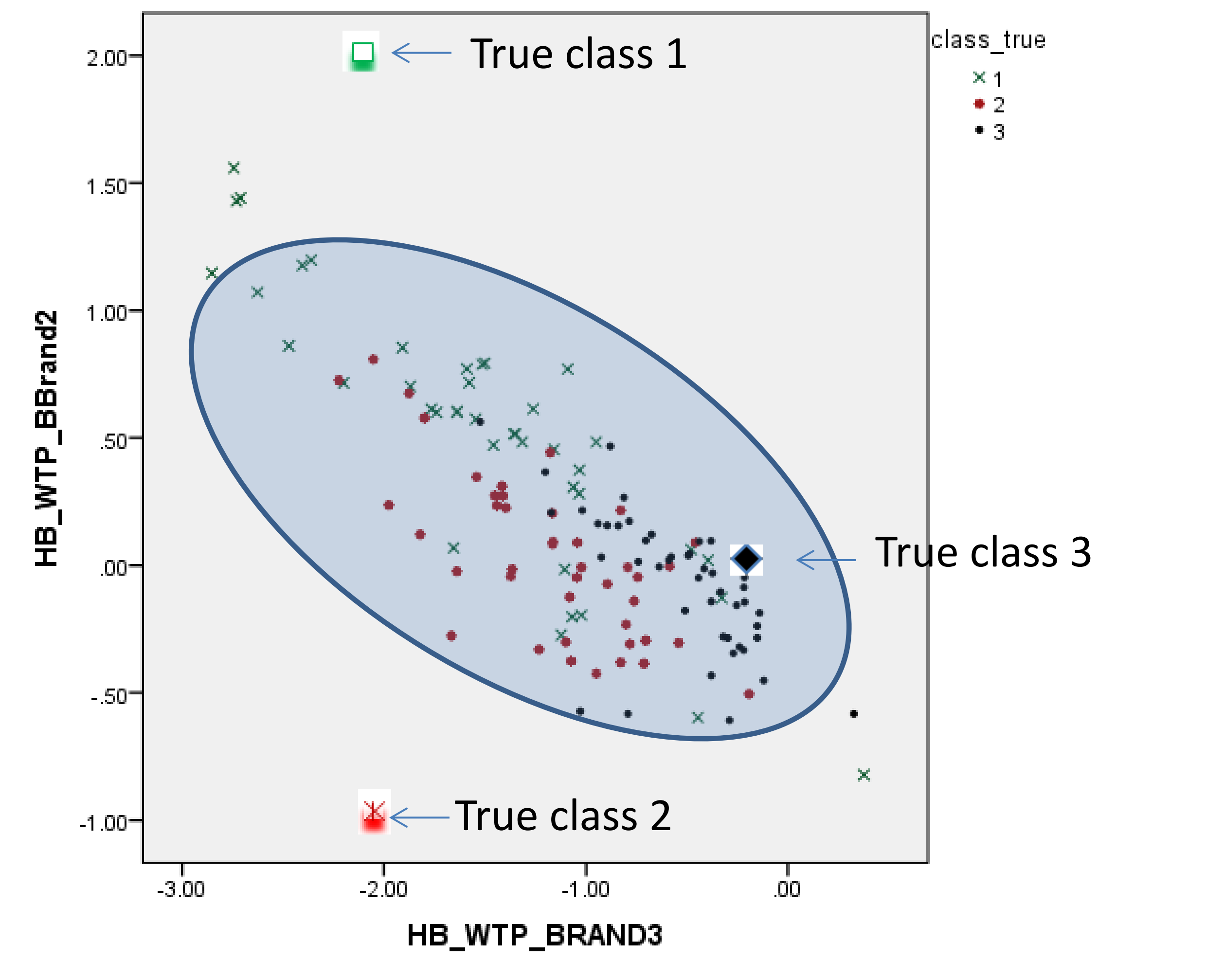




Overall 50% prefer brand B
Hit rate (in-sample) = 67%

Median Absolute Error (MAE) = 0.30
Segment Recovery: 65.2%

HB: MVN Yields Poor Segment Recovery



Attributes	True WTP		
Brand	Class 1	Class 2	Class 3
A	0	0	0
B	2	-1	0
C	-2	-2	0

Neither segments nor WTP parameters recovered well – 'Bayesian shrinkage' due to MVN assumption.

Regardless which clustering method used, segment recovery < 55% correct.

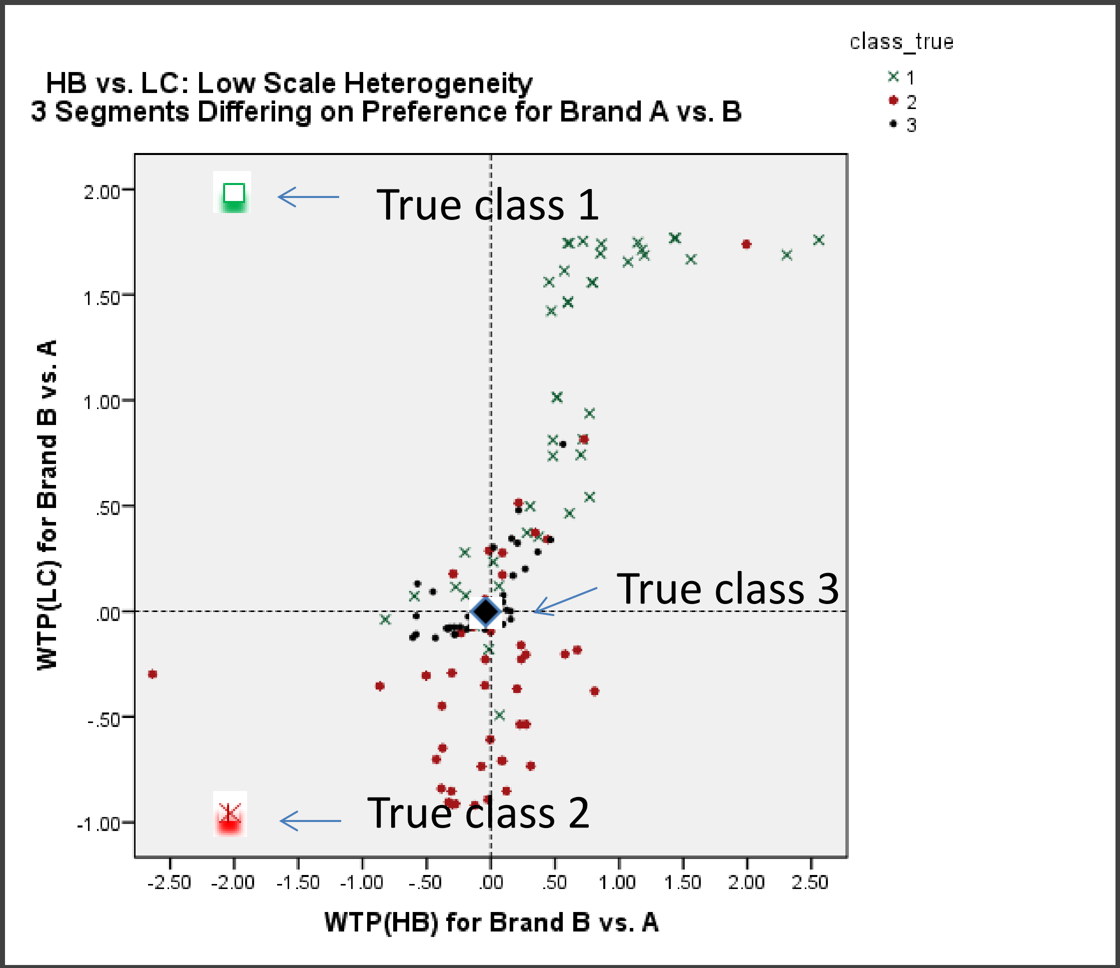
44% who prefer Brand B (Class 1) are mistakenly predicted WTP a premium for brand A over B!

HB predicts 58% prefer brand B while true population rate is 50%.

HB predicts 58% prefer brand B (true 50%)
Hit rate (in-sample) = 70% (overfitting)

MAE= 0.88 (more than twice LC)
Segment recovery (LC) = 54.6%

LC Much Better than HB in Segment Recovery



Attributes	True WTP		
Brand	Class 1	Class 2	Class 3
A	0	0	0
B	2	-1	0
C	-2	-2	0

Segments are not well separated – true WTP parameters not recovered well – substantial 'regression to mean'

Regardless which clustering method used, segment recovery is poor. Overall 58% prefer brand B while true population rate is 50%.

Note also that 44% of those who in truth prefer Brand B to A (Class 1) are predicted to pay a premium for brand A over B!

Summary of Results

Model		N=135 LL	BIC(LL)	Npar	Predicted Choice Hit Rate Holdout (In-sample)	Median error MAE(WTP_b2)	Segment Recovery
SALC/ continuous heterogeneity	**3cl/scfac**	-1002.1	2092.6	18	62% (68%)	0.33	69.2%
SALC/ discrete heterogeneity	**3cl/2scl**	-1003.0	2099.2	19	62% (68%)	0.41	69.4%
Traditional LC	**4class**	-1001.6	2116.1	23	62% (67%)	0.30	65.2%*
HB/ upper level model = MVN	**HB**	-1045.4	2188.8	20	58% (70%)	0.88	54.6%*

*Segment recovery based on 3-class model

SALC vs. G-MNL

- Hess and Rose (2012) suggest that G-MNL model (Fieberg, et. al.) paper is misguided in the interpretation of scale as distinct from preference heterogeneity.
- SALC resolves this criticism using a simple model structure to define the scale factor distinct from WTP or ratios of other part-worths.

MAE Analysis

- Median Absolute Errors (MAE) were calculated on WTP estimates
  - We use WTP instead of preferences because scale drops out
  - We use Median Absolute Error to avoid inflation issues due to a outliers
- Models tested
  - Traditional Latent Class Choice
  - Scale Adjusted Latent Class Choice
  - Mixed logit using Hierarchical Bayes with “default” priors

MAE (N = 135)

	Scale Adjusted Latent Class	Traditional Latent Class	HB
No Scale	N/A	0.39	0.53
Low Scale	0.40	0.34	0.63
Moderate Scale	0.55	0.69	0.73

- Marginal improvement in MAE with SALC as scale heterogeneity increases from Low to Moderate
- With Low scale, traditional LC does as well as SALC
- HB consistently has higher MAE but not substantially higher
- With moderate scale heterogeneity, HB and traditional LC perform similarly

MAE (N = 900)

	Scale Adjusted Latent Class	Traditional Latent Class	HB
No Scale	N/A	0.38	0.51
Low Scale	0.32	0.39	0.59
Moderate Scale	0.33	0.51	0.74

- With a larger sample size, the SALC consistently outperforms the other methods in the presence of scale heterogeneity
- SALC model unaffected by increased scale heterogeneity
- HB MAE consistent with the N = 135 case

MAE (N = 2700)

	Scale Adjusted Latent Class	Traditional Latent Class	HB
No Scale	N/A	0.26	0.49
Low Scale	0.26	0.21	0.54
Moderate Scale	0.28	0.29	0.70

- Traditional LC does okay relative to the SALC models
  - Larger sample results in more classes identified
  - Practical implications of larger classes
- Again, HB MAE is consistent with other sample sizes
- WTP from HB models are affected by scale heterogeneity

Classification Hit Rate Analysis

- Compared true class with modal assignments from LC, and with a variety of clustering techniques on HB posteriors.
- Techniques used with HB include latent cluster, k-means and hierarchical clustering methods
- Assumed 3 clusters even if the statistics suggested more/less clusters
- For the clusters using HB posteriors, we clustered on preference estimates and willingness-to-pay values

Classification Hit Rate – N = 135

- Increasing scale heterogeneity decreases classification hit rates across all methods with SALC models outperforming the other methods
- 2 stage clustering can do as well as traditional LC when there is no scale heterogeneity but suffers when scale is introduced
- Very little difference in 2-stage clustering methods on preferences
- Latent cluster does best with WTP

	No Scale	Low Scale	Moderate Scale
Scale Adjusted Latent Class *(Continuous scale class)*	N/A	69%	58%
Scale Adjusted Latent Class *(Discrete scale classes)*	N/A	69%	59%
Traditional Latent Class	80%	65%	51%
Latent Cluster (HB, wtp)	62%	55%	41%
Latent Cluster (HB, pref)	80%	53%	39%
K-Means (HB, wtp)	58%	36%	35%
K-Means (HB, pref)	81%	47%	40%
Hierarchical (HB, wtp)	53%	36%	35%
Hierarchical (HB, pref)	81%	53%	44%

Classification Hit Rate

- The results are similar across N = 900 and N = 2700

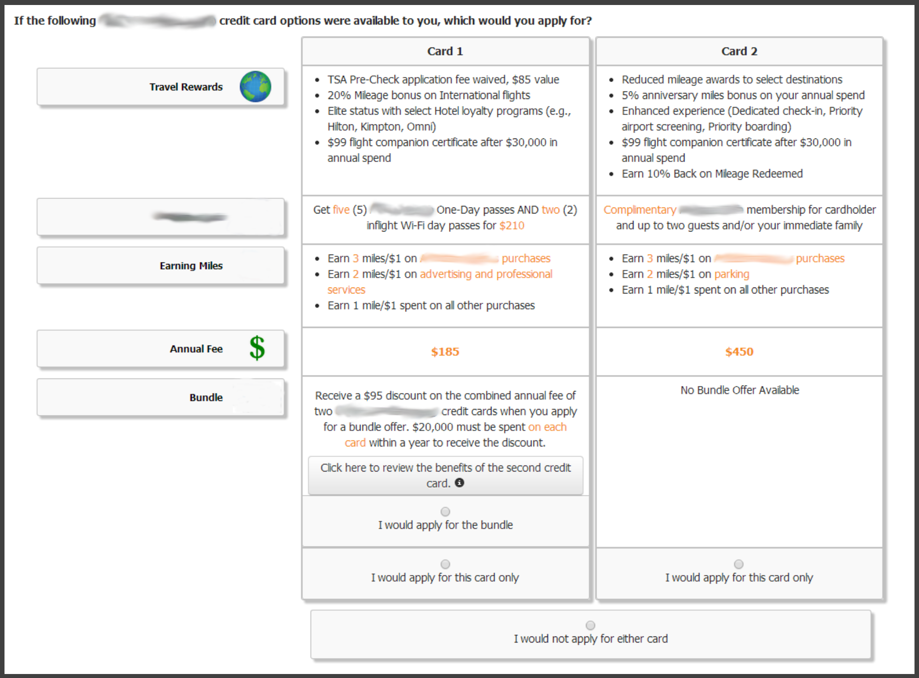
Why looking at real data is important

- Simulated populations assumed a true model that is latent class in nature
  - Intra-segment variation was due to scale differences only
- With real world data, the truth is not known
  - With this analysis, we want to determine which method does best in recovering the “truth”
  - We use a hold-out sample to accomplish this

Case Study with Real Data

- Data come from a 2014 credit card study
- 4,526 respondents based in the US
- Respondents were selected based on:
  - Interest in specific type of card
  - Ownership of card
  - Willingness to pay an annual fee
  - Specific age and income ranges
- Data come from 2 sources:
  - online panel and client's customer DB
- Respondents were allowed to complete the survey using laptops/desktops, tablets and smartphones

Case Study


If the following [illegible] credit card options were available to you, which would you apply for?

	Card 1	Card 2
Travel Rewards	• TSA Pre-Check application fee waived, $85 value • 20% Mileage bonus on International flights • Elite status with select Hotel loyalty programs (e.g., Hilton, Kimpton, Omni) • $99 flight companion certificate after $30,000 in annual spend	• Reduced mileage awards to select destinations • 5% anniversary miles bonus on your annual spend • Enhanced experience (Dedicated check-in, Priority airport screening, Priority boarding) • $99 flight companion certificate after $30,000 in annual spend • Earn 10% Back on Mileage Redeemed
[illegible]	Get five (5) [illegible] One-Day passes AND two (2) inflight Wi-Fi day passes for $210	Complimentary [illegible] membership for cardholder and up to two guests and/or your immediate family
Earning Miles	• Earn 3 miles/$1 on [illegible] purchases • Earn 2 miles/$1 on advertising and professional services • Earn 1 mile/$1 spent on all other purchases	• Earn 3 miles/$1 on [illegible] purchases • Earn 2 miles/$1 on parking • Earn 1 mile/$1 spent on all other purchases
Annual Fee	$185	$450
Bundle	Receive a $95 discount on the combined annual fee of two [illegible] credit cards when you apply for a bundle offer. $20,000 must be spent on each card within a year to receive the discount. Click here to review the benefits of the second credit card.	No Bundle Offer Available
	I would apply for the bundle	
	I would apply for this card only	I would apply for this card only

I would not apply for either card

- Respondents completed 8 stated choice experiments
- Up to 4 alternatives presented in an experiment

Analysis

- We compare SALC with traditional HB and LC in predictive accuracy using a holdout sample.
- Holdout sample includes:
  - A simple random sampling of one choice task for each respondent.
  - A simple random sampling of 10% of the respondents.

Case Study Results

Model	Log-Likelihood	# Param	BIC	In-Sample Hit Rate (Holdouts)	Out-of-sample (Probability of Choice)
Prior Variance 0.1	-23231	1484	58797	75.9% (76.3%)	67.1%
Prior Variance 2	-18164	1484	48663	79.9% (78.1%)	62.3%
Prior Variance 10	-18125	1484	48585	80.5% (78.6%)	62.3%
Prior Variance 0.1 (with scale)	-23264	1484	58862	75.9% (76.3%)	66.9%
Prior Variance 2 (with scale)	-18206	1484	48747	79.8% (78.6%)	62.3%
Prior Variance 10 (with scale)	-18116	1484	48568	80.6% (78.6%)	62.3%
LC (3-DFactors + 2 scale classes)	-16837	221	35535	83.8% (80.2%)	62.0%

References

- Fiebig, D. G., Keane, M., Louviere, J. J.,Wasi, N., 2010. The generalized multino-ial logit: accounting for scale and coecient heterogeneity. Marketing Science 29 (3), 393-421..
- Louviere, Jordan J., Thomas C. Eagle, 2006. "Confound it! That Pesky Little Scale Constant Messes Up," 2006 Sawtooth Software Conference Proceedings.
- Vermunt and Magidson (2007) User's manual and Technical Guide for LG-Syntax™ – the syntax module for Latent GOLD and LG Choice 4.5.
- Magidson, J., and Vermunt, J.K. (2007). Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. Sawtooth Software Conference Proceedings, October 2007, 139-154.
- Vermunt, J.K. (2013). Categorical response data. In: M.A. Scott, J.S. Simonoff, and B.D. Marx (Eds.), The SAGE handbook of multilevel modeling. Sage.
- Hess S. and J. Rose (2012) "Can scale and coecient heterogeneity be separated in random coefficients models?", Transportation.