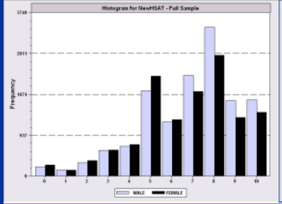
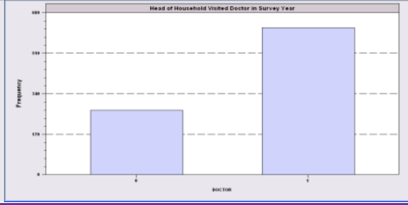
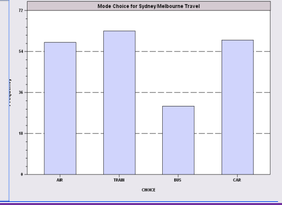
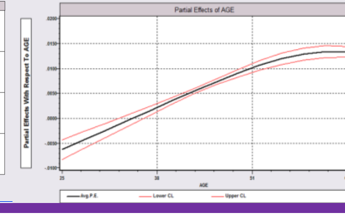
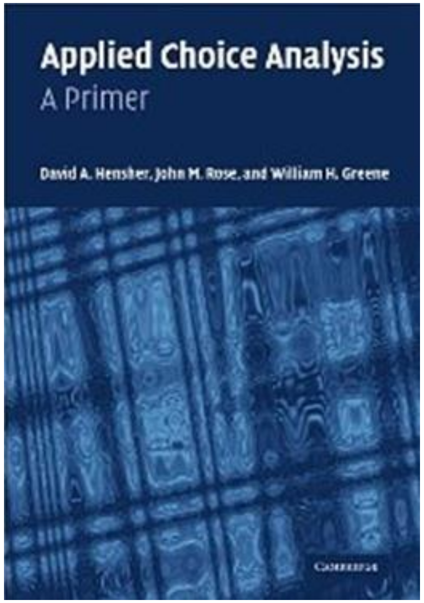
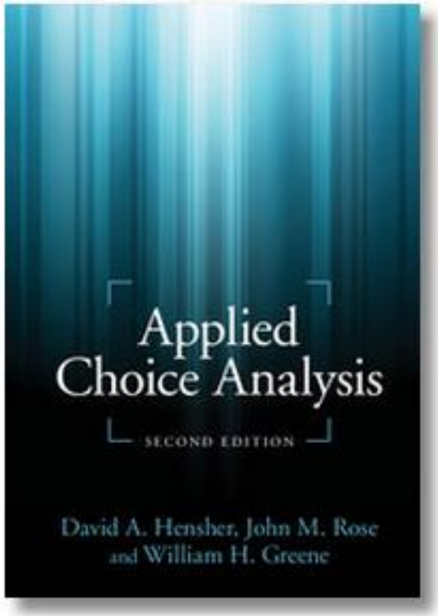
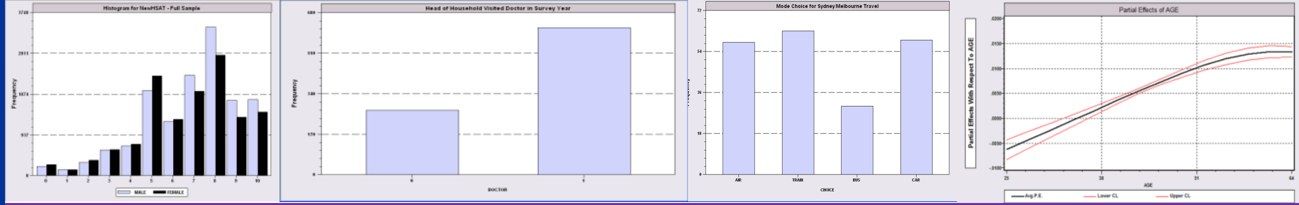





Discrete Choice Modeling

William Greene
Stern School of Business
New York University





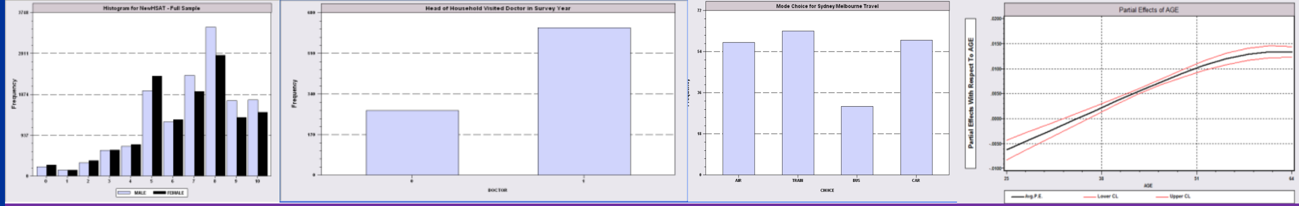


A Microeconomics Platform

- **Consumers Maximize Utility** (!!!)
- Fundamental Choice Problem: Maximize $U(x_1,x_2,\ldots)$ subject to prices and budget constraints
- A Crucial Result for the Classical Problem:
  - **Indirect Utility Function: V = V(p,I)**
  - Demand System of Continuous Choices

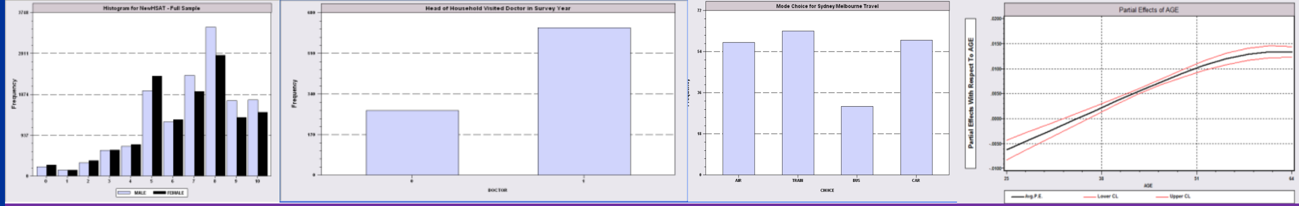
$$x_j^* = -\frac{\partial V(\mathbf{p},I)/\partial p_j}{\partial V(\mathbf{p},I)/\partial I}$$

  - Observed data usually consist of choices, prices, income
- **The Integrability Problem**: Utility is not revealed by demands



Implications for Discrete Choice Models

- Theory is silent about discrete choices
- Translation of utilities to discrete choice requires:
  - Well defined utility indexes: Completeness of rankings
  - Rationality: Utility maximization
  - Axioms of revealed preferences
- Consumers often act to simplify choice situations
- This allows us to build “models.”
  - What common elements can be assumed?
  - How can we account for heterogeneity?
- However, revealed choices do not reveal utility, only rankings which are scale invariant.



Multinomial Choice Among J Alternatives

- **Random Utility Basis**

$$U_{itj} = \alpha_{ij} + \beta_i'\mathbf{x}_{itj} + \gamma_{ij}\mathbf{z}_{it} + \varepsilon_{ijt}$$

$$i = 1,\ldots,N;\ \ j = 1,\ldots,J(i,t);\ \ t = 1,\ldots,T(i)$$

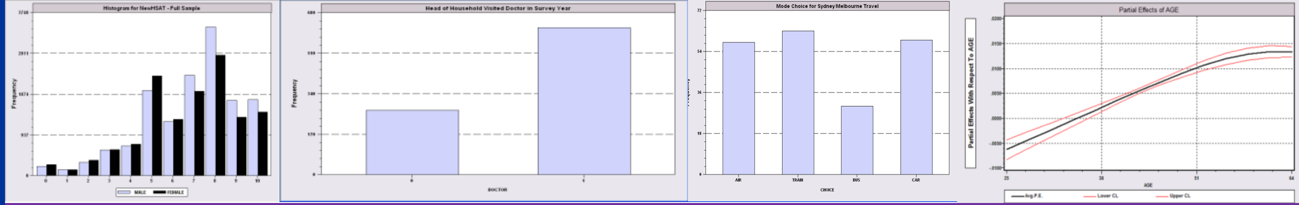
N individuals studied, J(i,t) alternatives in the choice set, T(i) [usually 1] choice situations examined.

- **Maximum Utility Assumption**

Individual i will Choose alternative j in choice setting t if and only if

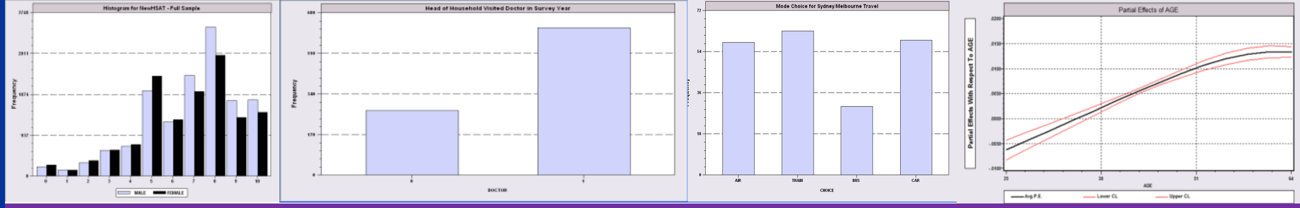
$$U_{itj} \geq U_{itk} \text{ for all } k \neq j.$$

- **Underlying assumptions**
  - Smoothness of utilities
  - Axioms of utility maximization: Transitive, Complete, Monotonic

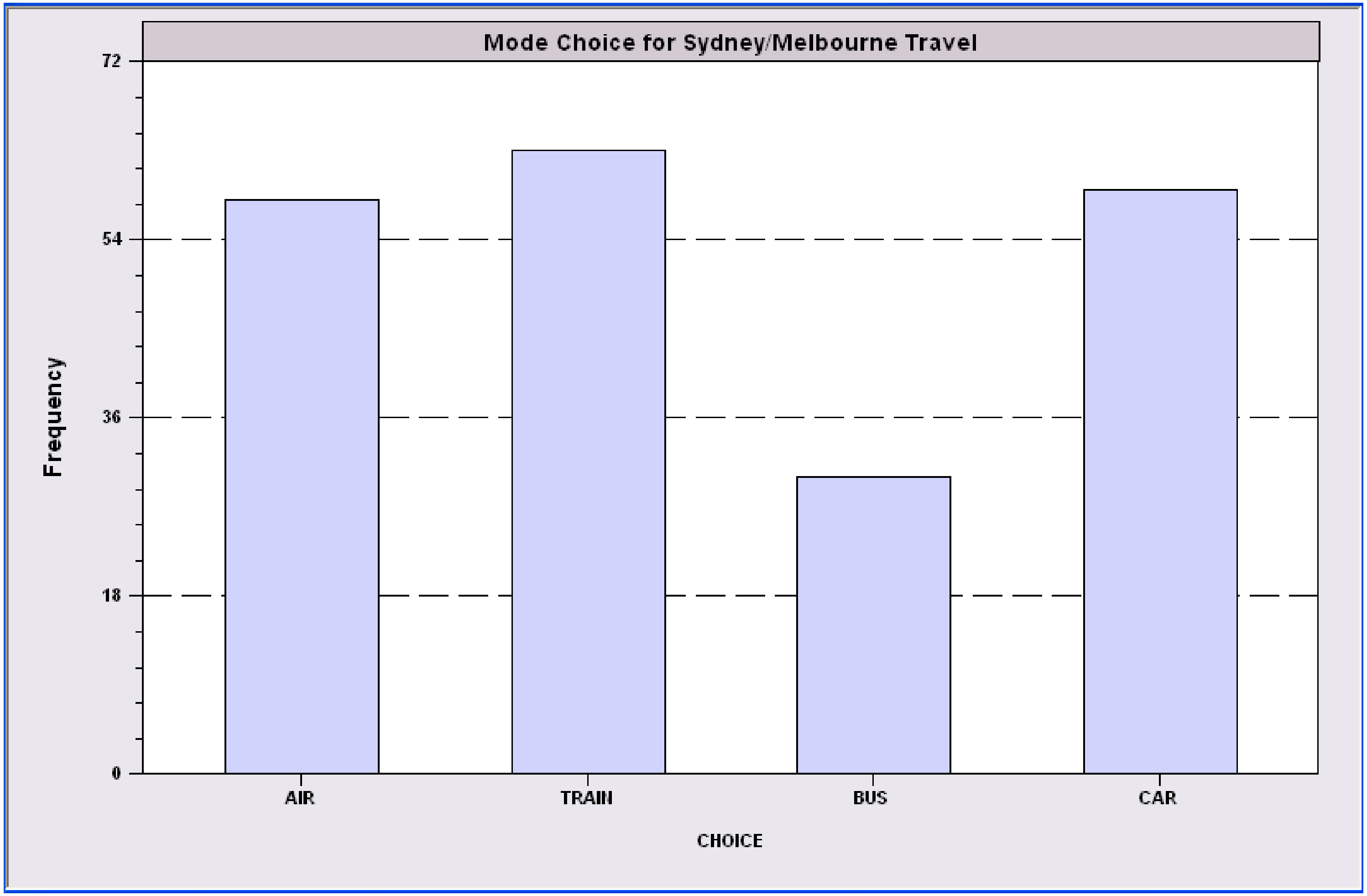


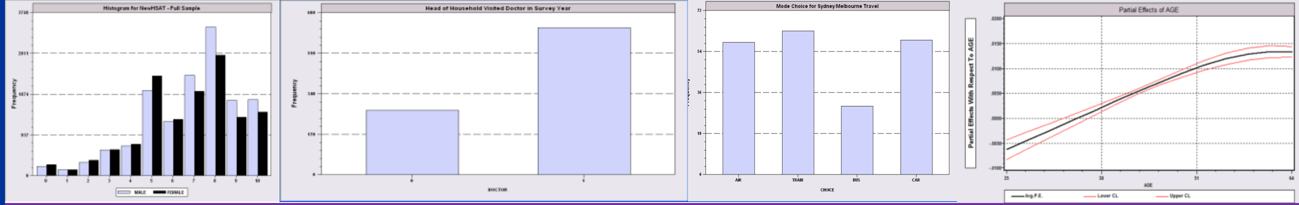
Features of Utility Functions

- The linearity assumption $U_{itj} = \alpha_{ij} + \boldsymbol{\beta}_i'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it} + \varepsilon_{ijt}$
  To be relaxed later: $U_{itj} = V(\mathbf{x}_{itj},\mathbf{z}_{it},\boldsymbol{\beta}_i) + \varepsilon_{ijt}$
- The choice set:
  - Individual (i) and situation (t) specific
  - Unordered alternatives j = 1,…,J(i,t)
- Deterministic ($\mathbf{x},\mathbf{z},\boldsymbol{\gamma}_j$) and random components ($\alpha_{ij},\boldsymbol{\beta}_i,\varepsilon_{ijt}$)
- Attributes of choices, $\mathbf{x}_{itj}$ and characteristics of the chooser, $\mathbf{z}_{it}$.
  - Alternative specific constants α_{ij} may vary by individual
  - Preference weights, $\boldsymbol{\beta}_i$ may vary by individual
  - Individual components, $\boldsymbol{\gamma}_j$ typically vary by choice, not by person
  - Scaling parameters, $\sigma_{ij} = \text{Var}[\varepsilon_{ijt}]$, subject to much modeling



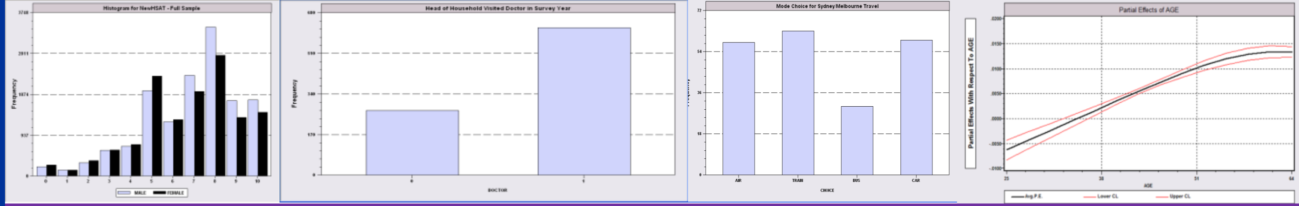
Unordered Choices of 210 Travelers





Data on Multinomial Discrete Choices

CHOICE		ATTRIBUTES				CHARACTERISTIC
MODE	TRAVEL	INVC	INVT	TTME	GC	HINC
AIR	.00000	59.000	100.00	69.000	70.000	35.000
TRAIN	.00000	31.000	372.00	34.000	71.000	35.000
BUS	.00000	25.000	417.00	35.000	70.000	35.000
CAR	1.0000	10.000	180.00	.00000	30.000	35.000
AIR	.00000	58.000	68.000	64.000	68.000	30.000
TRAIN	.00000	31.000	354.00	44.000	84.000	30.000
BUS	.00000	25.000	399.00	53.000	85.000	30.000
CAR	1.0000	11.000	255.00	.00000	50.000	30.000
AIR	.00000	127.00	193.00	69.000	148.00	60.000
TRAIN	.00000	109.00	888.00	34.000	205.00	60.000
BUS	1.0000	52.000	1025.0	60.000	163.00	60.000
CAR	.00000	50.000	892.00	.00000	147.00	60.000
AIR	.00000	44.000	100.00	64.000	59.000	70.000
TRAIN	.00000	25.000	351.00	44.000	78.000	70.000
BUS	.00000	20.000	361.00	53.000	75.000	70.000
CAR	1.0000	5.0000	180.00	.00000	32.000	70.000



Each person makes four choices from a choice set that includes either two or four alternatives.

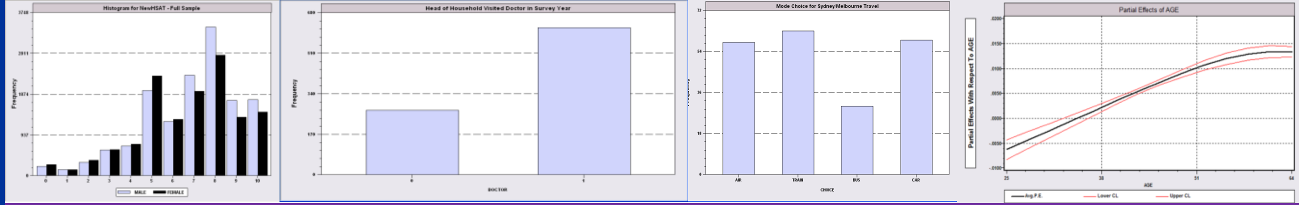
The first choice is the RP between two of the RP alternatives

The second-fourth are the SP among four of the six SP alternatives.

There are ten alternatives in total.

	ID	CITY	SPRP	SPEXP	ALTIJ	CHSNMODE	ALTMODE	SPCHOIC	CHOSEN	CSET
1 »	1000	1	1	0	1	11	2	0	1	2
2 »	1000	1	1	0	4	11	2	0	0	2
3 »	1000	1	2	1	5	0	0	5	1	4
4 »	1000	1	2	1	6	0	0	5	0	4
5 »	1000	1	2	1	8	0	0	5	0	4
6 »	1000	1	2	1	10	0	0	5	0	4
7 »	1000	1	2	2	5	0	0	10	0	4
8 »	1000	1	2	2	6	0	0	10	0	4
9 »	1000	1	2	2	9	0	0	10	0	4
10 »	1000	1	2	2	10	0	0	10	1	4
11 »	1000	1	2	3	5	0	0	8	0	4
12 »	1000	1	2	3	6	0	0	8	0	4
13 »	1000	1	2	3	7	0	0	8	0	4
14 »	1000	1	2	3	8	0	0	8	1	4

A Stated Choice Experiment with Variable Choice Sets



Stated Choice Experiment: Unlabeled Alternatives, One Observation

Brand 1
Brand 2
Brand 3
None

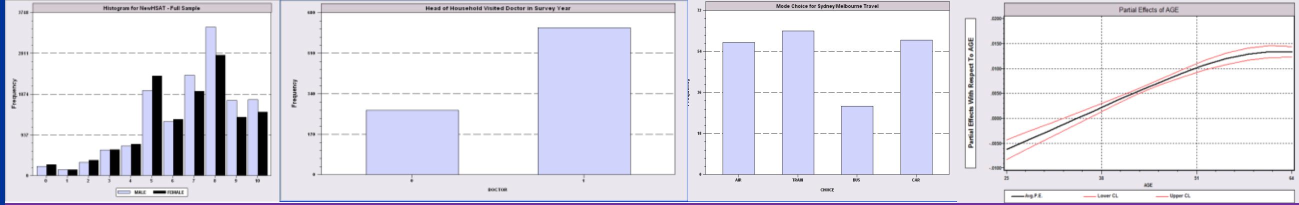
Brand 1
Brand 2
Brand 3
None

Brand 1
Brand 2
Brand 3
None

Brand 1
Brand 2
Brand 3
None

	ID	BRAND	CHOICE	FASH	QUAL	PRICE	PRICESQ	ASC4
1 »	1	1	0	0	0	0.12	0.0144	0
2 »	1	2	1	1	0	0.12	0.0144	0
3 »	1	3	0	0	1	0.08	0.0064	0
4 »	1	4	0	0	0	0	0	1
5 »	1	1	1	1	1	0.12	0.0144	0
6 »	1	2	0	0	1	0.12	0.0144	0
7 »	1	3	0	1	0	0.12	0.0144	0
8 »	1	4	0	0	0	0	0	1
9 »	1	1	0	0	1	0.08	0.0064	0
10 »	1	2	0	1	1	0.2	0.04	0
11 »	1	3	1	1	0	0.08	0.0064	0
12 »	1	4	0	0	0	0	0	1
13 »	1	1	0	0	0	0.08	0.0064	0
14 »	1	2	1	0	1	0.16	0.0256	0
15 »	1	3	0	1	1	0.2	0.04	0
16 »	1	4	0	0	0	0	0	1
17 »	1	1	1	0	0	0.04	0.0016	0
18 »	1	2	0	1	0	0.12	0.0144	0
19 »	1	3	0	1	0	0.08	0.0064	0
20 »	1	4	0	0	0	0	0	1
21 »	1	1	0	0	0	0.08	0.0064	0
22 »	1	2	0	0	1	0.12	0.0144	0
23 »	1	3	1	1	0	0.08	0.0064	0
24 »	1	4	0	0	0	0	0	1
25 »	1	1	0	1	1	0.2	0.04	0
26 »	1	2	1	0	0	0.08	0.0064	0
27 »	1	3	0	0	1	0.08	0.0064	0
28 »	1	4	0	0	0	0	0	1
29 »	1	1	0	0	1	0.08	0.0064	0
30 »	1	2	1	1	0	0.12	0.0144	0
31 »	1	3	0	0	0	0.04	0.0016	0
32 »	1	4	0	0	0	0	0	1

t=1
t=2
t=3
t=4
t=5
t=6
t=7
t=8



Unlabeled Choice Experiments

This an unlabelled choice experiment: Compare

Choice = (Air, Train, Bus, Car)

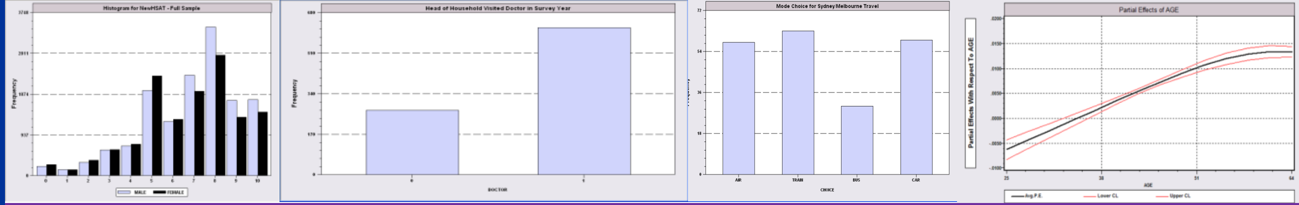
To

Choice = (Brand 1, Brand 2, Brand 3, None)

Brand 1 is only Brand 1 because it is first in the list.

What does it mean to substitute Brand 1 for Brand 2?

What does the own elasticity for Brand 1 mean?

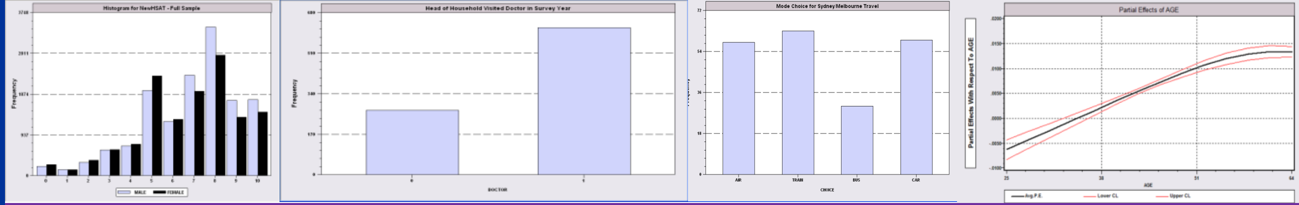


The Multinomial Logit (MNL) Model

- Independent extreme value (Gumbel):
  - $F(\varepsilon_{itj}) = \text{Exp}(-\text{Exp}(-\varepsilon_{itj}))$ (random part of each utility)
  - **Independence** across utility functions
  - **Identical variances** (means absorbed in constants)
  - **Same parameters** for all individuals (temporary)
- Implied probabilities for observed outcomes

$$P[\text{choice} = j \mid \mathbf{x}_{itj}, \mathbf{z}_{it}, i, t] = \text{Prob}[U_{i,t,j} > U_{i,t,k}],\ k = 1,...,J(i,t)$$

$$= \frac{\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it})}{\sum_{j=1}^{J(i,t)} \exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it})}$$



Multinomial Choice Models

Multinomial logit model depends on characteristics

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

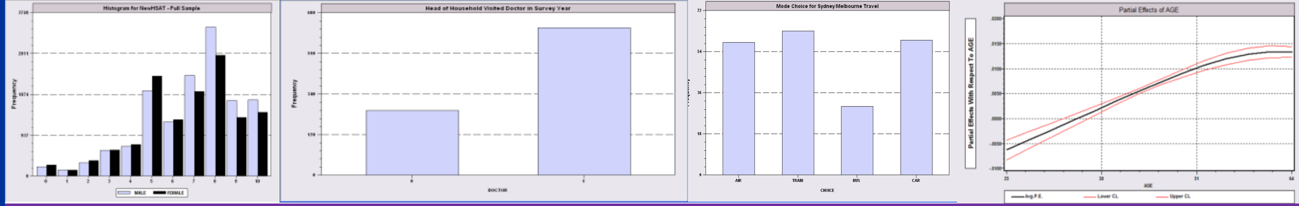
Conditional logit model depends on attributes

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

THE multinomial logit model accommodates both.

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

There is no meaningful distinction.



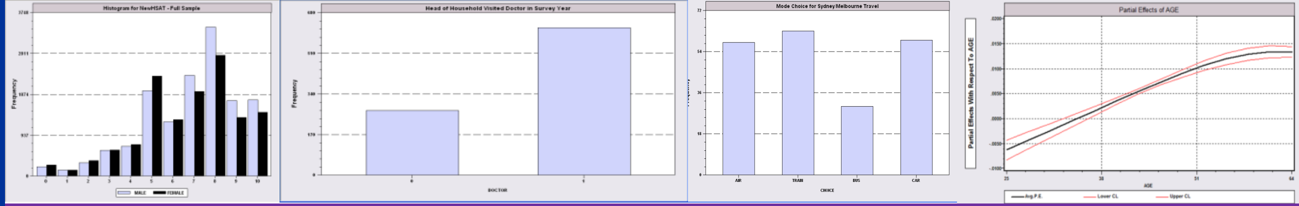
Specifying the Probabilities

- Choice specific attributes (**X**) vary by choices, multiply by generic coefficients. E.g., TTME=terminal time, GC=generalized cost of travel mode
- Generic characteristics (Income, constants) must be interacted with choice specific constants.
- Estimation by maximum likelihood; d_{ij} = 1 if person i chooses j

$$P[\text{choice}=j\,|\,\mathbf{x}_{itj},\mathbf{z}_{it},i,t]=\text{Prob}[U_{i,t,j}>U_{i,t,k}],\ k=1,\dots,J(i,t)$$

$$=\frac{\exp(\alpha_j+\boldsymbol{\beta}'\mathbf{x}_{itj}+\boldsymbol{\gamma}_j'\mathbf{z}_{it})}{\sum_{j=1}^{J(i,t)}\exp(\alpha_j+\boldsymbol{\beta}'\mathbf{x}_{itj}+\boldsymbol{\gamma}_j'\mathbf{z}_{it})}$$

$$\log L=\sum_{i=1}^{N}\sum_{j=1}^{J(i)}d_{ij}\log P_{ij}$$



Using the Model to Measure Consumer Surplus

$$\text{Consumer Surplus} = \frac{\text{Maximum}_j(U_j)}{\text{Marginal Utility of Income}}$$

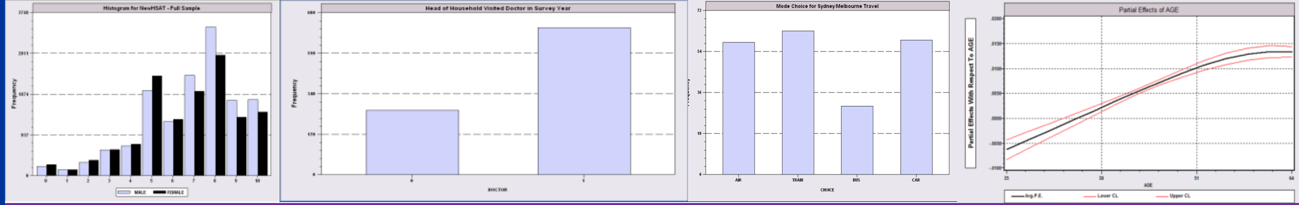
Utility and marginal utility are not observable

For the multinomial logit model (only),

$$E[CS] = \frac{1}{MU_I}\log\left(\sum_{j=1}^{J(i,t)} \exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it})\right) + C$$

Where U_j = the utility of the indicated alternative and C is the constant of integration.

The log sum is the "inclusive value."



Measuring the Change in Consumer Surplus

$$E[CS \mid \text{Scenario A}] = \frac{1}{MU_I}\log\left(\sum_{j=1}^{J(i,t)} \exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it}) \mid A\right) + C$$

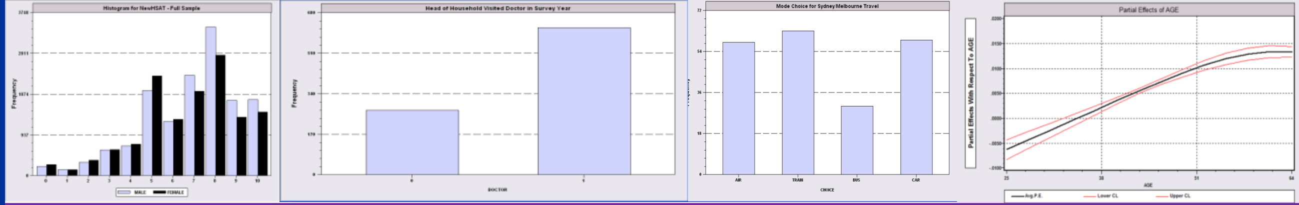
$$E[CS \mid \text{Scenario B}] = \frac{1}{MU_I}\log\left(\sum_{j=1}^{J(i,t)} \exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it}) \mid B\right) + C$$

MU_I and the constant of integration do not change under scenarios.

Change in expected consumer surplus from a policy (scenario) change

$$E[CS \mid \text{Scenario A}] - E[CS \mid \text{Scenario B}]$$

$$= \frac{1}{MU_I}\left[\log \frac{\sum_{j=1}^{J(i,t)} \{\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it}) \mid A\}}{\sum_{j=1}^{J(i,t)} \{\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it}) \mid B\}}\right]$$



Willingness to Pay

Generally a ratio of coefficients

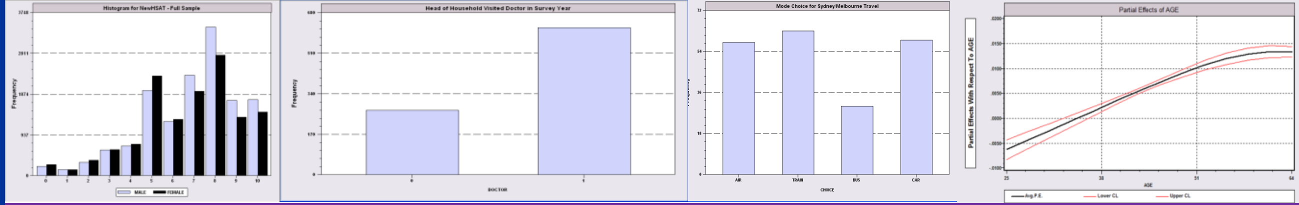
$$\text{WTP} = \frac{\beta(\text{Attribute Level})}{\beta(\text{Income})}$$

Use negative of cost coefficient as a proxu for MU of income

$$\text{WTP} = \frac{\text{negative } \beta(\text{Attribute Level})}{\beta(\text{cost})}$$

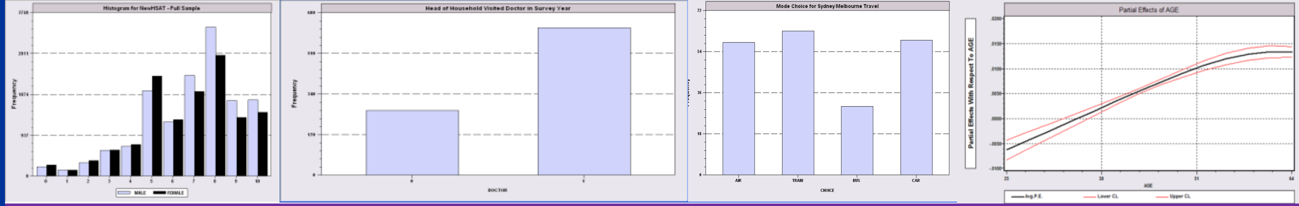
Measurable using model parameters

Ratios of possibly random parameters can produce wild and unreasonable values. We will consider a different approach later.



Observed Data

- Types of Data
  - Individual choice
  - Market shares – consumer markets
  - Frequencies – vote counts
  - Ranks – contests, preference rankings
- Attributes and Characteristics
  - Attributes are features of the choices such as price
  - Characteristics are features of the chooser such as age, gender and income.
- Choice Settings
  - Cross section
  - Repeated measurement (panel data)
    - Stated choice experiments
    - Repeated observations – THE scanner data on consumer choices

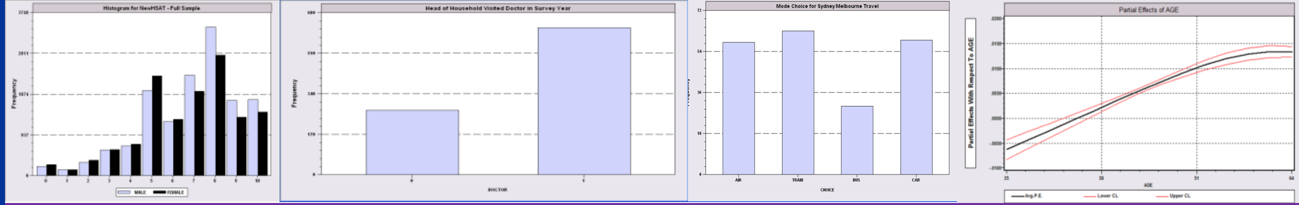


Choice Based Sampling

- Over/Underrepresenting alternatives in the data set

Choice	Air	Train	Bus	Car
True	0.14	0.13	0.09	0.64
Sample	0.28	0.30	0.14	0.28

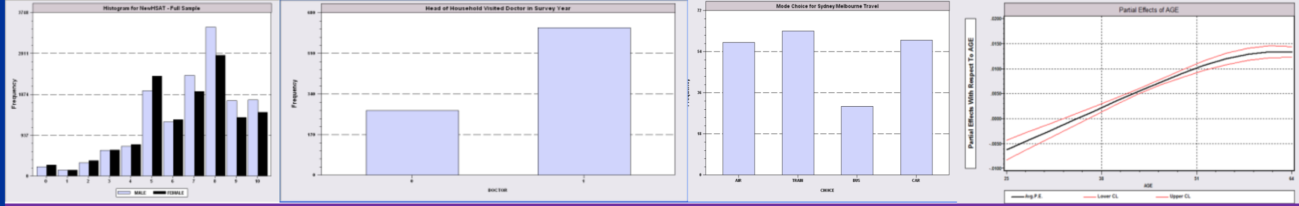
- May cause biases in parameter estimates. (Possibly constants only)
- Certainly causes biases in estimated variances
  - Weighted log likelihood, weight = π_j / F_j for all i.
  - Fixup of covariance matrix – use “sandwich” estimator. Using weighted Hessian and weighted BHHH in the center of the sandwich



Data on Discrete Choices

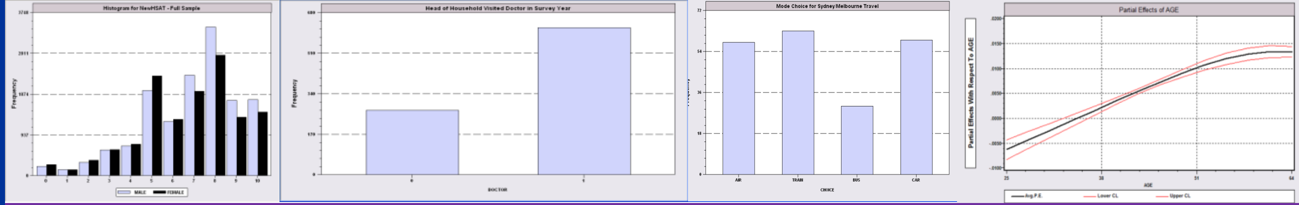
CHOICE		ATTRIBUTES				CHARACTERISTIC
MODE	TRAVEL	INVC	INVT	TTME	GC	HINC
AIR	.00000	59.000	100.00	69.000	70.000	35.000
TRAIN	.00000	31.000	372.00	34.000	71.000	35.000
BUS	.00000	25.000	417.00	35.000	70.000	35.000
CAR	1.0000	10.000	180.00	.00000	30.000	35.000
AIR	.00000	58.000	68.000	64.000	68.000	30.000
TRAIN	.00000	31.000	354.00	44.000	84.000	30.000
BUS	.00000	25.000	399.00	53.000	85.000	30.000
CAR	1.0000	11.000	255.00	.00000	50.000	30.000
AIR	.00000	127.00	193.00	69.000	148.00	60.000
TRAIN	.00000	109.00	888.00	34.000	205.00	60.000
BUS	1.0000	52.000	1025.0	60.000	163.00	60.000
CAR	.00000	50.000	892.00	.00000	147.00	60.000
AIR	.00000	44.000	100.00	64.000	59.000	70.000
TRAIN	.00000	25.000	351.00	44.000	78.000	70.000
BUS	.00000	20.000	361.00	53.000	75.000	70.000
CAR	1.0000	5.0000	180.00	.00000	32.000	70.000

This is the ‘long form.’ In the ‘wide form,’ all data for the individual appear on a single ‘line’. The ‘wide form’ is unmanageable for models of any complexity and for stated preference applications.



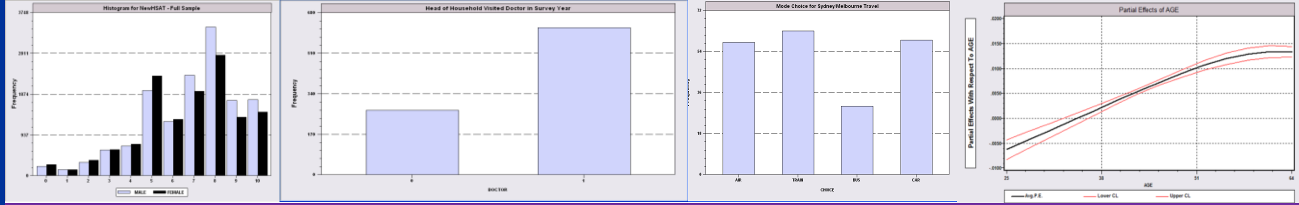
An Estimated MNL Model

```
-----------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                   Choice
Log likelihood function          -199.97662
Estimation based on N =     210, K =     5
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              1.95216      409.95325
Fin.Smpl.AIC     1.95356      410.24736
Bayes IC         2.03185      426.68878
Hannan Quinn     1.98438      416.71880
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .2953 .2896
Chi-squared[ 2]          =    167.56429
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
      GC|    -.01578***         .00438      -3.601   .0003
    TTME|    -.09709***         .01044      -9.304   .0000
   A_AIR|    5.77636***         .65592       8.807   .0000
 A_TRAIN|    3.92300***         .44199       8.876   .0000
   A_BUS|    3.21073***         .44965       7.140   .0000
--------+--------------------------------------------------
```



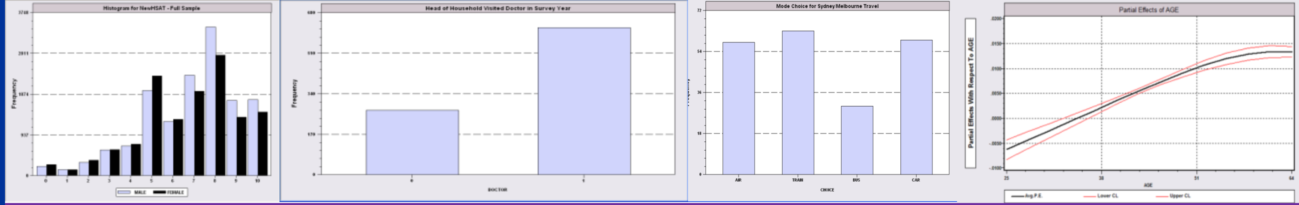
Estimated MNL Model

```
-----------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                Choice
Log likelihood function       -199.97662
Estimation based on N =     210, K =   5
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              1.95216      409.95325
Fin.Smpl.AIC     1.95356      410.24736
Bayes IC         2.03185      426.68878
Hannan Quinn     1.98438      416.71880
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .2953 .2896
Chi-squared[ 2]            =    167.56429
Prob [ chi squared > value ] =    .00000
Response data are given as ind. choices
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
      GC|   -.01578***       .00438        -3.601   .0003
    TTME|   -.09709***       .01044        -9.304   .0000
   A_AIR|   5.77636***       .65592         8.807   .0000
 A_TRAIN|   3.92300***       .44199         8.876   .0000
   A_BUS|   3.21073***       .44965         7.140   .0000
--------+--------------------------------------------------
```



Model Fit Based on Log Likelihood

- Three sets of predicted probabilities
  - No model: $P_{ij} = 1/J$ (.25)
  - Constants only: $P_{ij} = (1/N)\Sigma_i d_{ij}$
    (58,63,30,59)/210=.286,.300,.143,.281
    Constants only model matches sample shares
  - Estimated model: Logit probabilities
- Compute log likelihood
- Measure improvement in log likelihood with Pseudo R-squared = 1 – $\text{LogL}/\text{LogL}_0$ ("Adjusted" for number of parameters in the model.)



Fit the Model with Only ASCs

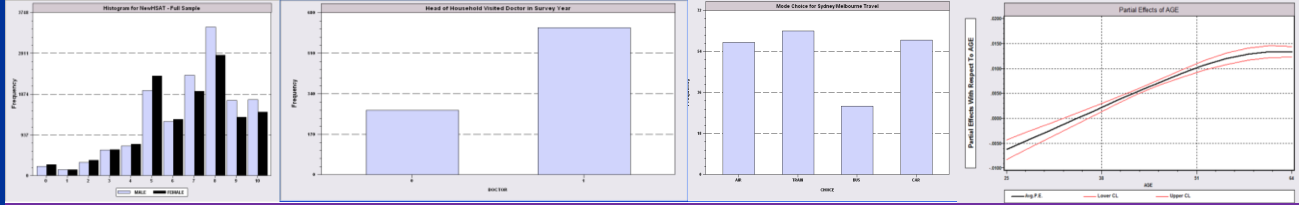
If the choice set varies across observations, this is the only way to obtain the restricted log likelihood.

```
-----------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                Choice
Log likelihood function       -283.75877
Estimation based on N =     210, K =   3
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              2.73104      573.51754
Fin.Smpl.AIC     2.73159      573.63404
Bayes IC         2.77885      583.55886
Hannan Quinn     2.75037      577.57687
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .0000-.0048
Response data are given as ind. choices
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
   A_AIR|    -.01709          .18491        -.092    .9263
 A_TRAIN|     .06560          .18117         .362    .7173
   A_BUS|    -.67634***       .22424       -3.016    .0026
--------+--------------------------------------------------
```

If the choice set is fixed at J, then

$$\text{logL} = \sum_{j=1}^{J} N_l \log\left(\frac{N_j}{N}\right)$$

$$= \sum_{j=1}^{J} N_l \log P_j$$

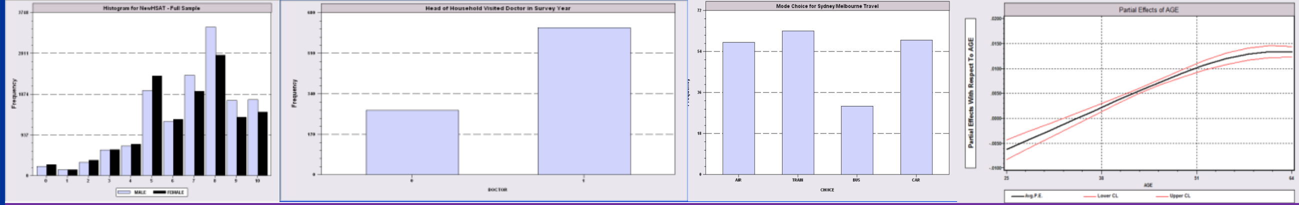


Estimated MNL Model

```
-----------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                    Choice
Log likelihood function           -199.97662
Estimation based on N =     210, K =     5
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              1.95216      409.95325
Fin.Smpl.AIC     1.95356      410.24736
Bayes IC         2.03185      426.68878
Hannan Quinn     1.98438      416.71880
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .2953 .2896
Chi-squared[ 2]          =    167.56429
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=    210, skipped    0 obs
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
      GC|    -.01578***       .00438       -3.601   .0003
    TTME|    -.09709***       .01044       -9.304   .0000
   A_AIR|    5.77636***       .65592        8.807   .0000
 A_TRAIN|    3.92300***       .44199        8.876   .0000
   A_BUS|    3.21073***       .44965        7.140   .0000
--------+--------------------------------------------------
```

$$\text{Pseudo } R^2 = 1 - \frac{\log L}{\log L_0}.$$

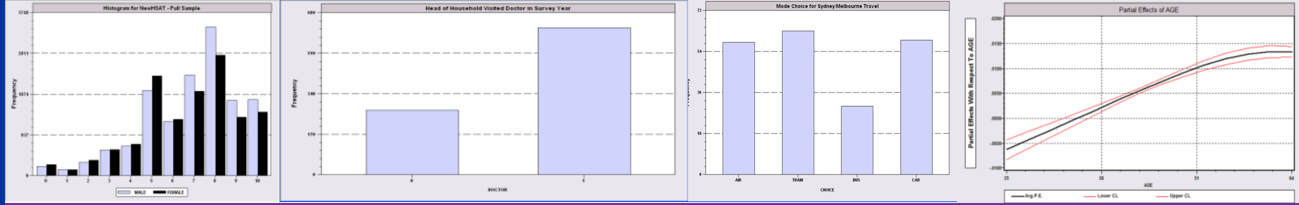
$$\text{Adjusted Pseudo } R^2 = 1 - \left(\frac{N(J-1)}{N(J-1)-K}\right)\left(\frac{\log L}{\log L_0}\right).$$



Model Fit Based on Predictions

- N_j = actual number of choosers of "j."
- $Nfit_j = \Sigma_i$ Predicted Probabilities for "j"
- Cross tabulate:
  Predicted vs. Actual, cell prediction is cell probability
  Predicted vs. Actual, cell prediction is the cell with the largest probability

$$N_{jk} = \Sigma_i d_{ij} \times \text{Predicted } P(i,k)$$



Fit Measures Based on Crosstabulation

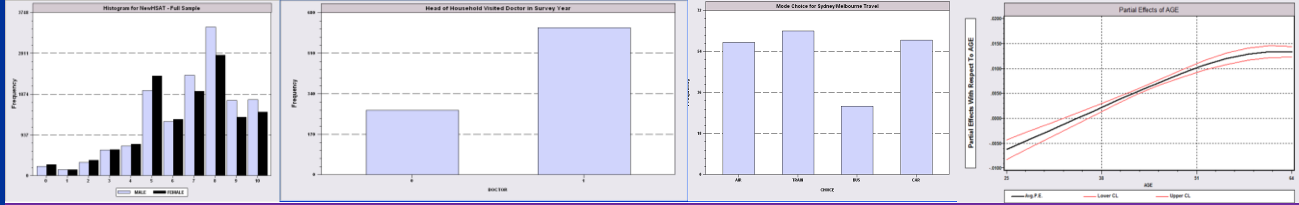
```
+----------------------------------------------------------+
| Cross tabulation of actual choice vs. predicted P(j)     |
| Row indicator is actual, column is predicted.            |
| Predicted total is F(k,j,i)=Sum(i=1,...,N) P(k,j,i).     |
| Column totals may be subject to rounding error.          |
+----------------------------------------------------------+
```

NLOGIT Cross Tabulation for 4 outcome Multinomial Choice Model

	AIR	TRAIN	BUS	CAR	Total
AIR	32	8	5	13	58
TRAIN	8	37	5	14	63
BUS	3	5	15	6	30
CAR	15	13	6	26	59
Total	58	63	30	59	210

NLOGIT Cross Tabulation for 4 outcome Constants Only Choice Model

	AIR	TRAIN	BUS	CAR	Total
AIR	16	17	8	16	58
TRAIN	17	19	9	18	63
BUS	8	9	4	8	30
CAR	16	18	8	17	59
Total	58	63	30	59	210



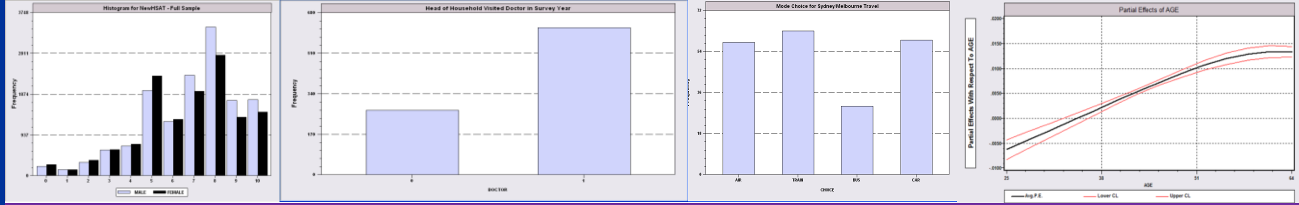
Partial effects:

Change in attribute "k" of alternative "m" on the probability that the individual makes choice "j"

j = Train

$$\frac{\partial \text{Prob}(j)}{\partial x_{m,k}} = \frac{\partial P_j}{\partial x_{m,k}} = P_j[\mathbf{1}(j=m) - P_m]\beta_k$$

m = Car **k = Price**



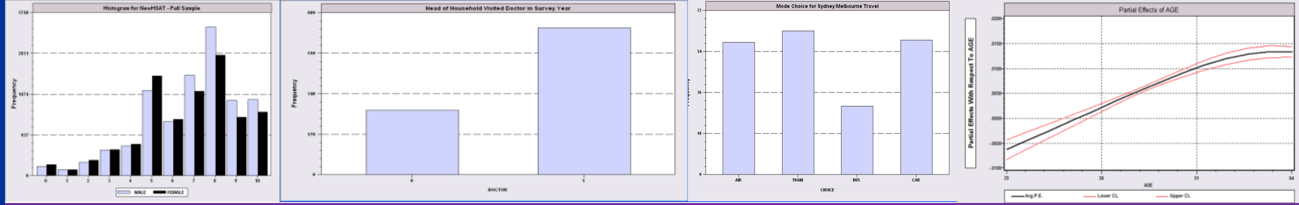
Partial effects : **k = Price**

Own effects : **j = Train**

$$\frac{\partial \text{Prob}(j)}{\partial x_{j,k}} = \frac{\partial P_j}{\partial x_{j,k}} = P_j[1-P_j]\beta_k$$

Cross effects : **j = Train** **m = Car**

$$\frac{\partial \text{Prob}(j)}{\partial x_{m,k}} = \frac{\partial P_j}{\partial x_{m,k}} = -P_j P_m \beta_k$$

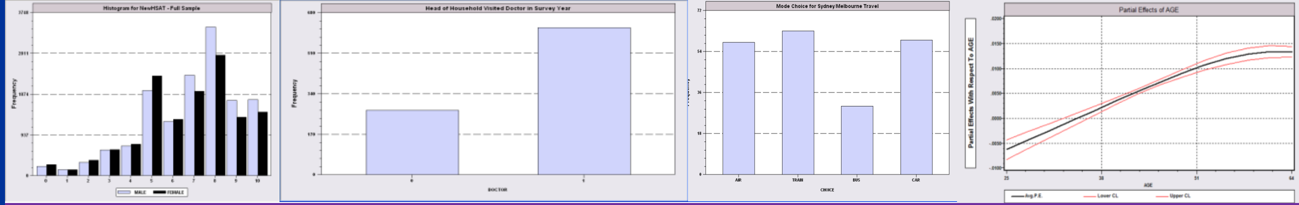


Elasticities for proportional changes:

$$\frac{\partial \log Prob(j)}{\partial \log x_{m,k}} = \frac{\partial \log P_j}{\partial \log x_{m,k}} = \frac{x_{m,k}}{P_j} P_j [\mathbf{1}(j=m) - P_m]\beta_k$$

$$= [\mathbf{1}(j=m) - P_m]\, x_{m,k}\, \beta_k$$

Note the elasticity is the same for all j. This is a consequence of the IIA assumption in the model specification made at the outset.



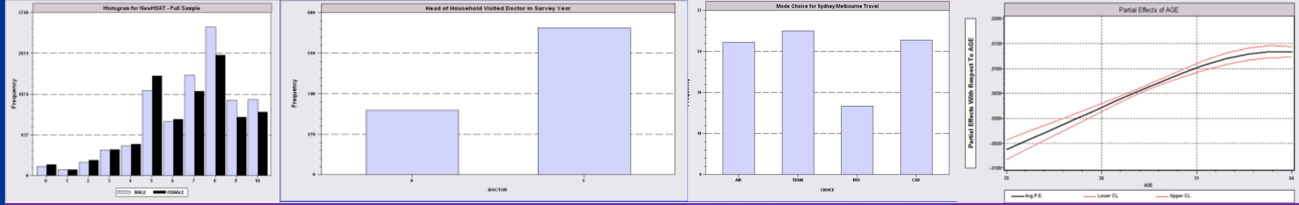

```
+---------------------------------------------------+
| Elasticity              averaged over observations.|
| Attribute is INVT      in choice AIR               |
|                                  Mean     St.Dev   |
| *      Choice=AIR               -.2055     .0666   |
|        Choice=TRAIN              .0903     .0681   |
|        Choice=BUS                .0903     .0681   |
|        Choice=CAR                .0903     .0681   |
+---------------------------------------------------+
| Attribute is INVT      in choice TRAIN             |
|        Choice=AIR                .3568     .1231   |
| *      Choice=TRAIN             -.9892     .5217   |
|        Choice=BUS                .3568     .1231   |
|        Choice=CAR                .3568     .1231   |
+---------------------------------------------------+
| Attribute is INVT      in choice BUS               |
|        Choice=AIR                .1889     .0743   |
|        Choice=TRAIN              .1889     .0743   |
| *      Choice=BUS              -1.2040     .4803   |
|        Choice=CAR                .1889     .0743   |
+---------------------------------------------------+
| Attribute is INVT      in choice CAR               |
|        Choice=AIR                .3174     .1195   |
|        Choice=TRAIN              .3174     .1195   |
|        Choice=BUS                .3174     .1195   |
| *      Choice=CAR               -.9510     .5504   |
+---------------------------------------------------+
| Effects on probabilities of all choices in model: |
| * = Direct Elasticity effect of the attribute.    |
+---------------------------------------------------+
```

Note the effect of IIA on the cross effects.

← Own effect

← Cross effects

Elasticities are computed for each observation; the mean and standard deviation are then computed across the sample observations.



Use Krinsky and Robb to compute standard errors for Elasticities

```
+---------------------------------------------------+
| Elasticity                 averaged over observations.|
| Effects on probabilities of all choices in model: |
| * = Direct Elasticity effect of the attribute.    |
+---------------------------------------------------+
```

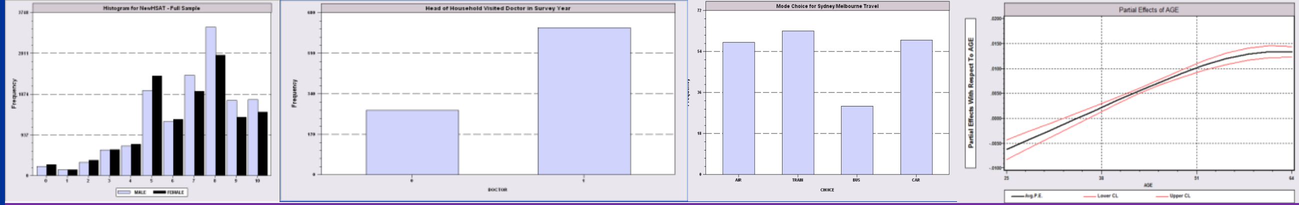
Average elasticity of prob(alt) wrt INVT in AIR

Choice	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
AIR	-1.33631***	.14024	-9.53	.0000	-1.61119	-1.06144
TRAIN	.53493***	.04870	10.98	.0000	.43948	.63039
BUS	.53493***	.04870	10.98	.0000	.43948	.63039
CAR	.53493***	.04870	10.98	.0000	.43948	.63039

***, **, * ==> Significance at 1%, 5%, 10% level.
Model was estimated on Nov 28, 2014 at 06:49:18 PM

Average elasticity of prob(alt) wrt INVT in AIR

Choice	Average Elasticity	Sample Standard Deviation	Sample Minimum	Sample Maximum
AIR	-1.33631	.05020	-4.601438	-.01069
TRAIN	.53493	.04388	.001027	3.54061
BUS	.53493	.04388	.001027	3.54061
CAR	.53493	.04388	.001027	3.54061



Analyzing the Behavior of Market Shares to Examine Discrete Effects

- Scenario: What happens to the number of people who make specific choices if a particular attribute changes in a specified way?

- Fit the model first, then using the identical model setup, add

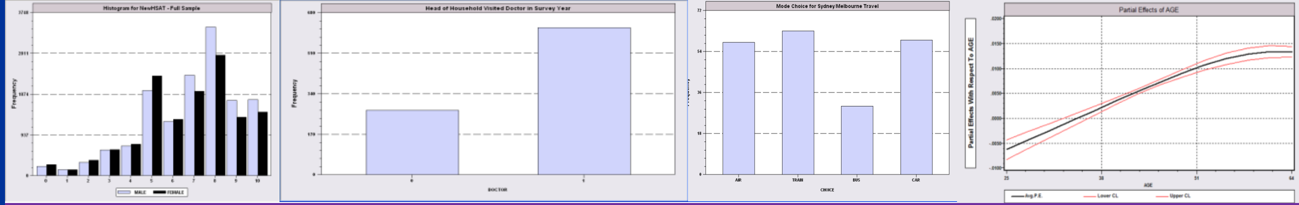
  ; Simulation = list of choices to be analyzed

  ; Scenario = Attribute (in choices) = type of change

- For the CLOGIT application

  ; Simulation = * ? This is ALL choices

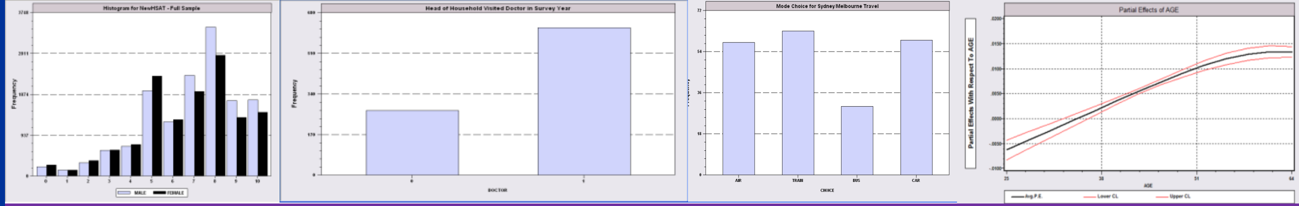
  ; Scenario: GC(car)=[*]1.25$ Car_GC rises by 25%



Model Simulation

```
+------------------------------------------------+
| Discrete Choice (One Level) Model              |
| Model Simulation Using Previous Estimates      |
| Number of observations             210         |
+------------------------------------------------+
+---------------------------------------------------------+
|Simulations of Probability Model                         |
|Model: Discrete Choice (One Level) Model                  |
|Simulated choice set may be a subset of the choices.     |
|Number of individuals is the probability times the       |
|number of observations in the simulated sample.          |
|Column totals may be affected by rounding error.         |
|The model used was simulated with    210 observations.|
+---------------------------------------------------------+
-------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected            Change type           Value
---------  ------------------------------   ------------------- ---------
GC         CAR                              Scale base by value     1.250
-------------------------------------------------------------------------
The simulator located    209 observations for this scenario.
Simulated Probabilities (shares) for this scenario:
+----------+--------------+--------------+------------------+
|Choice    |     Base     |   Scenario   | Scenario - Base  |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+--------------+--------------+------------------+
|AIR       | 27.619    58 | 29.592    62 |  1.973%        4 |
|TRAIN     | 30.000    63 | 31.748    67 |  1.748%        4 |
|BUS       | 14.286    30 | 15.189    32 |   .903%        2 |
|CAR       | 28.095    59 | 23.472    49 | -4.624%      -10 |
|Total     |100.000   210 |100.000   210 |   .000%        0 |
+----------+--------------+--------------+------------------+
```

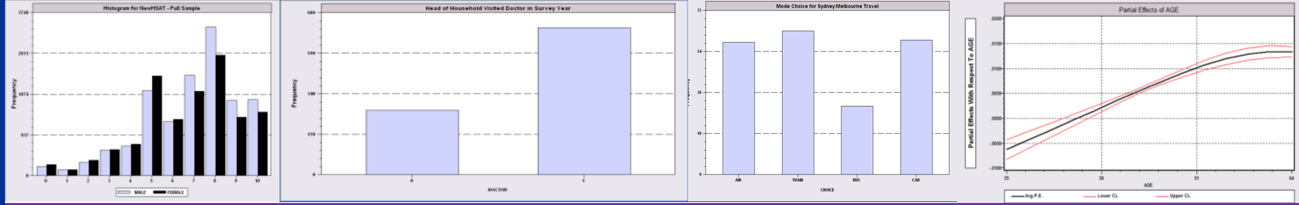
Changes in the predicted market shares when GC_CAR increases by 25%.



More Complicated Model Simulation

In vehicle cost of CAR falls by 10%
Market is limited to ground (Train, Bus, Car)

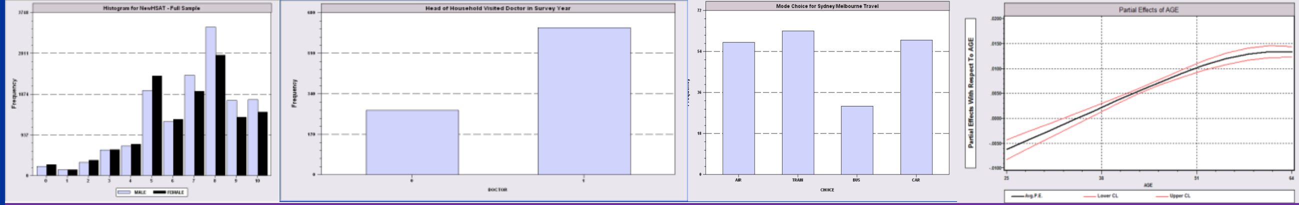
```
CLOGIT ; Lhs = Mode
       ; Choices = Air,Train,Bus,Car
       ; Rhs = TTME,INVC,INVT,GC
       ; Rh2 = One ,Hinc
       ; Simulation = TRAIN,BUS,CAR
       ; Scenario: GC(car)=[*].9$
```



Model Estimation Step

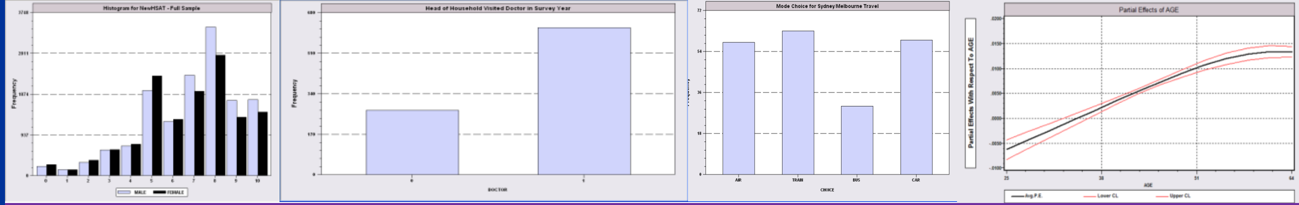
```
-----------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                  Choice
Log likelihood function         -172.94366
Estimation based on N =     210, K =   10
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3905 .3807
Chi-squared[ 7]          =     221.63022
Prob [ chi squared > value ] =    .00000
Response data are given as ind. choices
Number of obs.=    210, skipped    0 obs
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
    TTME|    -.10289***        .01109       -9.280    .0000
    INVC|    -.08044***        .01995       -4.032    .0001
    INVT|    -.01399***        .00267       -5.240    .0000
      GC|     .07578***        .01833        4.134    .0000
   A_AIR|    4.37035***       1.05734        4.133    .0000
AIR_HIN1|     .00428           .01306         .327    .7434
 A_TRAIN|    5.91407***        .68993        8.572    .0000
TRA_HIN2|    -.05907***        .01471       -4.016    .0001
   A_BUS|    4.46269***        .72333        6.170    .0000
BUS_HIN3|    -.02295           .01592       -1.442    .1493
--------+--------------------------------------------------
```

Alternative specific constants and interactions of ASCs and Household Income



Model Simulation Step

```
+---------------------------------------------+
| Discrete Choice (One Level) Model           |
| Model Simulation Using Previous Estimates   |
| Number of observations               210    |
+---------------------------------------------+
+------------------------------------------------------+
|Simulations of Probability Model                      |
|Model: Discrete Choice (One Level) Model              |
|Simulated choice set may be a subset of the choices.  |
|Number of individuals is the probability times the    |
|number of observations in the simulated sample.       |
|The model used was simulated with    210 observations.|
+------------------------------------------------------+
-------------------------------------------------------------------------
Specification of scenario 1 is:
Attribute  Alternatives affected            Change type                Value
---------  -------------------------------  -------------------  ---------
INVC       CAR                              Scale base by value       .900
-------------------------------------------------------------------------
The simulator located    210 observations for this scenario.
Simulated Probabilities (shares) for this scenario:
+----------+--------------+--------------+------------------+
|Choice    |     Base     |   Scenario   | Scenario - Base  |
|          |%Share Number |%Share Number |ChgShare ChgNumber|
+----------+--------------+--------------+------------------+
|TRAIN     | 37.321    78 | 35.854    75 | -1.467%       -3 |
|BUS       | 19.805    42 | 18.641    39 | -1.164%       -3 |
|CAR       | 42.874    90 | 45.506    96 |  2.632%        6 |
|Total     |100.000   210 |100.000   210 |   .000%        0 |
+----------+--------------+--------------+------------------+
```



Willingness to Pay

$U(alt) = a_j + b_{INCOME}$*INCOME + $b_{Attribute}$*Attribute + …

WTP = MU(Attribute)/MU(Income)

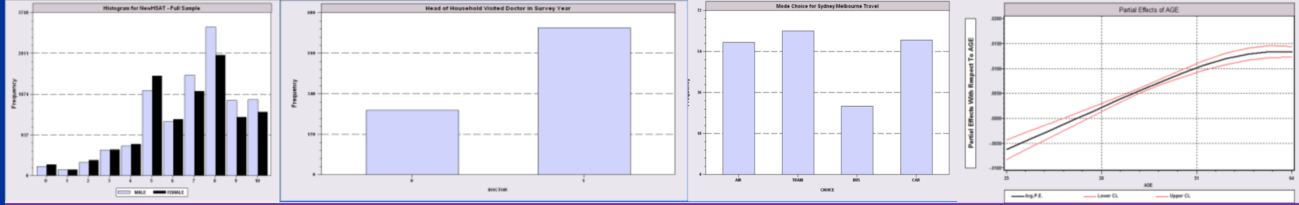
When MU(Income) is not available, an approximation often used is –MU(Cost).

U(Air,Train,Bus,Car)

$= \alpha_{alt} + \beta_{cost}\,\text{Cost} + \beta_{INVT}\,\text{INVT} + \beta_{TTME}\,\text{TTME} + \varepsilon_{alt}$

WTP for less in vehicle time $= -\beta_{INVT} / \beta_{COST}$

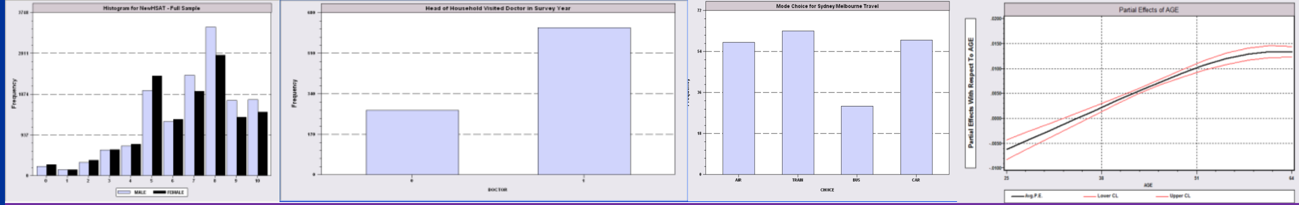
WTP for less terminal time $= -\beta_{TIME} / \beta_{COST}$



WTP from CLOGIT Model

```
-----------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                Choice
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
      GC|    -.00286          .00610         -.469   .6390
    INVT|    -.00349***       .00115        -3.037   .0024
    TTME|    -.09746***       .01035        -9.414   .0000
    AASC|    4.05405***       .83662         4.846   .0000
    TASC|    3.64460***       .44276         8.232   .0000
    BASC|    3.19579***       .45194         7.071   .0000
--------+--------------------------------------------------
WALD ; fn1=WTP_INVT=b_invt/b_gc ; fn2=WTP_TTME=b_ttme/b_gc$
-----------------------------------------------------------
WALD procedure.
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
WTP_INVT|    1.22006         2.88619          .423   .6725
WTP_TTME|    34.0771        73.07097          .466   .6410
--------+--------------------------------------------------
```

Very different estimates suggests this might not be a very good model.



Estimation in WTP Space

Problem with WTP calculation: Ratio of two estimates that are asymptotically normally distributed may have infinite variance.

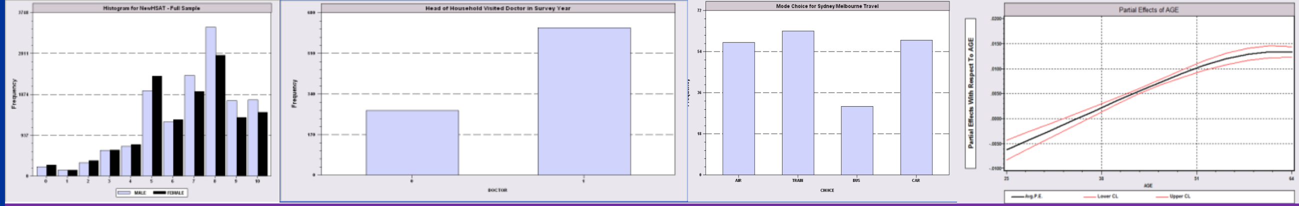
- Sample point estimates may be reasonable
- Inference - confidence intervals - may not be possible.

WTP estimates often become unreasonable in random parameter models in which parameters vary across individuals.

Estimation in WTP Space

$$
\begin{aligned}
U(Air) &= \alpha + \beta_{COST}COST + \beta_{TIME}TIME + \beta_{attr}Attr + \varepsilon \\
&= \alpha + \beta_{COST}\left[COST + \frac{\beta_{TIME}}{\beta_{COST}}TIME + \frac{\beta_{attr}}{\beta_{COST}}Attr\right] + \varepsilon \\
&= \alpha + \beta_{COST}\left[COST + \theta_{TIME}TIME + \theta_{attr}Attr\right] + \varepsilon
\end{aligned}
$$

For a simple MNL the transformation is 1:1. Results will be identical to the original model. In more elaborate, RP models, results change.



The I.I.D Assumption

$$U_{itj} = \alpha_{ij} + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}'\mathbf{z}_{it} + \varepsilon_{ijt}$$

$F(\varepsilon_{itj}) = \text{Exp}(-\text{Exp}(-\varepsilon_{itj}))$ (random part of each utility)

Independence across utility functions

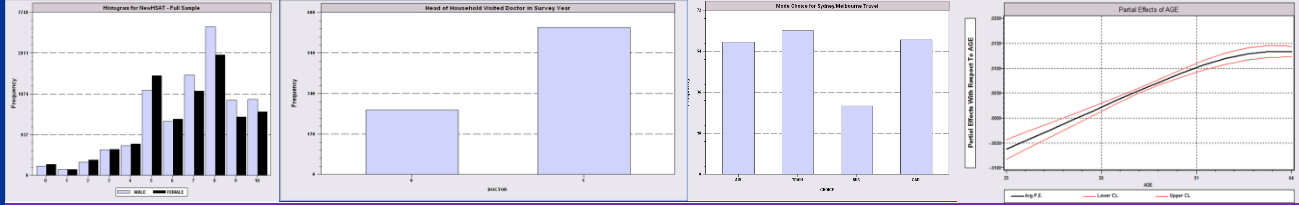
Identical variances (means absorbed in constants)

Restriction on equal scaling may be inappropriate

Correlation across alternatives may be suppressed

Equal cross elasticities is a substantive restriction

Behavioral implication of IID is independence from irrelevant alternatives. If an alternative is removed, probability is spread equally across the remaining alternatives. This is unreasonable

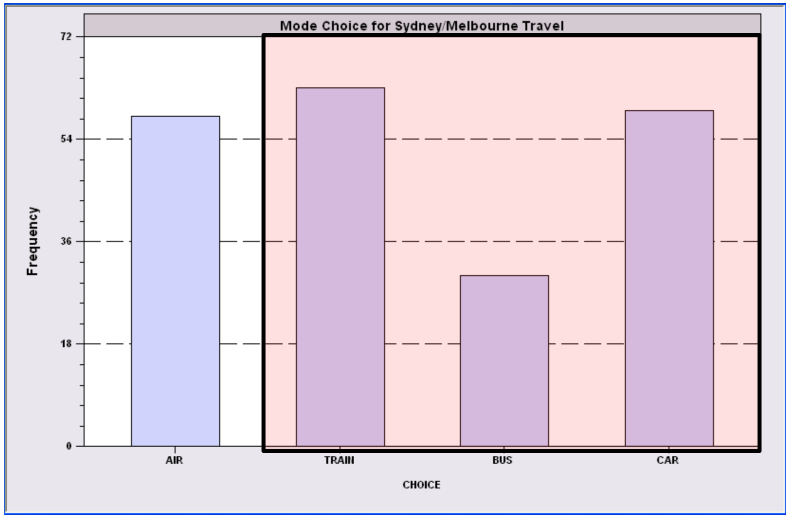


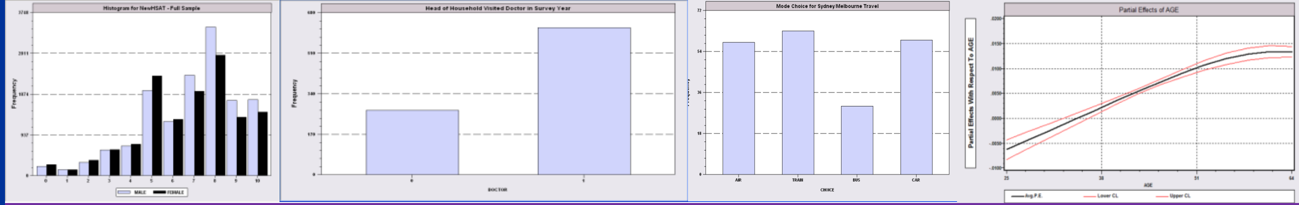
IIA Implication of IID

$$\text{Prob(train)} = \frac{\exp[U(train)]}{\exp[U(air)]+\exp[U(train)]+\exp[U(bus)]+\exp[U(car)]}$$

$$\text{Prob(train|train,bus,car)} = \frac{\exp[U(train)]}{\exp[U(train)]+\exp[U(bus)]+\exp[U(car)]}$$

Air is in the choice set, probabilities are independent from air if air is not in the condition. This is a testable behavioral assumption.





Behavioral Implication of IIA

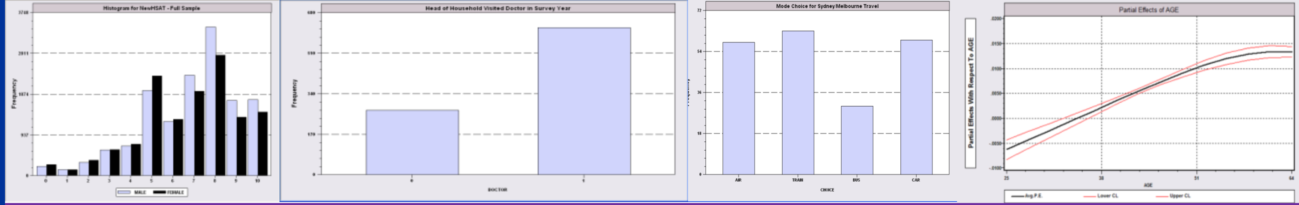
```
+---------------------------------------------------+
| Elasticity               averaged over observations.|
| Attribute is INVT      in choice AIR                |
|                                  Mean      St.Dev   |
| *      Choice=AIR               -.2055      .0666   |
|        Choice=TRAIN              .0903      .0681   |
|        Choice=BUS                .0903      .0681   |
|        Choice=CAR                .0903      .0681   |
+---------------------------------------------------+
| Attribute is INVT      in choice TRAIN              |
|        Choice=AIR                .3568      .1231   |
| *      Choice=TRAIN             -.9892      .5217   |
|        Choice=BUS                .3568      .1231   |
|        Choice=CAR                .3568      .1231   |
+---------------------------------------------------+
| Attribute is INVT      in choice BUS                |
|        Choice=AIR                .1889      .0743   |
|        Choice=TRAIN              .1889      .0743   |
| *      Choice=BUS              -1.2040      .4803   |
|        Choice=CAR                .1889      .0743   |
+---------------------------------------------------+
| Attribute is INVT      in choice CAR                |
|        Choice=AIR                .3174      .1195   |
|        Choice=TRAIN              .3174      .1195   |
|        Choice=BUS                .3174      .1195   |
| *      Choice=CAR               -.9510      .5504   |
+---------------------------------------------------+
| Effects on probabilities of all choices in model: |
| * = Direct Elasticity effect of the attribute.    |
+---------------------------------------------------+
```

Own effect

Cross effects

Note the effect of IIA on the cross effects.

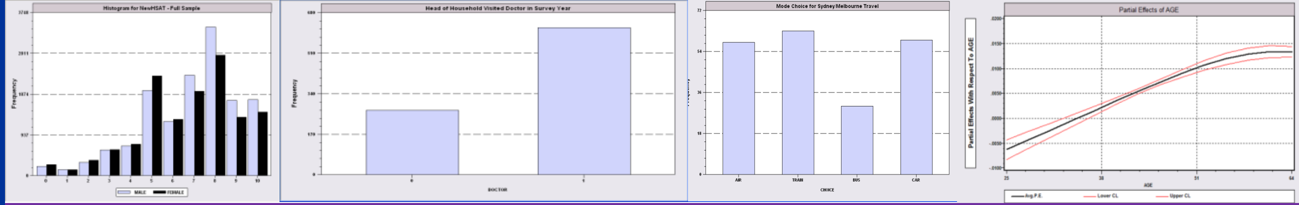
Elasticities are computed for each observation; the mean and standard deviation are then computed across the sample observations.



A Hausman and McFadden Test for IIA

- Estimate full model with “irrelevant alternatives”
- Estimate the short model eliminating the irrelevant alternatives
  - Eliminate individuals who chose the irrelevant alternatives
  - Drop attributes that are constant in the surviving choice set.
- Do the coefficients change? Under the IIA assumption, they should not.
  - Use a Hausman test:
  - Chi-squared, d.f. Number of parameters estimated

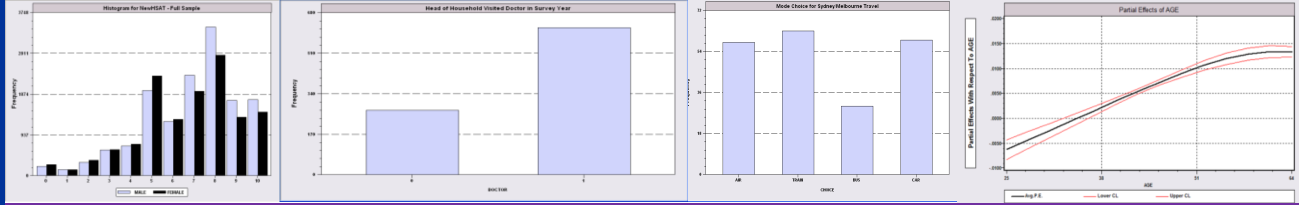
$$H = \left(\mathbf{b}_{\text{short}} - \mathbf{b}_{\text{full}}\right)'\left[\mathbf{V}_{\text{short}} - \mathbf{V}_{\text{full}}\right]^{-1}\left(\mathbf{b}_{\text{short}} - \mathbf{b}_{\text{full}}\right)$$



IIA Test for Choice AIR

```
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
 GC      |    .06929537       .01743306     3.975   .0001
 TTME    |   -.10364955       .01093815    -9.476   .0000
 INVC    |   -.08493182       .01938251    -4.382   .0000
 INVT    |   -.01333220       .00251698    -5.297   .0000
 AASC    |   5.20474275       .90521312     5.750   .0000
 TASC    |   4.36060457       .51066543     8.539   .0000
 BASC    |   3.76323447       .50625946     7.433   .0000
+--------+--------------+----------------+--------+--------+
 GC      |    .53961173       .14654681     3.682   .0002
 TTME    |   -.06847037       .01674719    -4.088   .0000
 INVC    |   -.58715772       .14955000    -3.926   .0001
 INVT    |   -.09100015       .02158271    -4.216   .0000
 TASC    |   4.62957401       .81841212     5.657   .0000
 BASC    |   3.27415138       .76403628     4.285   .0000
Matrix IIATEST  has  1 rows and  1 columns.
               1
        +--------------
       1|   33.78445         Test statistic
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =       9.487729     Critical value
```

IIA is rejected



Alternative to Utility Maximization (!) Minimizing Random Regret

The random regret model begins from an assumption that when choosing between alternatives, decision makers seek to minimize anticipated random regret, where random regret consists of the sum of the familiar iid extreme value and a regret function defined below. Systematic regret for choice *i*, is R_i, which consists of the sum of the binary regrets associated with bilateral comparisons of the attributes of the chosen alternative and the available alternatives. (See Chorus (2010), and Chorus, Greene and Hensher (2011).)

Attribute level regret for the *k*th attribute for alternative *i* compared to available alternative *j* is

$$R_{ij}(k) = \log\{1+\exp[\beta_k(x_{jk}-x_{ik})]\}.$$

Systematic regret for choice *i* is the sum over the available alternatives of the systematic regret,

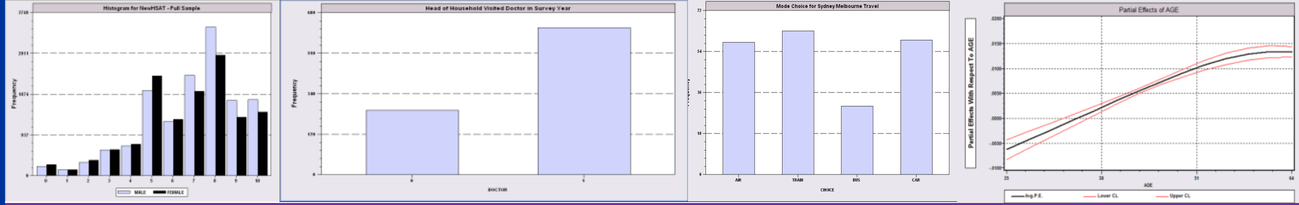
$$R_i = \sum_{j\neq i} \sum_{k=1}^{K} \log\{1+\exp[\beta_k(x_{jk}-x_{ik})]\}.$$

Random regret for alternative *i* is $R_i + \varepsilon_i$. Minimization of regret is equivalent to maximization of the negative of regret. This produces the familiar form for the probability,

$$P_i = \frac{\exp(-R_i)}{\sum_{j=1}^{J} \exp(-R_j)}.$$

We also consider a hybrid form, in which some attributes are treated in random regret form and others are contributors to random utility. The result is

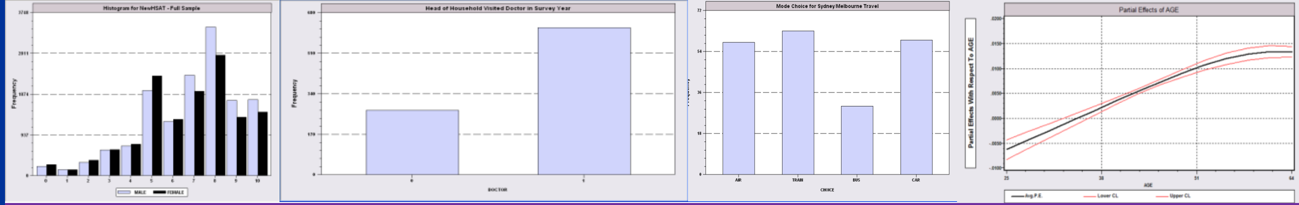
$$R_i = \sum_{k=1}^{K} \beta_k x_{ik} - \sum_{j\neq i} \sum_{k=1}^{K} \log\{1+\exp[\beta_k(x_{jk}-x_{ik})]\}.$$



RUM vs. Random Regret

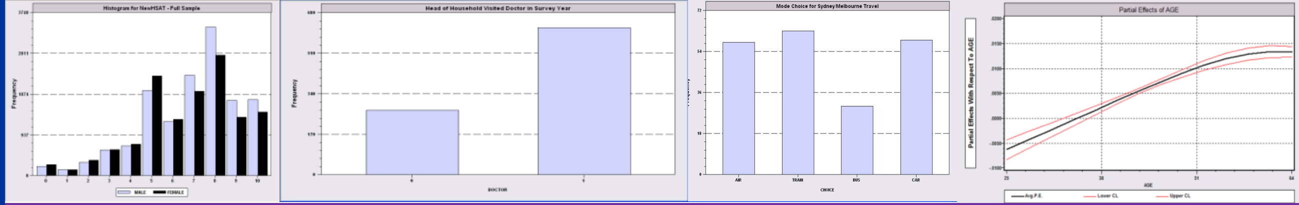
```
--------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                  Choice
Log likelihood function         -182.33831
Estimation based on N =     210, K =     8
Inf.Cr.AIC  =    380.7 AIC/N =       1.813
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3574  .3492
--------+-----------------------------------
        |                  Standard
    MODE|  Coefficient       Error          z
--------+-----------------------------------
      GC|     .07560***      .01825      4.14
    TTME|   -.10290***       .01099     -9.37
    INVT|   -.01435***       .00265     -5.41
    INVC|   -.08952***       .01995     -4.49
    AASC|   4.06574***      1.05260      3.86
    TASC|   4.27393***       .51214      8.35
    BASC|   3.71445***       .50856      7.30
   HINCA|     .02364**       .01155      2.05
--------+-----------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%,
--------------------------------------------

Elasticity wrt change of X in row choice on P
--------+-----------------------------------
GC      |      AIR     TRAIN       BUS      CAR
--------+-----------------------------------
     AIR|   5.4152   -2.3448   -2.3448  -2.3448
   TRAIN|  -2.3946    7.4483   -2.3946  -2.3946
     BUS|  -1.1512   -1.1512    7.5620  -1.1512
     CAR|  -1.9584   -1.9584   -1.9584   5.2548
```

```
--------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                  Choice
Log likelihood function         -173.31398
Estimation based on N =     210, K =     8
Inf.Cr.AIC  =    362.6 AIC/N =       1.727
Model estimated: Sep 15, 2011, 06:18:41
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3892  .3814
>>> Random Regret Form of MNL Model <<<
--------+-----------------------------------
        |                  Standard
    MODE|  Coefficient       Error          z
--------+-----------------------------------
      GC|     .02634***      .00458      5.75
    TTME|   -.03606***       .00426     -8.46
    INVT|   -.00877***       .00121     -7.28
        |Attributes Attended to in Random Uti
    INVC|   -.05957***       .01049     -5.68
    AASC|   1.85720**        .86496      2.15
    TASC|   2.59183***       .33957      7.63
    BASC|   1.99911***       .33786      5.92
   HINCA|     .02048**       .01021      2.01
--------+-----------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%,
--------------------------------------------

Elasticity wrt change of X in row choice on P
--------+-----------------------------------
GC      |      AIR     TRAIN       BUS      CAR
--------+-----------------------------------
     AIR|   1.6493   -1.0544   -1.0544  -1.0544
   TRAIN|   -.6910    2.7384    -.6910   -.6910
     BUS|   -.4518    -.4518    2.5840   -.4518
     CAR|   -.4492    -.4492    -.4492   2.0639
```



Fixed Effects Multinomial Logit:

Application of Minimum Distance Estimation



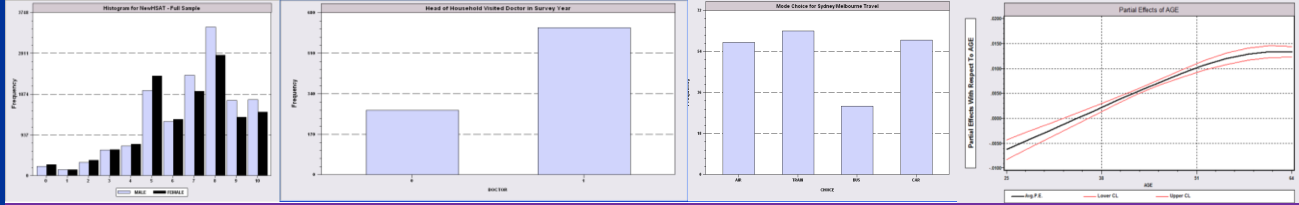
Binary Logit Conditional Probabiities

$$\text{Prob}(y_{it}=1\,|\,\mathbf{x}_{it}) = \frac{e^{\alpha_i+\mathbf{x}_{it}'\boldsymbol{\beta}}}{1+e^{\alpha_i+\mathbf{x}_{it}'\boldsymbol{\beta}}}.$$

$$\text{Prob}\left(Y_{i1}=y_{i1},Y_{i2}=y_{i2},\ldots,Y_{iT_i}=y_{iT_i}\,\middle|\,\sum_{t=1}^{T_i}y_{it}\right)$$

$$=\frac{\exp\left(\sum_{t=1}^{T_i}y_{it}\mathbf{x}_{it}'\boldsymbol{\beta}\right)}{\sum_{\Sigma_t d_{it}=S_i}\exp\left(\sum_{t=1}^{T_i}d_{it}\mathbf{x}_{it}'\boldsymbol{\beta}\right)}=\frac{\exp\left(\sum_{t=1}^{T_i}y_{it}\mathbf{x}_{it}'\boldsymbol{\beta}\right)}{\sum_{\text{All }\binom{T_i}{S_i}\text{ different ways that }\Sigma_t d_{it}\text{ can equal }S_i}\exp\left(\sum_{t=1}^{T_i}d_{it}\mathbf{x}_{it}'\boldsymbol{\beta}\right)}.$$

Denominator is summed over all the different combinations of T_i values of y_{it} that sum to the same sum as the observed $\Sigma_{t=1}^{T_i}y_{it}$. If S_i is this sum, there are $\binom{T}{S_i}$ terms. May be a huge number. An algorithm by Krailo and Pike makes it simple.



Example: Seven Period Binary Logit

y	x				
DOCTOR	AGE	EDUC	HSAT	INCOME	MARRIED
1	54	9	4	.08300	1
0	55	9	5	.09650	1
0	56	9	8	.12000	1
0	57	9	6	.13000	1
1	58	9	6	.11560	1
1	61	9	3	.10640	1
1	64	9	4	.09700	1

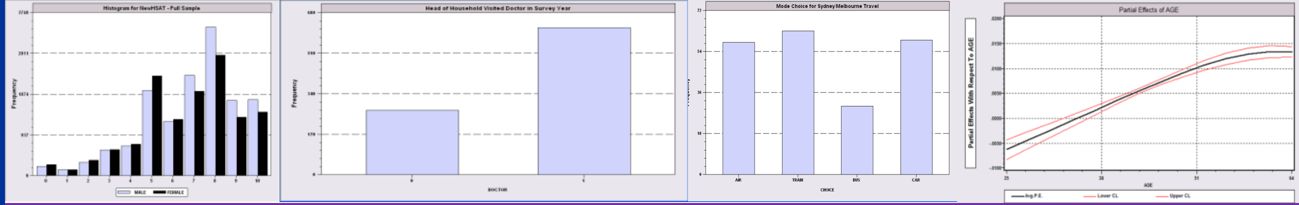
$$\text{Prob}[\mathbf{y} = (1,0,0,0,1,1,1)|\mathbf{X}_i] =$$

$$\frac{\exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_1)}{1+\exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_1)} \times \frac{1}{1+\exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_2)} \times \ldots \times \frac{\exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_7)}{1+\exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_7)}$$

There are 35 different sequences of y_{it} (permutations) that sum to 4.

For example, $y^*_{it|p=1}$ might be (1,1,1,1,0,0,0). Etc.

$$\text{Prob}[y=(1,0,0,0,1,1,1)|\mathbf{X}_i, \Sigma_{t=1}^{7} y_{it} = 7] = \frac{\exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{7} y_{it}\mathbf{x}_{it}\right]}{\sum_{p=1}^{35} \exp\left[\boldsymbol{\beta}'\Sigma_{t=1}^{7} y^*_{it|p}\mathbf{x}_{it}\right]}$$

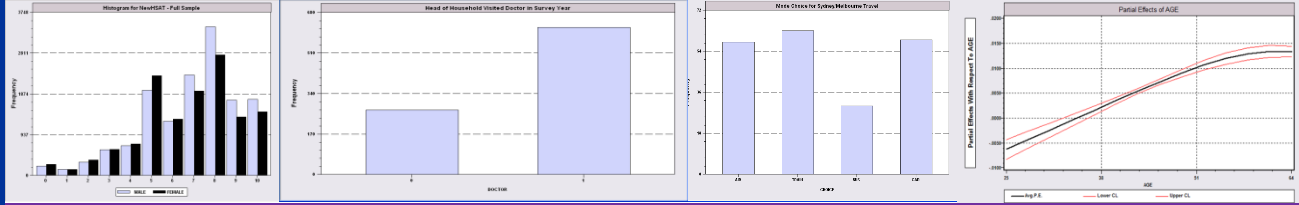


Panel Data Binomial Logit Model						DOCTOR		
Number of individuals					=	887		
Number of periods					=	7		
Conditioning event = the sum of						DOCTOR		
Observed distribution of sums by periods								
Sum	0	1	2	3	4	5	6	7
Number	48	73	82	100	115	116	151	202
Pct.	5.4	8.2	9.2	11.3	13.0	13.1	17.0	22.8

```
Logit Model for Panel Data
Dependent variable                  DOCTOR
Log likelihood function        -1712.75034
Estimation based on N =    6209, K =     5
Inf.Cr.AIC  =    3435.5 AIC/N =       .553
Fixed Effect Logit Model for Panel Data
```

DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
AGE	.05744***	.01128	5.09	.0000	.03534	.07954
EDUC	.07263	.17827	.41	.6837	-.27677	.42203
HSAT	-.25515***	.02396	-10.65	.0000	-.30211	-.20820
INCOME	.02831	.30633	.09	.9264	-.57208	.62870
MARRIED	.04337	.19974	.22	.8281	-.34810	.43485

***, **, * ==> Significance at 1%, 5%, 10% level.
Model was estimated on Jan 20, 2015 at 03:05:01 PM

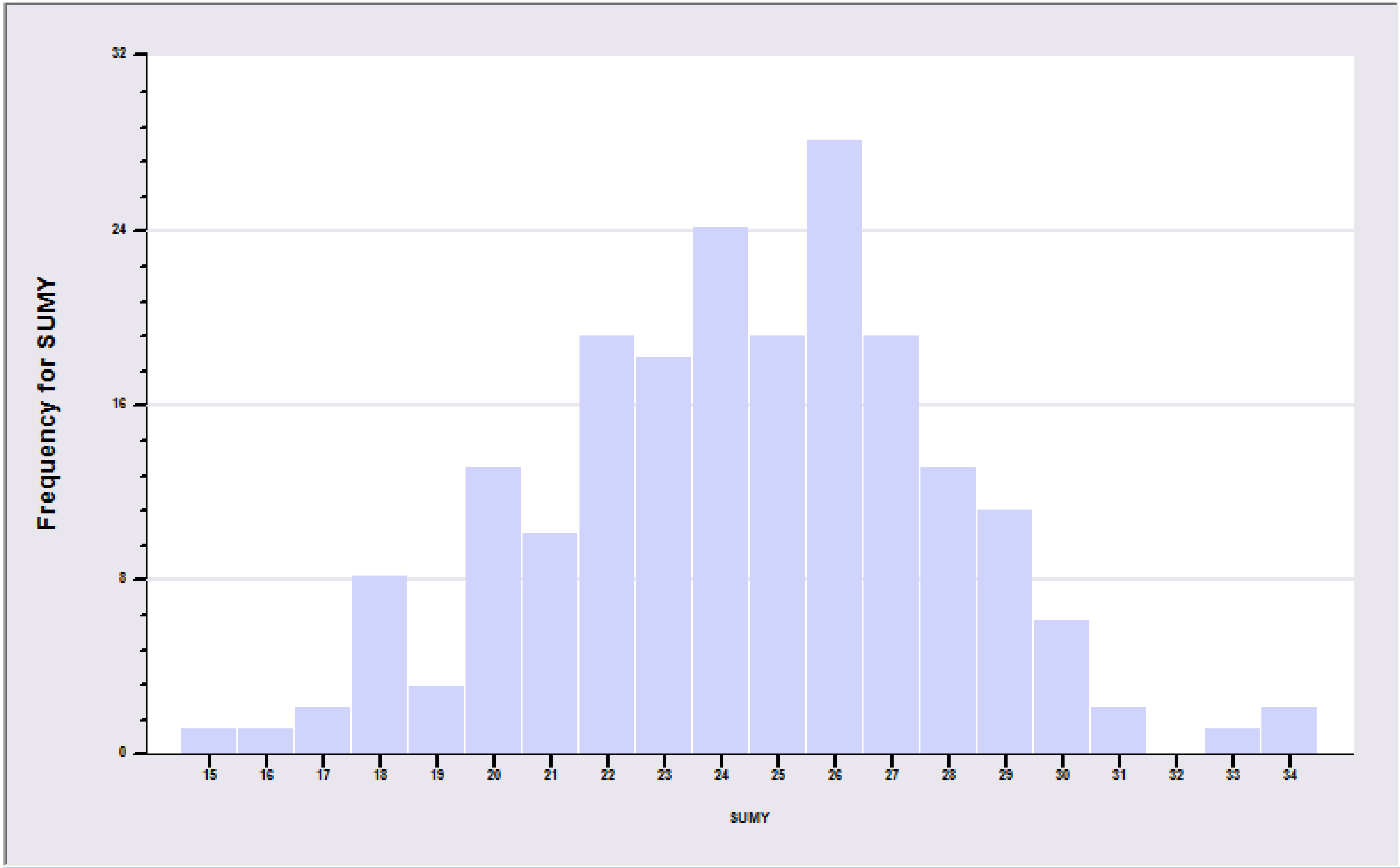


With T = 50, the number of permutations of sequences of y ranging from sum = 0 to sum = 50 ranges from 1 for 0 and 50, to 2.3×10^{12} for 15 or 35 up to a maximum of 1.3×10^{14} for sum =25. These are the numbers of terms that must be summed for a model with T = 50. In the application below, the sum ranges from 15 to 35.

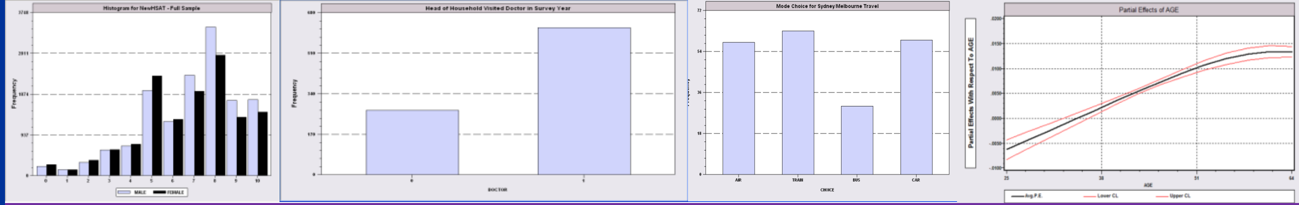
Untitled 1 *

Insert Name:

```
rows      ; 10000$
create    ; i=trn(50,0)$  (1,1,...,2,2,...,3,3,....)
create    ; t=trn(-50,0)$ (1,2,...,50,1,2,...,50...)
create    ; x1=rnn(0,1);x2=rnn(0,1) $
setpanel  ; group=i;pds=ti $
create    ; y=(.5*x1+.5*x2+rnn(0,1))>0$
create    ; sumy=groupsums(y,pds=50)$
histogram; if[t=1];rhs=sumy$
timer $
logit     ; lhs=y;rhs=x1,x2;panel;pds=50$
```

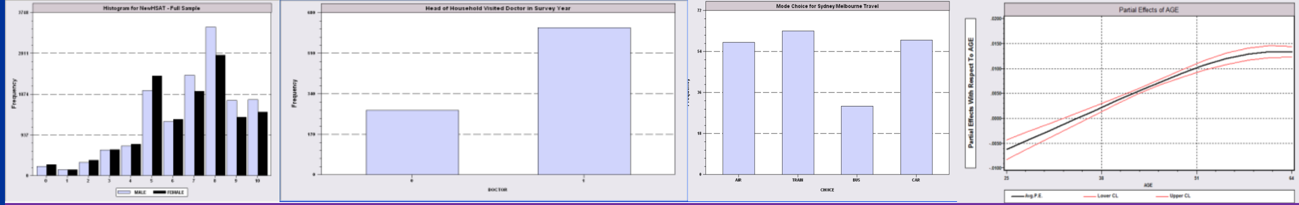


The sample is 200 individuals each observed 50 times.




```
-----------------------------------------------------------
Panel Data Binomial Logit Model                          Y
Number of individuals                       =      200
Number of periods                           =       50
Conditioning event = the sum of                          Y
-----------------------------------------------------------

Normal exit:   6 iterations. Status=0, F=    5213.705

-----------------------------------------------------------------------------
Logit Model for Panel Data
Dependent variable                      Y
Log likelihood function      -5213.70525
Estimation based on N =  10000, K =    2
Inf.Cr.AIC  =  10431.4 AIC/N =     1.043
Fixed Effect Logit Model for Panel Data
--------+--------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
       Y|   Coefficient     Error        z   |z|>Z*          Interval
--------+--------------------------------------------------------------------
      X1|      .84332***   .02600    32.44  .0000      .79236      .89428
      X2|      .82151***   .02601    31.59  .0000      .77054      .87248
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Jan 20, 2015 at 03:23:46 PM
-----------------------------------------------------------------------------

Elapsed time:     0 hours,  0 minutes,   .359 seconds.
```

The data are generated from a probit process with b1 = b2 = .5. But, it is fit as a logit model. The coefficients obey the familiar relationship, 1.6*probit.



Multinomial Logit Model: J+1 choices including a base choice.

y_{itj} = 1 if individual i makes choice j in period t

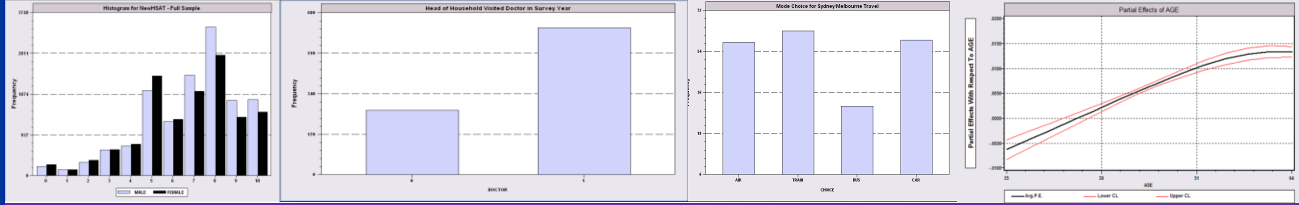
$$\text{Prob}(y_{itj}=1\,|\,\mathbf{x}_{itj}) = \frac{e^{\alpha_{ij}+\mathbf{x}'_{itj}\boldsymbol{\beta}}}{1+\Sigma_{m=1}^{J} e^{\alpha_{im}+\mathbf{x}'_{itm}\boldsymbol{\beta}}},\ j=1,\ldots,J.$$

$$\text{Prob}(y_{it0}=1\,|\,\mathbf{x}_{it0}) = \frac{1}{1+\Sigma_{m=1}^{J} e^{\alpha_{im}+\mathbf{x}'_{itm}\boldsymbol{\beta}}}.$$

The probability attached to the sequence of choices is remarkably complicated.

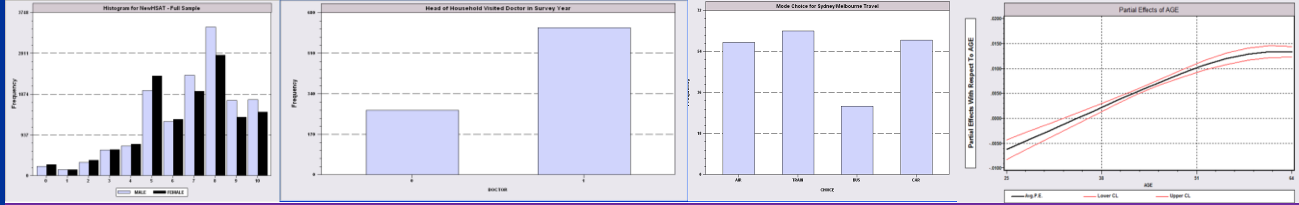
$$= \frac{\prod_{j=1}^{J} \exp\left(\sum_{t=1}^{T_i} y_{itj}\mathbf{x}'_{itj}\boldsymbol{\beta}\right)}{\sum_{j=1}^{J} \sum_{\Sigma_t d_{itj}=S_{ij}} \exp\left(\sum_{t=1}^{T_i} d_{it}\mathbf{x}'_{it}\boldsymbol{\beta}\right)} = \frac{\prod_{j=1}^{J} \exp\left(\sum_{t=1}^{T_i} y_{itj}\mathbf{x}'_{itj}\boldsymbol{\beta}\right)}{\sum_{j=1}^{J} \sum_{\text{All}\binom{T_i}{S_{ij}}\text{ different ways that }\Sigma_t d_{itj}\text{ can equal }S_{ij}} \exp\left(\sum_{t=1}^{T_i} d_{itj}\mathbf{x}'_{itj}\boldsymbol{\beta}\right)}.$$

Denominator is summed over all the different combinations of T_i values of y_{itj} that sum to the same sum as the observed $\Sigma_{t=1}^{T_i} y_{it}$. If S_{ij} is this sum, there are $\binom{T}{S_{ij}}$ terms. May be a huge number. Larger yet by summing over choices.



Estimation Strategy

- Conditional ML of the full MNL model. Impressively complicated.
- A Minimum Distance (MDE) Strategy
  - Each alternative treated as a binary choice vs. the base provides an estimator of β
    - Select subsample that chose either option j or the base
    - Estimate β using this binary choice setting
    - This provides J different estimators of the same β
  - Optimally combine the different estimators of β



Minimum Distance Estimation

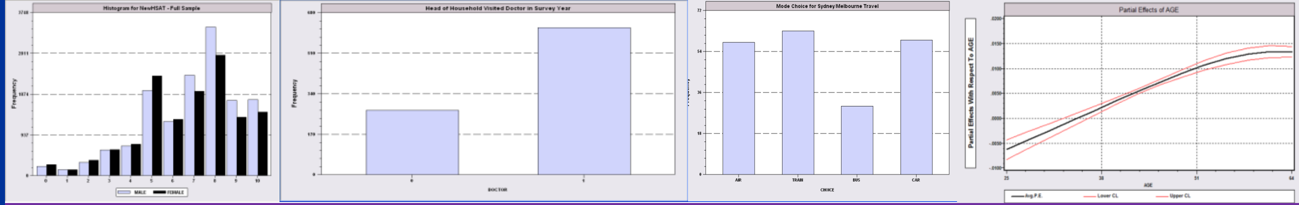
There are J estimators $\hat{\boldsymbol{\beta}}_j$ of the same parameter vector, $\hat{\boldsymbol{\beta}}$.

Each estimator is consistent and asymptotically normal.

Estimated covariance matrices $\hat{\mathbf{V}}_j$. How to combine the estimators?

$$\text{MDE: Minimize wrt } \hat{\beta}_* \ \ q = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_* \\ \hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_* \\ \vdots \\ \hat{\boldsymbol{\beta}}_J - \hat{\boldsymbol{\beta}}_* \end{bmatrix}' \mathbf{W} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_* \\ \hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_* \\ \vdots \\ \hat{\boldsymbol{\beta}}_J - \hat{\boldsymbol{\beta}}_* \end{bmatrix}$$

What to use for the weighting matrix **W**? Any positive definite matrix will do.

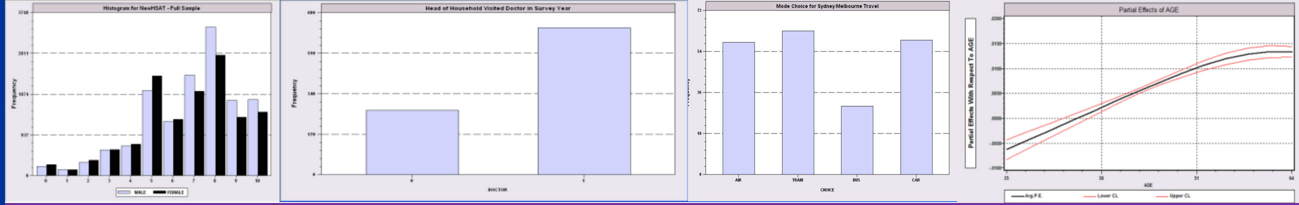


MDE Estimation

Estimated covariance matrices $\hat{\mathbf{V}}_j$. How to combine the estimators?

$$\text{MDE: Minimize wrt } \hat{\boldsymbol{\beta}}_* \ \ q = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_* \\ \hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_* \\ \vdots \\ \hat{\boldsymbol{\beta}}_J - \hat{\boldsymbol{\beta}}_* \end{bmatrix}' \mathbf{W} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_* \\ \hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_* \\ \vdots \\ \hat{\boldsymbol{\beta}}_J - \hat{\boldsymbol{\beta}}_* \end{bmatrix}. \text{ Propose a GLS approach}$$

$$\mathbf{W} = \mathbf{A}^{-1} = \begin{bmatrix} \hat{\mathbf{V}}_1 & 0 & \cdots & 0 \\ 0 & \hat{\mathbf{V}}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\mathbf{V}}_J \end{bmatrix}^{-1}$$



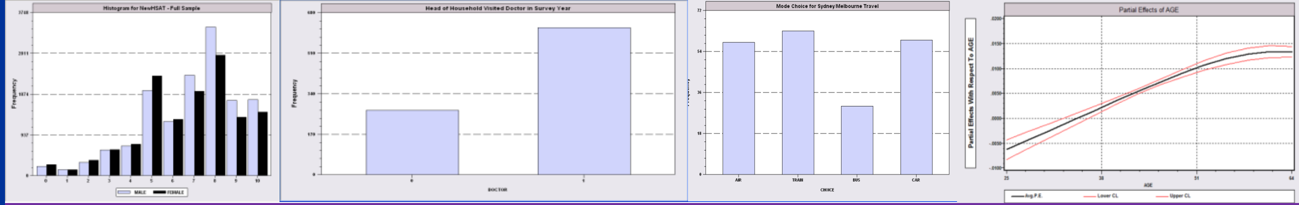
MDE Estimation

$$\text{MDE: Minimize wrt } \hat{\boldsymbol{\beta}}_* \ \ q = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_* \\ \hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_* \\ \vdots \\ \hat{\boldsymbol{\beta}}_J - \hat{\boldsymbol{\beta}}_* \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{V}}_1 & 0 & \cdots & 0 \\ 0 & \hat{\mathbf{V}}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\mathbf{V}}_J \end{bmatrix}^{-1} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_* \\ \hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_* \\ \vdots \\ \hat{\boldsymbol{\beta}}_J - \hat{\boldsymbol{\beta}}_* \end{bmatrix}.$$

$$\text{The solution is } \hat{\boldsymbol{\beta}}_* = \left[\hat{\mathbf{V}}_1^{-1} + \hat{\mathbf{V}}_2^{-1} + \ldots + \hat{\mathbf{V}}_J^{-1}\right]^{-1} \left[\hat{\mathbf{V}}_1^{-1}\hat{\boldsymbol{\beta}}_1 + \hat{\mathbf{V}}_2^{-1}\hat{\boldsymbol{\beta}}_2 + \ldots + \hat{\mathbf{V}}_J^{-1}\hat{\boldsymbol{\beta}}_J\right]$$

$$= \left[\sum_{j=1}^{J} \hat{\mathbf{V}}_j^{-1}\right]^{-1} \left[\sum_{j=1}^{J} \hat{\mathbf{V}}_j^{-1}\hat{\boldsymbol{\beta}}_j\right]$$

$$= \sum_{j=1}^{J} \mathbf{H}_j \hat{\boldsymbol{\beta}}_j \ \ \text{where} \ \ \sum_{j=1}^{J} \mathbf{H}_j = \mathbf{I}$$

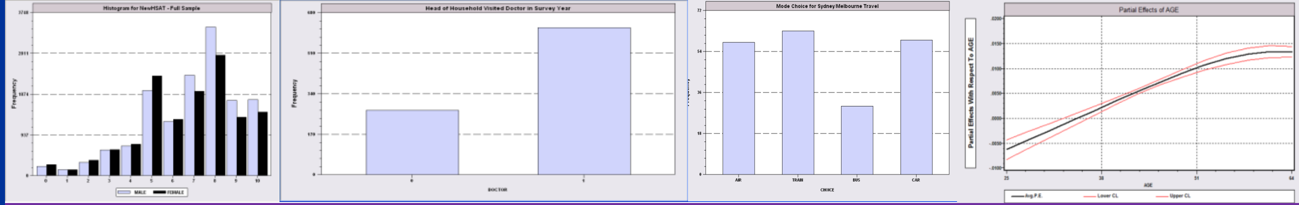


Implementation of a multinomial logit model with fixed effects

Klaus Pforr

Mannheim Centre for European Social Research (MZES)
University of Mannheim
klaus.pforr@mzes.uni-mannheim.de

July 1, 2011,
Ninth German Stata Users Group Meeting, Bamberg

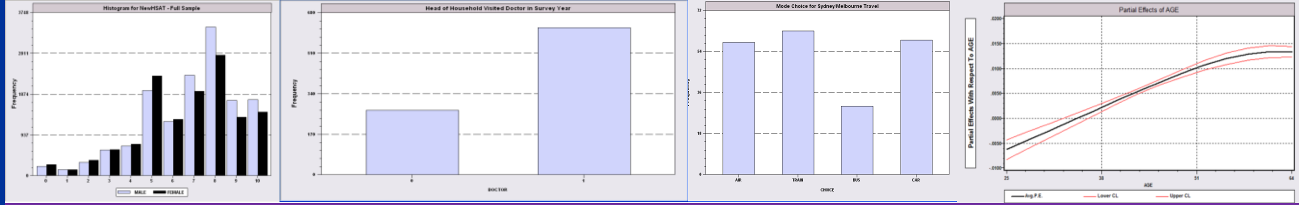


Examples: Simulated data

Performance with more alternatives

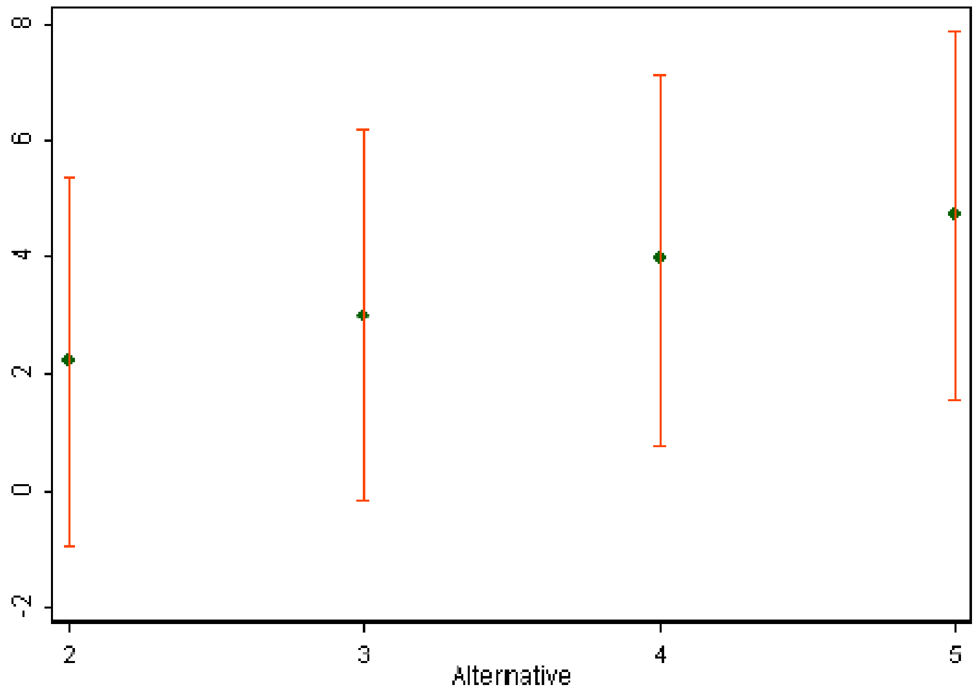
Simulated data

- N=1000, T=5, J=5
- Unobs. het. α_{ij}: over all i random draw $(\alpha_{i1}, \ldots, \alpha_{i5})$ from uniform distribution over 4-simplex Δ^4.
- Error ε_{itj}: over all i and t, for each j indep. draws from Gumbel-distribution $(E(\varepsilon_{itj}) = \gamma, Var(\varepsilon_{itj}) = \pi/\sqrt{6})$.
- Indep. variable: x correlated with α
  - $x_{it} = u_{it} + \alpha_{i2}$,
  - u_{it} drawn from uniform distribution.
- Coefficients $\beta_2 = 2, \beta_3 = 3, \beta_4 = 4, \beta_5 = 5$.

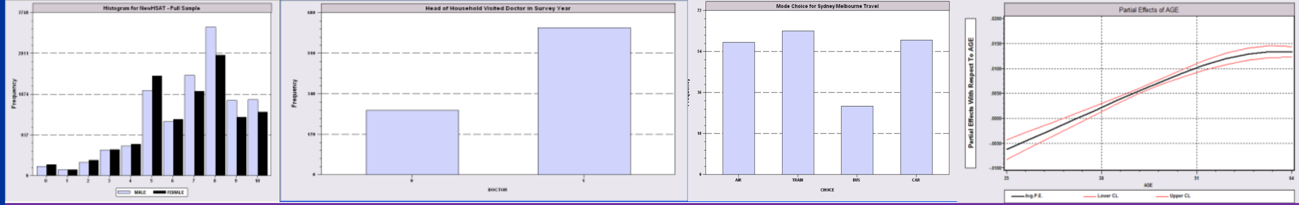


Examples: Simulated data (cont.)

Results



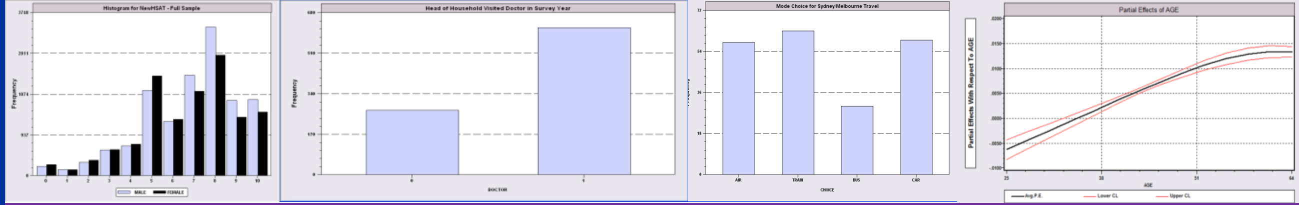
informative observations: N=3405; speed: 20.83 sec.



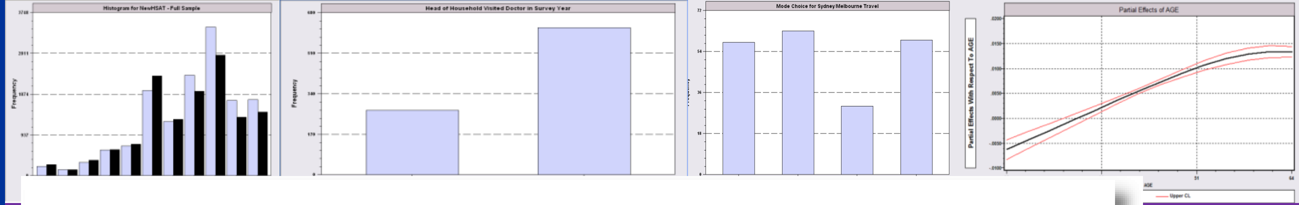
FEMLtest.lim *

Insert Name:

```
rows;25000$
sample;1-25000$
create;x1=rnn(0,1);x2=rnn(0,1);x3=(rnu(0,1)>.4)-.6$
create;e=rnu(0,1);u=-log(-log(e))$
create;f=x1+x2+x3+u$
create;j=trn(-5,0)$
create;choice=groupmax(f,pds=5)$
create;choice=(choice=j)$
timer $
femlogit ;lhs=choice
         ;rhs=x1,x2,x3
         ;choices=alt1,alt2,alt3,alt4,alt5
         ;base=alt5
         ;pds=5; mde $
```




```
Normal exit:   6 iterations. Status=0, F=    958.5621
Normal exit:   6 iterations. Status=0, F=    896.4347
Normal exit:   5 iterations. Status=0, F=    1011.594
Normal exit:   6 iterations. Status=0, F=    941.2443

-----------------------------------------------------------------------------
Fixed Effects  Multinomial Logit  Model
Dependent variable                 CHOICE
Log likelihood function       -3807.83527
Restricted log likelihood     -5033.94353
Chi squared [   3](P= .000)    2452.21653
Significance level                 .00000
McFadden Pseudo R-squared        .2435681
Estimation based on N =   25000, K =    3
Inf.Cr.AIC  =    7621.7 AIC/N =      .305
Estimator is Minimum Distance Wtd. Avrg
--------+--------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
  CHOICE|  Coefficient     Error       z     |z|>Z*          Interval
--------+--------------------------------------------------------------------
      X1|      .56724***    .01818    31.21  .0000      .53162     .60287
      X2|      .54600***    .01803    30.28  .0000      .51066     .58135
      X3|      .45942***    .03346    13.73  .0000      .39384     .52500
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Jan 29, 2015 at 09:23:04 PM
-----------------------------------------------------------------------------

Elapsed time:      0 hours,   0 minutes,     .344 seconds.
```

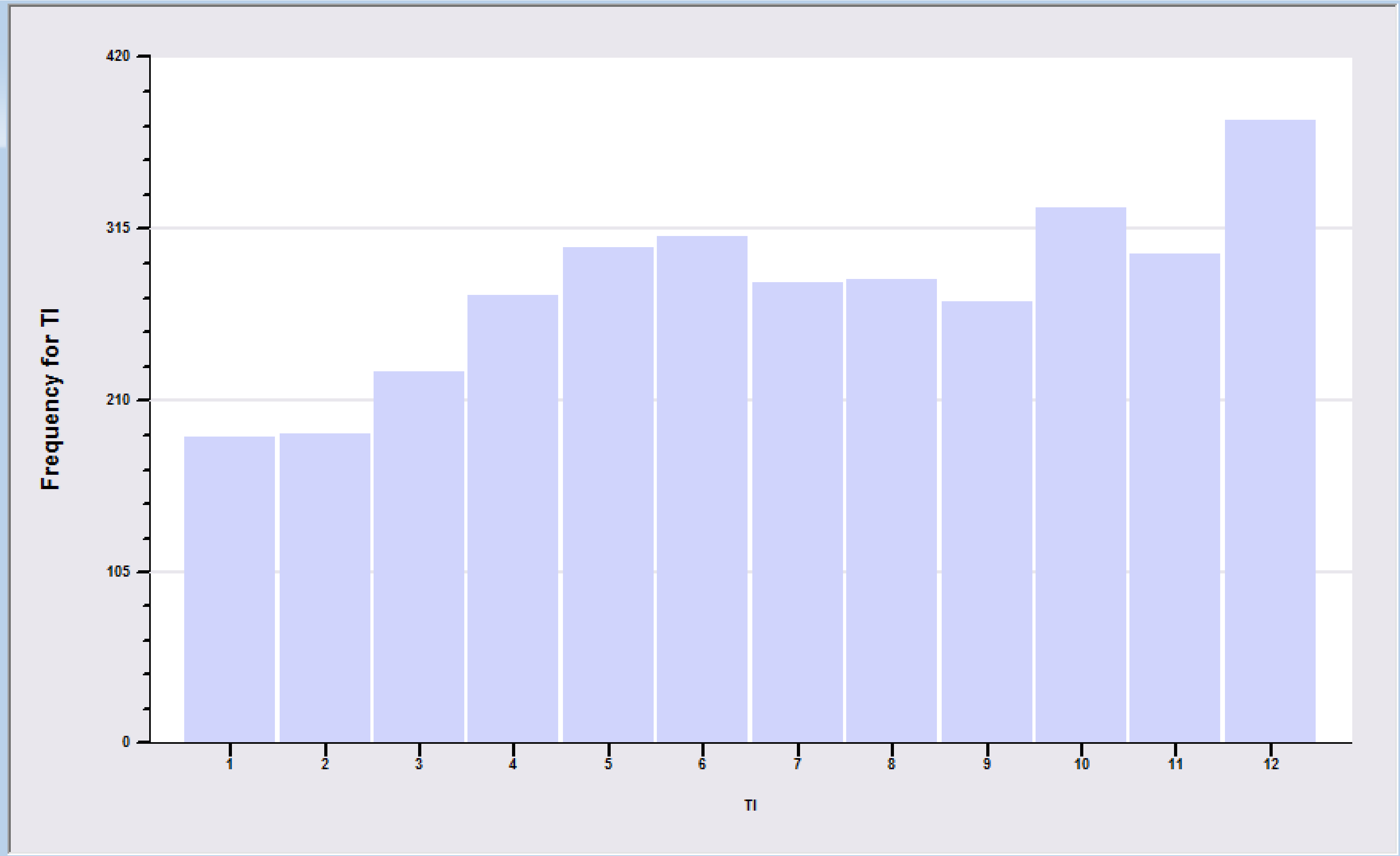


How precise and how fast is it?

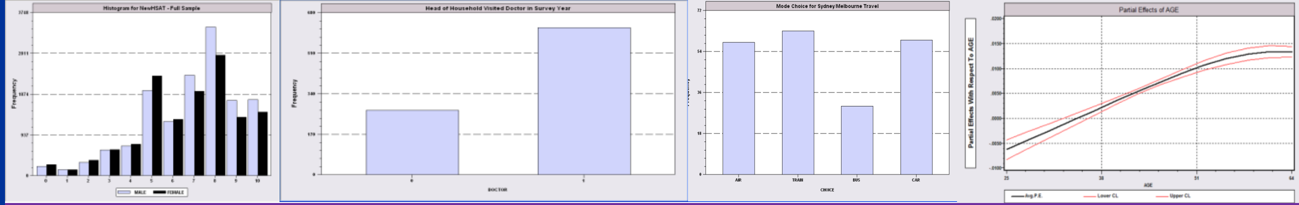
Comparison with `clogit` for $J = 2$.

- Data used: `http://www.stata-press.com/data/r11/union.dta`
- Relative difference of coefficients: 9.078e-16.
- Speed: `clogit`: 2.42 sec., `femlogit`: 101.58 sec..

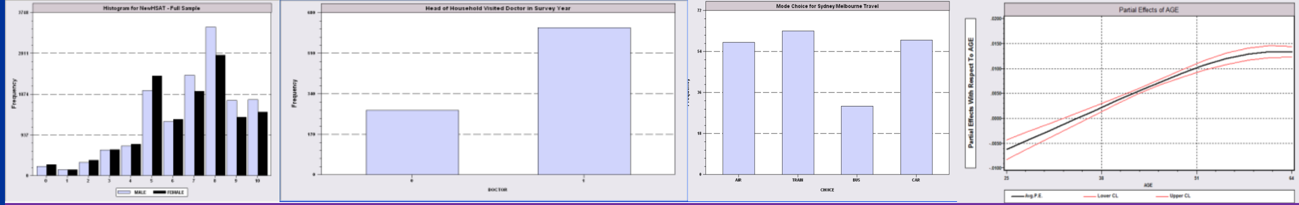
```
. femlogit union age grade not_smsa south black, group(idcode) b(0)

note: 2744 groups (14165 obs) dropped because of all positive or
      all negative outcomes.
note: black omitted because of no within-group variance.

Iteration 0:   log likelihood = -4521.3385
Iteration 1:   log likelihood = -4516.1404
Iteration 2:   log likelihood = -4516.1385
Iteration 3:   log likelihood = -4516.1385

                                                 Number of obs   =      12035
                                                 Wald chi2(4)    =          .
Log likelihood = -4516.1385                      Prob > chi2     =          .

------------------------------------------------------------------------------
       union |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0170301    .004146     4.11   0.000     .0089042    .0251561
       grade |   .0853572   .0418781     2.04   0.042     .0032777    .1674368
    not_smsa |   .0083678   .1127963     0.07   0.941    -.2127088    .2294445
       south |   -.748023   .1251752    -5.98   0.000    -.9933619   -.5026842
       black |  (omitted)
------------------------------------------------------------------------------
```



420
315
210
105
0
Frequency for TI
1
2
3
4
5
6
7
8
9
10
11
12
TI

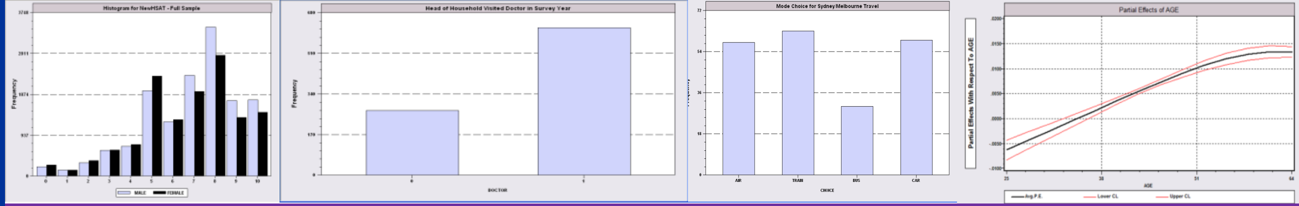



```
|-> femlogit;lhs=union;choices=union,notunion
    ;rhs=age,grade,not_smsa,south,black
    ;pds=ti
    ;base=notunion;list;mde$
-----------------------------------------------------------------------------
Fixed Effects  Multinomial Logit  Model
Dependent variable                UNION
Log likelihood function      -4516.13849
Restricted log likelihood    -4550.18592
Chi squared [   4](P= .000)     68.09487
Significance level               .00000
McFadden Pseudo R-squared      .0074826
Estimation based on N =   52400, K =   4
Inf.Cr.AIC   =     9040.3 AIC/N =    .173
Estimator is Conditional Max Likelihood
--------+--------------------------------------------------------------------
        |                Standard            Prob.        95% Confidence
   UNION|  Coefficient     Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
     AGE|    .01703***     .00415     4.11  .0000        .00890      .02516
   GRADE|    .08536**      .04188     2.04  .0415        .00328      .16744
NOT_SMSA|    .00837        .11280      .07  .9409      -.21271      .22944
   SOUTH|   -.74802***     .12518    -5.98  .0000      -.99336     -.50268
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Jan 29, 2015 at 08:44:46 PM
-----------------------------------------------------------------------------
Elapsed time:      0 hours,   0 minutes,     .297 seconds.
```

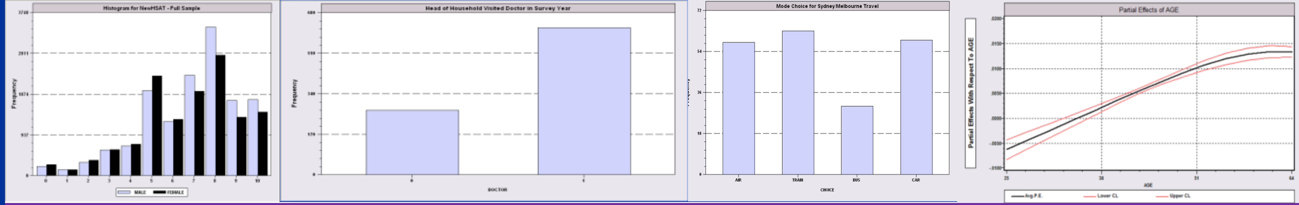


Why a 500 fold increase in speed?

- MDE is much faster
- Not using Krailo and Pike, or not using efficiently
- Numerical derivatives for an extremely messy function (increase the number of function evaluations by at least 5 times)



Rank Data and Best/Worst



Internet Dating Survey - 14 / 26

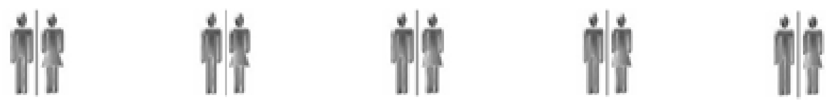
An example

We will show you 5 profiles of people that you might consider contacting.

We will then ask you to tell us which profile represents the best candidate and which represents the worst.

We will then get you to tell us of the three remaining profiles, which is the best and which is the worst.

If you were looking through a dating website and considered contacting among the five people shown based on the descriptions listed, which profile represents the best candidate and which represents the worst? And then which is the best and which is the worst of the three remaining profiles?



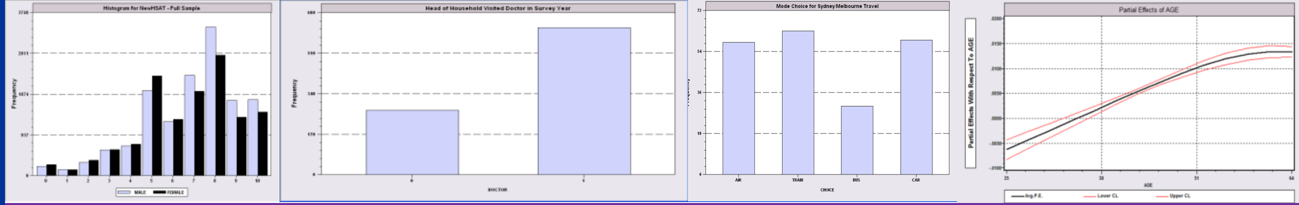
	Person A	Person B	Person C	Person D	Person E
Drinking Habit	Non drinker	Casual drinker	Moderate drinker	Casual drinker	Moderate drinker
Smoking Habit	Ex smoker	Smoker	Non smoker	Ex smoker	Smoker
Children	Single parent	None currently	Single parent	Single parent	None currently
Job	White Collar	Blue Collar	Blue Collar	Unemployed	White Collar
Looks	Above average	Below average	Above average	Below average	Average
Cost to contact	$20	$15	$10	$15	$10
Which profile do you consider to be the best and which is the worst?	Best			Worst	
Of the remaning profiles, which profile is the best and which is the worst?		Worst	Best		

In the above example, we found Person A to be the best candidate for contacting and person D the worst, and Of the three remaining profiles, we believe that Person C is the best remaining profile and Person B the worst.

You will be shown nine scenarios similar to the above one. Each scenario will show the profiles of different potential contacts.

Please make sure that you understand the task before proceeding. Once you go to the next screen, you will not be able to go back.

Back | Next



Rank Data and Exploded Logit

Resp	Set	RespSet	Explode	Altij	Altn	Cset	Choice	Drink	Smoke	Child	Job	Looks	Cost
1	1	1	1	1	1	5	1	0	1	1	0	2	20
1	1	1	1	2	2	5	0	1	2	0	1	0	15
1	1	1	1	3	3	5	0	2	0	1	1	2	10
1	1	1	1	4	4	5	0	1	1	1	2	0	15
1	1	1	1	5	5	5	0	2	2	0	0	1	10
1	2	1	2	2	7	4	0	1	2	0	1	0	15
1	2	1	2	3	8	4	1	2	0	1	1	2	10
1	2	1	2	4	9	4	0	1	1	1	2	0	15
1	2	1	2	5	10	4	0	2	2	0	0	1	10
1	3	1	3	2	12	3	0	1	2	0	1	0	15
1	3	1	3	4	14	3	0	1	1	1	2	0	15
1	3	1	3	5	15	3	1	2	2	0	0	1	10
1	4	1	4	2	17	2	1	1	2	0	1	0	15
1	4	1	4	4	19	2	0	1	1	1	2	0	15

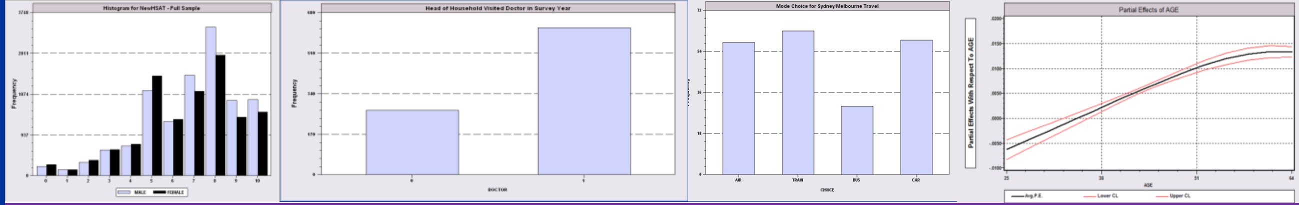
Alt 1 is the best overall

Alt 3 is the best among remaining alts 2,3,4,5

Alt 5 is the best among remaining alts 2,4,5

Alt 2 is the best among remaining alts 2,4

Alt 4 is the worst.

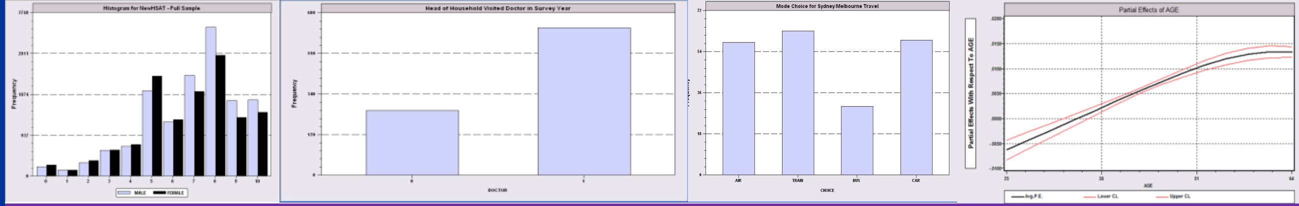


Exploded Logit

U[j] = jth favorite alternative among 5 alternatives

U[1] = the choice made if the individual indicates only the favorite

$$\text{Prob}\{j = [1],[2],[3],[4],[5]\} = \text{Prob}\{[1]|\text{choice set} = [1]...[5]\} \times$$
$$\text{Prob}\{[2]|\text{choice set} = [2]...[5]\} \times$$
$$\text{Prob}\{[3]|\text{choice set} = [3]...[5]\} \times$$
$$\text{Prob}\{[4]|\text{choice set} = [4],[5]\} \times$$
$$1$$



Exploded Logit

U[j] = jth favorite alternative among 5 alternatives

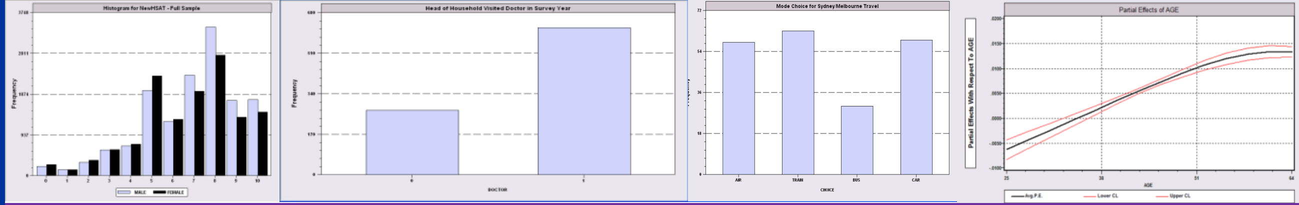
U[1] = the choice made if the individual indicates only the favorite

Individual ranked the alternatives 1,3,5,2,4

Prob{This set of ranks}

$$= \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_1)}{\sum_{j=1,2,3,4,5}\exp(\boldsymbol{\beta}'\mathbf{x}_j)} \times \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_3)}{\sum_{j=2,3,4,5}\exp(\boldsymbol{\beta}'\mathbf{x}_j)} \times$$

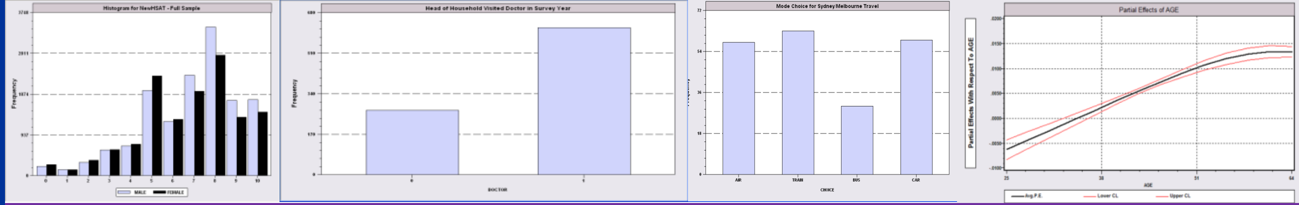
$$\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_5)}{\sum_{j=2,4,5}\exp(\boldsymbol{\beta}'\mathbf{x}_j)} \times \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_2)}{\sum_{j=2,4}\exp(\boldsymbol{\beta}'\mathbf{x}_j)} \times 1$$



Best Worst

- Individual simultaneously ranks best and worst alternatives.
- Prob(alt j) = best = $\exp[U(j)] / \Sigma_m \exp[U(m)]$
- Prob(alt k) = worst = $\exp[-U(k)] / \Sigma_m \exp[-U(m)]$

Resp	Set	Altij	Cset	Bestworst	AirNZ	Delta	Emirates	JetStar	Qantas	Singapore	United	Choice
1	1	1	4	1	0	0	0	0	0	1	0	0
1	1	2	4	1	0	0	1	0	0	0	0	0
1	1	3	4	1	0	0	0	0	1	0	0	0
1	1	4	4	1	0	0	0	0	0	0	0	1
1	1	1	4	-1	0	0	0	0	0	-1	0	0
1	1	2	4	-1	0	0	-1	0	0	0	0	1
1	1	3	4	-1	0	0	0	0	-1	0	0	0
1	1	4	4	-1	0	0	0	0	0	0	0	0



Case 1 involves respondents being shown subsets of alternatives/brands and being asked out of the subset shown, which alternative/brand is best and which is worst. Note that unlike discrete choice experiments, the alternatives/brands are not represented as bundles of attributes.

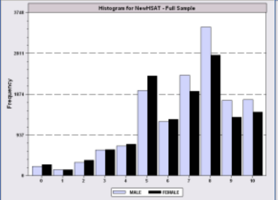
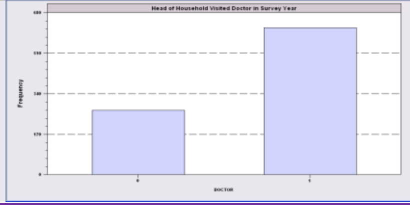
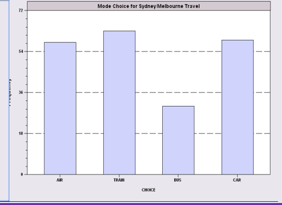
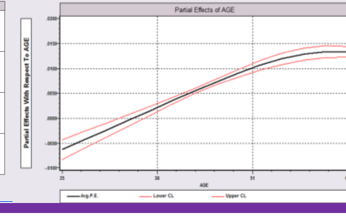
Consider an example of quadruples, selected from 1. Air NZ, 2. Delta, 3. Emirates, 4. Jetstar, 5. Qantas, 6. Singapore, 7. United, and 8. Virgin. An example of a choice question is show below.

Best worst scaling (Case 1)

Best	Attribute	Worst
○	Singapore	○
○	Emirates	○
○	Qantas	○
○	Virgin	○

Figure 1: Example B/W Case 1 task

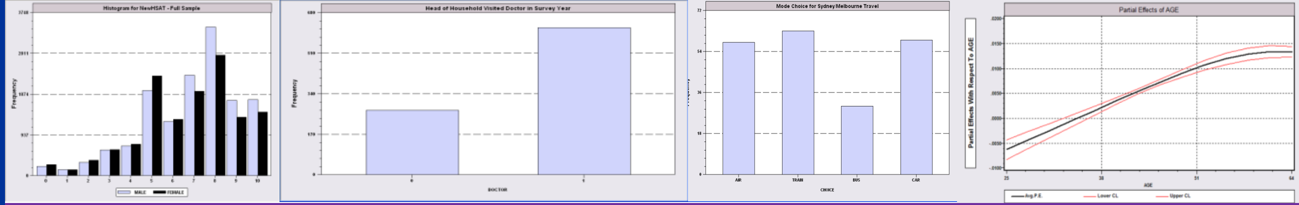
The data is set up as per a normal DCE where the attributes are dummy codes for the alternatives shown. Each task however is repeated, once for best and once for worst. For worst, the coding is the same, however -1 is used instead of 1. An example is presented in the table below, where the first task is an example of the above task.

Choices

Table 1: Example B/W Case 1 task data set up 1

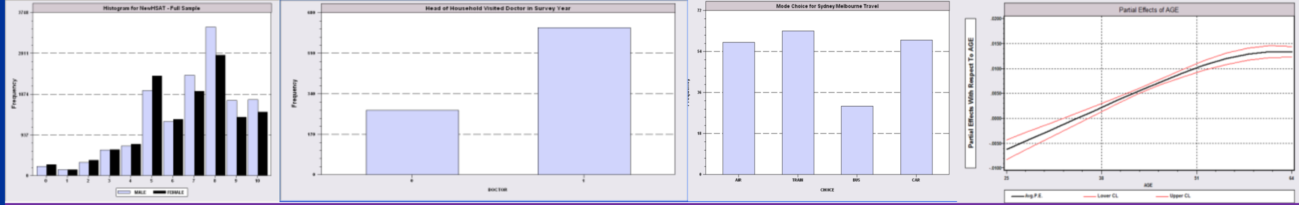
Resp	Set	Altij	Cset	Bestworst	AirNZ	Delta	Emirates	JetStar	Qantas	Singapore	United	Choice
1	1	1	4	1	0	0	0	0	0	1	0	0
1	1	2	4	1	0	0	1	0	0	0	0	0
1	1	3	4	1	0	0	0	0	1	0	0	0
1	1	4	4	1	0	0	0	0	0	0	0	1
1	1	1	4	-1	0	0	0	0	0	-1	0	0
1	1	2	4	-1	0	0	-1	0	0	0	0	1
1	1	3	4	-1	0	0	0	0	-1	0	0	0
1	1	4	4	-1	0	0	0	0	0	0	0	0
1	2	1	4	1	1	0	0	0	0	0	0	0
1	2	2	4	1	0	0	1	0	0	0	0	0
1	2	3	4	1	0	0	0	0	1	0	0	0
1	2	4	4	1	0	0	0	0	0	0	1	1
1	2	1	4	-1	-1	0	0	0	0	0	0	0
1	2	2	4	-1	0	0	-1	0	0	0	0	1
1	2	3	4	-1	0	0	0	0	-1	0	0	0
1	2	4	4	-1	0	0	0	0	0	0	-1	0



Best

Table 1: Example B/W Case 1 task data set up 1

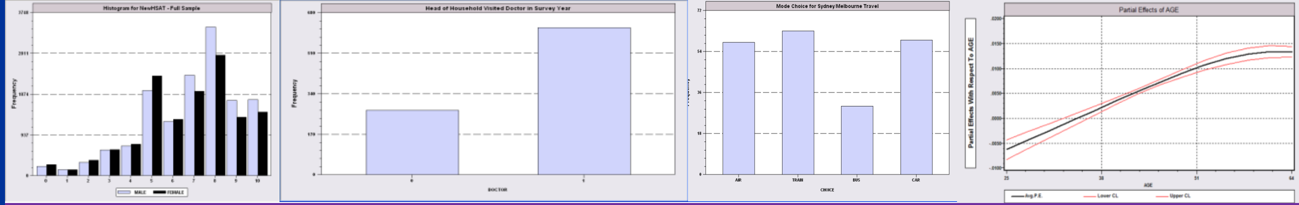
Resp	Set	Altij	Cset	Bestworst	AirNZ	Delta	Emirates	JetStar	Qantas	Singapore	United	Choice
1	1	1	4	1	0	0	0	0	0	1	0	0
1	1	2	4	1	0	0	1	0	0	0	0	0
1	1	3	4	1	0	0	0	0	1	0	0	0
1	1	4	4	1	0	0	0	0	0	0	0	1
1	2	1	4	1	1	0	0	0	0	0	0	0
1	2	2	4	1	0	0	1	0	0	0	0	0
1	2	3	4	1	0	0	0	0	1	0	0	0
1	2	4	4	1	0	0	0	0	0	0	1	1



Worst

Table 1: Example B/W Case 1 task data set up 1

Resp	Set	Altij	Cset	Bestworst	AirNZ	Delta	Emirates	JetStar	Qantas	Singapore	United	Choice
1	1	1	4	-1	0	0	0	0	0	-1	0	0
1	1	2	4	-1	0	0	-1	0	0	0	0	1
1	1	3	4	-1	0	0	0	0	-1	0	0	0
1	1	4	4	-1	0	0	0	0	0	0	0	0
1	2	1	4	-1	-1	0	0	0	0	0	0	0
1	2	2	4	-1	0	0	-1	0	0	0	0	1
1	2	3	4	-1	0	0	0	0	-1	0	0	0
1	2	4	4	-1	0	0	0	0	0	0	-1	0



Case 2 differs to case 1 in that the method concentrates on attributes, not alternatives. Consider an example with four attributes, seat pitch, entertainment, alcohol payment and stop over. The attribute levels of the four levels are given as:

Table 3: Example B/W Case 2 task attribute levels

Attribute	Level 1	Level 2	Level 3
Seat Pitch	28 inches	30 inches	32 inches
Entertainment	Single cabin screen	Limited movies	Full entertainment
Alcohol payment	Pay for alcohol	Free alcohol	
Stop over	No stop over	3 hour stop over	5 hour stop over

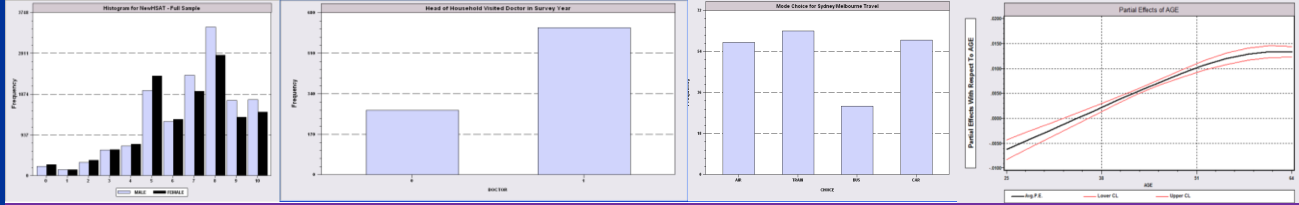
An example of a choice question is show below.

Best worst scaling (Case 2)

Best	Attribute	Worst
○	Seat pitch 30'	○
○	Limited movies	○
○	Pay for alcohol	○
○	5 hour stopover	○

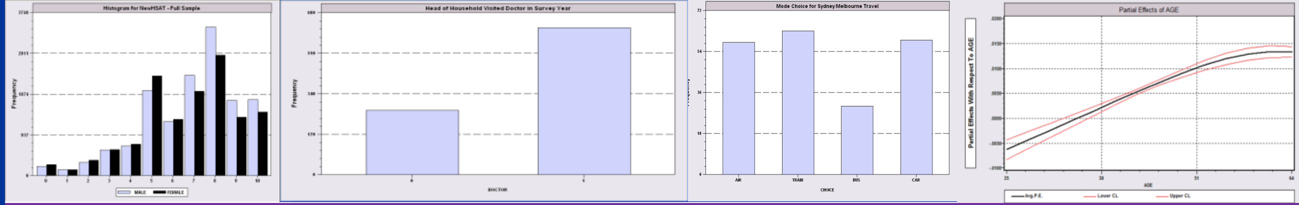
Figure 2: Example B/W Case 2 task

The data is set up as per a normal DCE where the attributes are dummy codes of the attribute levels. Each task however is repeated, once for best and once for worst. For worst, the coding is the same, however -1 is used instead of 1. An example is presented in the table below, where the first task is an example of the above task.



Uses the result that if U(i,j) is the lowest utility, -U(i,j) is the highest.

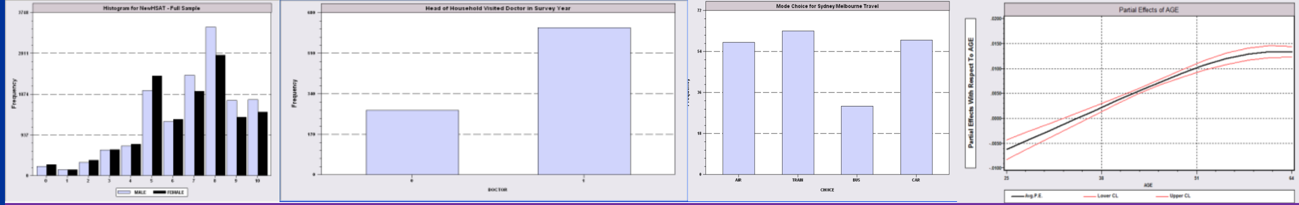
Resp	Set	Altij	Altn	Cset	Bestworst	Inch28	Inch30	CabScr	LimMov	Pay	Hour1	Hour3	Choice
1	1	1	1	4	1	0	1	0	0	0	0	0	0
1	1	2	2	4	1	0	0	0	1	0	0	0	0
1	1	3	3	4	1	0	0	0	0	0	0	0	1
1	1	4	4	4	1	0	0	0	0	0	0	0	0
1	1	1	5	4	-1	0	-1	0	0	0	0	0	0
1	1	2	6	4	-1	0	0	0	-1	0	0	0	1
1	1	3	7	4	-1	0	0	0	0	0	0	0	0
1	1	4	8	4	-1	0	0	0	0	0	0	0	0
1	2	1	1	4	1	1	0	0	0	0	0	0	0
1	2	2	2	4	1	0	0	1	0	0	0	0	0
1	2	3	3	4	1	0	0	0	0	1	0	0	1
1	2	4	4	4	1	0	0	0	0	0	0	1	0
1	2	1	5	4	-1	-1	0	0	0	0	0	0	1
1	2	2	6	4	-1	0	0	-1	0	0	0	0	0
1	2	3	7	4	-1	0	0	0	0	-1	0	0	0
1	2	4	8	4	-1	0	0	0	0	0	0	-1	0




```
nlogit
;lhs=choice,cset,altij
;choices=A,B,C,D
;model:
U(A) = Seat + in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 /
U(B) = Scrn + in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 /
U(C) = Alco + in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 /
U(D) =        in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 $
```

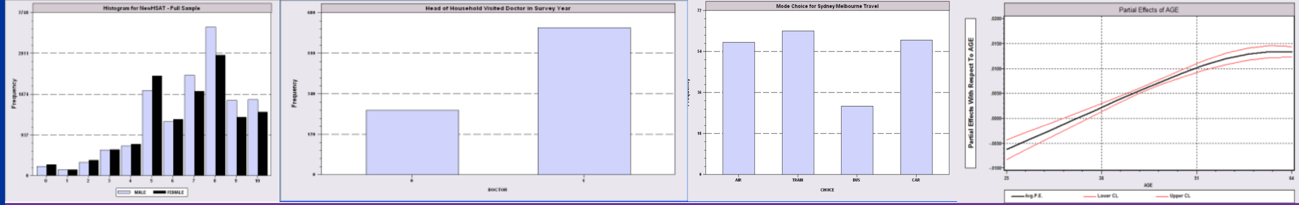
Note that you could potentially run any model for this data. For example, if one wanted to test for differences in scale between the best and worst alternatives, one could use the NL model (note however that you need the altij variable to take different values for best and worst now – see altn in the two examples above).

Uses the result that if U(i,j) is the lowest utility, -U(i,j) is the highest.



Nested Logit Approach.

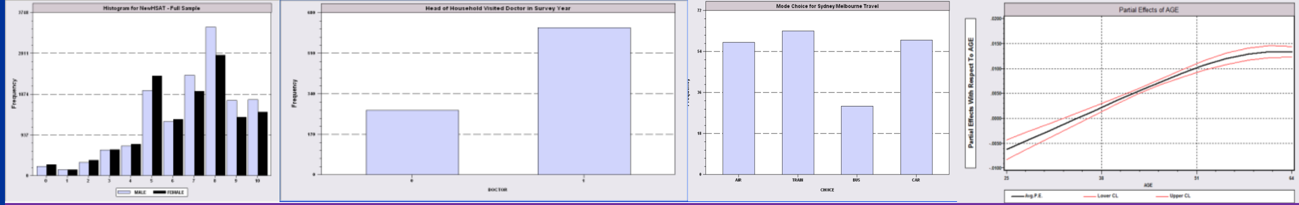
Resp	Set	Altij	Altn	Cset	Bestworst	Inch28	Inch30	CabScr	LimMov	Pay	Hour1	Hour3	Choice
1	1	1	1	4	1	0	1	0	0	0	0	0	0
1	1	2	2	4	1	0	0	0	1	0	0	0	0
1	1	3	3	4	1	0	0	0	0	0	0	0	1
1	1	4	4	4	1	0	0	0	0	0	0	0	0
1	1	1	5	4	-1	0	-1	0	0	0	0	0	0
1	1	2	6	4	-1	0	0	0	-1	0	0	0	1
1	1	3	7	4	-1	0	0	0	0	0	0	0	0
1	1	4	8	4	-1	0	0	0	0	0	0	0	0
1	2	1	1	4	1	1	0	0	0	0	0	0	0
1	2	2	2	4	1	0	0	1	0	0	0	0	0
1	2	3	3	4	1	0	0	0	0	1	0	0	1
1	2	4	4	4	1	0	0	0	0	0	0	1	0
1	2	1	5	4	-1	-1	0	0	0	0	0	0	1
1	2	2	6	4	-1	0	0	-1	0	0	0	0	0
1	2	3	7	4	-1	0	0	0	0	-1	0	0	0
1	2	4	8	4	-1	0	0	0	0	0	0	-1	0



Nested Logit Approach – Different Scaling for Worst

```
nlogit
;lhs=choice,cset,altn
;choices=Ab,Bb,Cb,Db, Aw,Bw,Cw,Dw
;tree =Bst(Ab,Bb,Cb,Db),Wst(Aw,Bw,Cw,Dw)
;ru1
;ivset:(Bst)=[1.0]
;model:
U(Ab) = Seat + in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 /
U(Bb) = Scrn + in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 /
U(Cb) = Alco + in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 /
U(Db) =        in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 /
U(Aw) = Seat + in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 /
U(Bw) = Scrn + in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 /
U(Cw) = Alco + in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 /
U(Dw) =        in28*inch28 + in30*inch30 + scr*cabscr + limmov*limmov + pay*pay + h1*hour1 + h3*hour3 $
```

8 choices are two blocks of 4. Best in one brance, worst in the second branch



Internet Dating Survey - 14 / 26

An example

We will show you 5 profiles of people that you might consider contacting.

We will then ask you to tell us which profile represents the best candidate and which represents the worst.

We will then get you to tell us of the three remaining profiles, which is the best and which is the worst.

If you were looking through a dating website and considered contacting among the five people shown based on the descriptions listed, which profile represents the best candidate and which represents the worst? And then which is the best and which is the worst of the three remaining profiles?

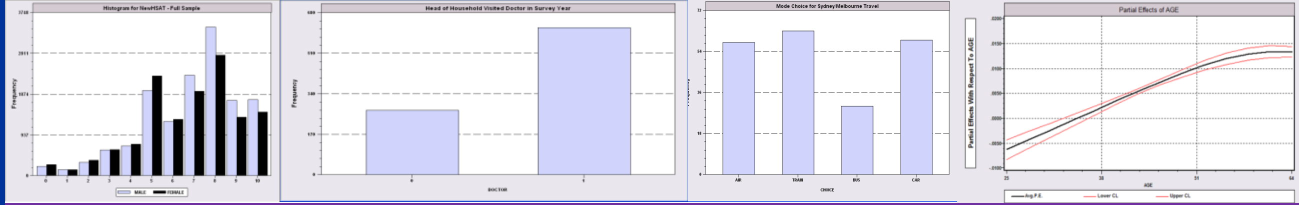
	Person A	Person B	Person C	Person D	Person E
Drinking Habit	Non drinker	Casual drinker	Moderate drinker	Casual drinker	Moderate drinker
Smoking Habit	Ex smoker	Smoker	Non smoker	Ex smoker	Smoker
Children	Single parent	None currently	Single parent	Single parent	None currently
Job	White Collar	Blue Collar	Blue Collar	Unemployed	White Collar
Looks	Above average	Below average	Above average	Below average	Average
Cost to contact	$20	$15	$10	$15	$10
Which profile do you consider to be the best and which is the worst?	Best			Worst	
Of the remaning profiles, which profile is the best and which is the worst?		Worst	Best		

In the above example, we found Person A to be the best candidate for contacting and person D the worst, and Of the three remaining profiles, we believe that Person C is the best remaining profile and Person B the worst.

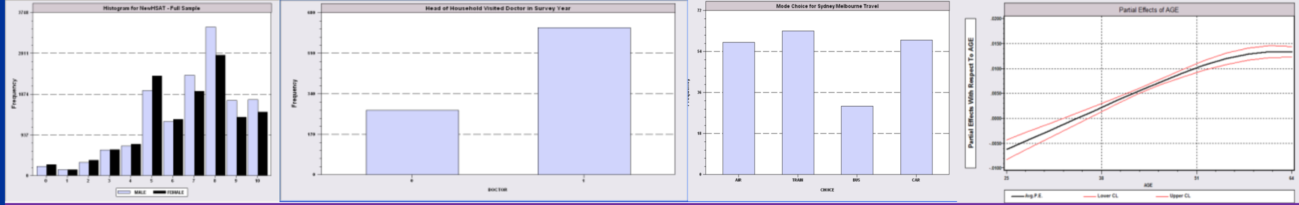
You will be shown nine scenarios similar to the above one. Each scenario will show the profiles of different potential contacts.

Please make sure that you understand the task before proceeding. Once you go to the next screen, you will not be able to go back.

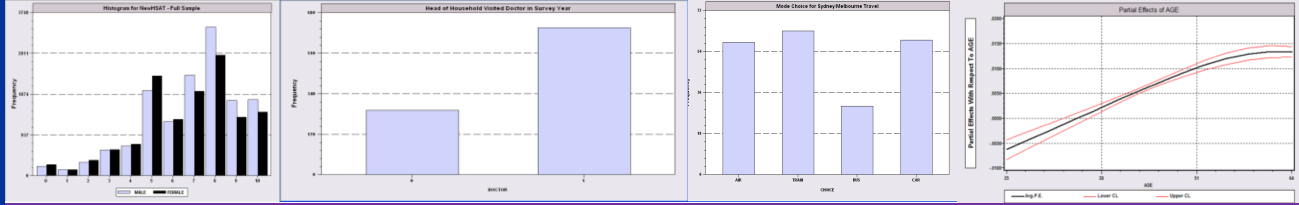
Back | Next



Resp	Set	RespSet	Explode	Altij	Altn	Cset	Choice	Drink	Smoke	Child	Job	Looks	Cost
1	1	1	1	1	1	5	1	0	1	1	0	2	20
1	1	1	1	2	2	5	0	1	2	0	1	0	15
1	1	1	1	3	3	5	0	2	0	1	1	2	10
1	1	1	1	4	4	5	0	1	1	1	2	0	15
1	1	1	1	5	5	5	0	2	2	0	0	1	10
1	2	1	2	2	7	4	0	1	2	0	1	0	15
1	2	1	2	3	8	4	1	2	0	1	1	2	10
1	2	1	2	4	9	4	0	1	1	1	2	0	15
1	2	1	2	5	10	4	0	2	2	0	0	1	10
1	3	1	3	2	12	3	0	1	2	0	1	0	15
1	3	1	3	4	14	3	0	1	1	1	2	0	15
1	3	1	3	5	15	3	1	2	2	0	0	1	10
1	4	1	4	2	17	2	1	1	2	0	1	0	15
1	4	1	4	4	19	2	0	1	1	1	2	0	15



Resp	Bestworst	Explode	Altij	Altn	Cset	Choice	Drink	Smoke	Child	Job	Looks	Cost
1	1	1	1	1	5	1	0	1	1	0	2	20
1	1	1	2	2	5	0	1	2	0	1	0	15
1	1	1	3	3	5	0	2	0	1	1	2	10
1	1	1	4	4	5	0	1	1	1	2	0	15
1	1	1	5	5	5	0	2	2	0	0	1	10
1	-1	2	1	6	5	0	0	-1	-1	0	-2	-20
1	-1	2	2	7	5	0	-1	-2	0	-1	0	-15
1	-1	2	3	8	5	0	-2	0	-1	-1	-2	-10
1	-1	2	4	9	5	1	-1	-1	-1	-2	0	-15
1	-1	2	5	10	5	0	-2	-2	0	0	-1	-10
1	1	3	2	12	4	0	1	2	0	1	0	15
1	1	3	3	13	4	1	2	0	1	1	2	10
1	1	3	4	14	4	0	1	1	1	2	0	15
1	1	3	5	15	4	0	2	2	0	0	1	10
1	-1	4	1	16	4	0	0	-1	-1	0	-2	-20
1	-1	4	2	17	4	1	-1	-2	0	-1	0	-15
1	-1	4	3	18	4	0	-2	0	-1	-1	-2	-10
1	-1	4	5	20	4	0	-2	-2	0	0	-1	-10
1	1	5	2	22	3	0	1	2	0	1	0	15
1	1	5	4	24	3	0	1	1	1	2	0	15
1	1	5	5	25	3	1	1	1	1	2	0	15
1	-1	6	1	26	3	0	0	-1	-1	0	-2	-20
1	-1	6	3	28	3	0	-2	0	-1	-1	-2	-10
1	-1	6	5	30	3	1	-2	-2	0	0	-1	-10
1	1	7	2	32	2	1	1	2	0	1	0	15
1	1	7	4	34	2	0	1	1	1	2	0	15
1	-1	8	1	35	2	0	0	-1	-1	0	-2	-20
1	-1	8	3	37	2	1	-2	0	-1	-1	-2	-10



Nonlinear Utility Functions

Generalized (in functional form) multinomial logit model

$$U(i,j) = V_j(x_{ij}, z_i, \beta) + \varepsilon_{ij}$$ (Utility function may vary by choice.)

$F(\varepsilon_{ij}) = \exp(-\exp(-(\varepsilon_{ij}))$ - the standard IID assumptions for MNL

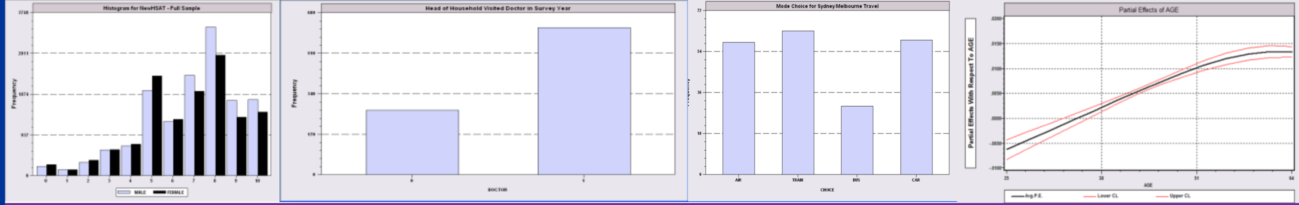
$$\text{Prob}(i,j) = \frac{\exp\left[V_j(x_{ij}, z_i, \beta)\right]}{\sum_{m=1}^{J} \exp\left[V_m(x_{im}, z_i, \beta)\right]}$$

Estimation problem is more complicated in practical terms

Large increase in model flexibility.

Note: Coefficients are no longer generic.

$$\text{WTP}(i,k \mid j) = -\frac{\partial V_j(x_{ij}, z_i, \beta) / \partial x_{i,j}(k)}{\partial V_j(x_{ij}, z_i, \beta) / \partial \text{Cost}}$$

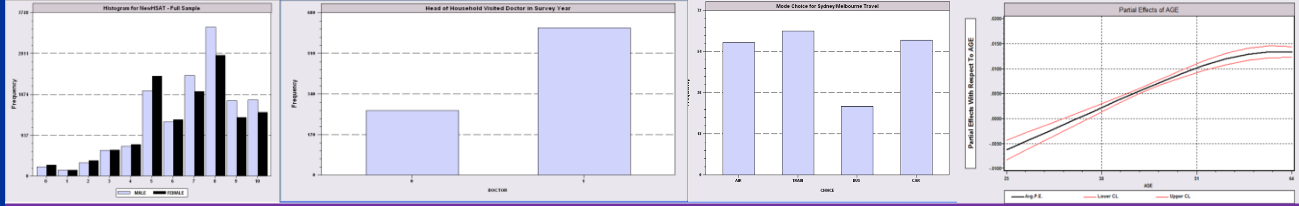


Assessing Prospect Theoretic Functional Forms and Risk in a Nonlinear Logit Framework: Valuing Reliability Embedded Travel Time Savings

David Hensher
The University of Sydney, ITLS
William Greene
Stern School of Business, New York University

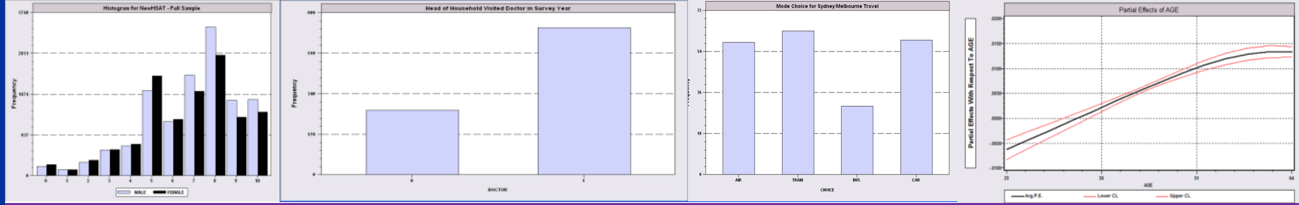
8th Annual Advances in Econometrics Conference
Louisiana State University
Baton Rouge, LA
November 6-8, 2009

Hensher, D., Greene, W., "Embedding Risk Attitude and Decisions Weights in Non-linear Logit to Accommodate Time Variability in the Value of Expected Travel Time Savings," *Transportation Research Part B*



Prospect Theory

- Marginal value function for an attribute (outcome) $v(x_m)$ = subjective value of attribute
- Decision weight $w(p_m)$ = impact of a probability on utility of a prospect
- Value function $V(x_m,p_m) = v(x_m)w(p_m)$ = value of a prospect that delivers outcome x_m with probability p_m
- We explore functional forms for $w(p_m)$ with implications for decisions

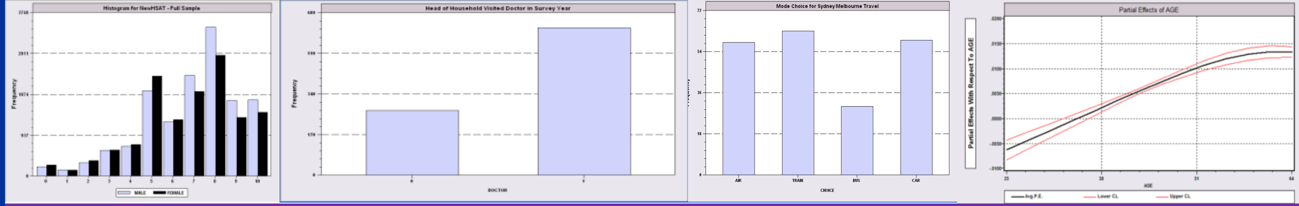


An Application of Valuing Reliability (due to Ken Small)

PLEASE CIRCLE EITHER CHOICE **A** OR CHOICE **B**

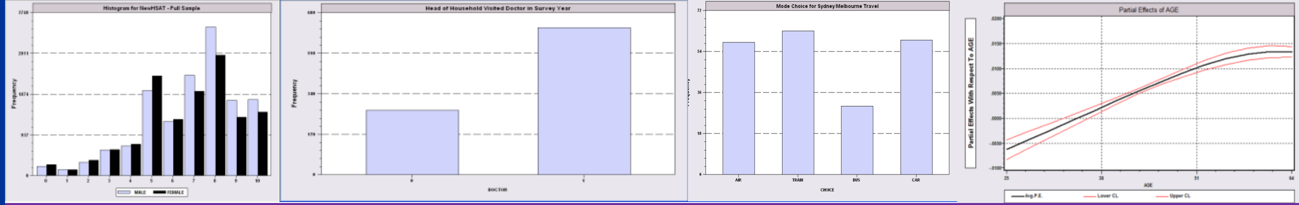
Choice A	Choice B
Average Travel Time **9 minutes**	Average Travel Time **9 minutes**
You have an equal chance of arriving at any of the following times:	You have an equal chance of arriving at any of the following times:
7 minutes early	**3 minutes early**
4 minutes early	**3 minutes ~~early~~ late**
1 minute early	**2 minute early**
5 minutes late	**2 minutes ~~early~~ late**
9 minutes late	**On time**
Your cost: **$0.25**	Your cost: **$1.50**

Choice A **Choice B**



Stated Choice Survey

- **Trip Attributes in Stated Choice Design**
  - Routes A and B
  - Free flow travel time
  - Slowed down travel time
  - Stop/start/crawling travel time
  - Minutes arriving earlier than expected
  - Minutes arriving later than expected
  - Probability of arriving earlier than expected
  - Probability of arriving at the time expected
  - Probability of arriving later than expected
  - Running cost
  - Toll Cost
- **Individual Characteristics: Age, Income, Gender**



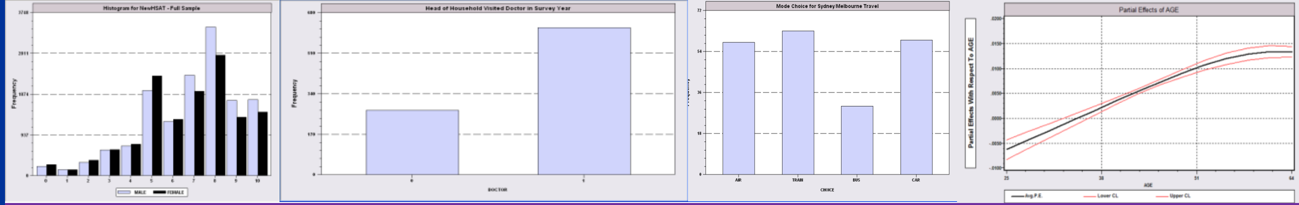
Value and Weighting Functions

Value Function: $V(x) = \dfrac{x^{1-\alpha}}{1-\alpha}$

Weighting Functions:

$$\text{Model 1} = \frac{p_m^{\gamma}}{[p_m^{\gamma} + (1-p_m)^{\gamma}]^{\frac{1}{\gamma}}} \qquad \text{Model 2} = \frac{\tau P_m^{\gamma}}{[\tau P_m^{\gamma} + (1-p_m)^{\gamma}]}$$

$$\text{Model 3} = \exp(-\tau(-\ln p_m)^{\gamma}) \qquad \text{Model 4} = \exp(-(-\ln p_m)^{\gamma})$$

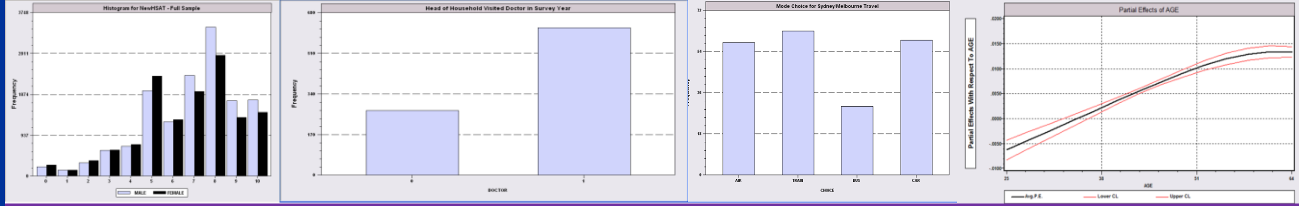


Choice Model

$$
\begin{aligned}
U(j) = \; & \beta_{ref} + \beta_{cost}Cost + \beta_{Age}Age + \beta_{Toll}TollASC \\
& + \beta_{curr} \quad w(p_{curr})v(t_{curr}) \\
& + \beta_{late} \quad w(p_{late})\, v(t_{late}) \\
& + \beta_{early} \quad w(p_{early})v(t_{early}) \; + \; \varepsilon_j
\end{aligned}
$$

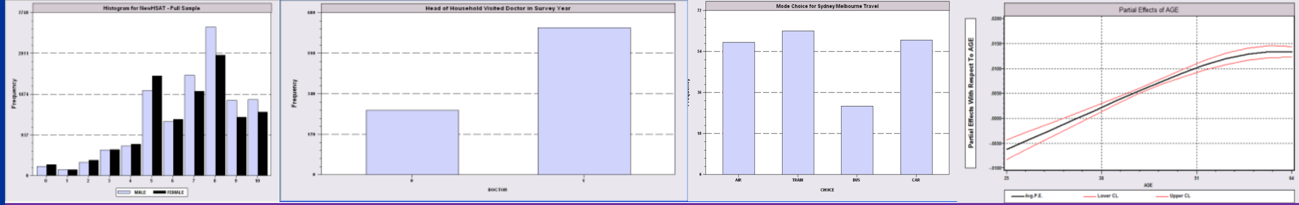
Constraint: $\beta_{curr} = \beta_{late} = \beta_{early}$

$$
\begin{aligned}
U(j) = \; & \beta_{ref} + \beta_{cost}Cost + \beta_{Age}Age + \beta_{Toll}TollASC \\
& + \beta[w(p_{curr})v(t_{curr}) + w(p_{late})v(t_{late}) + w(p_{early})v(t_{early})] \\
& + \; \varepsilon_j
\end{aligned}
$$



Application

- 2008 study undertaken in Australia
  - toll vs. free roads
  - stated choice (*SC*) experiment involving two *SC* alternatives (i.e., route A and route B) pivoted around the knowledge base of travellers (i.e., the current trip).
- 280 Individuals
- 32 Choice Situations (2 blocks of 16)



Data

Descriptive socioeconomic statistics

Purpose	Statistic	Gender (1=female)	Income	Age
Commuter	Mean	0.575	$67,145	42.52
	Std. Deviation	0.495	$36,493	14.25

Descriptive statistics for costs by segment

	All times of day		Peak		Off-Peak	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Running costs	$3.15	$2.56	$3.58	$3.01	$2.92	$2.26
Toll costs	$1.41	$1.50	$1.40	$1.50	$1.41	$1.51

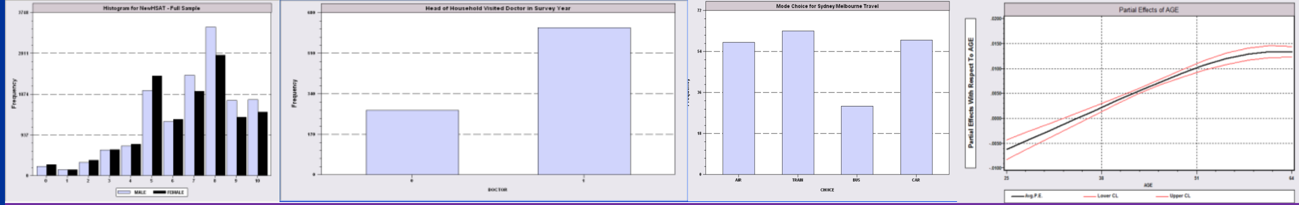
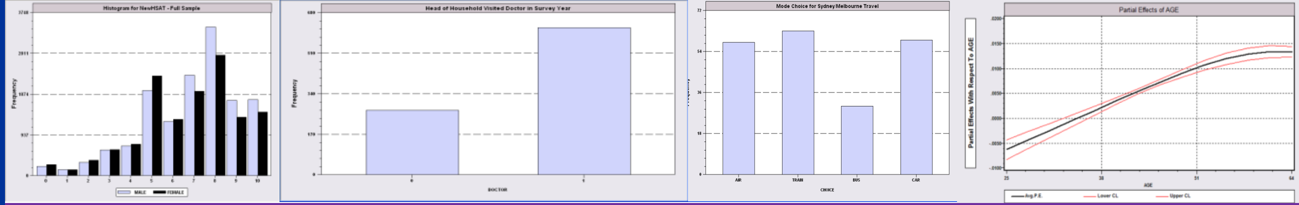
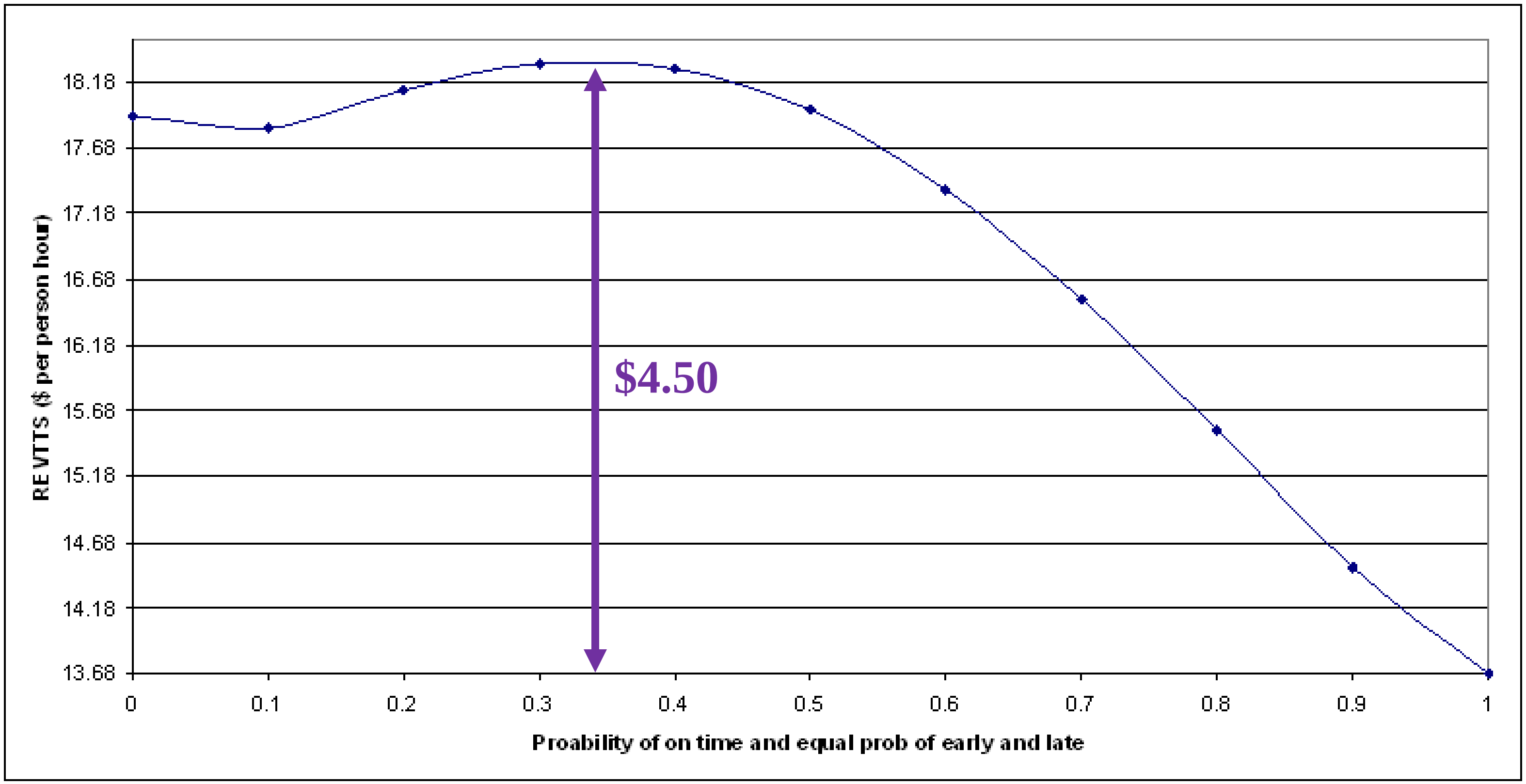


Table 7: Non-linear probability weighting function with uncertainty attitude
(All models are estimated using Nlogit5)

Variable	Model 1 (M1)	Model 2 (M2)	Model 3 (M3)	Model 4 (M4)
Decision Weight	$w(p_m)=\frac{p_m^\gamma}{[p_m^\gamma+(1-p_m)^\gamma]^{\frac{1}{\gamma}}}$	$w(p_m)=\frac{\tau p_m^\gamma}{[\tau p_m^\gamma+(1-p_m)^\gamma]}$	$w(p_m)=\exp(-\tau(-\ln p_m)^\gamma)$	$w(p_m)=\exp(-(-\ln p_m)^\gamma)$
Reference constant	0.5017 (4.12)	0.5318 (4.32)	0.5311 (4.33)	0.4933 (4.05)
Alpha (α)	0.3834 (3.41)	0.2670 (2.21)	0.2729 (2.26)	0.2288 (1.85)
Gamma (γ)	0.7641 (3.31)	1.2549 (6.43)	1.4185 (7.79)	1.1638 (5.82)
On-time/Early/Late (mins)	-0.2966 (-2.43)	-0.1532 (-2.1)	-0.1620 (-2.07)	-0.1742 (-2.16)
Cost ($)	-0.2612 (-12.2)	-0.2602 (-12.2)	-0.2607 (-12.2)	-0.2609 (-12.2)
Tollasc	-0.2815 (-3.02)	-0.2727 (-2.88)	-0.2711 (-2.87)	-0.3022 (-3.21)
Tau (τ)	-	1.9487 (7.00)	0.7304 (9.56)	-
Age (years)	0.0054 (2.11)	0.0053 (2.12)	0.0054 (2.13)	0.0052 (2.05)
No. of observations		4,480		
Information Criterion : AIC	6850.86	6829.54	6829.91	6864.23
Log-likelihood	-3418.43	-3406.77	-3406.96	-3425.11
REVTTS	**18.15 (3.21)**	**17.63 (2.28)**	**17.60 (2.31)**	**17.47 (1.80)**



Reliability Embedded Value of Travel Time Savings in Au$/hr



REVTTS Distribution given Probability of being on-time (Model 2) with equal residual probability of being early and late