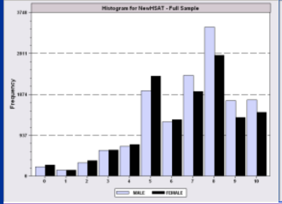
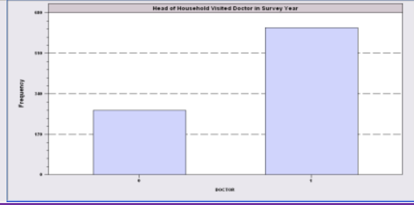
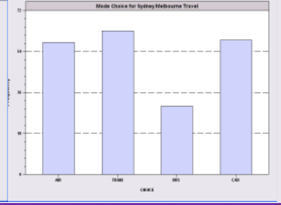
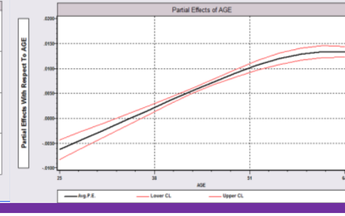
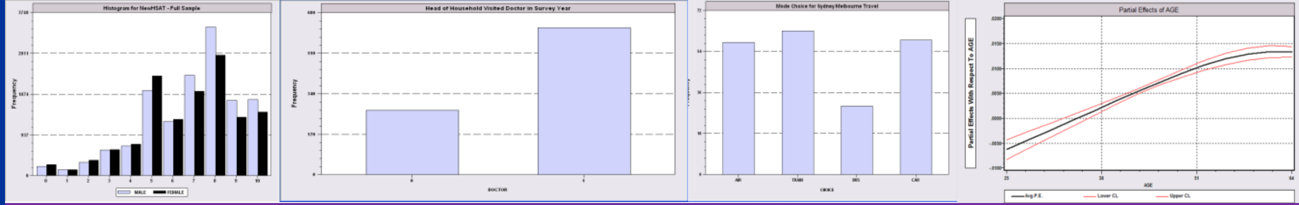





Discrete Choice Modeling

William Greene
Stern School of Business
New York University



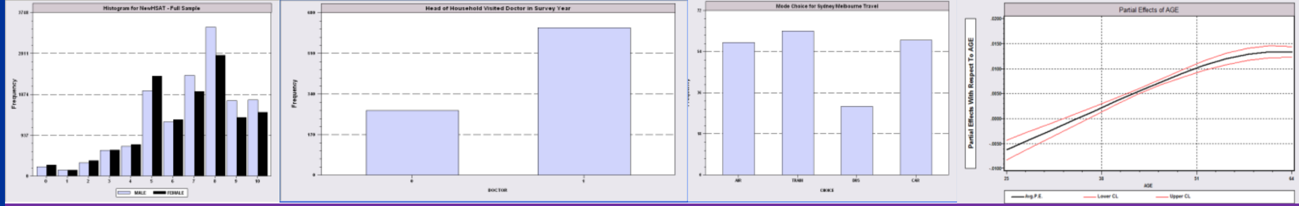
Discrete Parameter Heterogeneity
Latent Classes

Discrete unobservable partition of the population into Q classes

Discrete approximation to a continuous distribution of parameters across individuals

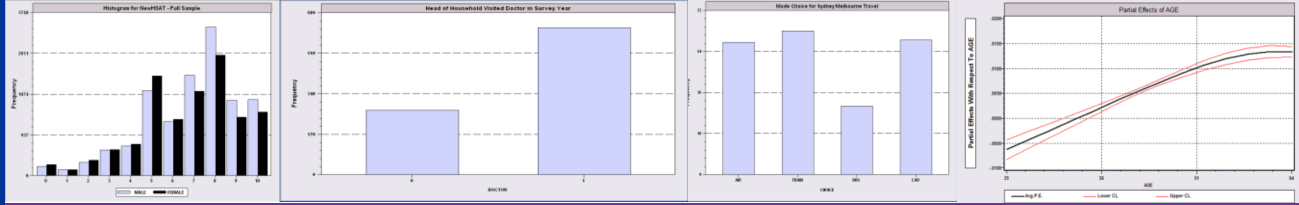
$$\text{Prob}[\boldsymbol{\beta} = \boldsymbol{\beta}_q \mid \mathbf{w}_i] = \pi_{iq}, \; q = 1,\ldots,Q$$

$$\pi_{iq} = \frac{\exp(\boldsymbol{\theta}_q' \mathbf{w}_i)}{\sum_{q=1}^{Q} \exp(\boldsymbol{\theta}_q' \mathbf{w}_i)}$$



Latent Class Probabilities

- Ambiguous – Classical Bayesian model?
  - The randomness of the class assignment is from the point of view of the observer, not a natural process governed by a discrete distribution.
- Equivalent to random parameters models with discrete parameter variation
  - Using nested logits, etc. does not change this
  - Precisely analogous to continuous ‘random parameter’ models
- Not always equivalent – zero inflation models – in which classes have completely different models



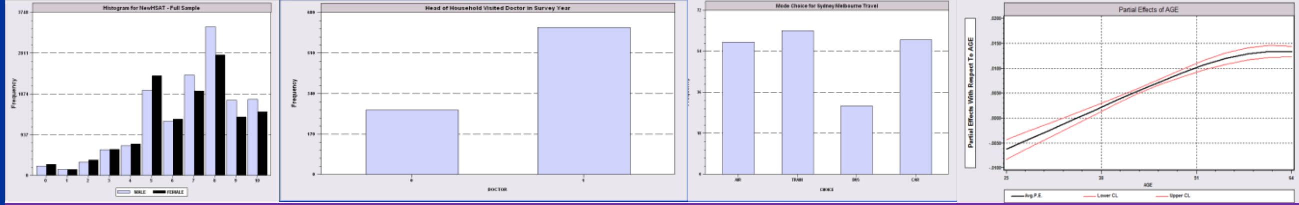
A Latent Class MNL Model

- Within a "class"

$$P[\text{choice } j \mid i,t,\text{class}=q] = \frac{\exp(\alpha_j + \boldsymbol{\beta}_q'\mathbf{x}_{itj} + \boldsymbol{\gamma}_{j,q}'\mathbf{z}_{it})}{\sum_{j=1}^{J(i)} \exp(\alpha_j + \boldsymbol{\beta}_q'\mathbf{x}_{itj} + \boldsymbol{\gamma}_{j,q}'\mathbf{z}_{it})}$$

- Class sorting is probabilistic (to the analyst) determined by individual characteristics

$$P[\text{class } q \mid i] = \frac{\exp(\boldsymbol{\theta}_q'\mathbf{w}_i)}{\sum_{c=1}^{Q} \exp(\boldsymbol{\theta}_c'\mathbf{w}_i)} = H_{iq}$$



Two Interpretations of Latent Classes

Heterogeneity with respect to 'latent' consumer classes

$$\Pr(\text{Choice}_i) = \sum_{q=1}^{Q} \Pr(\text{Choice}_i \mid \text{class} = q)\Pr(\text{Class} = q)$$

$$\Pr(\text{Choice}_i \mid \text{Class} = q) = \frac{\exp(\boldsymbol{\beta}_q'\mathbf{x}_{i,\text{choice}})}{\Sigma_{j=\text{choice}}\exp(\boldsymbol{\beta}_q'\mathbf{x}_{i,\text{choice}})}$$

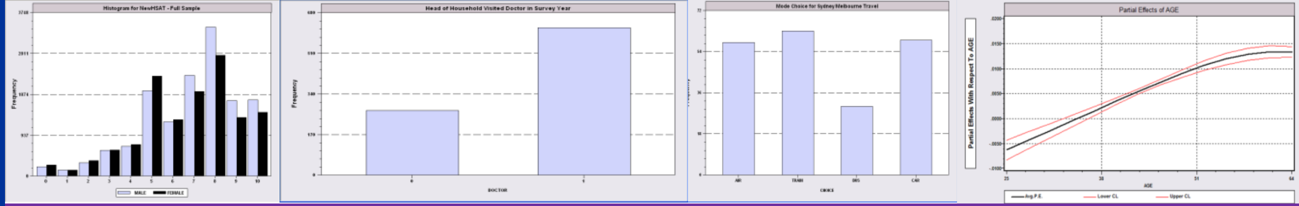
$$\Pr(\text{Class} = q \mid i) = F_{i,q} = \frac{\exp(\boldsymbol{\theta}_q'\mathbf{z}_i)}{\Sigma_{q=\text{classes}}\exp(\boldsymbol{\theta}_q'\mathbf{z}_i)}$$

Discrete random parameter variation

$$\Pr(\text{Choice}_i \mid \boldsymbol{\beta}_i) = \frac{\exp(\boldsymbol{\beta}_i'\mathbf{x}_{i,j})}{\Sigma_{j=\text{choice}}\exp(\boldsymbol{\beta}_i'\mathbf{x}_{i,j})}$$

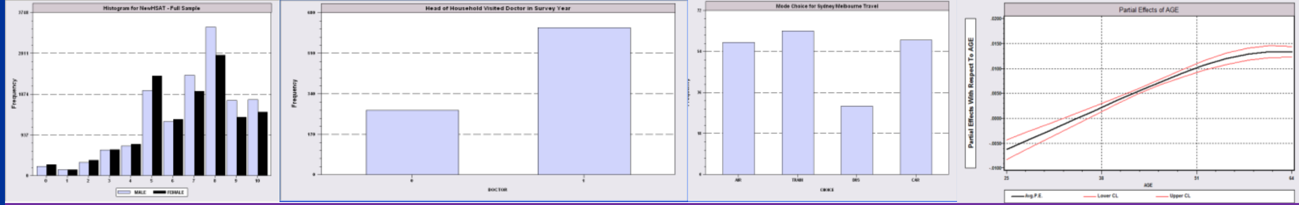
$$\Pr(\boldsymbol{\beta}_i = \boldsymbol{\beta}_q) = F_{i,q} = \frac{\exp(\boldsymbol{\theta}_q'\mathbf{z}_i)}{\Sigma_{q=\text{classes}}\exp(\boldsymbol{\theta}_q'\mathbf{z}_i)}, q = 1,\ldots,Q$$

$$\Pr(\text{Choice}_i) = \sum_{q=1}^{Q} \Pr(\text{choice} \mid \boldsymbol{\beta}_i = \boldsymbol{\beta}_q)\Pr(\boldsymbol{\beta}_i = \boldsymbol{\beta}_q)$$



Estimates from the LCM

- Taste parameters within each class $\boldsymbol{\beta}_q$
- Parameters of the class probability model, $\boldsymbol{\theta}_q$
- For each person:
  - Posterior estimates of the class they are in q|i
  - Posterior estimates of their taste parameters $E[\beta_q|i]$
  - Posterior estimates of their behavioral parameters, elasticities, marginal effects, etc.



Using the Latent Class Model

Computing posterior (individual specific) class probabilities

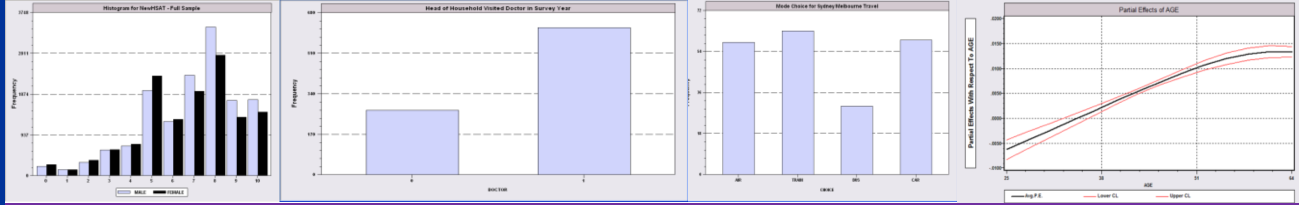
$$\hat{F}_{q|i} = \frac{\hat{P}_{i|q}\hat{F}_{iq}}{\sum_{q=1}^{Q}\hat{P}_{i|q}\hat{H}_{iq}} \quad \text{(posterior)} \qquad \text{Note } \hat{F}_{q|i} \text{ vs. } \hat{F}_{iq}$$

$$\hat{F}_{iq} = \text{estimated prior class probability}$$

$$\hat{P}_{i|q} = \text{estimated choice probability for the choice made, given the class}$$

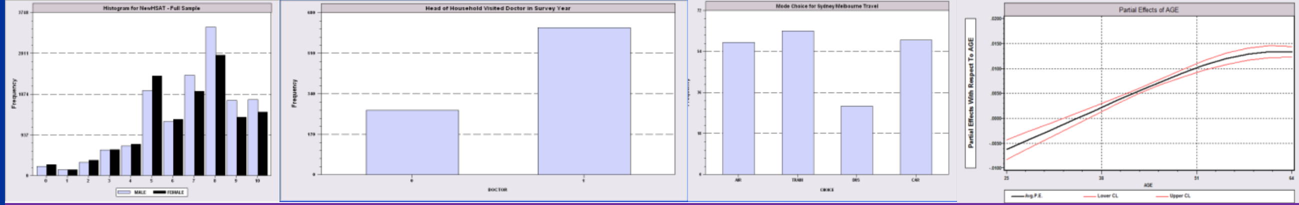
Computing posterior (individual specific) taste parameters

$$\hat{\boldsymbol{\beta}}_i = \sum_{q=1}^{Q}\hat{F}_{q|i}\,\hat{\boldsymbol{\beta}}_q$$

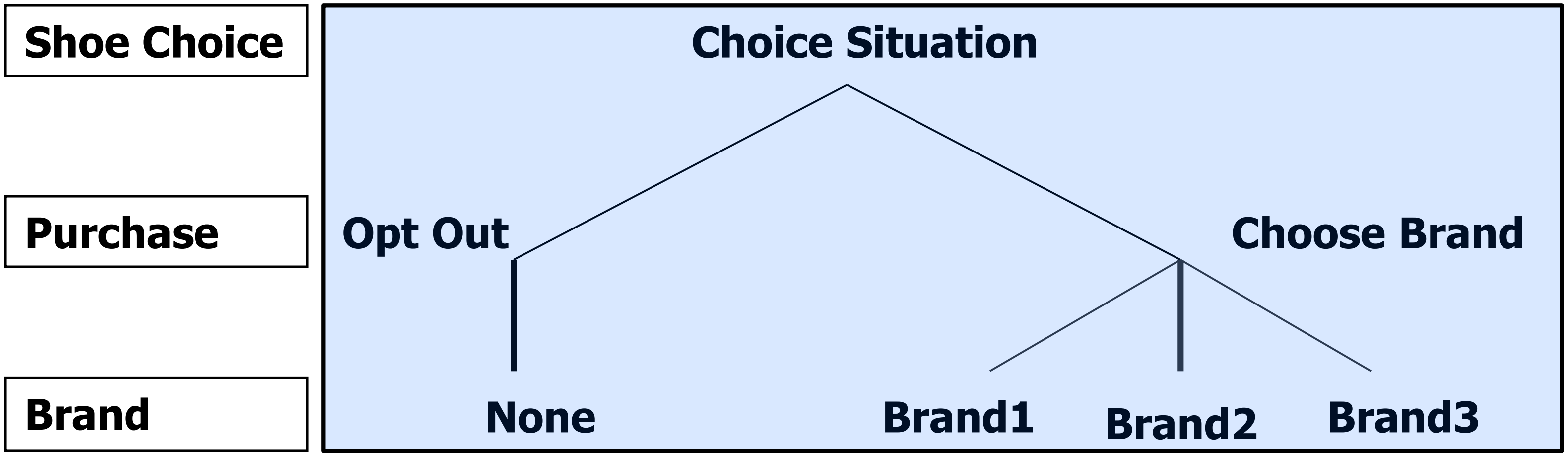


Application: Shoe Brand Choice

- Simulated Data: Stated Choice, 400 respondents, 8 choice situations, 3,200 observations
- 3 choice/attributes + NONE
  - Fashion = High / Low
  - Quality = High / Low
  - Price = 25/50/75,100 coded 1,2,3,4
- Heterogeneity: Sex (Male=1), Age (<25, 25-39, 40+)
- Underlying data generated by a 3 class latent class process (100, 200, 100 in classes)
- Thanks to www.statisticalinnovations.com (Latent Gold)

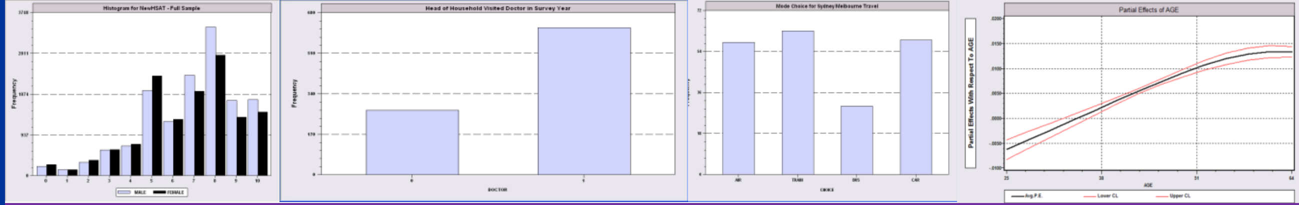


Degenerate Branches



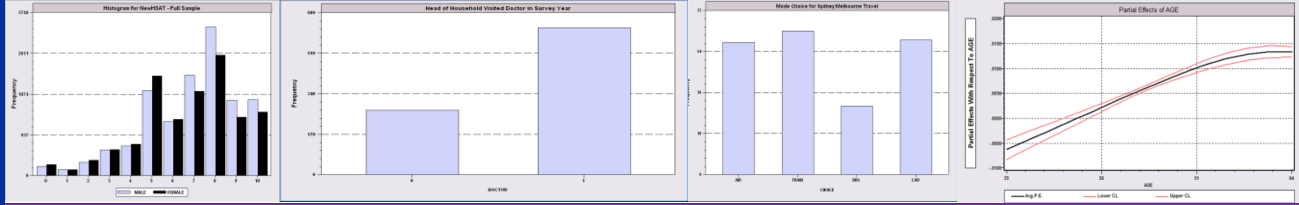
$$U(\text{Brand } j) = \beta_1\text{Fashion} + \beta_2\text{Quality} + \beta_3\text{Price} + \varepsilon_{\text{Brand}}$$

$$U(\text{None}) = \beta_0 + \varepsilon_{\text{None}}$$



One Class MNL Estimates

```
-----------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                  Choice
Log likelihood function        -4158.50286
Estimation based on N =    3200, K =    4
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only  -4391.1804  .0530 .0510
Response data are given as ind. choices
Number of obs.=  3200, skipped    0 obs
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
  FASH|1|    1.47890***       .06777        21.823   .0000
  QUAL|1|    1.01373***       .06445        15.730   .0000
 PRICE|1|   -11.8023***       .80406       -14.678   .0000
  ASC4|1|     .03679          .07176          .513   .6082
--------+--------------------------------------------------
```



Application: Brand Choice

True underlying model is a three class LCM

NLOGIT

; Lhs=choice

; Choices=Brand1,Brand2,Brand3,None

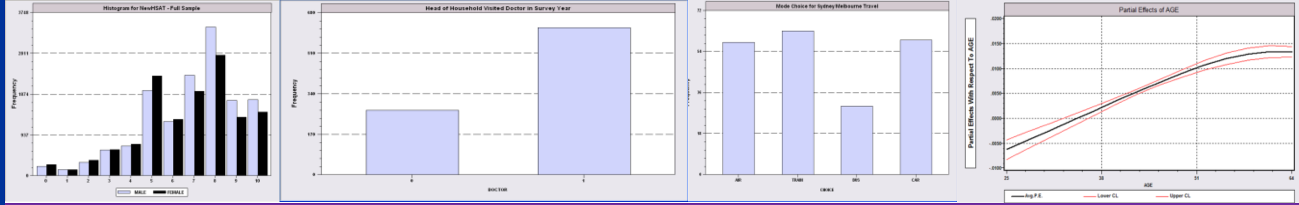
; Rhs = Fash,Qual,Price,ASC4

; LCM=Male,Age25,Age39

; Pts=3

; Pds=8

; Parameters (Save posterior results) $

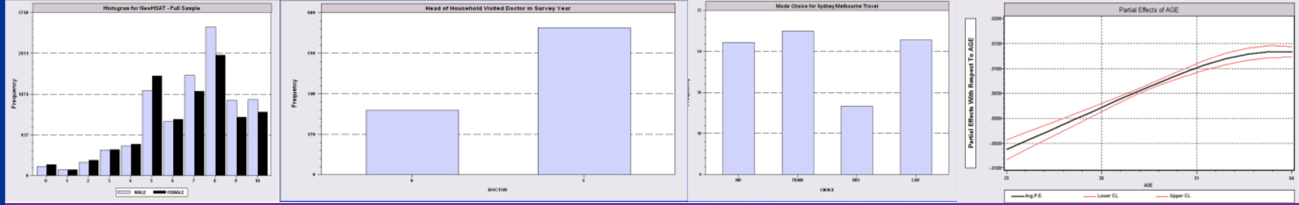


Three Class LCM

```
Normal exit from iterations. Exit status=0.
-----------------------------------------------------------
Latent Class Logit Model
Dependent variable                     CHOICE
Log likelihood function           -3649.13245
Restricted log likelihood         -4436.14196
Chi squared [  20 d.f.]            1574.01902
Significance level                     .00000
McFadden Pseudo R-squared            .1774085
Estimation based on N =   3200, K =   20
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients -4436.1420  .1774 .1757
Constants only  -4391.1804  .1690 .1673
At start values -4158.5428  .1225 .1207
Response data are given as ind. choices
Number of latent classes =              3
Average Class Probabilities
     .506  .239  .256
LCM model with panel has     400 groups
Fixed number of obsrvs./group=          8
Number of obs.=  3200, skipped    0 obs
--------+--------------------------------------------------
```

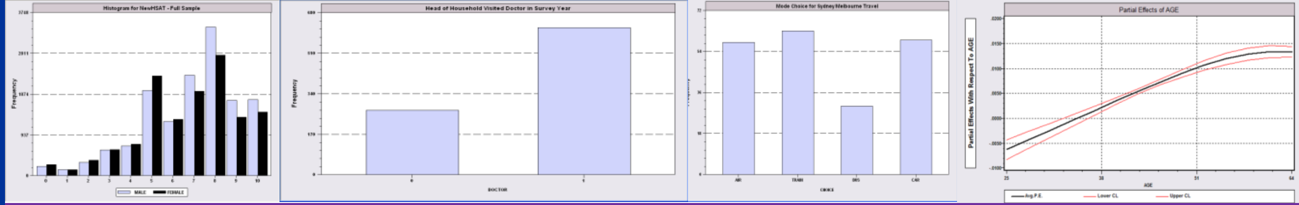
LogL for one class MNL = -4158.503

Based on the LR statistic it would seem unambiguous to reject the one class model. The degrees of freedom for the test are uncertain, however.



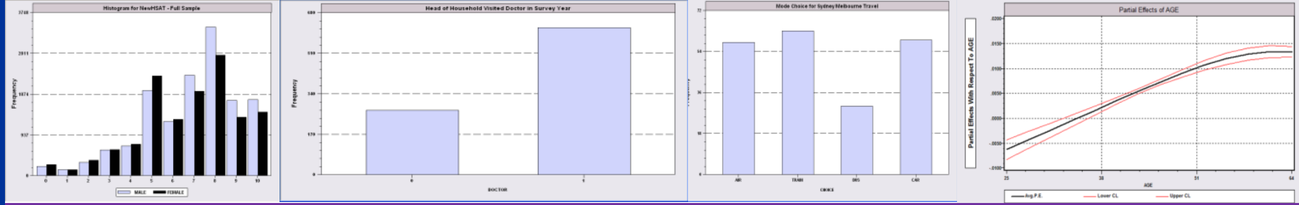
Estimated LCM: Utilities

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
	Utility parameters in latent class -->> 1			
FASH|1	3.02570***	.14549	20.796	.0000
QUAL|1	-.08782	.12305	-.714	.4754
PRICE|1	-9.69638***	1.41267	-6.864	.0000
ASC4|1	1.28999***	.14632	8.816	.0000
	Utility parameters in latent class -->> 2			
FASH|2	1.19722***	.16169	7.404	.0000
QUAL|2	1.11575***	.16356	6.821	.0000
PRICE|2	-13.9345***	1.93541	-7.200	.0000
ASC4|2	-.43138**	.18514	-2.330	.0198
	Utility parameters in latent class -->> 3			
FASH|3	-.17168	.16725	-1.026	.3047
QUAL|3	2.71881***	.17907	15.183	.0000
PRICE|3	-8.96483***	1.93400	-4.635	.0000
ASC4|3	.18639	.18412	1.012	.3114



Estimated LCM: Class Probability Model

```
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
        |This is THETA(01) in class probability model.
Constant|    -.90345**        .37612       -2.402   .0163
   _MALE|1|   .64183*         .36245        1.771   .0766
  _AGE25|1|  2.13321***       .32096        6.646   .0000
  _AGE39|1|   .72630*         .43511        1.669   .0951
        |This is THETA(02) in class probability model.
Constant|     .37636          .34812        1.081   .2796
   _MALE|2| -2.76536***       .69325       -3.989   .0001
  _AGE25|2|  -.11946          .54936        -.217   .8279
  _AGE39|2|  1.97657***       .71684        2.757   .0058
        |This is THETA(03) in class probability model.
Constant|       .000       ......(Fixed Parameter)......
   _MALE|3|     .000       ......(Fixed Parameter)......
  _AGE25|3|     .000       ......(Fixed Parameter)......
  _AGE39|3|     .000       ......(Fixed Parameter)......
--------+--------------------------------------------------
```

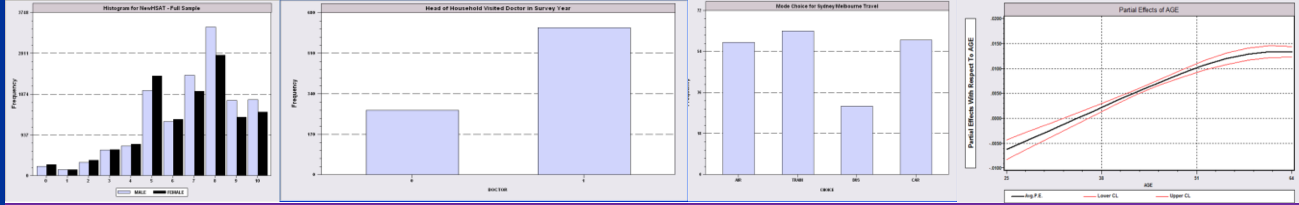


Estimated LCM: Conditional Parameter Estimates

Matrix - BETA_I

[400, 4] Cell: 1.22418

	1	2	3	4
1	1.22418	1.09802	-13.8718	-0.405949
2	-0.15862	2.70354	-9.01192	0.180574
3	-0.00214476	2.52027	-9.58031	0.109885
4	-0.169643	2.71642	-8.97222	0.185475
5	-0.170067	2.71692	-8.97068	0.185666
6	2.54209	0.230583	-10.8165	0.834913
7	-0.0906979	2.62398	-9.25879	0.149855
8	-0.137504	2.67879	-9.08889	0.170972
9	-0.154206	2.69835	-9.02826	0.178509
10	0.138079	2.35607	-10.0894	0.0466067
11	-0.0295339	2.55235	-9.48087	0.122245
12	0.0689965	2.43696	-9.83858	0.0777801
13	0.525364	1.90255	-11.4952	-0.128137
14	-0.133549	2.67423	-9.10241	0.169384
15	-0.170907	2.7179	-8.96763	0.186045
16	-0.170993	2.71801	-8.96731	0.186084
17	-0.161914	2.70737	-9.00028	0.181987
18	-0.00529745	2.52397	-9.56886	0.111309
19	-0.0527786	2.58006	-9.39075	0.134078
20	-0.157123	2.70176	-9.01767	0.179825
21	1.25046	1.08392	-13.7739	-0.372533
22	0.0293861	2.48339	-9.69433	0.0957607
23	-0.0685934	2.59811	-9.33889	0.139916
24	-0.0866831	2.61927	-9.2734	0.148036

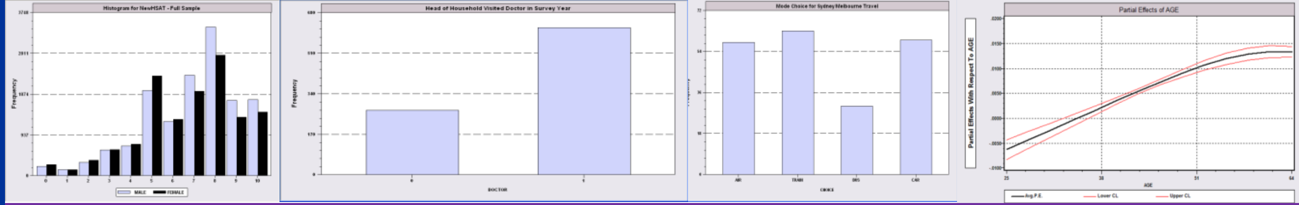


Estimated LCM: Conditional (Posterior) Class Probabilities

Matrix - CLASSP_I

[400, 3] Cell: 0.0147633

	1	2	3
1	0.0147633	0.985211	2.60918e-005
2	2.89729e-005	0.009471	0.9905
3	2.1798e-007	0.123846	0.876154
4	5.98856e-008	0.00148665	0.998513
5	8.85439e-008	0.00117665	0.998823
6	0.735593	0.264301	0.000105599
7	3.02646e-006	0.0591501	0.940847
8	2.67545e-007	0.0249636	0.975036
9	5.82809e-008	0.0127631	0.987237
10	1.6727e-006	0.226278	0.77372
11	2.16528e-007	0.103838	0.896162
12	4.70195e-007	0.175815	0.824184
13	1.52577e-005	0.509164	0.490821
14	7.7743e-005	0.0276719	0.97225
15	2.59739e-009	0.000563339	0.999437
16	2.19284e-008	0.000499939	0.9995
17	4.34144e-009	0.00713275	0.992867
18	7.19395e-007	0.121542	0.878458
19	0.000527523	0.0856256	0.913847
20	1.22675e-007	0.0106322	0.989368
21	0.0325424	0.962881	0.0045771
22	4.16132e-005	0.146783	0.853175
23	1.7085e-005	0.0752649	0.924718
24	2.25175e-008	0.0620899	0.93791
25	2.21755e-006	0.0477313	0.952267
26	0.00115813	0.173652	0.82519
27	1.52652e-006	0.0861543	0.913844



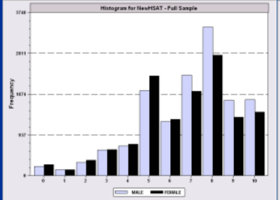
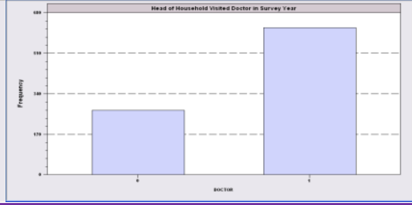
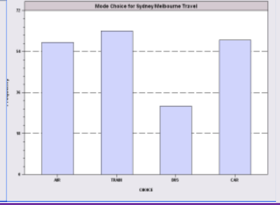
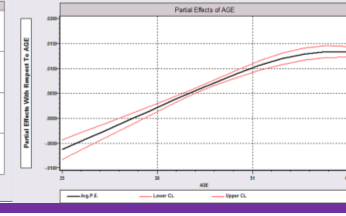
Average Estimated Class Probabilities

MATRIX ; list ; 1/400 * classp_i'1$

Matrix Result has 3 rows and 1 columns.

```
                 1
      +--------------
     1|        .50555
     2|        .23853
     3|        .25593
```

This is how the data were simulated. Class probabilities are .5, .25, .25. The model ‘worked.’

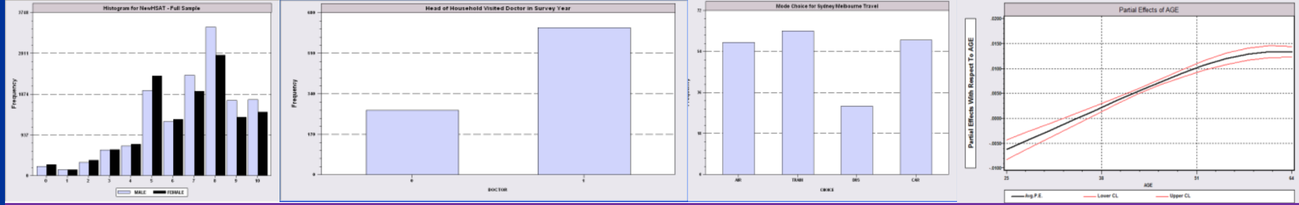





Elasticities

```
+--------------------------------------------------------+
| Elasticity             averaged over observations.|
| Effects on probabilities of all choices in model: |
| * = Direct Elasticity effect of the attribute.    |
| Attribute is PRICE    in choice BRAND1            |
|                                 Mean      St.Dev  |
| *      Choice=BRAND1           -.8010      .3381  |
|        Choice=BRAND2            .2732      .2994  |
|        Choice=BRAND3            .2484      .2641  |
|        Choice=NONE              .2193      .2317  |
+--------------------------------------------------------+
| Attribute is PRICE    in choice BRAND2            |
|        Choice=BRAND1            .3106      .2123  |
| *      Choice=BRAND2          -1.1481      .4885  |
|        Choice=BRAND3            .2836      .2034  |
|        Choice=NONE              .2682      .1848  |
+--------------------------------------------------------+
| Attribute is PRICE    in choice BRAND3            |
|        Choice=BRAND1            .3145      .2217  |
|        Choice=BRAND2            .3436      .2991  |
| *      Choice=BRAND3           -.6744      .3676  |
|        Choice=NONE              .3019      .2187  |
+--------------------------------------------------------+
```

Elasticities are computed by averaging individual elasticities computed at the expected (posterior) parameter vector.

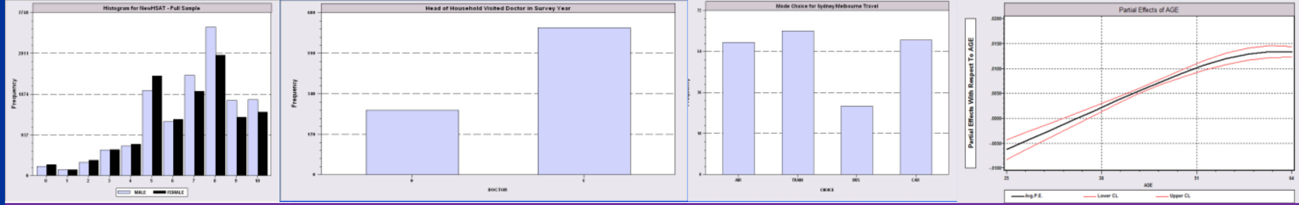
This is an **unlabeled choice experiment**. It is not possible to attach any significance to the fact that the elasticity is different for Brand1 and Brand 2 or Brand 3.



Application: Long Distance Drivers' Preference for Road Environments

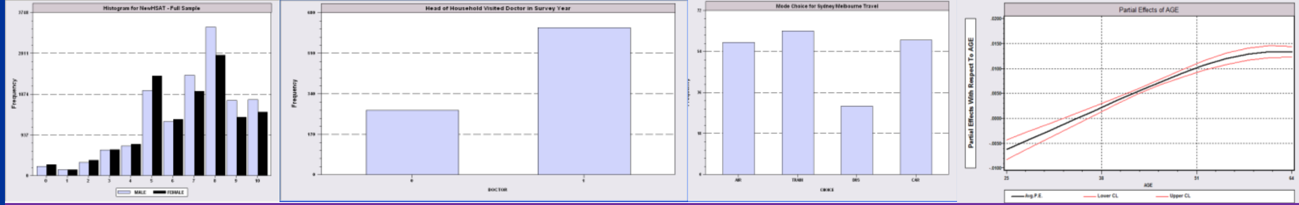
- New Zealand survey, 2000, 274 drivers
- Mixed revealed and stated choice experiment
- 4 Alternatives in choice set
  - The current road the respondent is/has been using;
  - A hypothetical 2-lane road;
  - A hypothetical 4-lane road with no median;
  - A hypothetical 4-lane road with a wide grass median.
- 16 stated choice situations for each with 2 choice profiles
  - choices involving all 4 choices
  - choices involving only the last 3 (hypothetical)

Hensher and Greene, A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit – Transportation Research B, 2003



Attributes

- Time on the open road which is free flow (in minutes);
- Time on the open road which is slowed by other traffic (in minutes);
- Percentage of total time on open road spent with other vehicles close behind (ie tailgating) (%);
- Curviness of the road (A four-level attribute - almost straight, slight, moderate, winding);
- Running costs (in dollars);
- Toll cost (in dollars).

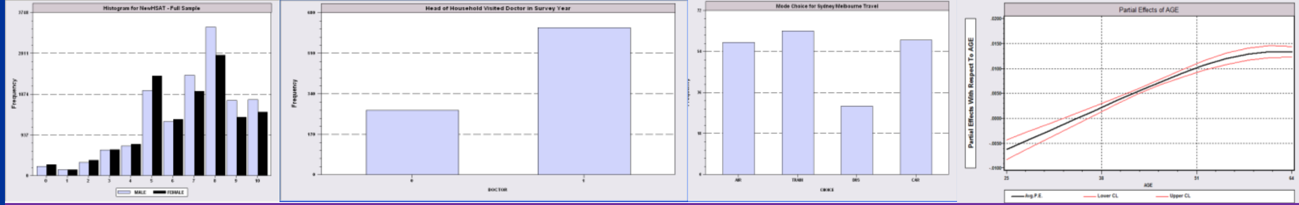


Experimental Design

The four levels of the six attributes chosen are:

- Free Flow Travel Time: -20%, -10%, +10%, +20%
- Time Slowed Down: -20%, -10%, +10%, +20%
- Percent of time with vehicles close behind: -50%, -25%, +25%, +50%
- Curviness: almost, straight, slight, moderate, winding
- Running Costs: -10%, -5%, +5%, +10%
- Toll cost for car and double for truck if trip duration is:

1 hours or less	0,	0.5,	1.5,	3
Between 1 hour and 2.5 hours	0,	1.5,	4.5,	9
More than 2.5 hours	0,	2.5,	7.5,	15

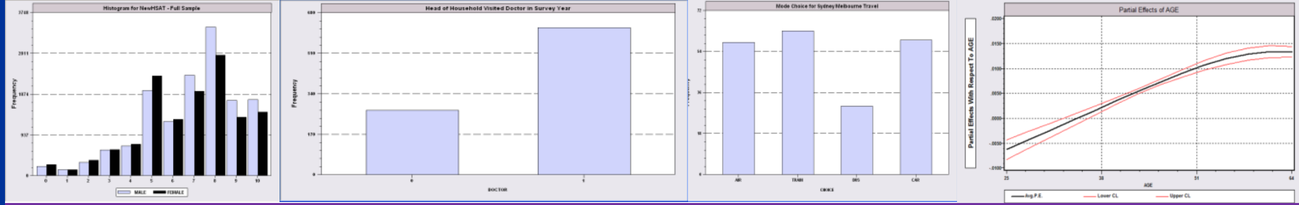


Estimated Latent Class Model

Table 1 Estimated Discrete Choice Models (*t* ratios in parentheses)

Attribute	Alternative	MNL	LCM		
			Class 1	Class 2	Class 3
Travel time[a]	**2 Lane (2L)**	**-.00541 (-5.9)**	**-.00885 (-3.0)**	**-.0090 (-6.9)**	**-.0051 (-5.5)**
Travel time	**4 Lane w/out Median (4NM)**	**-.00475 (-5.3)**	**-.01119 (-4.4)**	**-.0068 (-6.6)**	**-.0063 (-6.2)**
Travel time	**4 Lane with Med. (4WM)**	**-.00375 (-4.4)**	**-.00348 (-1.4)**	**-.0062 (-5.6)**	**-.00424 (-4.3)**
Percent time being tailgated (%)	**All**	**-.01061 (-6.1)**	**-.00976 (-2.6)**	**-.0308 (-15.2)**	**-.0039 (-1.6)**
Total trip cost (toll plus running cost)[b]	**All**	**-.1292 (-25.70**	**-.1565 (-15.9)**	**-.0741 (-13.6)**	**-.2447 (-43.3)**
4NM constant	**4NM**	**.22029 (2.8)**	**2.0259 (7.9)**	**.9637 (8.6)**	**-.3533 (-4.9)**
4WM constant	**4WM**	**.72072 (10.3)**	**3.0696 (12.9)**	**.6770 (5.3)**	**-.2886 (-4.8)**
Latent class Probability		-	**.31722 (10.5)**	**.2703 (8.4)**	**.4124 (12.3)**
Log-likelihood[c]	**-4095.2**		**-3532.9**		
Pseudo-R^2	**.0999**		**.2645**		

[a]Travel time is in minutes. [b]Cost is in dollars [c]4384 observations, 16 observations per person.

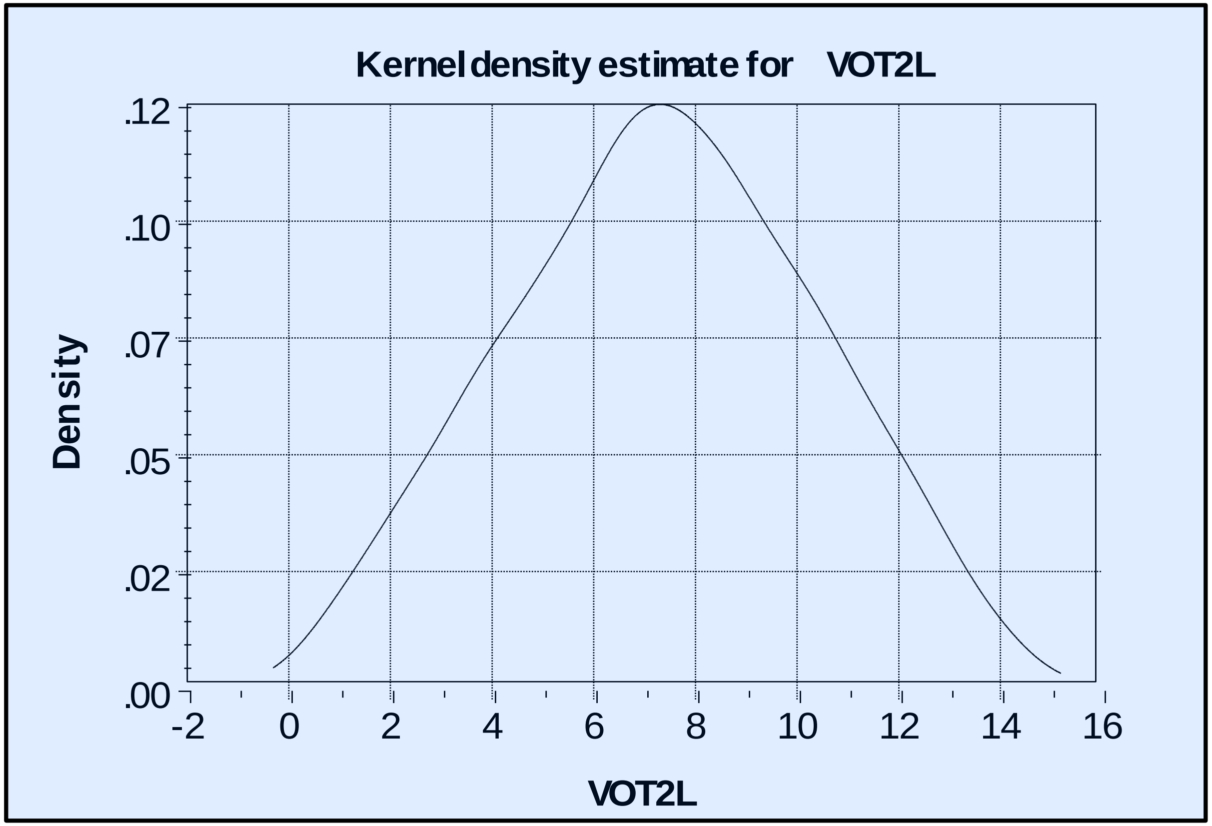


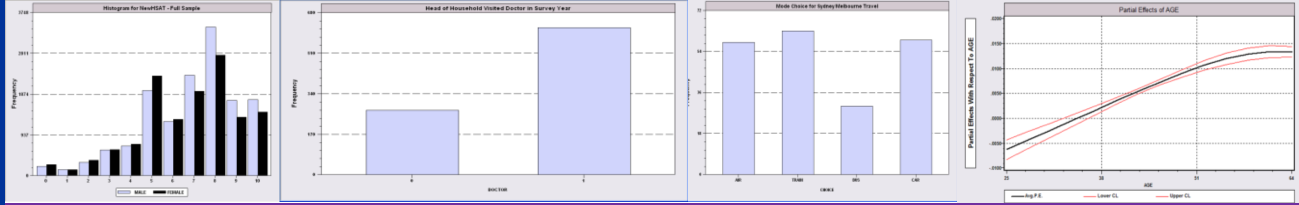
Estimated Value of Time Saved

Table 2 Implied Values of Travel Time Savings (Willingness to Pay) ($NZ per person hour) ns = not statistically significant.
Estimates in parenthesis for MLM are the standard deviation values

Alternative	MNL	LCM		
		Class 1	Class 2	Class 3
2 Lane (2L)	2.52	3.39	7.33	1.26
4 Lane w/out Median (4NM)	2.20	4.30	5.53	1.55
4 Lane with Median (4WM)	1.74	1.34 (ns)	5.02	1.04

Distribution of Parameters – Value of Time on 2 Lane Road





Decision Strategy in Multinomial Choice

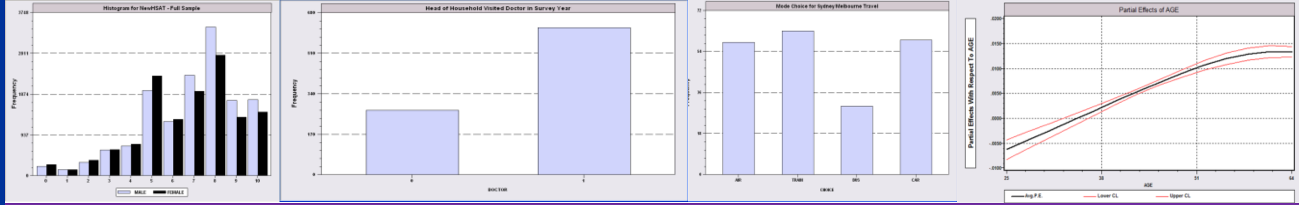
Choice Situation: Alternatives $A_1,\ldots,A_J$

Attributes of the choices: $x_1,\ldots,x_K$

Characteristics of the individual: $z_1,\ldots,z_M$

Random utility functions: $U(j|x,z) = U(x_{ij},z_j,\varepsilon_{ij})$

Choice probability model: $\text{Prob}(\text{choice}=j)=\text{Prob}(U_j > U_l)\ \forall\ l \neq j$



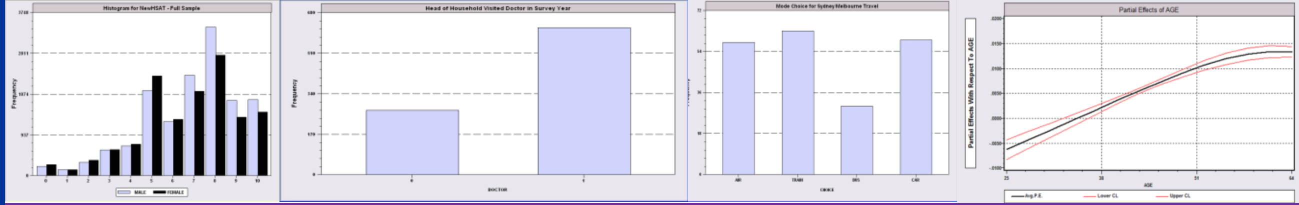
Multinomial Logit Model

$$\text{Prob}(\text{choice} = j) = \frac{\exp[\boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}_j'\mathbf{z}_i]}{\sum_{j=1}^{J} \exp[\boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}_j'\mathbf{z}_i]}$$

Behavioral model assumes

(1) Utility maximization (and the underlying micro- theory)

(2) **Individual pays attention to all attributes. That is the implication of the nonzero β.**



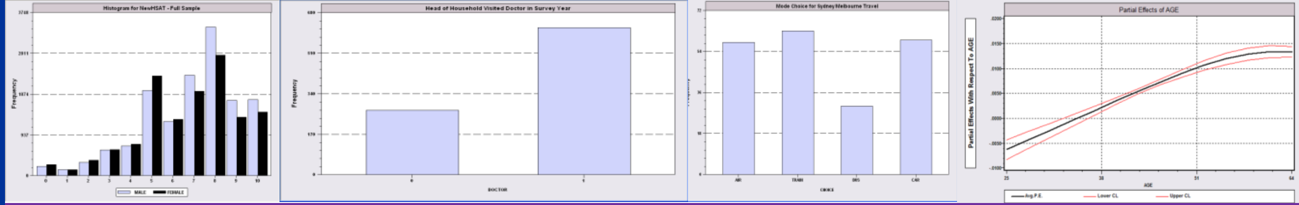
Individual Explicitly Ignores Attributes

Hensher, D.A., Rose, J. and Greene, W. (2005) The Implications on Willingness to Pay of Respondents Ignoring Specific Attributes (DoD#6) Transportation, 32 (3), 203-222.

Hensher, D.A. and Rose, J.M. (2009) Simplifying Choice through Attribute Preservation or Non-Attendance: Implications for Willingness to Pay, *Transportation Research Part E*, 45, 583-590.

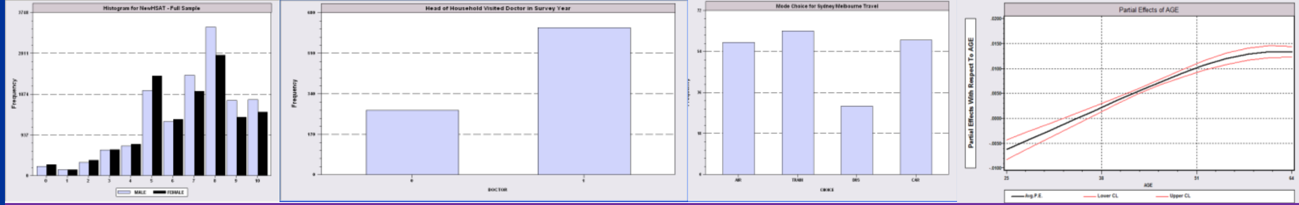
Rose, J., Hensher, D., Greene, W. and Washington, S. Attribute Exclusion Strategies in Airline Choice: Accounting for Exogenous Information on Decision Maker Processing Strategies in Models of Discrete Choice, *Transportmetrica, 2011*

Choice situations in which the individual explicitly states that they ignored certain attributes in their decisions.



Appropriate Modeling Strategy

- Fix ignored attributes at zero? Definitely not!
  - Zero is an unrealistic value of the attribute (price)
  - The probability is a function of $\mathbf{x}_{ij} - \mathbf{x}_{il}$, so the substitution distorts the probabilities
- Appropriate model: for that individual, the specific coefficient is zero – consistent with the utility assumption. A person specific, exogenously determined model
- Surprisingly simple to implement

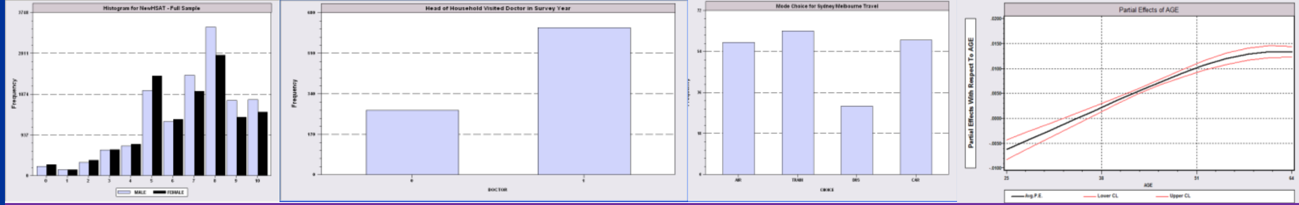


Choice Strategy Heterogeneity

- Methodologically, a rather minor point – construct appropriate likelihood given known information

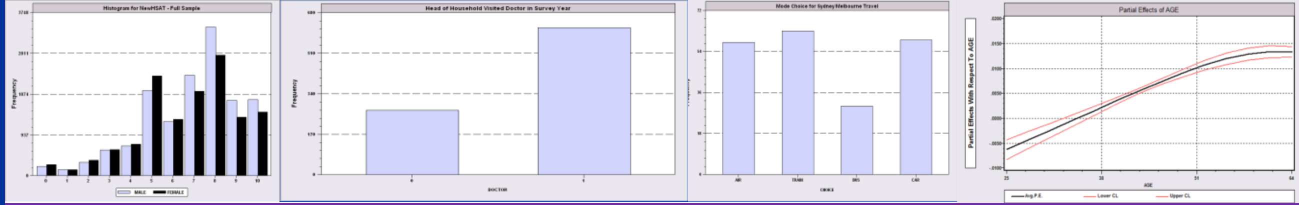
$$\log L = \sum_{m=1}^{M} \sum_{i \in M} \log L_i(\boldsymbol{\theta} \mid \text{data}, m)$$

- Not a latent class model. Classes are not latent.
- Not the ‘variable selection’ issue (the worst form of “stepwise” modeling)
- Familiar strategy gives the wrong answer.



Application: Sydney Commuters' Route Choice

- Stated Preference study – several possible choice situations considered by each person
- Multinomial and mixed (random parameters) logit
- Consumers included data on which attributes were ignored.
- Ignored attributes visibly coded as ignored are automatically treated by constraining **β=0** for that observation.



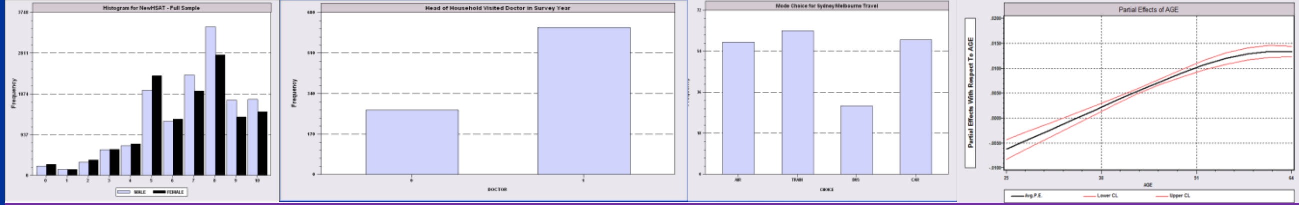
Data for Application of Information Strategy

Stated/Revealed preference study, Sydney car commuters. 500+ surveyed, about 10 choice situations for each.

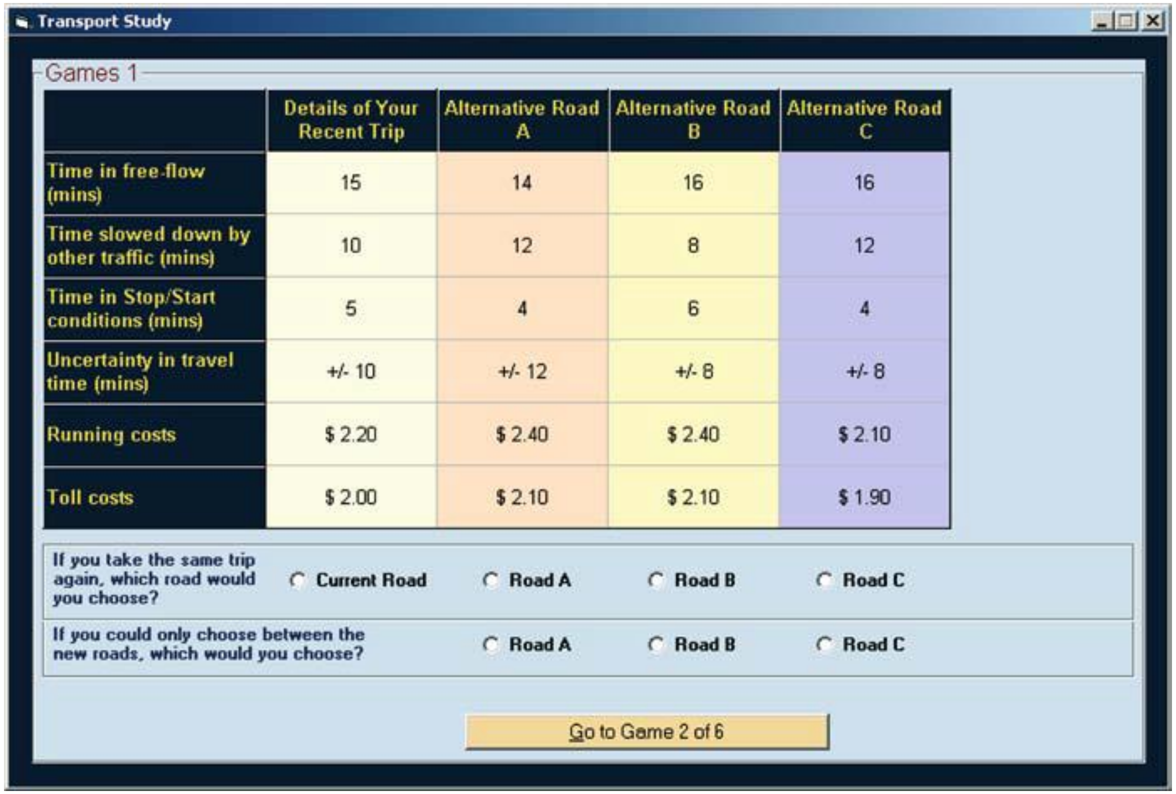
Existing route vs. 3 proposed alternatives.

Attribute design

- **Original: respondents presented with 3, 4, 5, or 6 attributes**
- **Attributes – four level design.**
  - **Free flow time**
  - **Slowed down time**
  - **Stop/start time**
  - **Trip time variability**
  - **Toll cost**
  - **Running cost**
- **Final: respondents use only some attributes and indicate when surveyed which ones they ignored**



Stated Choice Experiment



Transport Study

Games 1

	Details of Your Recent Trip	Alternative Road A	Alternative Road B	Alternative Road C
Time in free-flow (mins)	15	14	16	16
Time slowed down by other traffic (mins)	10	12	8	12
Time in Stop/Start conditions (mins)	5	4	6	4
Uncertainty in travel time (mins)	+/- 10	+/- 12	+/- 8	+/- 8
Running costs	$ 2.20	$ 2.40	$ 2.40	$ 2.10
Toll costs	$ 2.00	$ 2.10	$ 2.10	$ 1.90

If you take the same trip again, which road would you choose? ○ Current Road ○ Road A ○ Road B ○ Road C

If you could only choose between the new roads, which would you choose? ○ Road A ○ Road B ○ Road C

Go to Game 2 of 6

Ancillary questions: Did you ignore any of these attributes?

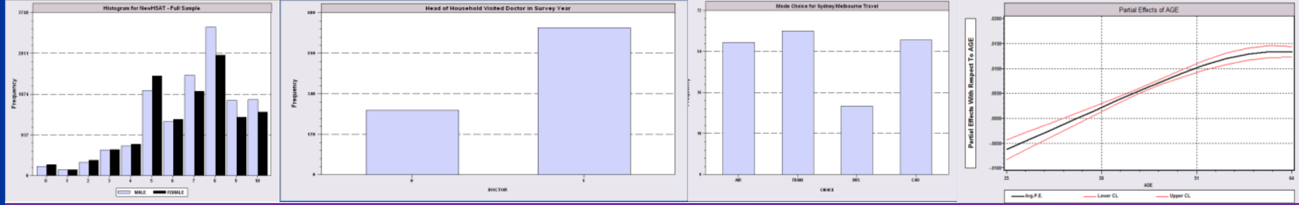
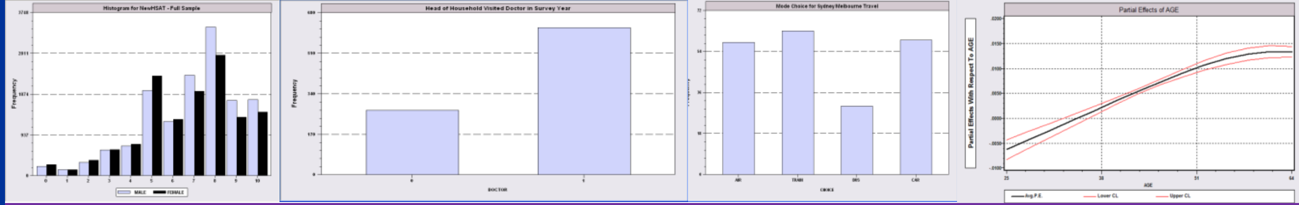


Table 2 Profile of Mixtures of Attributes Not Attended to by a Respondent

	Percentage Not Attended To
Individuals facing six attributes before exclusion (Free flow time, slowed down time, stop start time, uncertainty with run cost, toll cost):	
Uncertainty of time (in presence/absence of other attributes)	37.4
Slowed down time, stop start time	12.32
Slowed down time	9.08
Free flow time, slowed down time, stop start time	6.38
Slowed down time, toll cost	2.52
Slowed down time, stop start time, run cost	0.98
Free flow time, slowed down time, stop start time, run cost	0.49
Free flow time, slowed down time, stop start time, toll cost	0.49
Slowed down time, stop start time, toll cost	0.49
Free flow time, slowed down time, stop start time, run cost, toll cost	0.49
Individuals facing five attributes before exclusion (Free flow time, slowed down time, stop start time, uncertainty with total cost):	
Uncertainty of time (in presence/absence of other attributes)	31.5
Slowed down time, stop start time	7.93
Slowed down time	7.12
Free flow time, slowed down time, stop start time	4.77
Individuals facing four attributes before exclusion (Free flow time, slowed down time plus stop start time, total cost, uncertainty):	
Uncertainty of time (in presence/absence of other attributes)	27.6
Free flow time, slowed down time plus stop start time	8.65
Slowed down time plus stop start time	8.48
Free flow time	3.78
Free flow time, slowed down time plus stop start time, total cost	1.62
Free flow time, total cost	1.45
Slowed down time plus stop start time, total cost	1.45

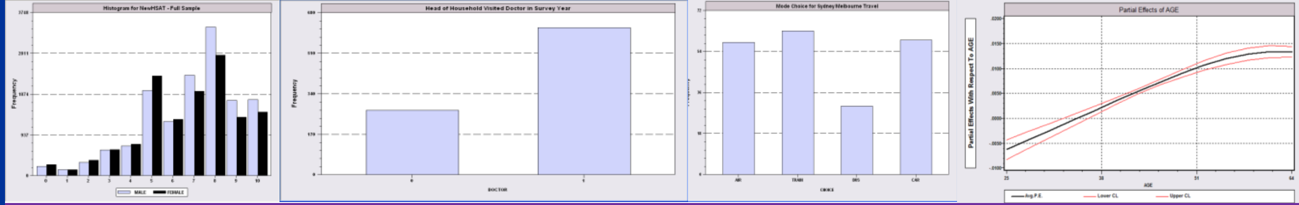


Individual Implicitly Ignores Attributes

Hensher, D.A. and Greene, W.H. (2010) Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification, *Empirical Economics* 39 (2), 413-426

Campbell, D., Hensher, D.A. and Scarpa, R. Non-attendance to Attributes in Environmental Choice Analysis: A Latent Class Specification, *Journal of Environmental Planning and Management*, proofs 14 May 2011.

Hensher, D.A., Rose, J.M. and Greene, W.H. Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design, 14 February 2011, *Transportation*, online 2 June 2001 DOI 10.1007/s11116-011-9347-8.



Stated Choice Experiment

Transport Study

Games 1

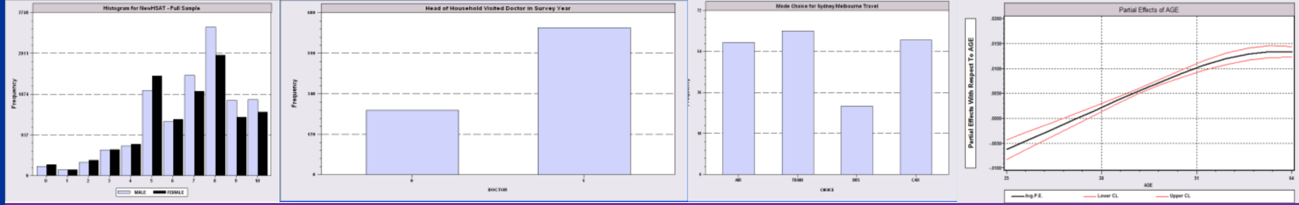
	Details of Your Recent Trip	Alternative Road A	Alternative Road B	Alternative Road C
Time in free-flow (mins)	15	14	16	16
Time slowed down by other traffic (mins)	10	12	8	12
Time in Stop/Start conditions (mins)	5	4	6	4
Uncertainty in travel time (mins)	+/- 10	+/- 12	+/- 8	+/- 8
Running costs	$ 2.20	$ 2.40	$ 2.40	$ 2.10
Toll costs	$ 2.00	$ 2.10	$ 2.10	$ 1.90

If you take the same trip again, which road would you choose? ○ Current Road ○ Road A ○ Road B ○ Road C

If you could only choose between the new roads, which would you choose? ○ Road A ○ Road B ○ Road C

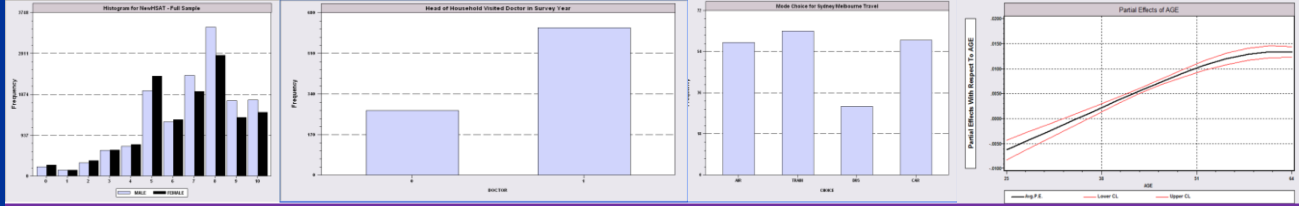
Go to Game 2 of 6

Individuals seem to be ignoring attributes. Unknown to the analyst



The 2^K model

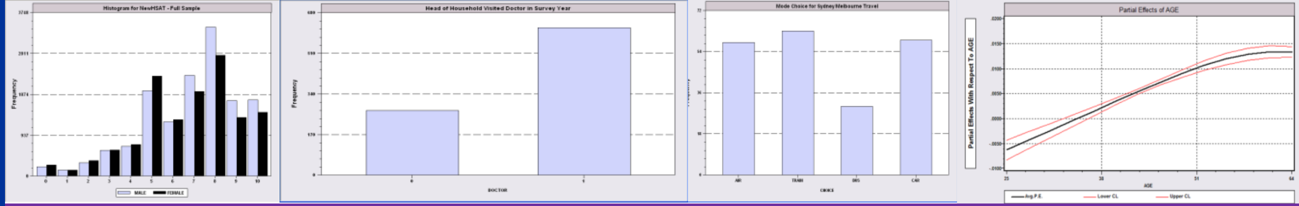
- The analyst believes some attributes are ignored. There is no indicator.
- Classes distinguished by which attributes are ignored
- Same model applies, now a latent class. For K attributes there are 2^K candidate coefficient vectors



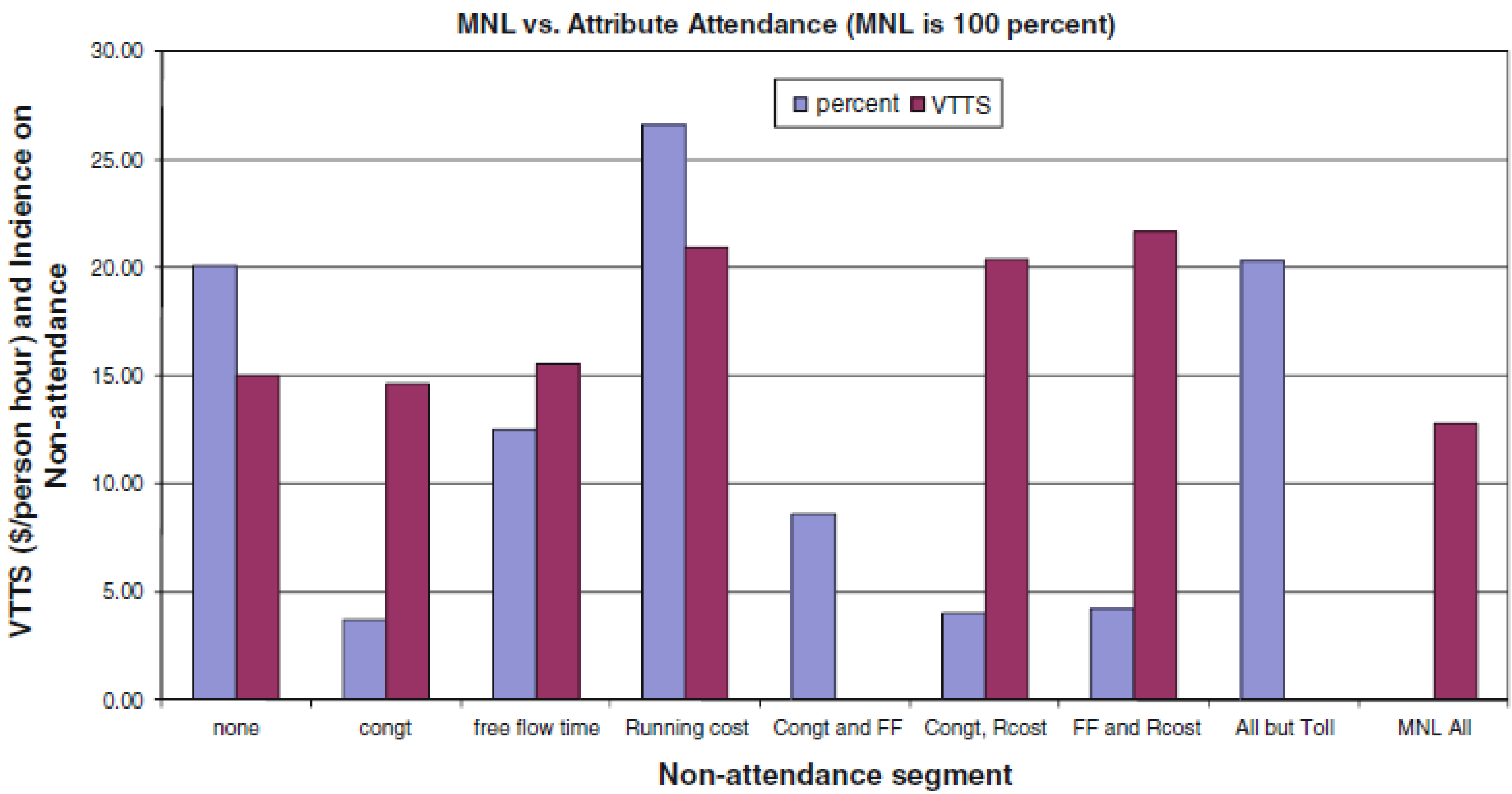
Latent Class Models with Cross Class Restrictions

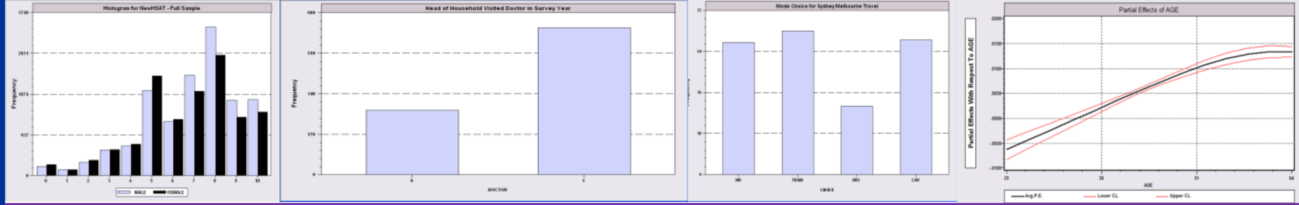
$$
\boldsymbol{\theta}=\left\{\begin{bmatrix}\text{Uncertainty} & \text{Toll Cost} & \text{Running Cost}\\ \beta_1 & \beta_2 & \beta_3\end{bmatrix}\begin{bmatrix}\text{Free Flow} & \text{Slowed} & \text{Start / Stop}\\ 0 & 0 & 0\\ \beta_4 & 0 & 0\\ 0 & \beta_5 & 0\\ 0 & 0 & \beta_6\\ \beta_4 & \beta_5 & 0\\ \beta_4 & 0 & \beta_6\\ 0 & \beta_5 & \beta_6\\ \beta_4 & \beta_5 & \beta_6\end{bmatrix}\right\}\begin{bmatrix}\text{Prior Probs}\\ \pi_1\\ \pi_2\\ \pi_3\\ \pi_4\\ \pi_5\\ \pi_6\\ \pi_7\\ 1-\Sigma_{j=1}^{7}\pi_j\end{bmatrix}
$$

- 8 Class Model: 6 structural utility parameters, 7 unrestricted prior probabilities.
- Reduced form has 8(6)+8 = 56 parameters. ($\pi_j = \exp(\alpha_j)/\Sigma_j\exp(\alpha_j)$, α_J = 0.)
- EM Algorithm: Does not provide any means to impose cross class restrictions.
- "Bayesian" MCMC Methods: May be possible to force the restrictions – it will not be simple.
- Conventional Maximization: Simple



Results for the 2^K model





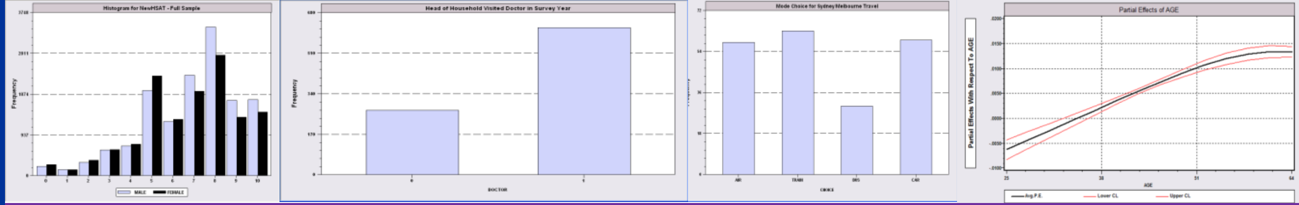
HEALTH ECONOMICS
Health Econ. **22**: 554–567 (2013)
Published online 20 April 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hec.2824

INVESTIGATING ATTRIBUTE NON-ATTENDANCE AND ITS CONSEQUENCES IN CHOICE EXPERIMENTS WITH LATENT CLASS MODELS

MYLENE LAGARDE*

Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London, UK

Based on the vector of coefficient estimates β_i representing taste intensities, the probability that respondents would prefer a new set of guidelines to manage malaria in pregnancy over the current ones can be simulated by computing the probability associated with the utility derived from the new guidelines.

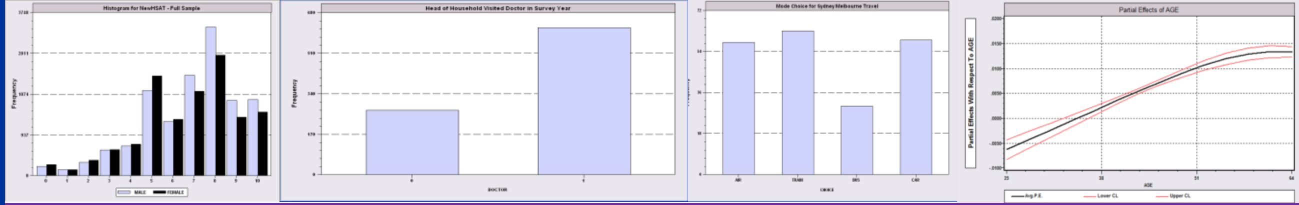


Choice Model with 6 Attributes

$$U_{Ji} = \text{WEIGH}_J * \beta_{\text{weigh}} + \text{ANEM}_J * \beta_{\text{ane}} + \text{DRUG}_J * \beta_{\text{drug}} + \text{BONUS}_J * \beta_{\text{bon}} + \text{WORK}_J * \beta_{\text{work}} + \text{TREAT}_J * \beta_{\text{treat}} + \varepsilon_i$$

Table I. Attributes and levels in the choice experiment

Attribute	Levels
The type of approach to managing malaria in pregnancy	▪ Preventive approach ▪ Curative approach (test and treat if parasite positive)
The anti-malarial drugs you have to prescribe to pregnant women	▪ SP (Fansidar) ▪ Artesunate-amodiaquine (AS–AQ)
Prevalence of anaemia for mothers treated with protocol	▪ 1% ▪ 15%
Prevalence of low birth weight among infants of mothers treated with the protocol	▪ 10% ▪ 15%
Staffing level for the ANC clinic	▪ Under-staffed ▪ Adequately staffed
The salary supplement included in the protocol	▪ GH. C10 ▪ GH. C20



Stated Choice Experiment

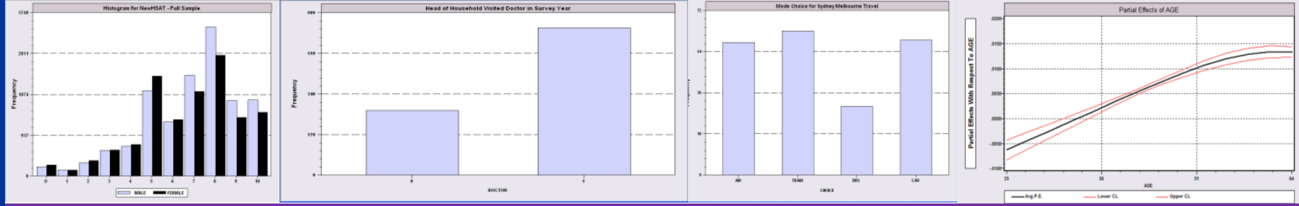
3. DATA

The data used in this study come from a DCE designed to elicit preferences regarding the introduction of new guidelines to managing malaria in pregnancy in Ghana (Lagarde *et al*., 2011). The choice experiment was designed after a series of focus group discussions and in-depth interviews with healthcare providers and a pilot study.

$$U_{Ji} = WEIGH_J * \beta_{weigh} + ANEM_J * \beta_{ane} + DRUG_J * \beta_{drug} + BONUS_J * \beta_{bon} + WORK_J * \beta_{work} + TREAT_J * \beta_{treat} + \varepsilon_i$$

Based on the vector of coefficient estimates β_i representing taste intensities, the probability that respondents would prefer a new set of guidelines to manage malaria in pregnancy over the current ones can be simulated by computing the probability associated with the utility derived from the new guidelines.

Having randomly selected 68 facilities in the Ashanti region in Ghana, all the staff in the ANC clinic present on the day of the data collection was interviewed. For more details about the study design, refer to Smith Paintain *et al*. (2011). Because each respondent answered a series of 16 choice sets, a total of 2,128 observations were used for model estimations. Ethical approval for this study was granted by the



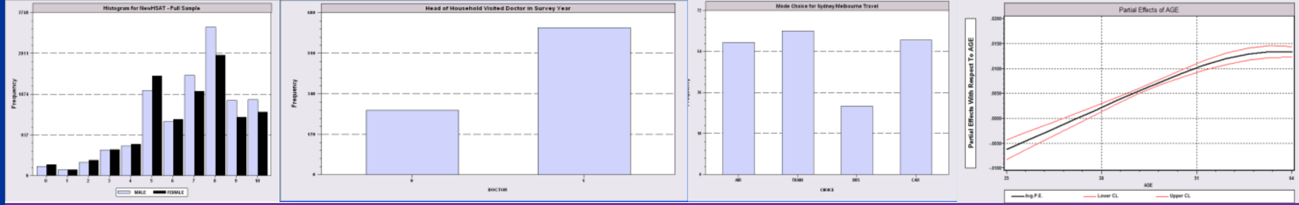
Latent Class Model – Prior Class Probabilities

vation is in which class, the model assumes that individuals belong to a certain group up to a probability. The logit choice probability function of choosing one particular alternative from J alternatives for an individual i belonging to a specific class q can be then written as

$$Pr\ (\mathrm{y}_{\mathrm{it}} = 1\ |class\ q) = \ \mathrm{P}_{it|}\mathrm{q} = \frac{e^{X'_{itj}\beta_q}}{\sum_{\mathrm{j}=1}^{J} e^{X'_{itj}\beta_q}} \tag{1}$$

The probability that an individual i belongs to class q (out of a total of Q classes) is given by

$$H_{iq} = \frac{e^{\theta_q}}{\sum_{\mathrm{q}=1}^{Q} e^{\theta_q}}, \mathrm{q} = 1, \ldots,\ \mathrm{Q}\ \text{and}\ \theta_{\mathrm{Q}} = 0 \tag{2}$$

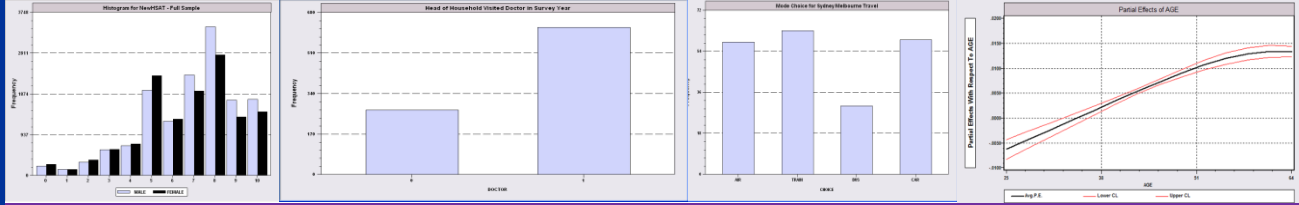


Latent Class Model – Posterior Class Probabilities

Having retrieved the parameter estimates, Bayes' formula can be applied to calculate the posterior estimates of the individual-specific class probabilities ($\hat{H}_{q|i}$) conditional on the observed sequence of T choices (Greene and Hensher, 2003):

$$\hat{H}_{q|}i = \frac{\hat{\mathrm{P}}_{i|}\mathrm{q}\hat{H}_{iq}}{\sum_{\mathbf{q}=1}^{Q}\hat{P}_{i|\mathrm{q}}\hat{H}_{iq}} \tag{4}$$

Although each class q can be defined by a vector β_q—see (1)—the analyst can decide to impose particular constraints on these parameters. In the present case, the objective is to test whether respondents have chosen to ignore certain attributes, which is equivalent to setting the coefficient(s) associated with (a) particular attribute(s) to zero (Hess and Rose, 2007; Hensher and Greene, 2010).

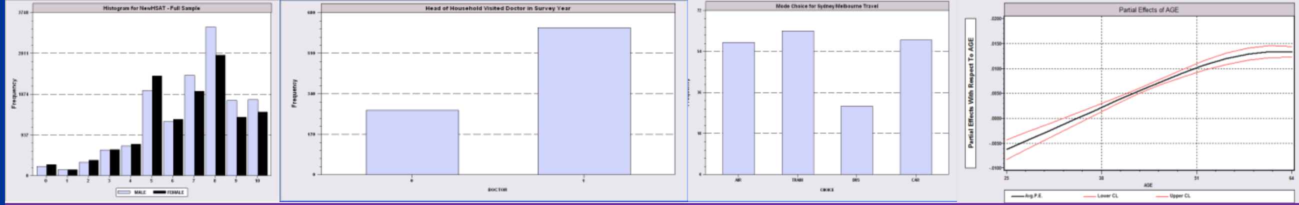


6 attributes implies 64 classes. Strategy to reduce the computational burden on a small sample

2.2. Analytical strategy

The first step of the analysis maps out the extent to which attribute non-attendance is an issue in the dataset. This is carried out by estimating six consecutive two-class LCMs where respondents are either assumed to have considered all attributes (class 1) or to have ignored one attribute (class 2). Following Scarpa *et al.* (2009) and Hensher and Greene (2010), the estimated parameters across the two classes are constrained to be equal to each other. This equality-constrained specification allows the estimation of a model where preferences across individuals can only differ in the information processing rule they use. For these six models, Equation (4) is used to retrieve the distribution of posterior individual probabilities that respondents belong to class 2, that is, ignore one attribute.

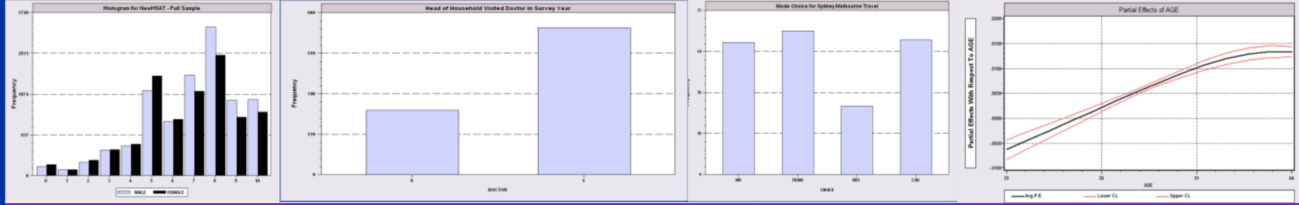
Then, a series of LCMs is estimated that aim to capture all ANA strategies that could have been adopted by respondents. In a DCE with k attributes, there are 2^k possible permutations of ANA strategies. Here, with six attributes, 64 classes would need to be estimated. Because class membership is defined at the individual level, one might end up with too few individuals in each class with a sample size of 132 individuals. To overcome this problem but still try to identify all ANA patterns used by respondents, a stepwise approach is proposed. Having identified all 64 possible response patterns in the present experiment, the first specification includes eight classes: one class that allows respondents to have not ignored any attribute (class 1), and seven others where only one attribute at a time can be ignored (classes 2–7). As in the previous step, all parameters are constrained to be equal to each other across all classes, forcing the analysis to focus only on ANA patterns. Based on the results



Posterior probabilities of membership in the nonattendance class for 6 models

Table II. Average proportion of respondents who ignored one attribute

Model	Attribute assumed to have been ignored in the second class	Average class membership (%)
Model 1	Low birth weight ignored	7.99
Model 2	Anaemia ignored	17.24
Model 3	Bonus ignored	21.75
Model 4	Workload ignored	46.92
Model 5	Drug ignored	62.57
Model 6	Treatment ignored	66.76



The EM Algorithm

Latent Class is a '**missing data**' model

$d_{i,q} = 1$ if individual i is a member of class q

If $d_{i,q}$ were observed, the complete data log likelihood would be

$$\log L_c = \sum_{i=1}^{N} \log\left\{\sum_{q=1}^{Q} d_{i,q}\left[\prod_{t=1}^{T_i} f(y_{i,t} \mid \text{data}_{i,t}, \text{class} = q)\right]\right\}$$

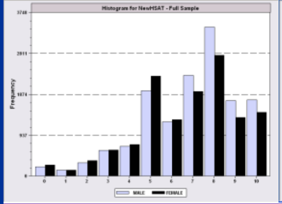
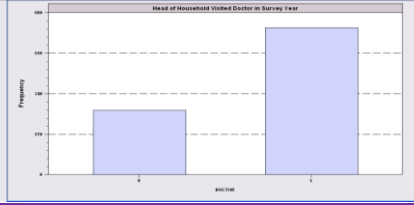
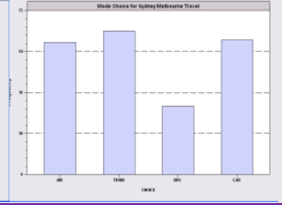
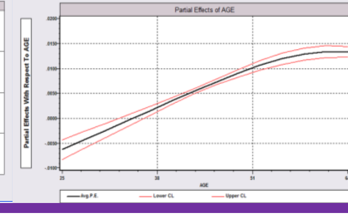
(Only one of the Q terms would be nonzero.)

Expectation - Maximization algorithm has two steps

(1) Expectation Step: Form the 'Expected log likelihood' given the data and a prior guess of the parameters.

(2) Maximize the expected log likelihood to obtain a new guess for the model parameters.

(E.g., http://crow.ee.washington.edu/people/bulyko/papers/em.pdf)

Implementing EM for LC Models

Given initial guesses $\pi_q^0 = \pi_1^0, \pi_2^0, ..., \pi_Q^0$, $\boldsymbol{\beta}_q^0 = \boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, ..., \boldsymbol{\beta}_Q^0$

E.g., use 1/Q for each π_j and the MLE of $\boldsymbol{\beta}$ from a one class model. (Must perturb each one slightly, as if all π_q are equal and all $\boldsymbol{\beta}_q$ are the same, the model will satisfy the FOC.)

(1) Compute $\hat{F}(q|i)$ = posterior class probabilities, using $\hat{\boldsymbol{\beta}}^0, \hat{\boldsymbol{\delta}}^0$

Reestimate each $\boldsymbol{\beta}_q$ using a weighted log likelihood

Maximize wrt $\boldsymbol{\beta}_q$ $\sum_{i=1}^{N} \hat{F}_{iq} \sum_{t=1}^{T_i} \log f(y_{it} \,|\, \mathbf{x}_{it}, \boldsymbol{\beta}_q)$

(2) Reestimate π_q by reestimating $\boldsymbol{\delta}$

$\hat{\pi}_q = (1/N)\Sigma_{i=1}^{N} \hat{F}(q|i)$ using old $\hat{\pi}$ and new $\boldsymbol{\beta}$

Now, return to step 1.

Iterate until convergence.