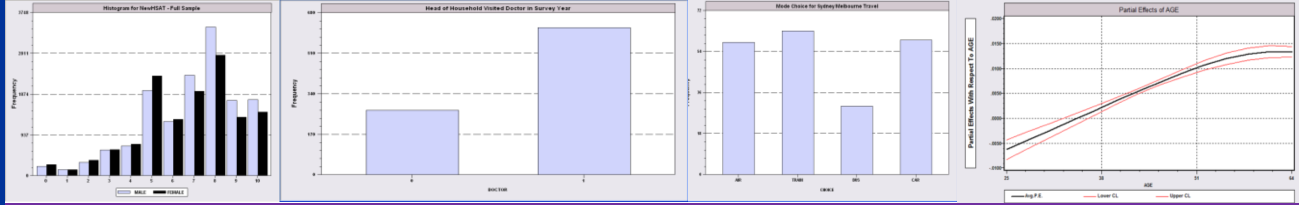
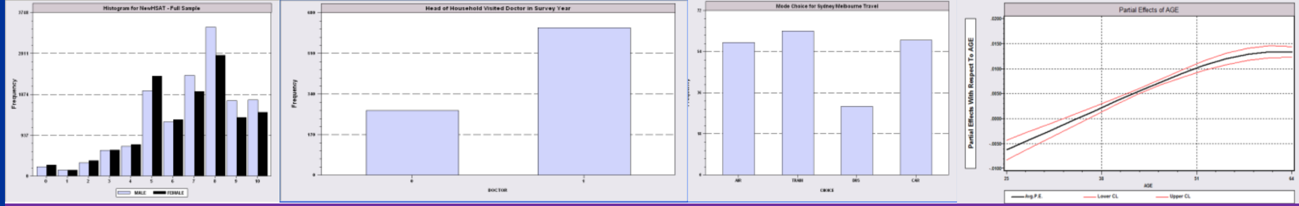


Discrete Choice Modeling

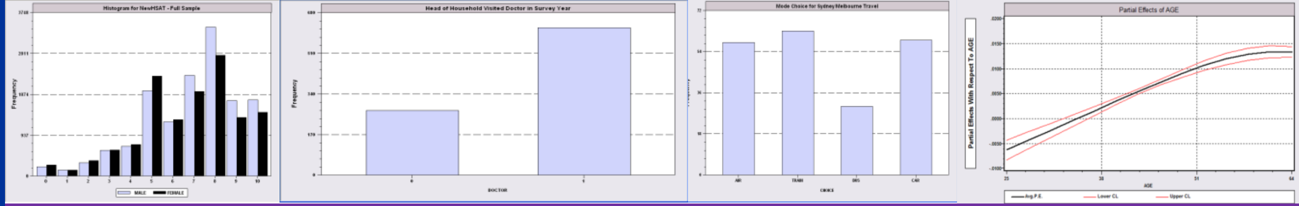
William Greene
Stern School of Business
New York University



What’s Wrong with the MNL Model?

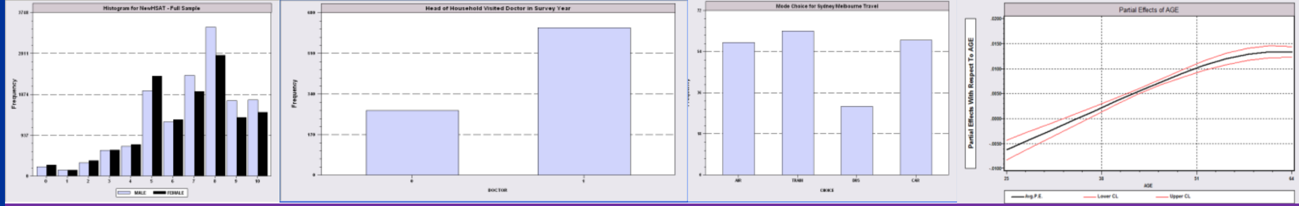
Insufficiently heterogeneous:

“… economists are often more interested in aggregate effects and regard heterogeneity as a statistical nuisance parameter problem which must be addressed but not emphasized. Econometricians frequently employ methods which do not allow for the estimation of individual level parameters.” (Allenby and Rossi, Journal of Econometrics, 1999)



Several Types of Heterogeneity

- **Differences across choice makers**
  - Observable: Usually demographics such as age, sex
  - Unobservable: Usually modeled as ‘random effects’
- **Choice strategy:** How consumers make decisions. (E.g., omitted attributes)
- **Preference Structure:** Model frameworks such as latent class structures
- **Preferences:** Model ‘parameters’
  - Discrete variation – latent class
  - Continuous variation – mixed models
  - Discrete-Continuous variation

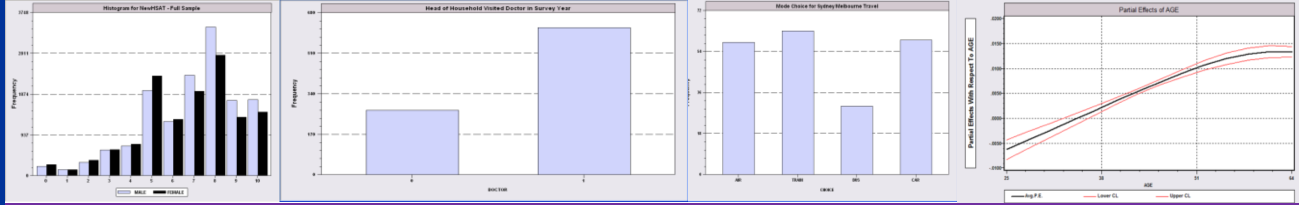


Heterogeneity in Choice Strategy

Consumers avoid ‘complexity’

- Lexicographic preferences eliminate certain choices ➔ choice set may be endogenously determined
- Simplification strategies may eliminate certain attributes

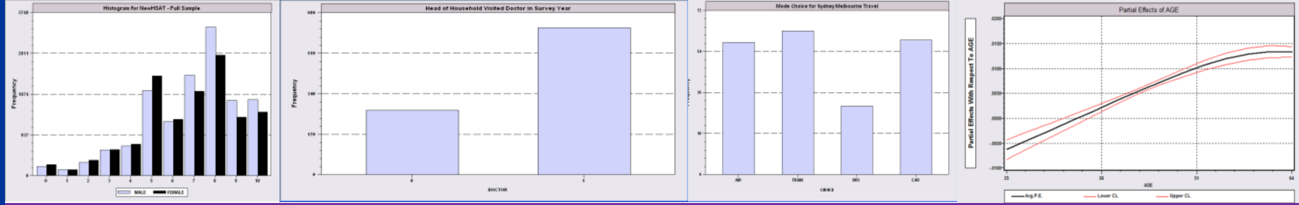
Information processing strategy is a source of heterogeneity in the model.



Accommodating Heterogeneity

Observed? Enter in the model in familiar (and unfamiliar) ways.

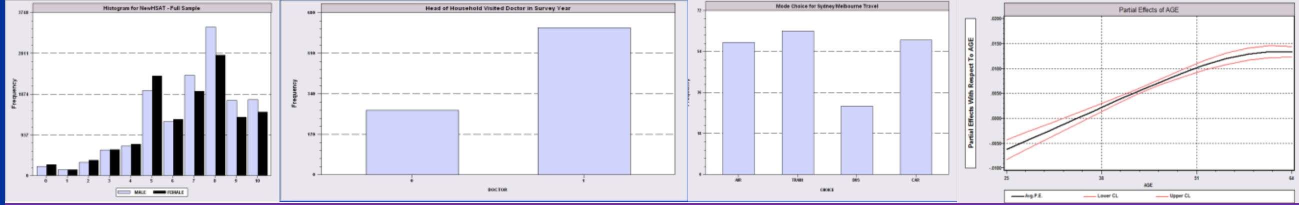
Unobserved? Takes the form of randomness in the model.



Heterogeneity and the MNL Model

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

- Limitations of the MNL Model:
  - IID → IIA
  - Fundamental tastes are the same across all individuals
- How to adjust the model to allow variation across individuals?
  - Full random variation
  - Latent grouping – allow <u>*some*</u> variation



Observable Heterogeneity in Utility Levels

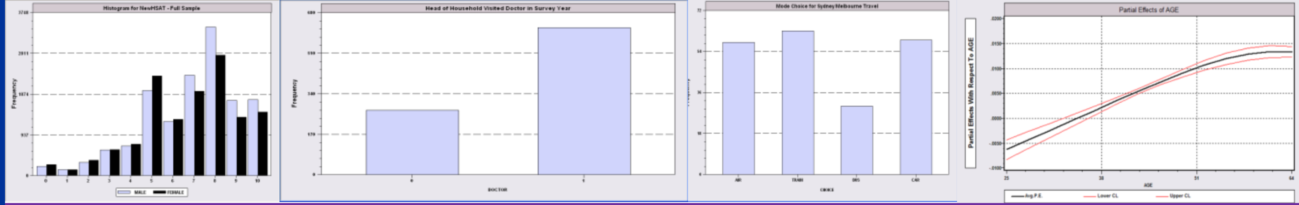
$$U_{ijt} = \alpha_j + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it} + \varepsilon_{ijt}$$

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

Choice, e.g., among brands of cars

x_{itj} = attributes: price, features

z_{it} = observable characteristics: age, sex, income



Observable Heterogeneity in Preference Weights

Hierarchical model - Interaction terms

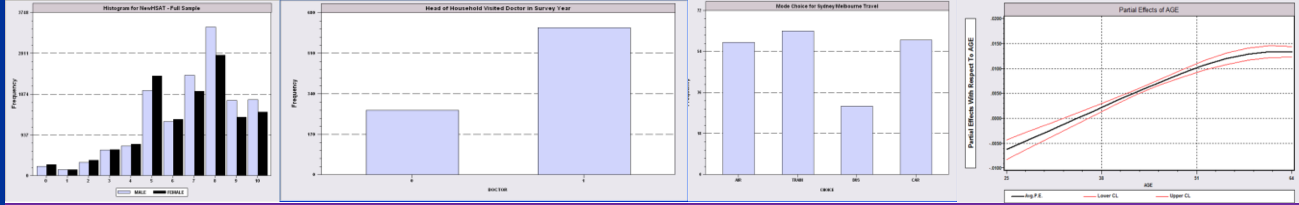
$$U_{ijt} = \alpha_j + \boldsymbol{\beta}_i'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it} + \varepsilon_{ijt}$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Phi}\mathbf{h}_i$$

Parameter heterogeneity is observable.

Each parameter $\beta_{i,k} = \beta_k + \boldsymbol{\varphi}_k'\mathbf{h}_i$

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



Heteroscedasticity in the MNL Model

- Motivation: Scaling in utility functions
- If ignored, distorts coefficients
- Random utility basis

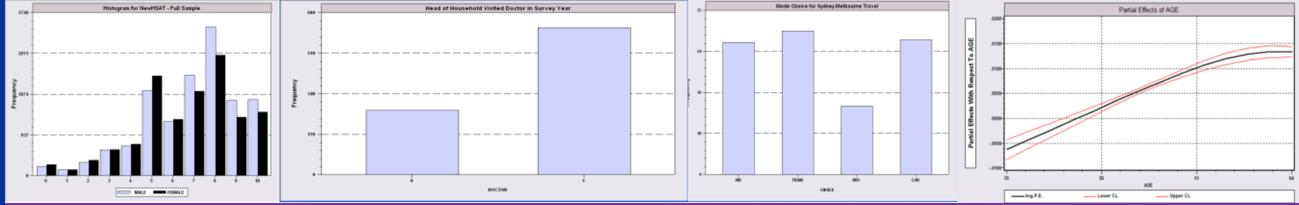
$$U_{ij} = \alpha_j + \boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}'\mathbf{z}_i + \sigma_j\varepsilon_{ij}$$

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

$F(\varepsilon_{ij}) = \text{Exp}(-\text{Exp}(-\varepsilon_{ij}))$ now scaled

- Extensions: Relaxes IIA

  Allows heteroscedasticity across choices and across individuals

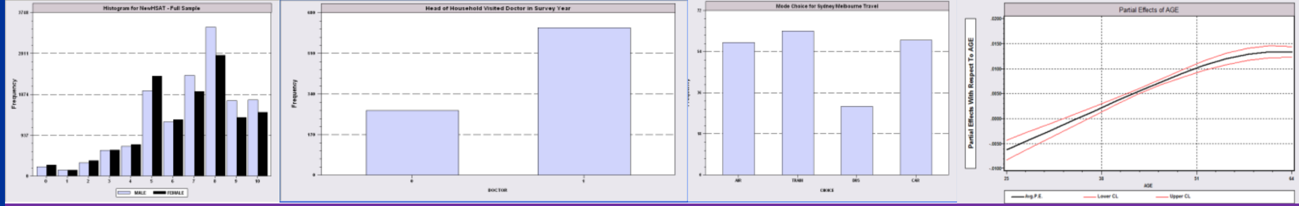


‘Quantifiable’ Heterogeneity in Scaling

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

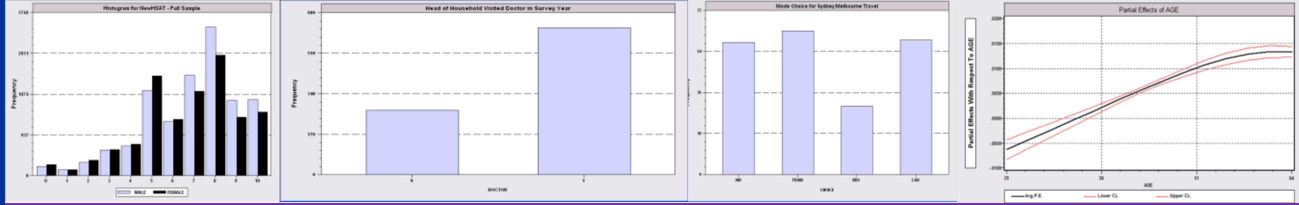
$$\text{Var}[\varepsilon_{ijt}] = \sigma_j^2 \exp(\boldsymbol{\delta}_j'\mathbf{w}_i), \sigma_1^2 = \pi^2/6$$

w_i = observable characteristics: age, sex, income, etc.

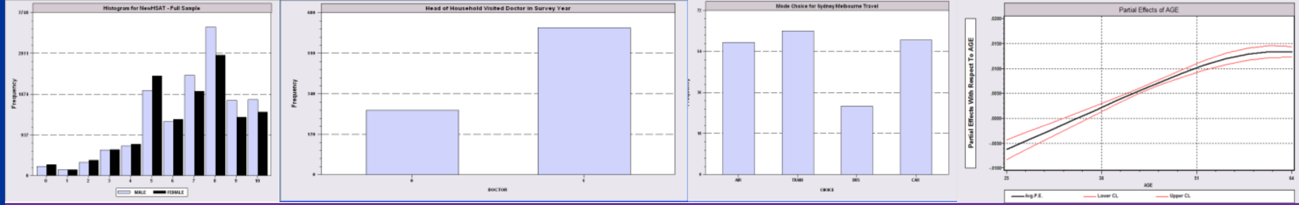


Modeling Unobserved Heterogeneity

- Latent class – Discrete approximation
- Mixed logit – Continuous
- Many extensions and blends of LC and RP

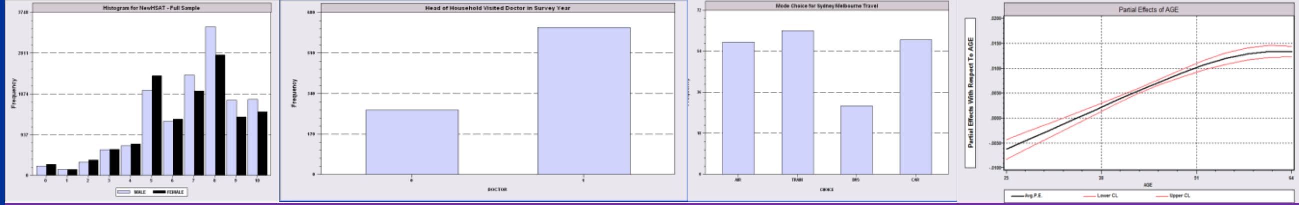


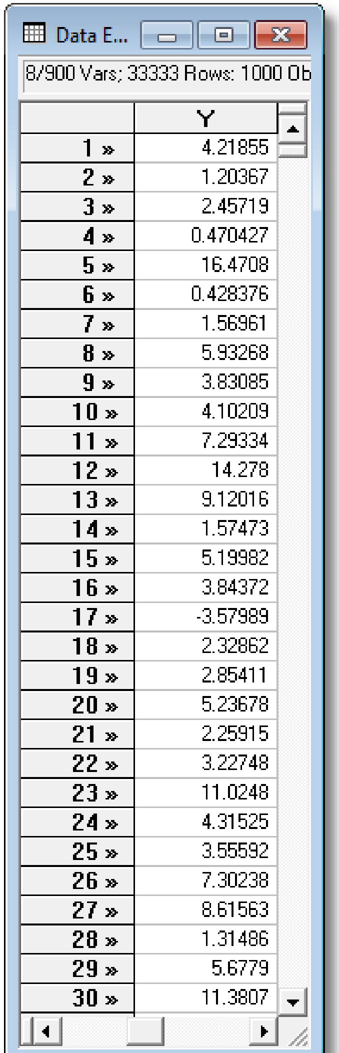
LATENT CLASS MODELS



The “Finite Mixture Model”

- An unknown parametric model governs an outcome y
  - F(y|**x**,θ)
  - This is *the* model
- We approximate F(y|**x**,θ) with a weighted sum of specified (e.g., normal) densities:
  - $F(y|\mathbf{x},\theta) \cong \Sigma_j \pi_j G(y|\mathbf{x},\beta)$
  - This is a search for functional form. With a sufficient number of (normal) components, we can approximate any density to any desired degree of accuracy. (McLachlan and Peel (2000))
  - There is no “mixing” ***process*** at work

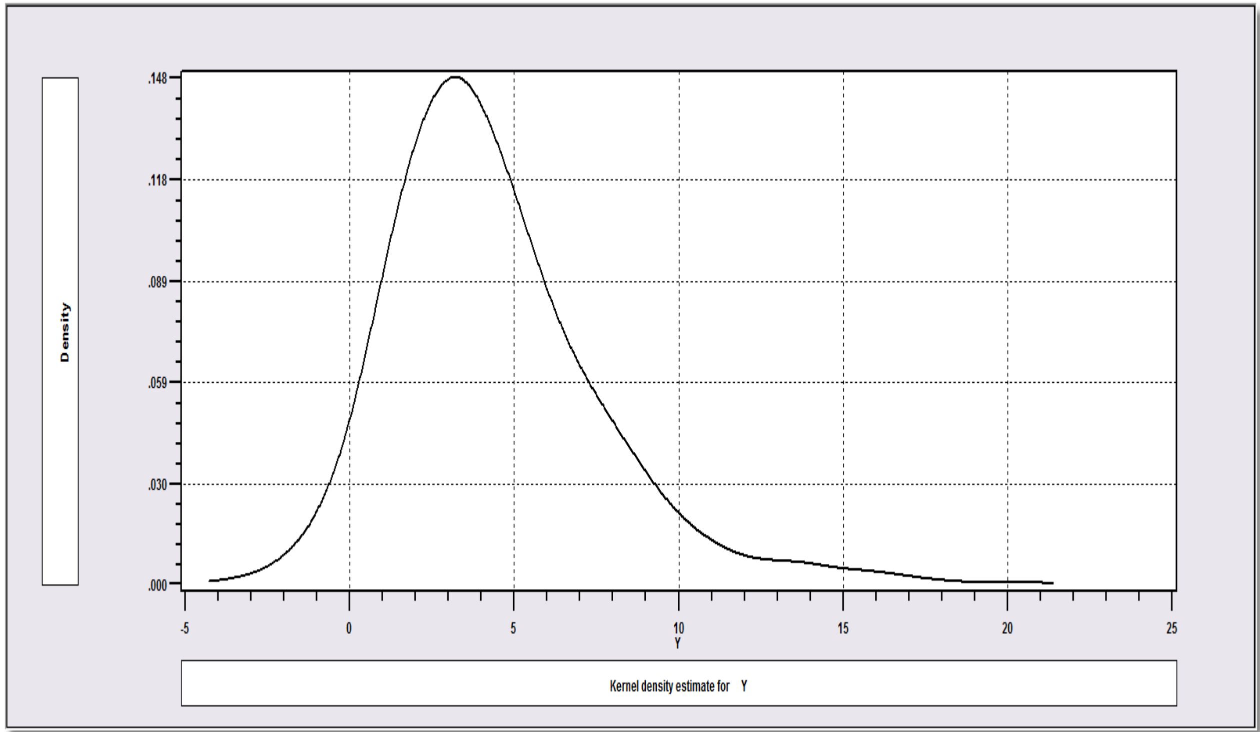




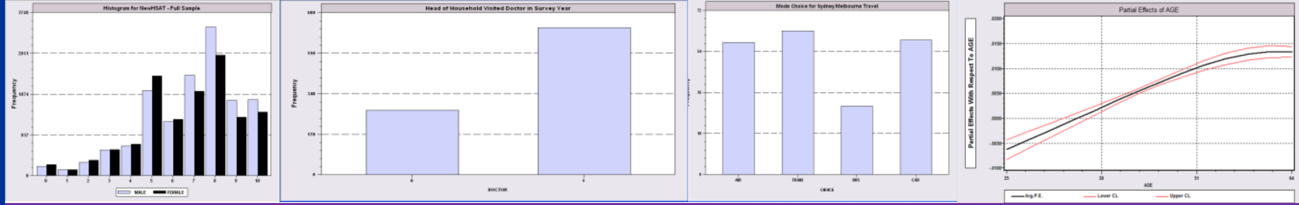
Data E...

8/900 Vars; 33333 Rows: 1000 Ob

	Y
1 »	4.21855
2 »	1.20367
3 »	2.45719
4 »	0.470427
5 »	16.4708
6 »	0.428376
7 »	1.56961
8 »	5.93268
9 »	3.83085
10 »	4.10209
11 »	7.29334
12 »	14.278
13 »	9.12016
14 »	1.57473
15 »	5.19982
16 »	3.84372
17 »	-3.57989
18 »	2.32862
19 »	2.85411
20 »	5.23678
21 »	2.25915
22 »	3.22748
23 »	11.0248
24 »	4.31525
25 »	3.55592
26 »	7.30238
27 »	8.61563
28 »	1.31486
29 »	5.6779
30 »	11.3807



Density? Note significant mass below zero. Not a gamma or lognormal or any other familiar density.



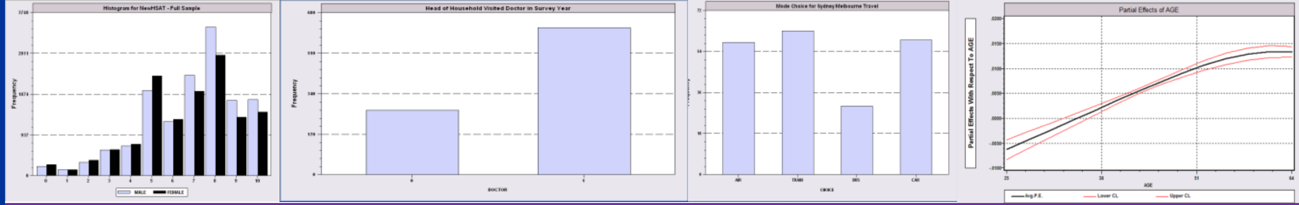
ML Mixture of Two Normal Densities

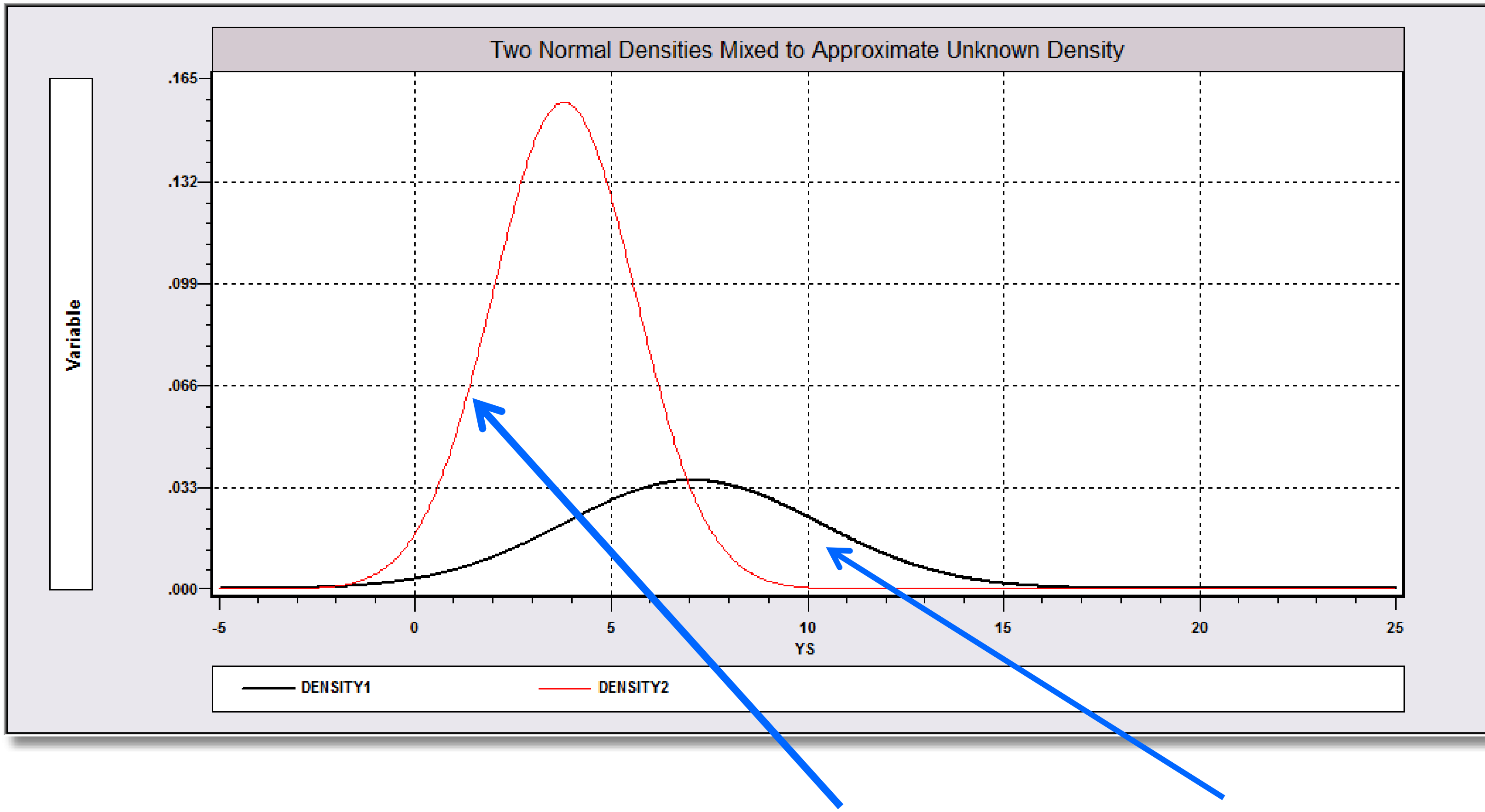
$$\text{LogL} = \sum_{i=1}^{1000} \log\left(\sum_{j=1}^{2} \pi_j \frac{1}{\sigma_j} \phi\left(\frac{y_i - \mu_j}{\sigma_j} \right) \right)$$

Maximum Likelihood Estimates

	Class 1		Class 2	
	Estimate	Std. Error	Estimate	Std. error
μ	7.05737	.77151	3.25966	.09824
σ	3.79628	.25395	1.81941	.10858
π	.28547	.05953	.71453	.05953

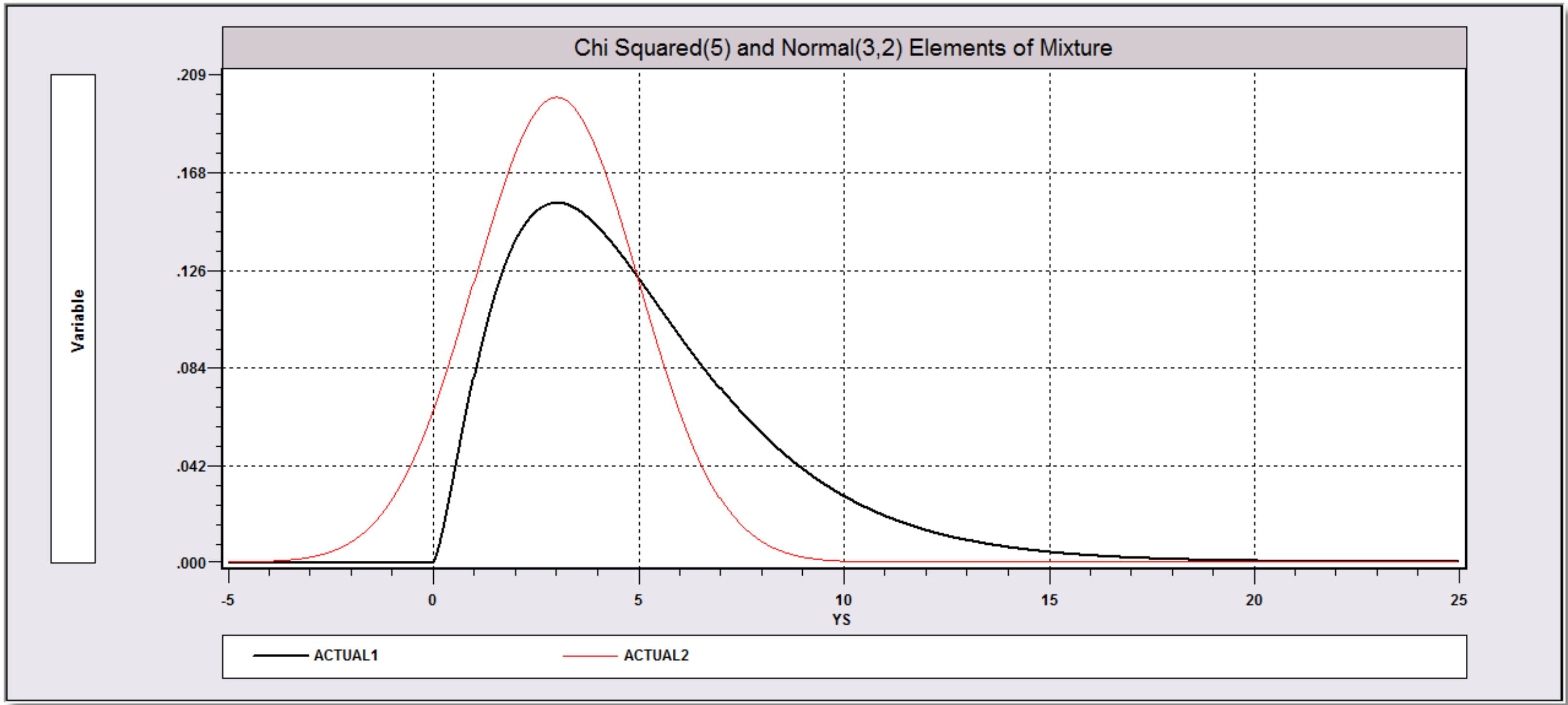
$$\hat{F}(y) = .28547\left[\frac{1}{3.79628} \phi\left(\frac{y - 7.05737}{3.79628} \right) \right] + .71453\left[\frac{1}{1.81941} \phi\left(\frac{y - 3.25966}{1.81941} \right) \right]$$



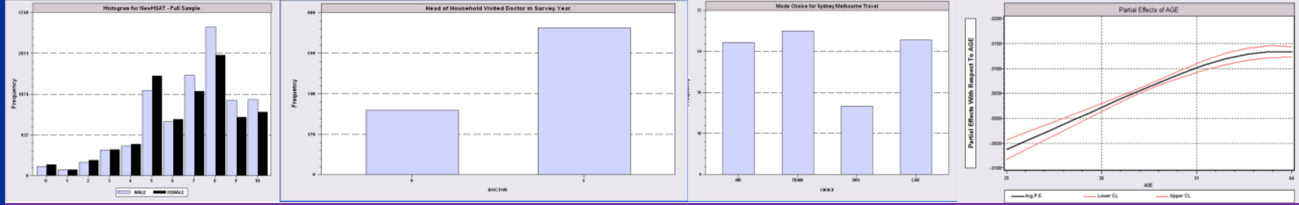


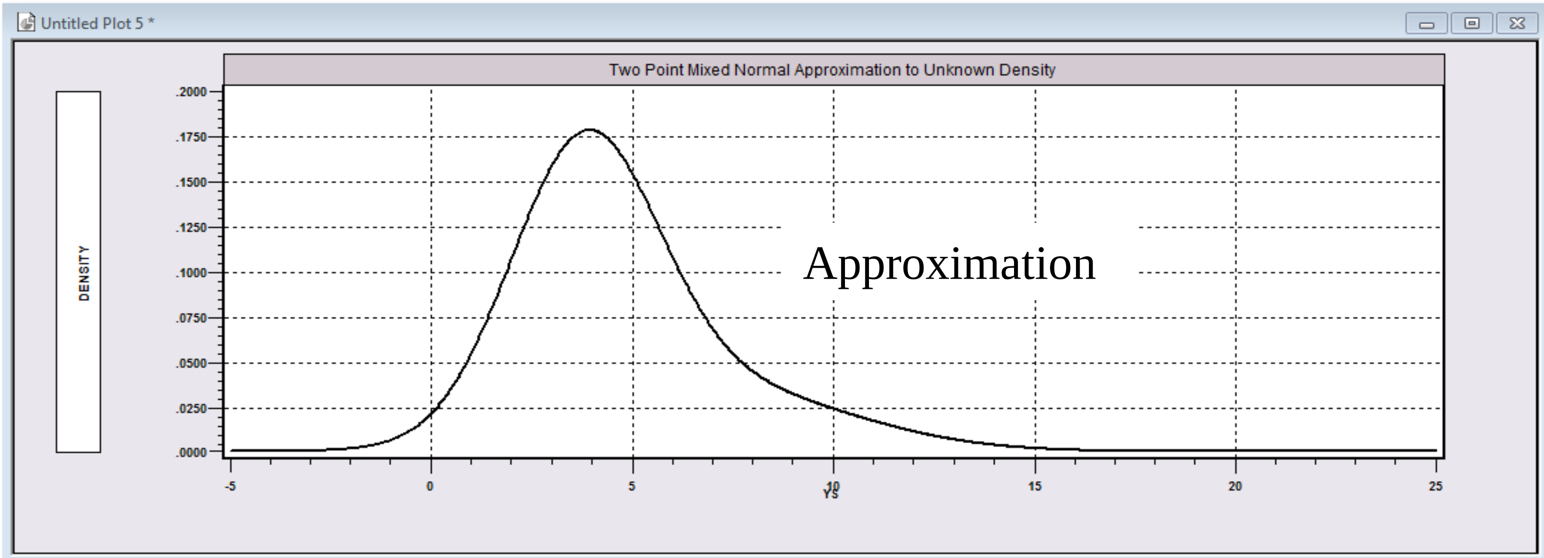
Mixing probabilities .715 and .285

The actual process is a mix of chi squared(5) and normal(3,2) with mixing probabilities .7 and .3.

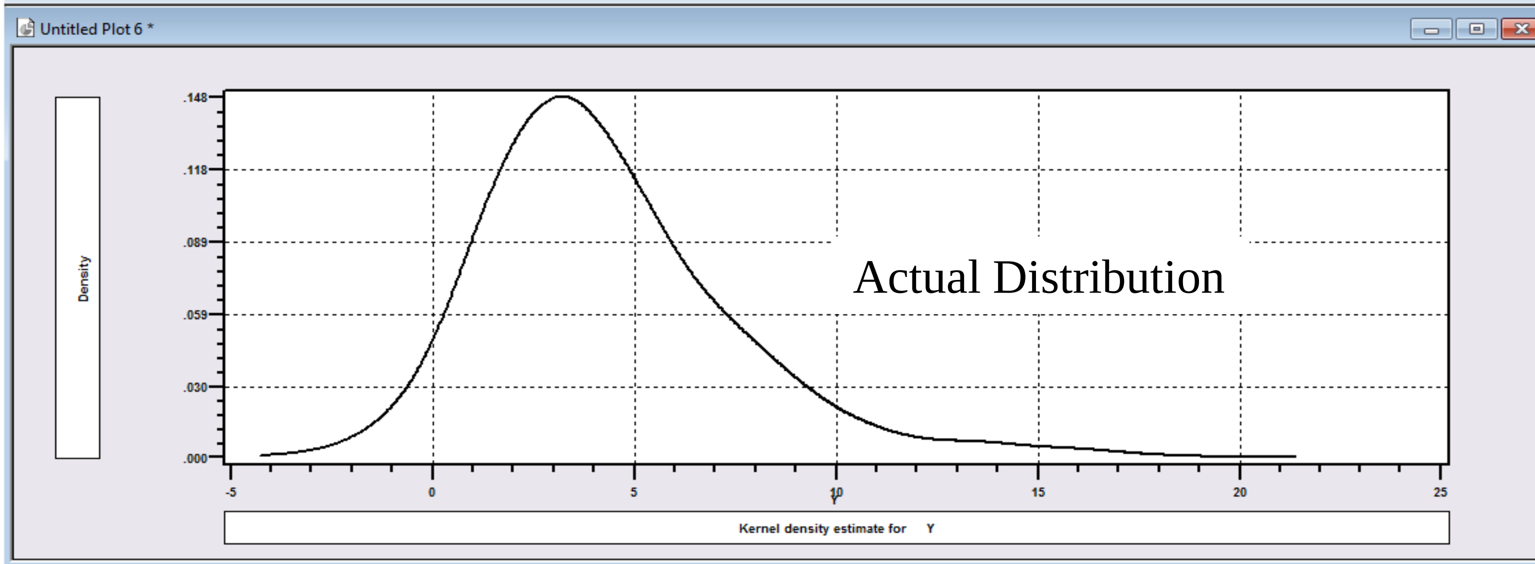


$$f(y) = .7\frac{.5^{2.5}\exp(-.5y)y^{1.5}}{\Gamma(2.5)} + .3\frac{1}{2}\phi\left(\frac{y-3}{2}\right)$$

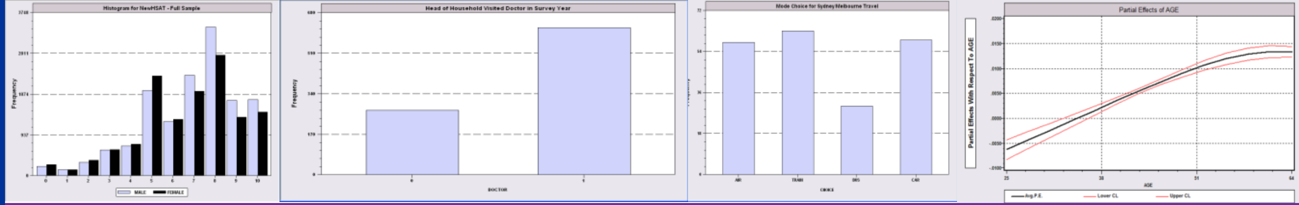




Untitled Plot 5 *
Two Point Mixed Normal Approximation to Unknown Density
DENSITY
.2000
.1750
.1500
.1250
.1000
.0750
.0500
.0250
.0000
-5
0
5
10
15
20
25
YS
Approximation

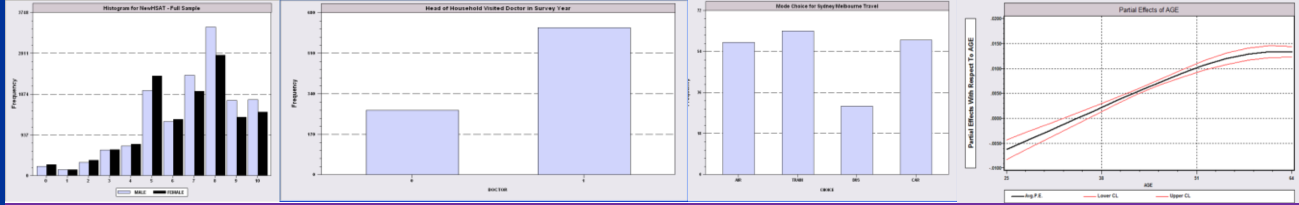


Untitled Plot 6 *
Density
.148
.118
.089
.059
.030
.000
-5
0
5
10
15
20
25
Y
Kernel density estimate for Y
Actual Distribution



Latent Classes

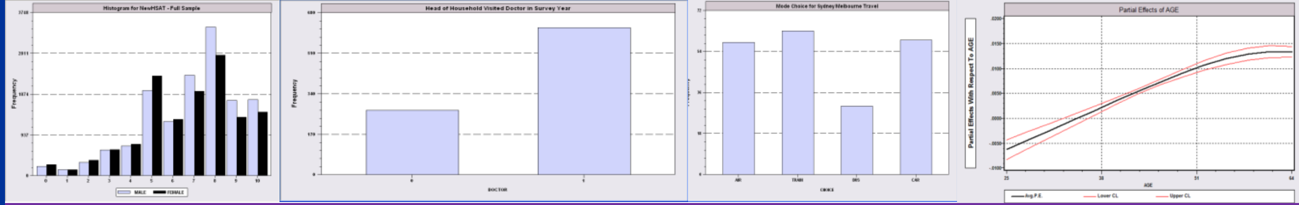
- Population contains a mixture of individuals of different types
- Common form of the generating mechanism within the classes
- Observed outcome y is governed by the **common process** F(y|**x**,θ_j)
- Classes are distinguished by the parameters, θ_j.



Published Online April 12 2007
Science 11 May 2007:
Vol. 316 no. 5826 pp. 889-894
DOI: 10.1126/science.1141634

REPORT

A Common Variant in the *FTO* Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity



The Latent Class “Model”

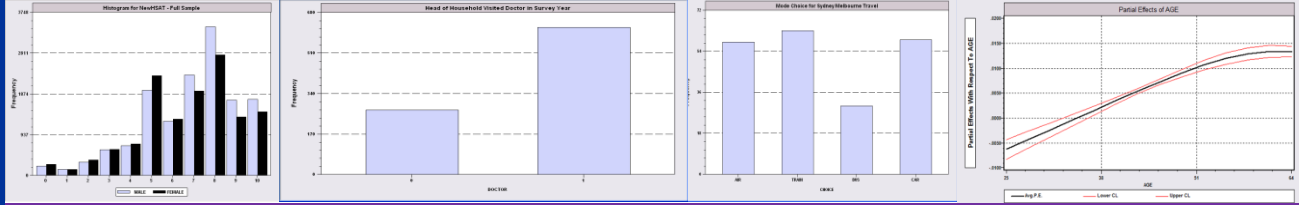
- Parametric Model:
  - $F(y|\mathbf{x},\theta)$
  - E.g., $y \sim N[\mathbf{x}'\beta, \sigma^2]$, $y \sim \text{Poisson}[\lambda=\exp(\mathbf{x}'\beta)]$, etc.

$$\text{Density } F(y|\mathbf{x},\boldsymbol{\Theta}) \cong \Sigma_j \pi_j F(y|\mathbf{x},\theta_j),$$

$$\boldsymbol{\Theta} = [\theta_1, \theta_2, \ldots, \theta_J, \pi_1, \pi_2, \ldots, \pi_J]$$

$$\Sigma_j \pi_j = 1$$

  - Generating mechanism for an individual drawn at random from the mixed population is $F(y|\mathbf{x},\boldsymbol{\Theta})$.
  - Class probabilities relate to a stable process governing the mixture of types in the population



Journal of Health Economics
Volume 28, Issue 2, March 2009, Pages 265-279

doi:10.1016/j.jhealeco.2008.11.002 | How to Cite or Link Using DOI
Permissions & Reprints

Health care utilisation in Europe: New evidence from the ECHP

Teresa Bago d'Uva[a, b], and **Andrew M. Jones**[c]

[a]Department of Applied Economics (Room H13-09), Erasmus School of Economics, PB 1738, 3000 DR Rotterdam, The Netherlands

[b]Netspar, The Netherlands

[c]Department of Economics and Related Studies, University of York, York, YO10 5DD, United Kingdom

Received 9 July 2007; revised 30 October 2008; accepted 7 November 2008. Available online 19 November 2008.

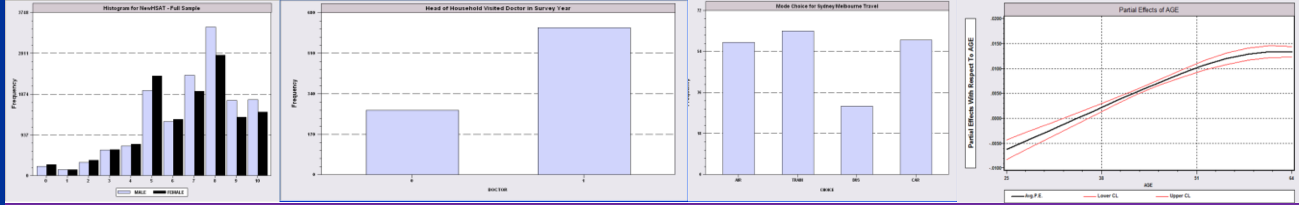
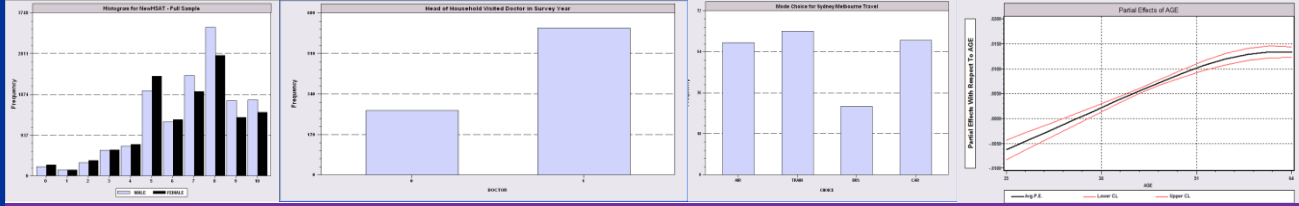


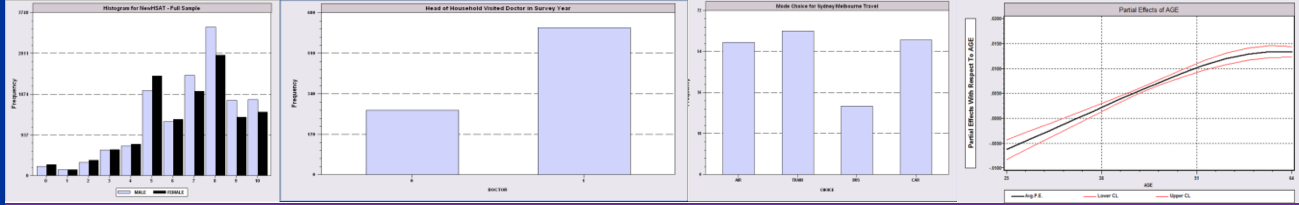
Table 8. Estimated income coefficients and elasticities for GP and specialist visits—country-specific LC hurdle models.

Country		**GPs**				**Specialists**			
		Low users		**High users**		**Low users**		**High users**	
		Estimated coefficient	**Estimated elasticity**	**Estimated coefficient**	**Estimated elasticity**	**Estimated coefficient**	**Estimated elasticity**	**Estimated coefficient**	**Estimated elasticity**
Austria	*P*(*Y*>0)	−0.051 (−1.467)	−0.012	−0.109 (−0.872)	−0.005	**0.191** (3.743)	**0.110**	**0.211** (3.556)	**0.030**
	E(*Y*|*Y*>0)	0.012 (0.693)	0.009	**0.039** (2.167)	**0.035**	0.014 (0.210)	0.006	**0.105** (3.858)	**0.070**
Belgium	*P*(*Y*>0)	0.035 (1.002)	0.008	**0.292** (4.004)	**0.010**	0.054 (1.399)	0.036	0.079 (1.348)	0.014
	E(*Y*|*Y*>0)	**−0.052** (−3.125)	**−0.037**	**−0.055** (−4.030)	**−0.050**	−0.112 (−1.611)	−0.052	−0.049 (−1.920)	−0.035
Denmark	*P*(*Y*>0)	0.083 (1.746)	0.033	**0.261** (2.302)	**0.023**	0.053 (0.738)	0.045	0.079 (1.123)	0.034
	E(*Y*|*Y*>0)	0.042 (0.992)	0.021	−0.030 (−1.009)	−0.024	−0.053 (−0.434)	−0.022	−0.082 (−1.120)	−0.050
Finland	*P*(*Y*>0)	0.054 (1.358)	0.024	−0.030 (−0.263)	−0.003	**0.203** (3.525)	**0.155**	0.167 (1.909)	0.041
	E(*Y*|*Y*>0)	0.007 (0.237)	0.004	−0.048 (−1.706)	−0.037	**−0.229** (−2.985)	**−0.090**	0.025 (0.487)	0.014
Greece	*P*(*Y*>0)	0.012 (0.565)	0.006	0.015 (0.447)	0.004	**0.184** (7.641)	**0.128**	**0.148** (5.413)	**0.060**
	E(*Y*|*Y*>0)	−0.024 (−1.864)	−0.015	**0.026** (1.967)	**0.020**	0.017 (0.878)	0.010	**0.067** (4.192)	**0.055**
Ireland	*P*(*Y*>0)	**0.164** (4.754)	**0.064**	0.026 (0.339)	0.003	**0.172** (3.274)	**0.152**	**0.313** (4.367)	**0.144**
	E(*Y*|*Y*>0)	**−0.095** (−3.865)	**−0.057**	**−0.049** (−2.528)	**−0.043**	0.063 (0.738)	0.027	−0.091 (−1.838)	−0.057
Italy	*P*(*Y*>0)	−0.001 (−0.054)	0.000	**0.116** (3.766)	**0.011**	**0.136** (6.251)	**0.105**	**0.190** (7.918)	**0.063**
	E(*Y*|*Y*>0)	**−0.044** (−4.944)	−0.031	**−0.024** (−2.691)	**−0.021**	**−0.084** (−2.787)	**−0.035**	0.000 (−0.026)	0.000
The Netherlands	*P*(*Y*>0)	**0.082** (2.897)	**0.035**	0.094 (1.739)	0.009	**0.071** (2.085)	**0.055**	−0.055 (−1.084)	−0.016
	E(*Y*|*Y*>0)	−0.037 (−1.484)	−0.019	**−0.085** (−5.446)	**−0.068**	**−0.250** (−4.377)	**−0.129**	−0.008 (−0.299)	−0.006
Portugal	*P*(*Y*>0)	**0.223** (10.888)	**0.104**	**0.243** (8.070)	**0.036**	**0.252** (9.190)	**0.198**	**0.295** (9.454)	**0.099**
	E(*Y*|*Y*>0)	**0.027** (2.302)	**0.018**	0.001 (0.078)	0.001	**−0.087** (−3.292)	**−0.045**	**0.041** (2.340)	**0.028**
Spain	*P*(*Y*>0)	−0.015 (−0.997)	−0.006	0.037 (1.261)	0.005	**0.112** (5.680)	**0.080**	**0.138** (5.189)	**0.042**
	E(*Y*|*Y*>0)	**−0.053** (−4.401)	**−0.034**	**−0.025** (−2.324)	**−0.021**	**−0.070** (−2.460)	**−0.033**	0.017 (1.026)	0.012

Notes: t-statistics of coefficients in parentheses. Coefficients in bold are those significant at 5%. Elasticities are calculated for each individual and averaged over the sample. Elasticities in bold correspond to significant coefficients.



RANDOM PARAMETER MODELS



A Recast Random Effects Model

$$U_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + u_i + \varepsilon_{it},\ u_i \sim N[0,\sigma_u]$$

$$\boxed{\alpha_i = \alpha + u_i}$$

T_i = observations on individual i

For each period, $y_{it} = 1[U_{it} > 0]$ (given u_i)

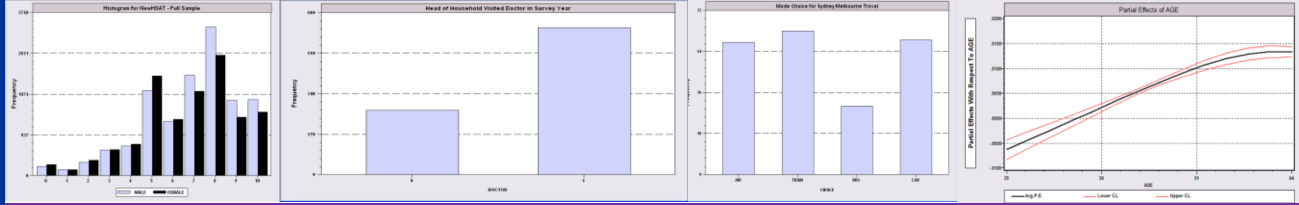
Joint probability for T_i observations is

$$\text{Prob}(y_{i1}, y_{i2},\ldots \mid u_i) = \prod_{t=1}^{T_i} \boxed{F(y_{it,}\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it})}$$

Write $u_i = \sigma_u v_i,\ v_i \sim N[0,1],\ \boxed{\alpha_i = \alpha + \sigma_u v_i}$

$$\log L \mid v_1,\ldots v_N = \sum_{i=i}^{N} \log\left[\prod_{t=1}^{T_i} \boxed{F(y_{it,}(\alpha + \sigma_u v_i) + \boldsymbol{\beta}'\mathbf{x}_{it})}\right]$$

It is not possible to maximize $\log L \mid v_1,\ldots v_N$ because of the unobserved random effects embedded in α_i.



A Computable Log Likelihood

The unobserved heterogeneity is averaged out of $\log L \mid \mathbf{v}$

$$E_{\mathbf{v}}[\log L \mid \mathbf{v}] \approx \sum_{i=1}^{N} \log \int_{\alpha_i=-\infty}^{\infty} \left[\prod_{t=1}^{T_i} F(y_{it}, \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it}) \right] f(\alpha_i)\, d\alpha_i$$

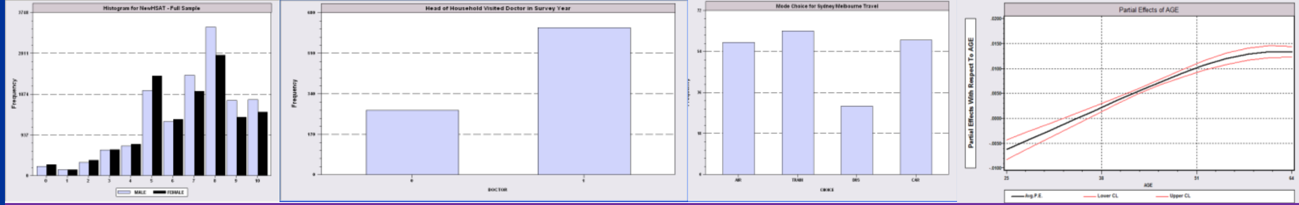
Maximize this function with respect to $\alpha, \boldsymbol{\beta}, \sigma_{\text{u}}$. $(\alpha_i = \alpha + \sigma_u v_i)$

How to compute the integral?

(1) Analytically? No, no formula exists.

(2) Approximately, using Gauss-Hermite quadrature

(3) Approximately using Monte Carlo simulation



Simulation

$$\log L = \sum_{i=1}^{N} \log \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T_i} F(y_{it}, \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it}) \right] \phi(\alpha_i) d\alpha_i$$

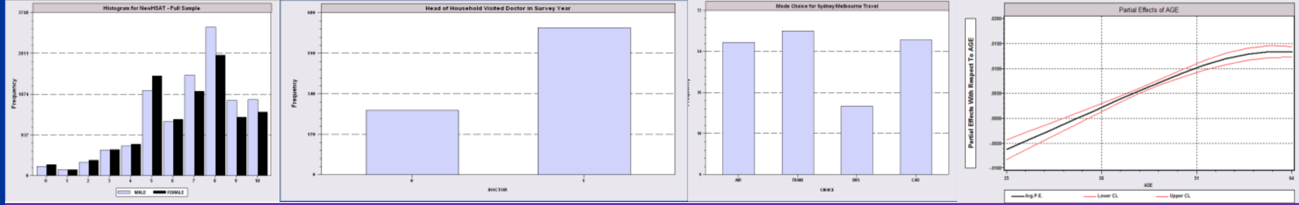
$$= \sum_{i=1}^{N} \log \int_{-\infty}^{\infty} g(\alpha_i) \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\alpha_i^2}{2} \right) d\alpha_i$$

This equals $\sum_{i=1}^{N} \log E[g(\alpha_i)]$

The expected value of the function of α_i can be approximated by drawing R random draws v_{ir} from the population N[0,1] and averaging the R functions of $\alpha_{ir} = \alpha + \sigma_u v_{ir}$. We maximize

$$\log L_S = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \left[\prod_{t=1}^{T_i} F(y_{it}, (\alpha + \sigma_u v_{ir}) + \boldsymbol{\beta}'\mathbf{x}_{it}) \right]$$

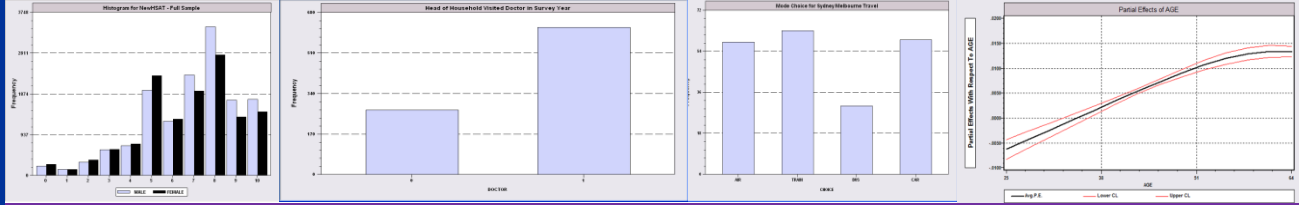
(We did this in part 4 for the random effects probit model.)



Random Effects Model: Simulation

```
----------------------------------------------------------------------
Random Coefficients  Probit   Model
Dependent variable               DOCTOR (Quadrature Based)
Log likelihood function    -16296.68110 (-16290.72192)
Restricted log likelihood  -17701.08500
Chi squared [   1 d.f.]      2808.80780
Simulation based on  50 Halton draws
--------+-------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+-------------------------------------------------
        |Nonrandom parameters
     AGE|      .02226***        .00081       27.365   .0000  ( .02232)
    EDUC|     -.03285***        .00391       -8.407   .0000  (-.03307)
  HHNINC|      .00673           .05105         .132   .8952  ( .00660)
        |Means for random parameters
Constant|     -.11873**         .05950       -1.995   .0460  (-.11819)
        |Scale parameters for dists. of random parameters
Constant|      .90453***        .01128       80.180   .0000
--------+-------------------------------------------------------------
```

Implied ρ from these estimates is $.90454^2/(1+.90453^2)$ = .449998.



The Entire Parameter Vector is Random

$$U_{it} = \boldsymbol{\beta}_i'\mathbf{x}_{it} + \varepsilon_{it},$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_\mathrm{i}, \quad \mathbf{u}_\mathrm{i} \sim N[\mathbf{0}, diag(\sigma_1,...,\sigma_K)]$$

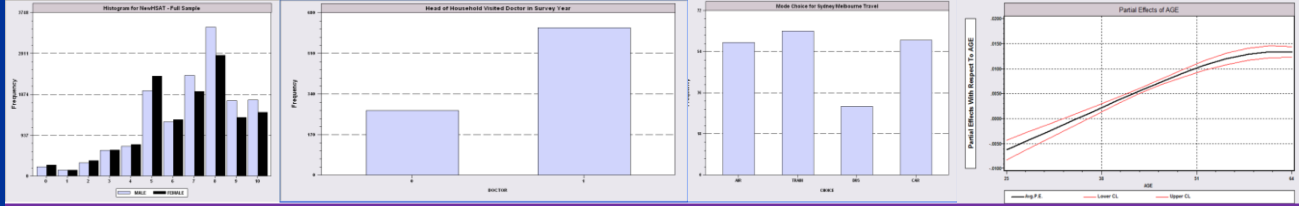
Joint probability for T_i observations is

$$\text{Prob}(y_{i1}, y_{i2},...\,|\,u_i) = \prod_{t=1}^{T_i} F(y_{it,}\boldsymbol{\beta}_i'\mathbf{x}_{it})$$

For convenience, write $u_{ik} = \sigma_k v_{ik}$, $v_{ik} \sim N[0,1]$, $\beta_{ik} = \beta_k + \sigma_k v_{ik}$

$$\log L\,|\,v_1,...v_N = \sum_{i=i}^{N} \log\left[\prod_{t=1}^{T_i} F(y_{it,}\boldsymbol{\beta}_i'\mathbf{x}_{it})\right]$$

It is not possible to maximize $\log L\,|\,\mathbf{v}_1,...\mathbf{v}_N$ because of the unobserved random effects embedded in $\boldsymbol{\beta}_i$.



Estimating the RPL Model

Estimation: β_1

$$\beta_{2it} = \beta_2 + \mathbf{\Delta z}_i + \mathbf{\Gamma v}_{i,t}$$

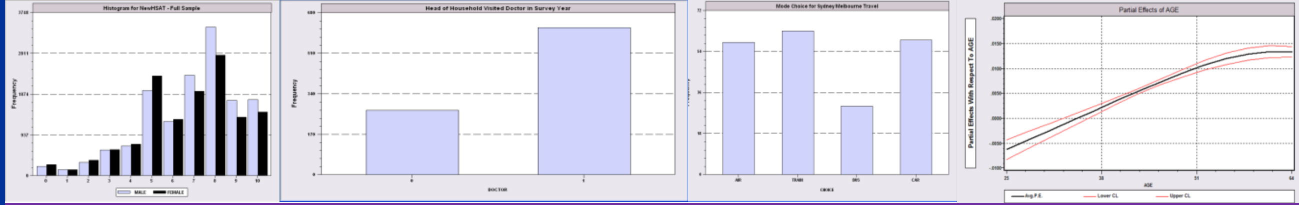
Uncorrelated: $\mathbf{\Gamma}$ is diagonal

Autocorrelated: $\mathbf{v}_{i,t} = \mathbf{R}\mathbf{v}_{i,t-1} + \mathbf{u}_{i,t}$

(1) Estimate "structural parameters"

(2) Estimate individual specific utility parameters

(3) Estimate elasticities, etc.



Classical Estimation Platform: The Likelihood

$$\text{Marginal}: f(\boldsymbol{\beta}_i \mid \text{data}, \boldsymbol{\Omega})$$

$$\text{Population Mean} = E[\boldsymbol{\beta}_i \mid \text{data}, \boldsymbol{\Omega}]$$

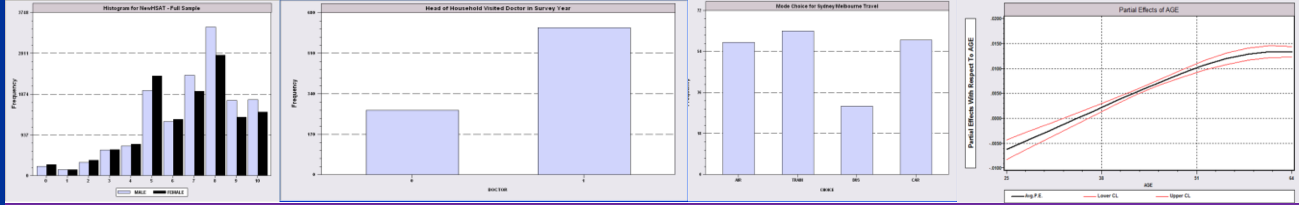
$$= \int_{\beta_i} \boldsymbol{\beta}_i f(\boldsymbol{\beta}_i \mid \boldsymbol{\Omega}) d\boldsymbol{\beta}_i$$

$$= \overline{\boldsymbol{\beta}} = \text{a subvector of } \boldsymbol{\Omega}$$

$$\hat{\boldsymbol{\Omega}} = \text{Argmax } L(\boldsymbol{\beta}_i, i = 1,\ldots,N \mid \text{data}, \boldsymbol{\Omega})$$

$$\text{Estimator} = \hat{\overline{\boldsymbol{\beta}}}$$

Expected value over all possible realizations of β_i. I.e., over all possible samples.

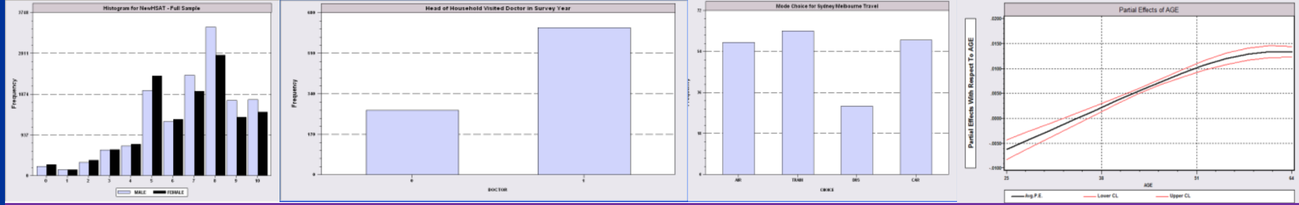


Simulation Based Estimation

- Choice probability = $P[\text{data} \mid \beta(\beta_1,\beta_2,\mathbf{\Delta},\mathbf{\Gamma},\mathbf{R},\mathbf{v}_{i,t})]$
- Need to integrate out the unobserved random term
- $E\{P[\text{data} \mid \beta(\beta_1,\beta_2,\mathbf{\Delta},\mathbf{\Gamma},\mathbf{R},\mathbf{v}_{i,t})]\}$

$$= \int_{\mathbf{v}} P[\ldots|\mathbf{v}_{i,t}]f(\mathbf{v}_{i,t})d\mathbf{v}_{i,t}$$

- Integration is done by simulation
  - Draw values of **v** and compute β then probabilities
  - Average many draws
  - Maximize the sum of the logs of the averages
  - (See Train[Cambridge, 2003] on simulation methods.)



Maximum Simulated Likelihood

True log likelihood

$$L_i(\boldsymbol{\beta}_i \mid \text{data}_i) = \prod_{t=1}^{T_i} f(\text{data}_i \mid \boldsymbol{\beta}_i)$$

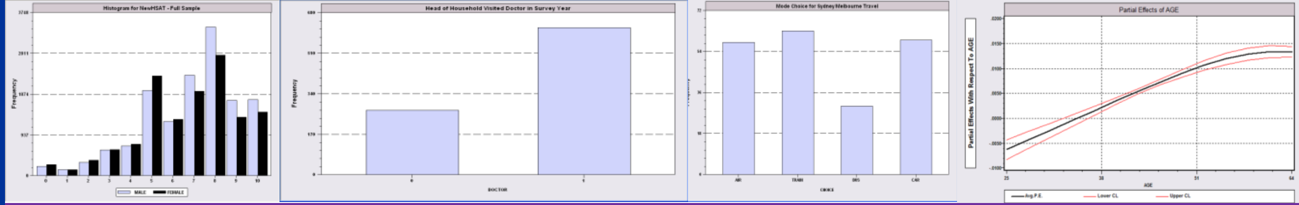
$$L_i(\boldsymbol{\Omega} \mid \text{data}_i) = \int_{\beta_i} \prod_{t=1}^{T_i} f(\text{data}_i \mid \boldsymbol{\beta}_i) f(\boldsymbol{\beta}_i \mid \boldsymbol{\Omega}) d\boldsymbol{\beta}_i$$

$$\log L = \sum_{i=1}^{N} \log \int_{\beta_i} L_i(\boldsymbol{\beta}_i \mid \text{data}_i) f(\boldsymbol{\beta}_i \mid \boldsymbol{\Omega}) d\boldsymbol{\beta}_i$$

Simulated log likelihood

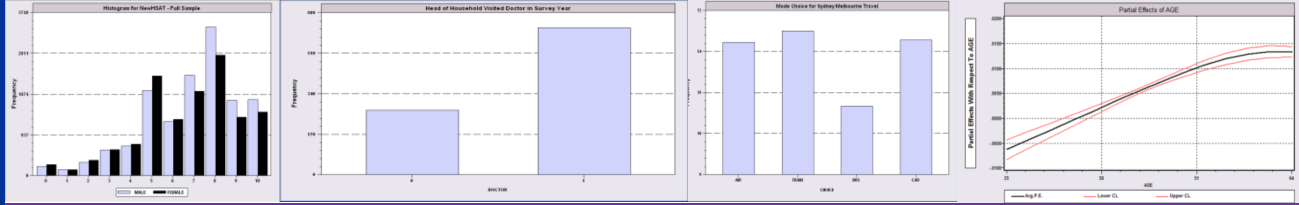
$$\log L_S = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} L_i(\boldsymbol{\beta}_{iR} \mid \text{data}_i, \boldsymbol{\Omega})$$

$$\hat{\boldsymbol{\Omega}} = \text{argmax}(\log L_S)$$




```
probit ; lhs = doctor
       ; rhs = one,age,educ,income
       ; panel ; rpm
       ; fcn=one(n),age(n),educ(n),income(n)
       ; halton ; pts=10 $
```

```
-----------------------------------------------------------------------------
Random Coefficients  Probit   Model
Dependent variable                DOCTOR
Log likelihood function     -16384.69537
Restricted log likelihood   -17701.07532
Chi squared [    4 d.f.]     2632.75991
Significance level                .00000
McFadden Pseudo R-squared       .0743672
Estimation based on N =   27326, K =   8
Inf.Cr.AIC  =  32785.4 AIC/N =     1.200
Unbalanced panel has   7293 individuals
PROBIT (normal)  probability model
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Means for random parameters
Constant|    -.14345**     .05804    -2.47  .0135     -.25721    -.02969
     AGE|     .02235***    .00081    27.65  .0000      .02076     .02393
    EDUC|    -.03259***    .00384    -8.49  .0000     -.04011    -.02507
  INCOME|     .03865       .05042      .77  .4433     -.06017     .13748
        |Scale parameters for dists. of random parameters
Constant|     .09551***    .00866    11.02  .0000      .07853     .11249
     AGE|     .02063***    .00026    78.57  .0000      .02011     .02114
    EDUC|     .01207***    .00075    16.20  .0000      .01061     .01353
  INCOME|     .03222       .02175     1.48  .1385     -.01041     .07486
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```



When Var[$\mathbf{u}_i$] is not a diagonal matrix. How to estimate a positive definite matrix, $\mathbf{\Omega}$.

$\mathbf{u}_i \sim N[\mathbf{0},\mathbf{\Omega}]$

Cholesky Decomposition : $\mathbf{\Omega} = \mathbf{LL}'$ where $\mathbf{L}$ is upper triangular

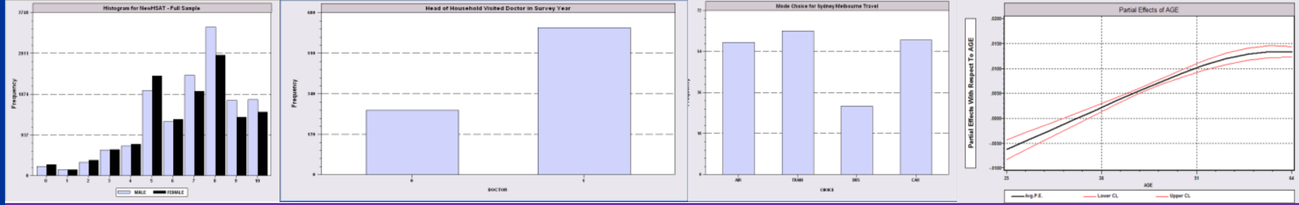
$\mathbf{u}_i = \mathbf{Lv}_i$ with $\mathbf{v}_i \sim N[\mathbf{0},\mathbf{I}]$

Convenient Refinement: $\mathbf{\Omega} = \mathbf{LL}' = (\mathbf{MS})(\mathbf{MS})'$ where the diagonal elements of $\mathbf{M}$ equal 1, and $\mathbf{S}$ is the diagonal matrix with free positive elements. (Cholesky values)

$\mathbf{u}_i = \mathbf{MSv}_i$

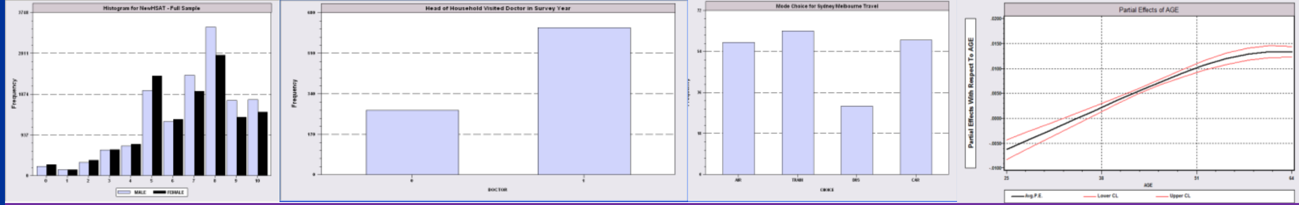
$\mathbf{M} = \mathbf{I}$ returns the original uncorrelated case.

We used the Cholesky decomposition in developing the Krinsky and Robb method for standard errors for partial effects, in Part 3.



```
Random Coefficients Probit   Model
Dependent variable                DOCTOR
Log likelihood function    -16349.68809
Restricted log likelihood  -17701.07532
Chi squared [  10 d.f.]      2702.77447
Significance level               .00000
McFadden Pseudo R-squared      .0763449
Estimation based on N =  27326, K =  14
Inf.Cr.AIC  =  32727.4 AIC/N =    1.198
Unbalanced panel has   7293 individuals
PROBIT (normal)  probability model
```

	DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
		Means for random parameters					
	Constant	-.22219***	.05983	-3.71	.0002	-.33944	-.10493
	AGE	.02218***	.00083	26.67	.0000	.02055	.02381
	EDUC	-.02318***	.00404	-5.74	.0000	-.03110	-.01526
	INCOME	-.02549	.05308	-.48	.6311	-.12953	.07855
S		Diagonal elements of Cholesky matrix					
	Constant	.10235*	.06112	1.67	.0940	-.01744	.22214
	AGE	.01119***	.00078	14.37	.0000	.00966	.01271
	EDUC	.00428**	.00173	2.47	.0134	.00089	.00766
	INCOME	.04191*	.02184	1.92	.0549	-.00089	.08472
M		Below diagonal elements of Cholesky matrix					
	lAGE_ONE	-.00131	.00081	-1.62	.1057	-.00290	.00028
	lEDU_ONE	.01903***	.00410	4.64	.0000	.01100	.02707
	lEDU_AGE	-.04614***	.00327	-14.11	.0000	-.05255	-.03974
	lINC_ONE	.36705***	.05258	6.98	.0000	.26400	.47010
	lINC_AGE	-.27716***	.06251	-4.43	.0000	-.39967	-.15466
	lINC_EDU	-.01902	.05020	-.38	.7048	-.11741	.07937



MSS′M′

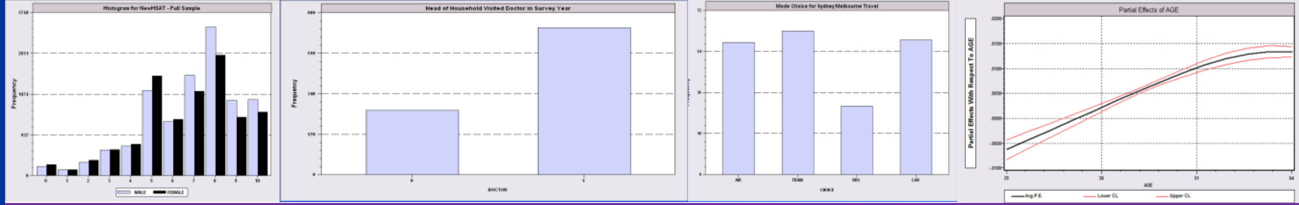
Implied covariance matrix of random parameters

Covariance matrix

	ONE	AGE	EDUC	INCOME
ONE	.1048E-01			
AGE	-.1341E-03	.1269E-03		
EDUC	.1948E-02	-.5412E-03	.2510E-02	
INCOME	.3757E-01	-.3582E-02	.1969E-01	.2137

Implied standard deviations of random parameters

S.D_Beta	1
1	.102350
2	.0112652
3	.0500983
4	.462238



Modeling Parameter Heterogeneity

Individual heterogeneity in the means of the parameters

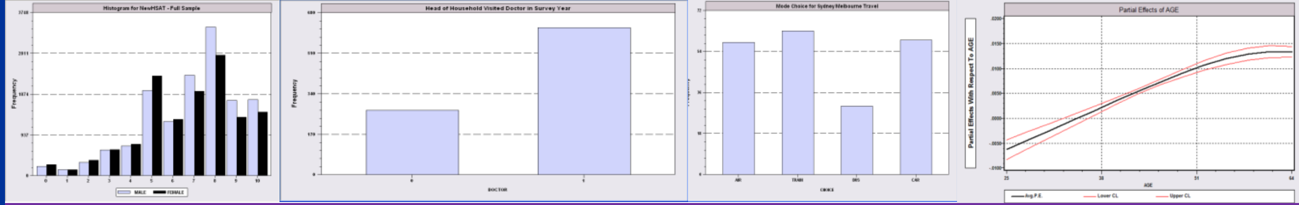
$$\boldsymbol{\beta}_i = \bar{\boldsymbol{\beta}} + \boldsymbol{\Delta}\mathbf{z}_i + \mathbf{u}_i$$

$$E[\mathbf{u}_i \mid \mathbf{X}_i, \mathbf{z}_i]$$

Heterogeneity in the variances of the parameters

$$\text{Var}[u_{i,k} \mid \text{data}_i] = \theta_k \exp(\mathbf{h}_i'\boldsymbol{\delta}_k)$$

Estimation by maximum simulated likelihood



A Hierarchical Probit Model

$$U_{it} = \beta_{1i} + \beta_{2i}Age_{it} + \beta_{3i}Educ_{it} + \beta_{4i}Income_{it} + \varepsilon_{it}.$$

$$\beta_{1i}=\beta_1+\Delta_{11}\ Female_i + \Delta_{12}\ Married_i + u_{1i}$$

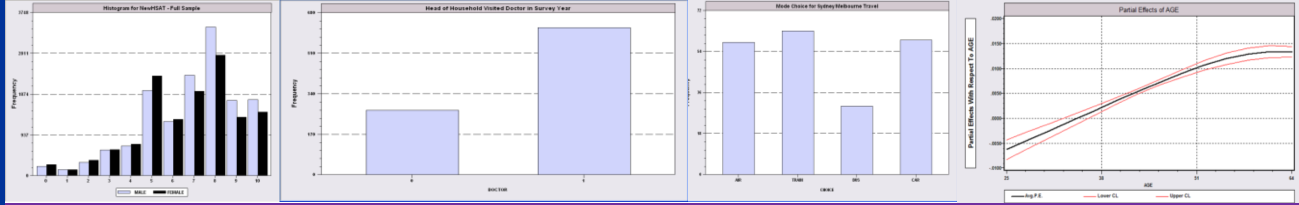
$$\beta_{2i}=\beta_2+\Delta_{21}\ Female_i + \Delta_{22}\ Married_i + u_{2i}$$

$$\beta_{3i}=\beta_3+\Delta_{31}\ Female_i + \Delta_{32}\ Married_i + u_{3i}$$

$$\beta_{4i}=\beta_4+\Delta_{41}\ Female_i + \Delta_{42}\ Married_i + u_{4i}$$

$$Y_{it} = 1[U_{it} > 0]$$

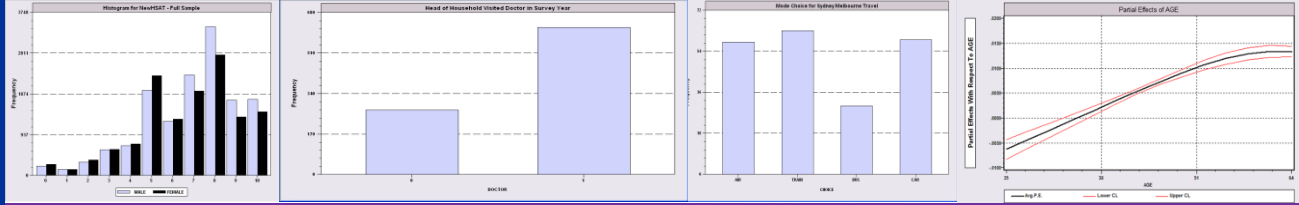
All random variables normally distributed.



Random Coefficients Probit Model
Dependent variable DOCTOR
Log likelihood function -16228.33601
Restricted log likelihood -17701.07532
Chi squared [18 d.f.] 2945.47863
Significance level .00000
McFadden Pseudo R-squared .0832006
Estimation based on N = 27326, K = 22
Inf.Cr.AIC = 32500.7 AIC/N = 1.189
Unbalanced panel has 7293 individuals
PROBIT (normal) probability model

DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Means for random parameters					
Constant	-.11521	.12228	-.94	.3461	-.35487	.12445
AGE	.01660***	.00173	9.61	.0000	.01321	.01998
EDUC	-.03299***	.00788	-4.19	.0000	-.04843	-.01754
INCOME	-.03627	.10540	-.34	.7308	-.24285	.17032
	Diagonal elements of Cholesky matrix					
Constant	.49489***	.06515	7.60	.0000	.36719	.62259
AGE	.02057***	.00064	31.94	.0000	.01931	.02183
EDUC	.06413***	.00208	30.81	.0000	.06005	.06821
INCOME	.03200	.02142	1.49	.1352	-.00998	.07399
	Below diagonal elements of Cholesky matrix					
lAGE_ONE	-.01351***	.00087	-15.44	.0000	-.01522	-.01179
lEDU_ONE	.02670***	.00440	6.07	.0000	.01808	.03532
lEDU_AGE	-.04147***	.00268	-15.46	.0000	-.04672	-.03621
lINC_ONE	-.39265***	.05705	-6.88	.0000	-.50446	-.28084
lINC_AGE	-.20638***	.05181	-3.98	.0001	-.30793	-.10483
lINC_EDU	.01241	.06018	.21	.8366	-.10553	.13035
	Heterogeneity in the means of random parameters					
cONE_FEM	.05724	.12313	.46	.6420	-.18409	.29856
cONE_MAR	-.49702***	.13211	-3.76	.0002	-.75595	-.23810
cAGE_FEM	-.00595***	.00163	-3.66	.0003	-.00914	-.00276
cAGE_MAR	.01330***	.00182	7.33	.0000	.00974	.01686
cEDU_FEM	.05860***	.00838	7.00	.0000	.04219	.07502
cEDU_MAR	-.00520	.00863	-.60	.5473	-.02212	.01173
cINC_FEM	.00600	.10167	.06	.9529	-.19326	.20527
cINC_MAR	-.03671	.11437	-.32	.7482	-.26086	.18745

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

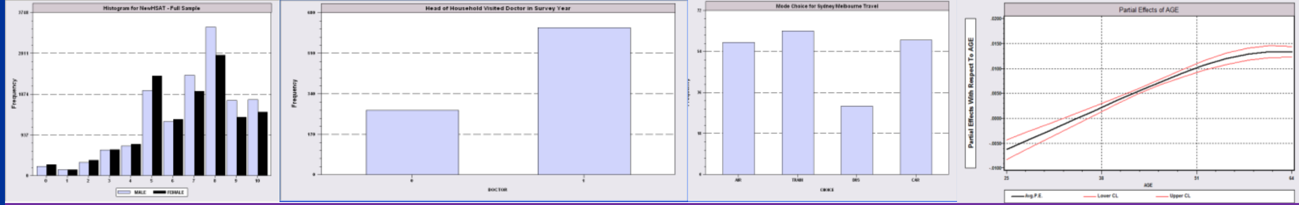


Simulating Conditional Means for Individual Parameters

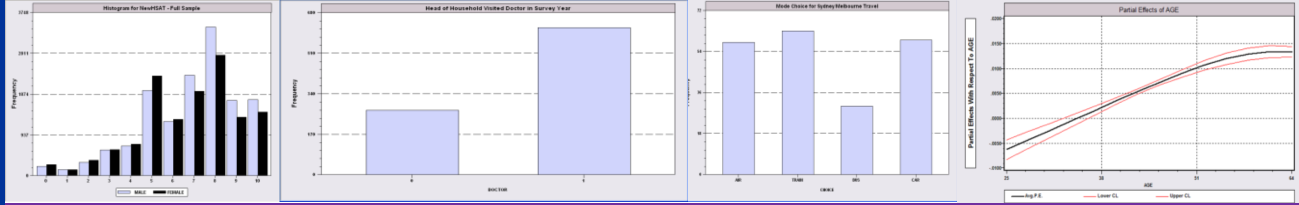
$$\hat{E}(\boldsymbol{\beta}_i \mid \mathbf{y}_i, \mathbf{X}_i) = \frac{\frac{1}{R}\sum_{r=1}^{R} (\hat{\boldsymbol{\beta}} + \hat{\mathbf{L}}\mathbf{w}_{i,r}) \prod_{t=1}^{T_i} \Phi\left((2y_{it} - 1)(\hat{\boldsymbol{\beta}} + \hat{\mathbf{L}}\mathbf{w}_{i,r})'\mathbf{x}_{it}\right)}{\frac{1}{R}\sum_{r=1}^{R} \prod_{t=1}^{T_i} \Phi\left((2y_{it} - 1)(\hat{\boldsymbol{\beta}} + \hat{\mathbf{L}}\mathbf{w}_{i,r})'\mathbf{x}_{it}\right)}$$

$$= \frac{1}{R}\sum_{r=1}^{R} \hat{Weight}_{ir} \hat{\boldsymbol{\beta}}_{ir}$$

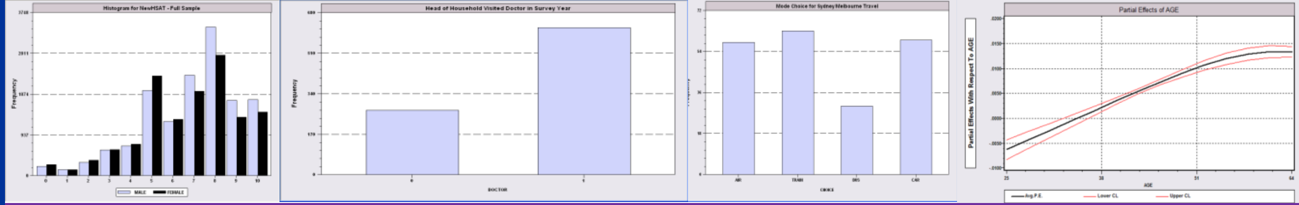
Posterior estimates of E[parameters(i) | Data(i)]




```
----------------------------------------------------------------------------
Probit   Regression Start Values for DOCTOR
Dependent variable               DOCTOR
Log likelihood function    -17422.71144
Estimation based on N =  27326, K =   7
Inf.Cr.AIC  =  34859.4 AIC/N =    1.276
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient     Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
     AGE|     .01189***    .00080    14.94  .0000      .01033     .01345
 MARRIED|     .07352***    .02064     3.56  .0004      .03306     .11398
  FEMALE|     .35591***    .01602    22.22  .0000      .32451     .38730
  HHKIDS|    -.15211***    .01833    -8.30  .0000    -.18803    -.11619
Constant|    -.12433**     .05815    -2.14  .0325    -.23830    -.01037
    EDUC|    -.01496***    .00358    -4.18  .0000    -.02197    -.00795
  INCOME|    -.13260***    .04657    -2.85  .0044    -.22387    -.04132
--------+-------------------------------------------------------------------
```




```
probit ; lhs = doctor
       ; rhs = one,age,educ,income,married,female,hhkids
       ; panel ; rpm
       ; fcn=one(n),educ(n),income(n) ; correlated
       ; halton ; pts=10
       ; par $
matrix ; beduc = beta_i(1:7293,2:2) $
kernel ; rhs = beduc
       ; title=Sample Conditional Means of Education Coefficients$
```



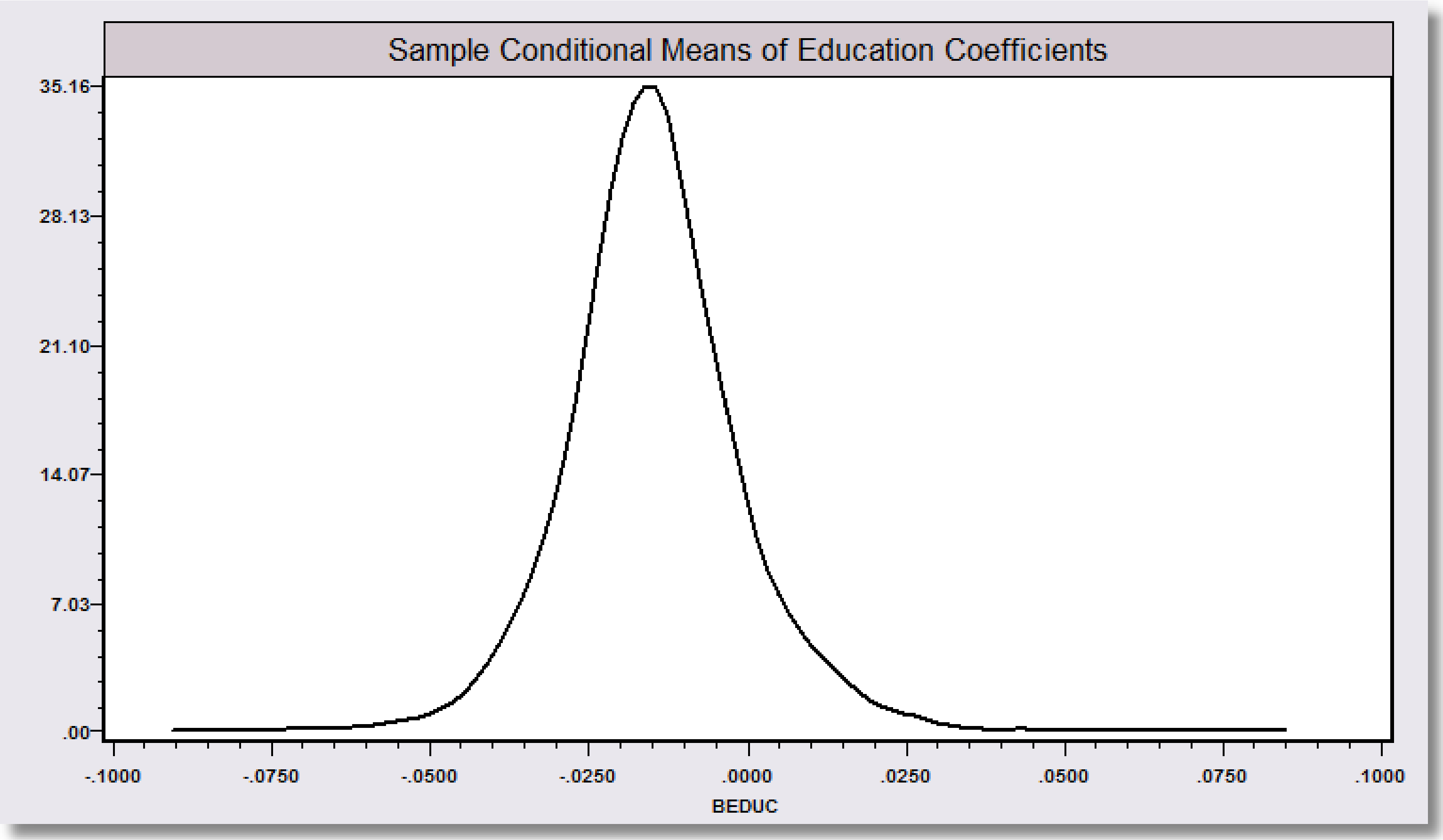

```
Implied covariance matrix of random parameters

Covariance  matrix
--------+-----------------------------------------
                  ONE          EDUC        INCOME
--------+-----------------------------------------
ONE            1.532
EDUC       -.4437E-01    .2526E-02
INCOME          .8081  -.3632E-01          .6045

Implied standard deviations of random parameters

S.D_Beta|              1
--------+--------------
       1|        1.23780
       2|       .0502610
       3|        .777470


Implied correlation matrix of random parameters


--------+------------------------------
Cor.Mat.|      ONE      EDUC     INCOME
--------+------------------------------
     ONE| 1.00000   -.71325    .83968
    EDUC| -.71325   1.00000   -.92948
  INCOME|  .83968   -.92948   1.00000
```

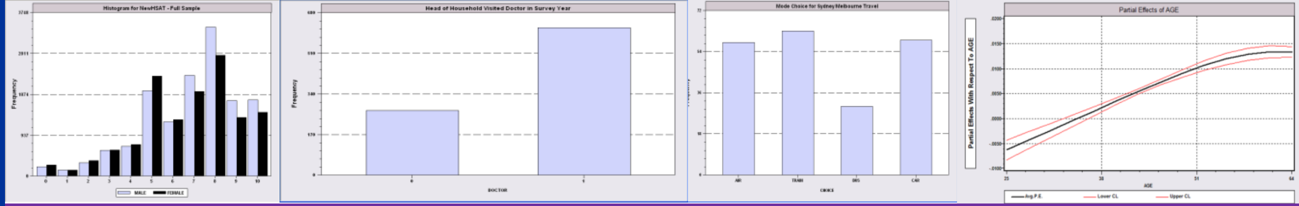
Matrix - BETA_I

[7293, 3] Cell: -0.0207049

	1	2	3
1	-0.696212	-0.0163842	0.00232707
2	-0.280491	-0.0201771	-0.126372
3	-1.13626	-0.0115655	0.128616
4	-0.0848325	-0.0059294	-0.0483072
5	0.00761816	-0.017374	-0.217985
6	0.509238	-0.0158648	-0.205521
7	-0.935034	-0.0012945	0.307114
8	0.125411	-0.0215151	-0.238285
9	0.433899	-0.0360507	-0.451719
10	-1.03039	-0.020177	0.00283152
11	-0.556545	-0.00872936	0.0862487
12	-0.954397	-0.0196685	0.0229384
13	-0.395515	-0.0165901	-0.0717262
14	-0.409222	-0.0491196	-0.418569
15	0.414609	-0.00622438	-0.144994
16	-1.65292	0.00581486	0.498509
17	-0.27507	-0.0114305	-0.00559716

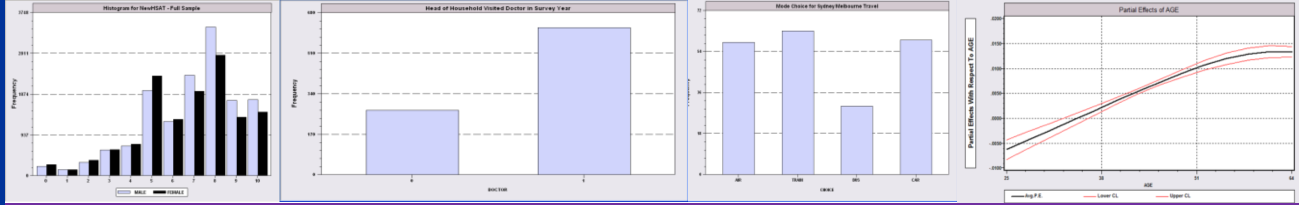
"Individual Coefficients"



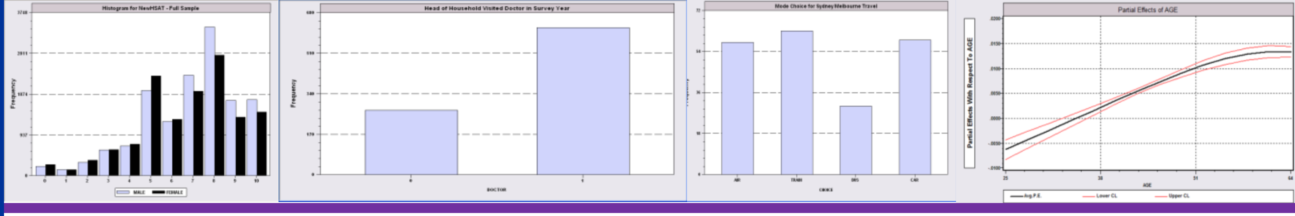


Programs differ on the models fitted, the algorithms, the paradigm, and the extensions provided to the simplest RPM, $\beta_i = \beta + \mathbf{w}_i$.

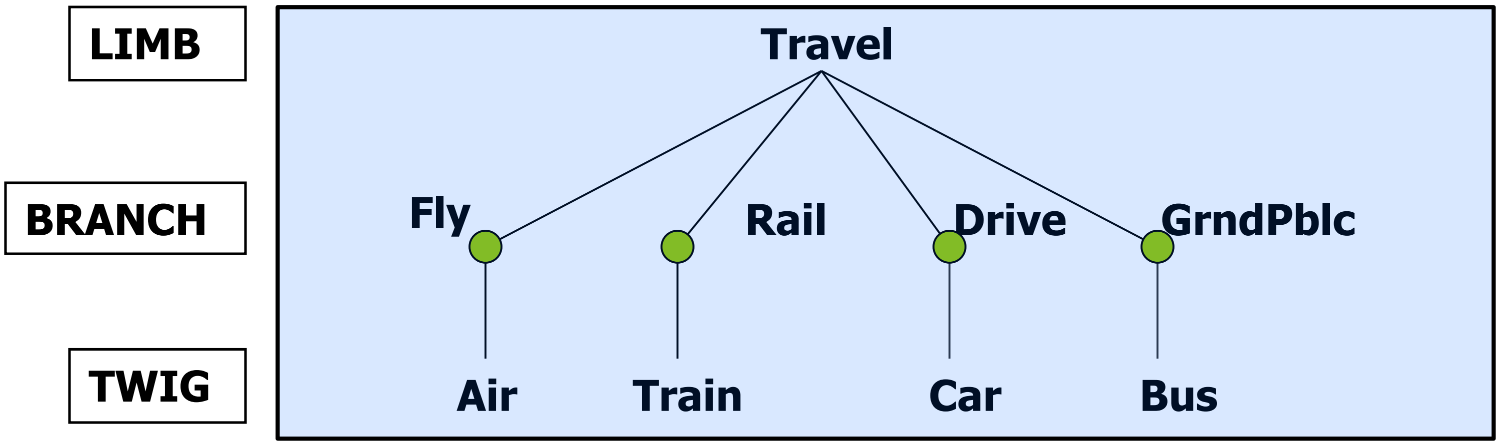
- WinBUGS:
  - MCMC
  - User specifies the model – constructs the Gibbs Sampler/Metropolis Hastings
- MLWin:
  - Linear and some nonlinear – logit, Poisson, etc.
  - Uses MCMC for MLE (noninformative priors)
- SAS: Proc Mixed.
  - Classical
  - Uses primarily a kind of GLS/GMM (method of moments algorithm for loglinear models)
- Stata: Classical
  - Several loglinear models – GLAMM. Mixing done by quadrature.
  - Maximum simulated likelihood for multinomial choice (Arne Hole, user provided)
- LIMDEP/NLOGIT
  - Classical
  - Mixing done by Monte Carlo integration – maximum simulated likelihood
  - Numerous linear, nonlinear, loglinear models
- Ken Train's Gauss Code, miscellaneous freelance R and Matlab code
  - Monte Carlo integration
  - Mixed Logit (mixed multinomial logit) model only (but free!)
- Biogeme
  - Multinomial choice models
  - Many experimental models (developer's hobby)

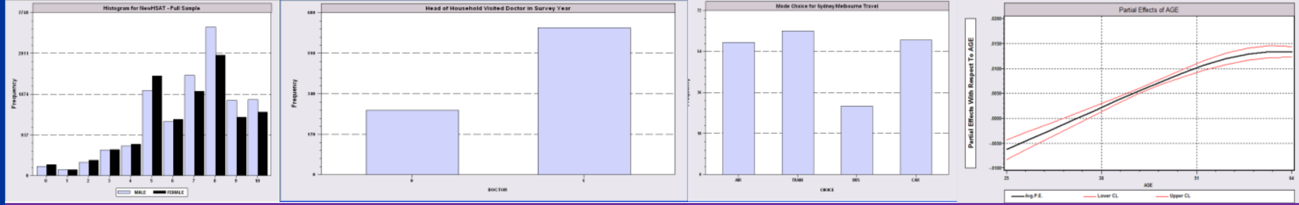


SCALING IN CHOICE MODELS



Using Degenerate Branches to Reveal Scaling

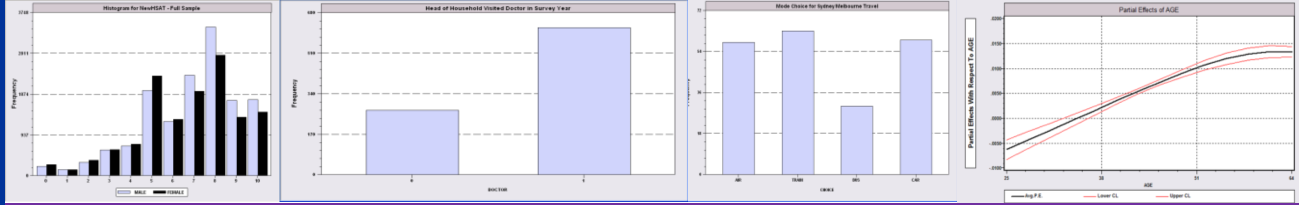




Scaling in Transport Modes

```
-----------------------------------------------------------
FIML Nested Multinomial Logit Model
Dependent variable                 MODE
Log likelihood function      -182.42834
The model has 2 levels.
Nested Logit form:IVparms=Taub|l,r,Sl|r
& Fr.No normalizations imposed a priori
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|      .09622**        .03875        2.483   .0130
    TTME|     -.08331***       .02697       -3.089   .0020
    INVT|     -.01888***       .00684       -2.760   .0058
    INVC|     -.10904***       .03677       -2.966   .0030
   A_AIR|     4.50827***      1.33062        3.388   .0007
 A_TRAIN|     3.35580***       .90490        3.708   .0002
   A_BUS|     3.11885**       1.33138        2.343   .0192
        |IV parameters, tau(b|l,r),sigma(l|r),phi(r)
     FLY|     1.65512**        .79212        2.089   .0367
    RAIL|      .92758***       .11822        7.846   .0000
LOCLMASS|     1.00787***       .15131        6.661   .0000
   DRIVE|     1.00000        ......(Fixed Parameter)......
--------+--------------------------------------------------
```

```
NLOGIT ; Lhs=mode
; Rhs=gc,ttme,invt,invc,one
; Choices=air,train,bus,car
; Tree=Fly(Air),
       Rail(train),
       LoclMass(bus),
       Drive(Car)
; ivset:(drive)=[1]$
```



A Model with Choice Heteroscedasticity

$$U(i,t,j) = \alpha_j + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it} \quad + \boxed{\sigma_j \varepsilon_{i,t,j}}$$

$$F(\varepsilon_{i,t,j}) = \exp(-\exp(-\varepsilon_{i,t,j}))$$

IID after scaling by a choice specific scale parameter

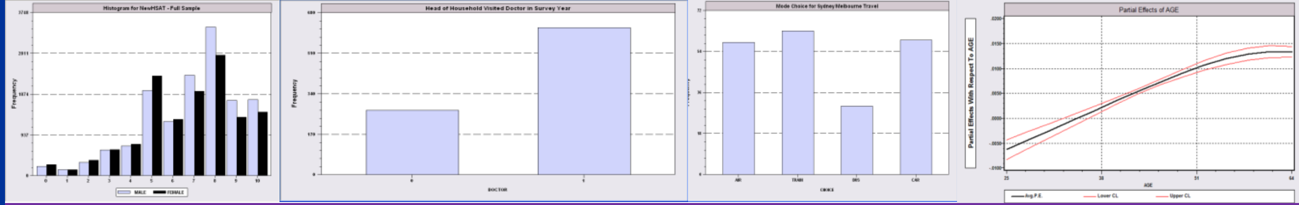
$$P[\text{choice} = j \mid \mathbf{x}_{itj}, \mathbf{z}_{it}, i, t] = \text{Prob}[U_{i,t,j} > U_{i,t,k}],\ k = 1,\ldots,J(i,t)$$

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

Normalization required as only ratios can be estimated;

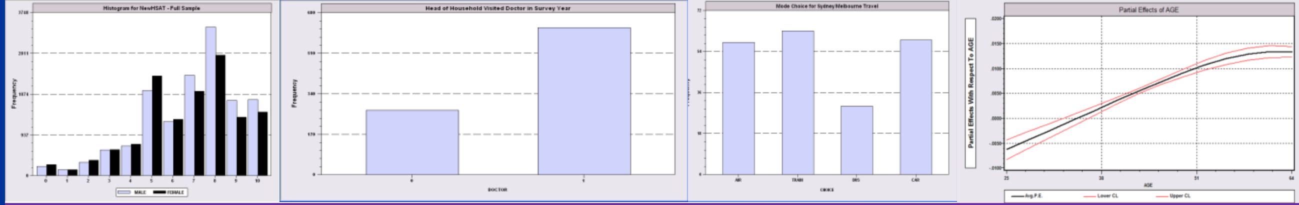
$$\sigma_j = 1 \text{ for one of the alternatives}$$

(Remember the integrability problem - scale is not identified.)



Heteroscedastic Extreme Value Model (1)

```
+---------------------------------------------+
| Start values obtained using MNL model       |
| Maximum Likelihood Estimates                |
| Log likelihood function       -184.5067     |
| Dependent variable                Choice    |
| Response data are given as ind. choice.     |
| Number of obs.=    210, skipped   0 bad obs. |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+
|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
```



Heteroscedastic Extreme Value Model (2)

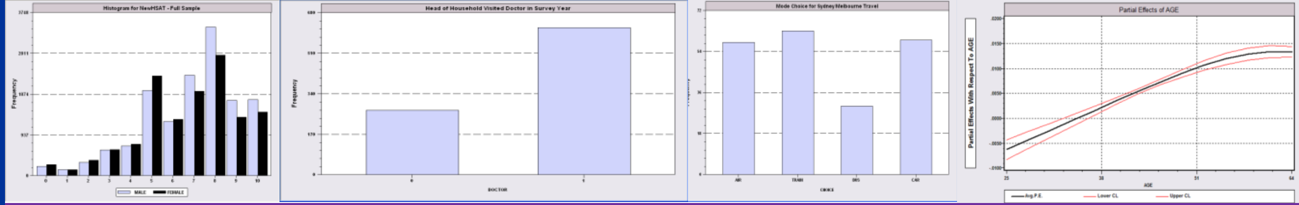
```
+---------------------------------------------+
| Heteroskedastic Extreme Value Model         |
| Log likelihood function        -182.4440    |
| Number of parameters                  10    |
| Restricted log likelihood      -291.1218    |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
---------+Attributes in the Utility Functions (beta)
 GC      |     .11903513       .06402510     1.859    .0630
 TTME    |    -.11525581       .05721397    -2.014    .0440
 INVC    |    -.15515877       .07928045    -1.957    .0503
 INVT    |    -.02276939       .01122762    -2.028    .0426
 AASC    |    4.69411460      2.48091789     1.892    .0585
 TASC    |    5.15629868      2.05743764     2.506    .0122
 BASC    |    5.03046595      1.98259353     2.537    .0112
---------+Scale Parameters of Extreme Value Distns Minus 1.0
 s_AIR   |    -.57864278       .21991837    -2.631    .0085
 s_TRAIN |    -.45878559       .34971034    -1.312    .1896
 s_BUS   |     .26094835       .94582863      .276    .7826
 s_CAR   |      .000000     ......(Fixed Parameter).......
---------+Std.Dev=pi/(theta*sqr(6)) for H.E.V. distribution.
 s_AIR   |    3.04385384      1.58867426     1.916    .0554
 s_TRAIN |    2.36976283      1.53124258     1.548    .1217
 s_BUS   |    1.01713111       .76294300     1.333    .1825
 s_CAR   |    1.28254980     ......(Fixed Parameter).......
```

(MNL logL was -184.5067)

```
 .06929537
-.10364955
-.08493182
-.01333220
5.20474275
4.36060457
3.76323447
```

Normalized for estimation

Structural parameters



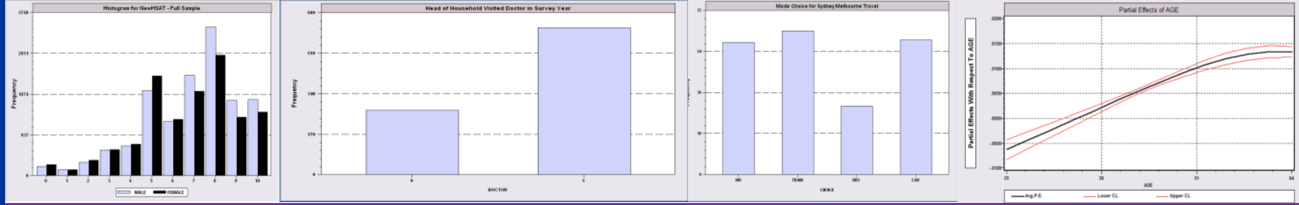
HEV Model - Elasticities

Elasticity averaged over observations.
Attribute is INVC in choice AIR
Effects on probabilities of all choices in model:
* = Direct Elasticity effect of the attribute.

		Mean	St.Dev
Attribute is INVC in choice AIR			
*	Choice=AIR	-4.2604	1.6745
	Choice=TRAIN	1.5828	1.9918
	Choice=BUS	3.2158	4.4589
	Choice=CAR	2.6644	4.0479
Attribute is INVC in choice TRAIN			
	Choice=AIR	.7306	.5171
*	Choice=TRAIN	-3.6725	4.2167
	Choice=BUS	2.4322	2.9464
	Choice=CAR	1.6659	1.3707
Attribute is INVC in choice BUS			
	Choice=AIR	.3698	.5522
	Choice=TRAIN	.5949	1.5410
*	Choice=BUS	-6.5309	5.0374
	Choice=CAR	2.1039	8.8085
Attribute is INVC in choice CAR			
	Choice=AIR	.3401	.3078
	Choice=TRAIN	.4681	.4794
	Choice=BUS	1.4723	1.6322
*	Choice=CAR	-3.5584	9.3057

Multinomial Logit

	Mean	St.Dev
INVC in AIR		
*	-5.0216	2.3881
	2.2191	2.6025
	2.2191	2.6025
	2.2191	2.6025
INVC in TRAIN		
	1.0066	.8801
*	-3.3536	2.4168
	1.0066	.8801
	1.0066	.8801
INVC in BUS		
	.4057	.6339
	.4057	.6339
*	-2.4359	1.1237
	.4057	.6339
INVC in CAR		
	.3944	.3589
	.3944	.3589
	.3944	.3589
*	-1.3888	1.2161



Variance Heterogeneity in MNL

We extend the HEV model by allowing variances to differ across individuals

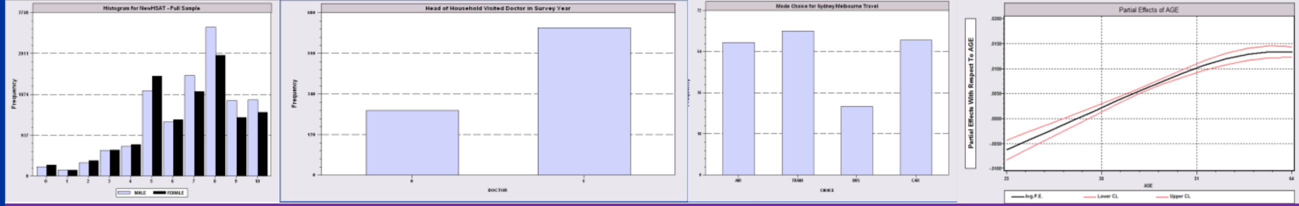
$$U(i,t,j) = \alpha_j + \boldsymbol{\beta}'\mathbf{x}_{itj} + \boldsymbol{\gamma}_j'\mathbf{z}_{it} + \sigma_{ij}\varepsilon_{i,t,j}$$

$$\sigma_{ij} = \exp(\theta_j + \boldsymbol{\delta}'\mathbf{w}_i). \quad \boldsymbol{\delta} = \mathbf{0} \text{ returns the HEV model}$$

$$F(\varepsilon_{i,t,j}) = \exp(-\exp(-\varepsilon_{i,t,j}))$$

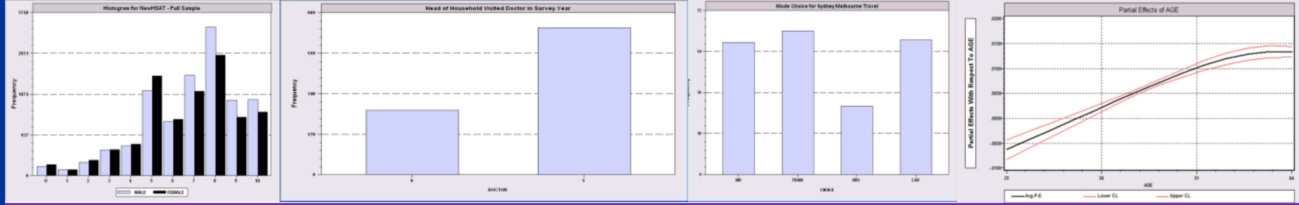
$$\theta_j = 0 \text{ for one of the alternatives}$$

Scaling now differs both across alternatives and across individuals



Application: Shoe Brand Choice

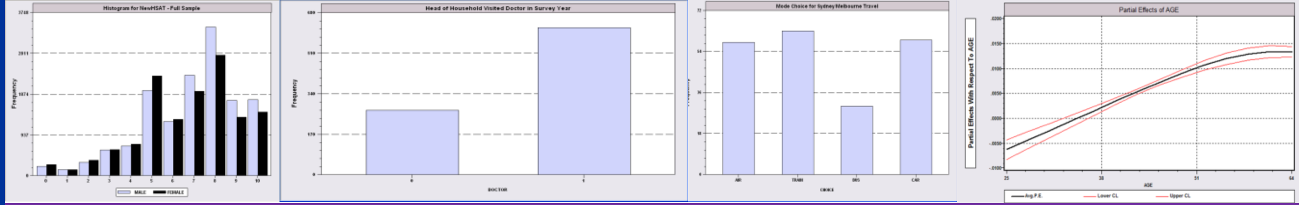
- Simulated Data: Stated Choice, 400 respondents, 8 choice situations, 3,200 observations
- 3 choice/attributes + NONE
  - Fashion = High / Low
  - Quality = High / Low
  - Price = 25/50/75,100 coded 1,2,3,4
- Heterogeneity: Sex, Age (<25, 25-39, 40+)
- Underlying data generated by a 3 class latent class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



Multinomial Logit Baseline Values

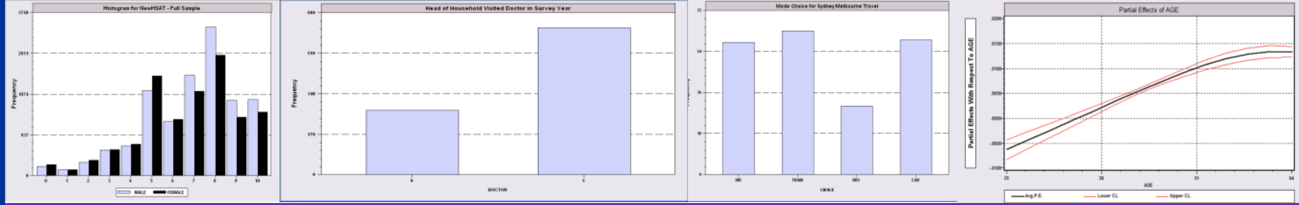
Discrete choice (multinomial logit) model
Number of observations 3200
Log likelihood function -4158.503
Number of obs.= 3200, skipped 0 bad obs.

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
FASH	1.47890473	.06776814	21.823	.0000
QUAL	1.01372755	.06444532	15.730	.0000
PRICE	-11.8023376	.80406103	-14.678	.0000
ASC4	.03679254	.07176387	.513	.6082



Multinomial Logit Elasticities

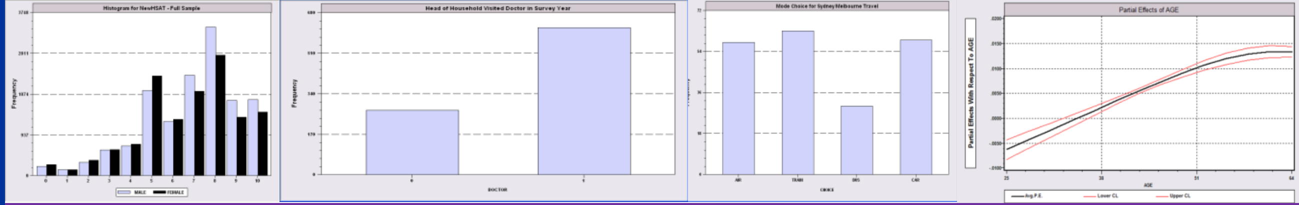
```
+---------------------------------------------------+
| Elasticity             averaged over observations.|
| Attribute is PRICE    in choice BRAND1            |
| Effects on probabilities of all choices in model: |
| * = Direct Elasticity effect of the attribute.    |
|                                   Mean    St.Dev  |
| *     Choice=BRAND1             -.8895    .3647   |
|       Choice=BRAND2              .2907    .2631   |
|       Choice=BRAND3              .2907    .2631   |
|       Choice=NONE                .2907    .2631   |
| Attribute is PRICE    in choice BRAND2            |
|       Choice=BRAND1              .3127    .1371   |
| *     Choice=BRAND2            -1.2216    .3135   |
|       Choice=BRAND3              .3127    .1371   |
|       Choice=NONE                .3127    .1371   |
| Attribute is PRICE    in choice BRAND3            |
|       Choice=BRAND1              .3664    .2233   |
|       Choice=BRAND2              .3664    .2233   |
| *     Choice=BRAND3             -.7548    .3363   |
|       Choice=NONE                .3664    .2233   |
+---------------------------------------------------+
```



HEV Model without Heterogeneity

```
+---------------------------------------------+
| Heteroskedastic Extreme Value Model         |
| Dependent variable               CHOICE     |
| Number of observations             3200     |
| Log likelihood function       -4151.611     |
| Response data are given as ind. choice.     |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
---------+Attributes in the Utility Functions (beta)
 FASH    |     1.57473345       .31427031     5.011   .0000
 QUAL    |     1.09208463       .22895113     4.770   .0000
 PRICE   |   -13.3740754       2.61275111    -5.119   .0000
 ASC4    |    -.01128916        .22484607     -.050   .9600
---------+Scale Parameters of Extreme Value Distns Minus 1.0
 s_BRAND1|      .03779175       .22077461      .171   .8641
 s_BRAND2|     -.12843300       .17939207     -.716   .4740
 s_BRAND3|      .01149458       .22724947      .051   .9597
 s_NONE  |       .000000     ......(Fixed Parameter).......
---------+Std.Dev=pi/(theta*sqr(6)) for H.E.V. distribution.
 s_BRAND1|     1.23584505       .26290748     4.701   .0000
 s_BRAND2|     1.47154471       .30288372     4.858   .0000
 s_BRAND3|     1.26797496       .28487215     4.451   .0000
 s_NONE  |     1.28254980    ......(Fixed Parameter).......
```

Essentially no differences in variances across choices



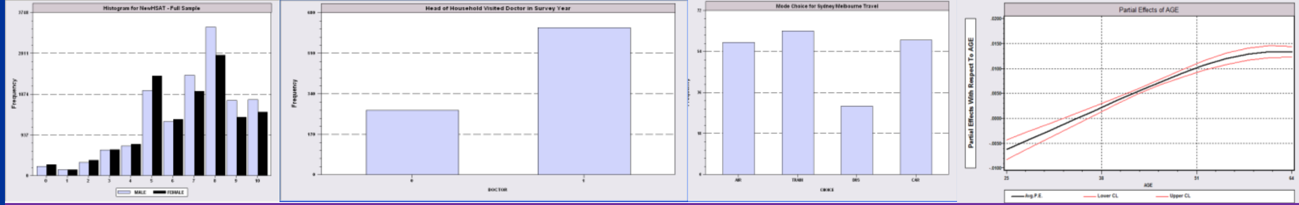
Homogeneous HEV Elasticities

Attribute is PRICE in choice BRAND1		Mean	St.Dev
*	Choice=BRAND1	-1.0585	.4526
	Choice=BRAND2	.2801	.2573
	Choice=BRAND3	.3270	.3004
	Choice=NONE	.3232	.2969
Attribute is PRICE in choice BRAND2			
	Choice=BRAND1	.3576	.1481
*	Choice=BRAND2	-1.2122	.3142
	Choice=BRAND3	.3466	.1426
	Choice=NONE	.3429	.1411
Attribute is PRICE in choice BRAND3			
	Choice=BRAND1	.4332	.2532
	Choice=BRAND2	.3610	.2116
*	Choice=BRAND3	-.8648	.4015
	Choice=NONE	.4156	.2436

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

Multinomial Logit

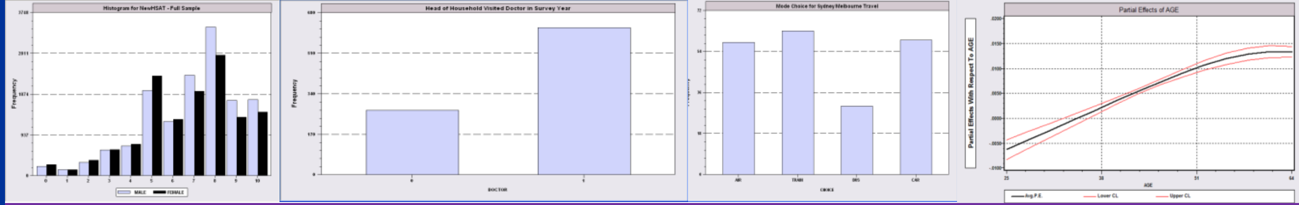
PRICE in choice BRAND1	Mean	St.Dev
*	-.8895	.3647
	.2907	.2631
	.2907	.2631
	.2907	.2631
PRICE in choice BRAND2		
	.3127	.1371
*	-1.2216	.3135
	.3127	.1371
	.3127	.1371
PRICE in choice BRAND3		
	.3664	.2233
	.3664	.2233
*	-.7548	.3363
	.3664	.2233



Heteroscedasticity Across Individuals

	Heteroskedastic Extreme Value Model	Homog-HEV	MNL
Log likelihood function	-4129.518[10]	-4151.611[7]	-4158.503[4]

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
Attributes in the Utility Functions (beta)				
FASH	1.01640726	.20261573	5.016	.0000
QUAL	.55668491	.11604080	4.797	.0000
PRICE	-7.44758292	1.52664112	-4.878	.0000
ASC4	.18300524	.09678571	1.891	.0586
Scale Parameters of Extreme Value Distributions				
s_BRAND1	.81114924	.10099174	8.032	.0000
s_BRAND2	.72713522	.08931110	8.142	.0000
s_BRAND3	.80084114	.10316939	7.762	.0000
s_NONE	1.00000000	......(Fixed Parameter).......		
Heterogeneity in Scales of Ext.Value Distns.				
MALE	.21512161	.09359521	2.298	.0215
AGE25	.79346679	.13687581	5.797	.0000
AGE39	.38284617	.16129109	2.374	.0176

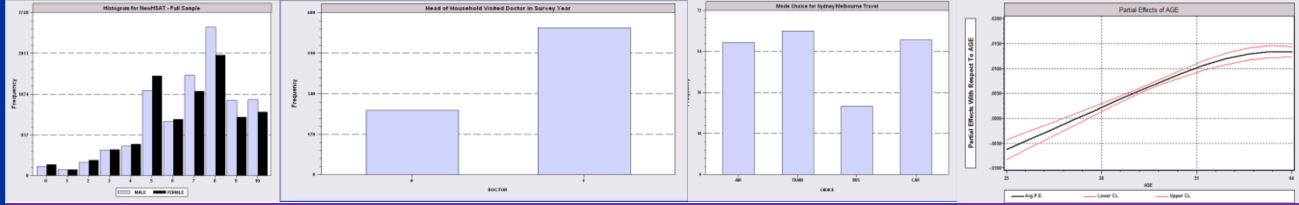


Variance Heterogeneity Elasticities

Attribute is PRICE		Mean	St.Dev
in choice BRAND1			
*	Choice=BRAND1	-.8978	.5162
	Choice=BRAND2	.2269	.2595
	Choice=BRAND3	.2507	.2884
	Choice=NONE	.3116	.3587
Attribute is PRICE in choice BRAND2			
	Choice=BRAND1	.2853	.1776
*	Choice=BRAND2	-1.0757	.5030
	Choice=BRAND3	.2779	.1669
	Choice=NONE	.3404	.2045
Attribute is PRICE in choice BRAND3			
	Choice=BRAND1	.3328	.2477
	Choice=BRAND2	.2974	.2227
*	Choice=BRAND3	-.7458	.4468
	Choice=NONE	.4056	.3025

Multinomial Logit

PRICE in choice BRAND1	Mean	St.Dev
*	-.8895	.3647
	.2907	.2631
	.2907	.2631
	.2907	.2631
PRICE in choice BRAND2		
	.3127	.1371
*	-1.2216	.3135
	.3127	.1371
	.3127	.1371
PRICE in choice BRAND3		
	.3664	.2233
	.3664	.2233
*	-.7548	.3363
	.3664	.2233



MARKETING SCIENCE
Vol. 29, No. 3, May–June 2010, pp. 393–421
ISSN 0732-2399 | EISSN 1526-548X | 10 | 2903 | 0393

informs®

DOI 10.1287/mksc.1090.0508

The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity

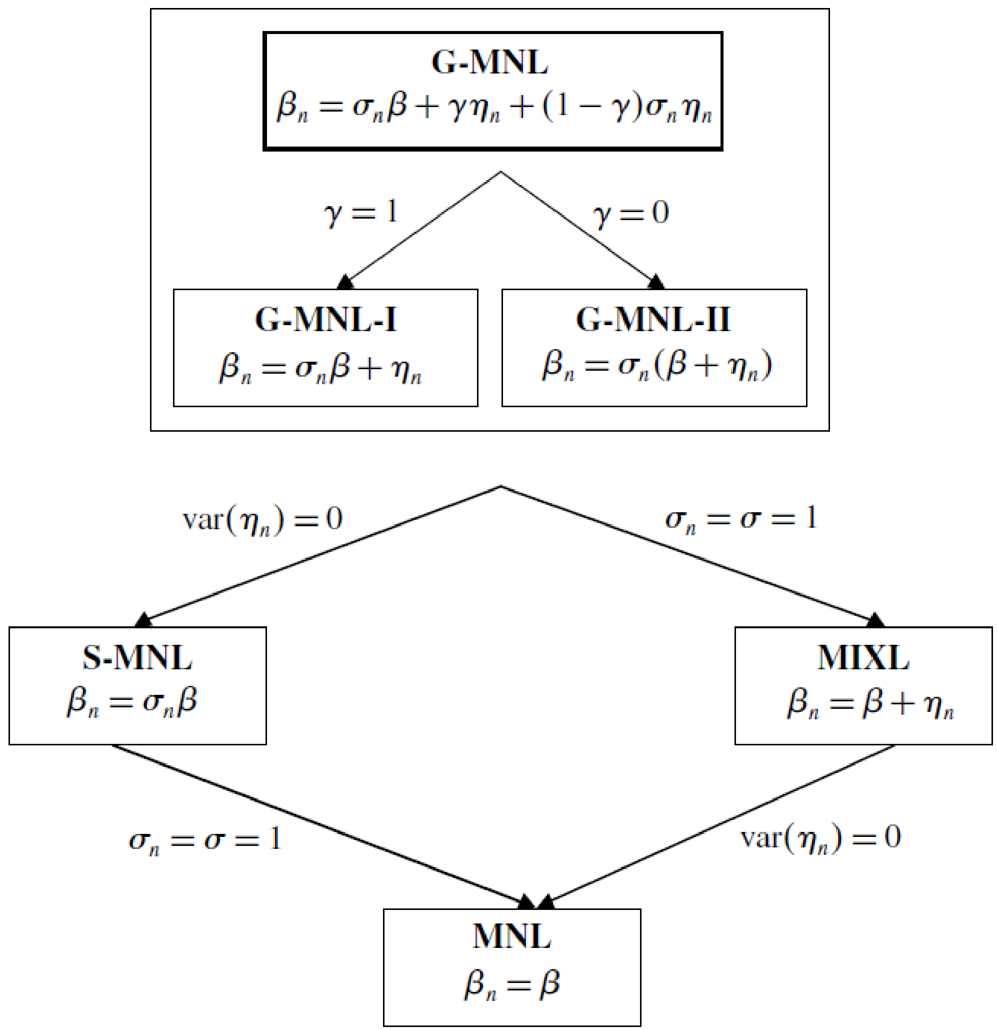
Denzil G. Fiebig
School of Economics, University of New South Wales, Sydney, New South Wales 2052, Australia, d.fiebig@unsw.edu.au

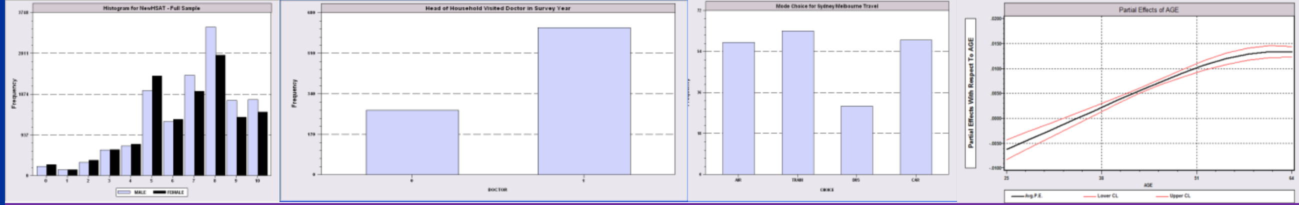
Michael P. Keane
University of Technology Sydney, Sydney, New South Wales 2007;
and Arizona State University, Tempe, Arizona 85287, michael.keane@uts.edu.au

Jordan Louviere
School of Marketing, Centre for the Study of Choice, University of Technology Sydney, Sydney, New South Wales 2007, Australia, jordan.louviere@uts.edu.au

Nada Wasi
School of Finance and Economics, Centre for the Study of Choice, University of Technology Sydney, Sydney, New South Wales 2007, Australia, nada.wasi@uts.edu.au

Figure 1 The G-MNL Model and Its Special Cases





Generalized Mixed Logit Model

$$U(i,t,j) = \boldsymbol{\beta}_i'\mathbf{x}_{i,t,j} + \text{Common effects} + \varepsilon_{i,t,j}$$

Random Parameters

$$\boldsymbol{\beta}_i = \sigma_i[\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{h}_i] + [\gamma + \sigma_i(1-\gamma)]\boldsymbol{\Gamma}_i\mathbf{v}_i$$

$$\boldsymbol{\Gamma}_i = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_i$$

$\boldsymbol{\Lambda}$ is a lower triangular matrix

with 1s on the diagonal (Cholesky matrix)

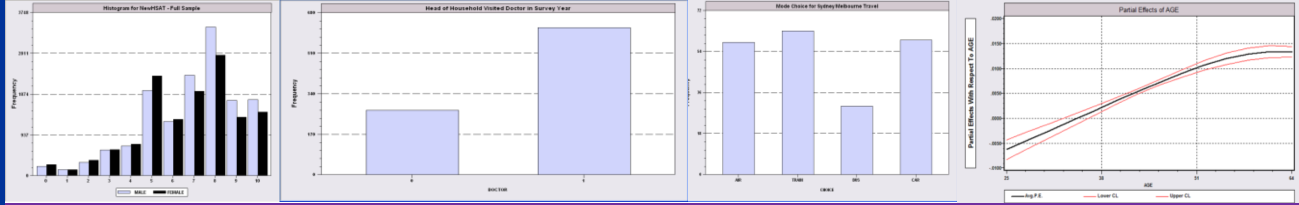
$\boldsymbol{\Sigma}_i$ is a diagonal matrix with $\varphi_k \exp(\boldsymbol{\psi}_k'\mathbf{h}_i)$

Overall preference scaling

$$\sigma_i = \exp(-\tau_i^2 / 2 + \tau_i w_i + \boldsymbol{\theta}'\mathbf{h}_i]$$

$$\tau_i = \exp(\boldsymbol{\lambda}'\mathbf{r}_i)$$

$$0 < \gamma < 1$$



Unobserved Heterogeneity in Scaling

HEV formulation: $U_{it,j} = \boldsymbol{\beta}'\mathbf{x}_{itj} + (1/\sigma_i)\varepsilon_{it,j}$

Generalized model with $\gamma = 1$ and $\boldsymbol{\Gamma} = [\mathbf{0}]$.

Produces a scaled multinomial logit model with $\boldsymbol{\beta}_i = \sigma_i\boldsymbol{\beta}$

$$\text{Prob(choice}_{it} = \text{j}) = \frac{\exp(\boldsymbol{\beta}_i'\mathbf{x}_{itj})}{\sum_{j=1}^{J_{it}}\exp(\boldsymbol{\beta}_i'\mathbf{x}_{itj})}, i = 1,\ldots,N,\ j = 1,\ldots,J_{it}, t = 1,\ldots,T_i$$

The random variation in the scaling is

$$\sigma_i = \exp(-\tau^2/2 + \tau w_i)$$

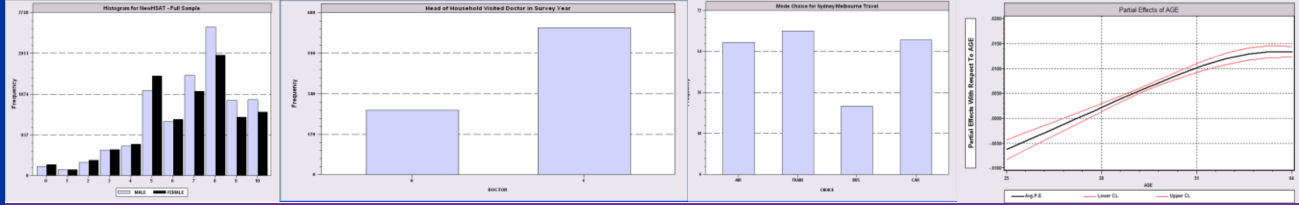
The variation across individuals may also be observed, so that

$$\sigma_i = \exp(-\tau^2/2 + \tau w_i + \boldsymbol{\delta}'\mathbf{z}_i)$$

Scaled MNL

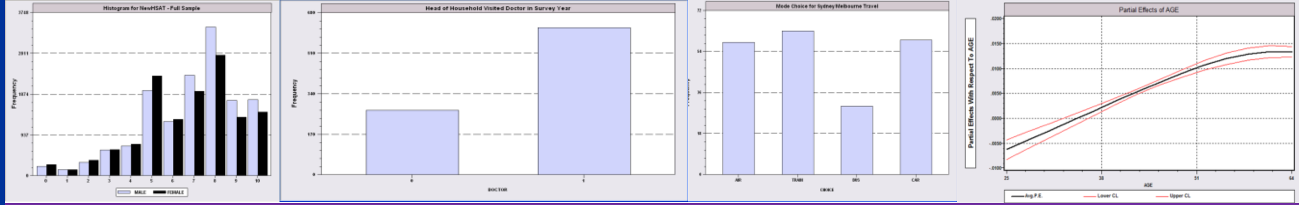
```
-----------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                  Choice
Log likelihood function         -172.94366
Estimation based on N =    210, K =   10
Inf.Cr.AIC  =     365.9 AIC/N =      1.742
Model estimated: Mar 17, 2012, 14:03:50
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
--------+--------------------------------------------------
        |                 Standard
    MODE|  Coefficient      Error              z
--------+--------------------------------------------------
      GC|     .07578***      .01833          4.13
    TTME|    -.10289***      .01109    |    -9.28
    INVC|    -.08044***      .01995         -4.03
    INVT|    -.01399***      .00267         -5.24
   A_AIR|    4.37035***     1.05734          4.13
AIR_HIN1|     .00428         .01306           .33
 A_TRAIN|    5.91407***      .68993          8.57
TRA_HIN2|    -.05907***      .01471         -4.02
   A_BUS|    4.46269***      .72333          6.17
BUS_HIN3|    -.02295         .01592         -1.44
--------+--------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%,
-----------------------------------------------------------


Elasticity wrt change of X in row choice on P
--------+--------------------------------------------------
GC      |      AIR     TRAIN       BUS       CAR
--------+--------------------------------------------------
     AIR|   5.4136   -2.3647   -2.3647   -2.3647
   TRAIN|  -2.4293    7.4368   -2.4293   -2.4293
     BUS|  -1.1513   -1.1513    7.5825   -1.1513
     CAR|  -1.9222   -1.9222   -1.9222    5.3080
```

```
-----------------------------------------------------------
Scaled Multinomial Logit Model
Dependent variable                         MODE
Log likelihood function          -164.66285
Restricted log likelihood        -291.12182
Chi squared [   11 d.f.]          252.91793
Significance level                   .00000
McFadden Pseudo R-squared           .4343851
Estimation based on N =     210, K =   11
Inf.Cr.AIC  =    351.3 AIC/N =      1.673
Response data are given as ind. choices
Replications for simulated probs. =   50
Halton sequences used for simulations
BHHH estimator used for asymp. variance
--------+--------------------------------------------------
        |                 Standard
    MODE|  Coefficient      Error              z
--------+--------------------------------------------------
        |Scaled parameters in utility functio
      GC|     .19759***      .05816          3.40
    TTME|    -.34526***      .04137         -8.35
    INVC|    -.21990***      .06310         -3.48
    INVT|    -.03438***      .00898         -3.83
   A_AIR|   16.1670***      3.37933          4.78
AIR_HIN1|     .03846         .03567          1.08
 A_TRAIN|   17.3511***      2.48222          6.99
TRA_HIN2|    -.13911***      .04336         -3.21
   A_BUS|   12.8715***      2.52222          5.10
BUS_HIN3|    -.02154         .07031          -.31
        |Variance parameter tau in GMX scale :
TauScale|    1.09975***      .09099         12.09
        |  Sample Mean      Sample Std.Dev.
Sigma(i)|     .98688        1.38648
--------+--------------------------------------------------
Elasticity wrt change of X in row choice on P
--------+--------------------------------------------------
GC      |      AIR     TRAIN       BUS       CAR
--------+--------------------------------------------------
     AIR|   4.0909   -1.1938   -1.1111   -1.5117
   TRAIN|  -1.3356    5.7020   -1.4903   -1.9511
     BUS|   -.9067   -1.0550    4.6017   -1.0007
     CAR|  -1.9987   -1.9777   -1.5731    3.9486
```



Observed and Unobserved Heterogeneity

```
-----------------------------------------------------------------------------
Scaled Multinomial Logit Model
Dependent variable               CHOICE
Log likelihood function     -4088.80526
Restricted log likelihood   -4436.14196
Chi squared [   7](P= .000)    694.67339
Significance level               .00000
McFadden Pseudo R-squared      .0782970
Estimation based on N =   3200, K =   7
Inf.Cr.AIC  =    8191.6 AIC/N =    2.560
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients -4436.1420  .0783 .0776
Constants only  -4391.1804  .0689 .0682
At start values -4155.7814  .0161 .0154
Response data are given as ind. choices
RPL model with panel has      400 groups
Fixed number of obsrvs./group=        8
BHHH estimator used for asymp. variance
The model has 1 levels.
Number of obs.=  3200, skipped    0 obs
--------+--------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
  CHOICE|  Coefficient     Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Attributes in the Utility Functions (beta)
    FASH|    3.08282***     .64663     4.77  .0000     1.81545    4.35019
    QUAL|     .00036        .09386      .00  .9969     -.18360     .18432
   PRICE|   -4.85162***     .09386   -51.69  .0000    -5.03558   -4.66767
    ASC4|    2.23299***     .06135    36.40  .0000     2.11276    2.35323
        |Variance parameter tau in scaled MNL model
TauScale|    1.99650***     .12248    16.30  .0000     1.75645    2.23656
        |Heterogeneity in tau(i)
 TauMALE|     .61759***     .02211    27.94  .0000      .57426     .66092
TauAGE39|    -.59159***     .04874   -12.14  .0000     -.68710    -.49607
        |  Sample Mean    Sample Std.Dev.
Sigma(i)|    1.04712       3.13241      .33  .7382    -5.09230    7.18653
--------+--------------------------------------------------------------------
```



Price Elasticities

SCALED MNL

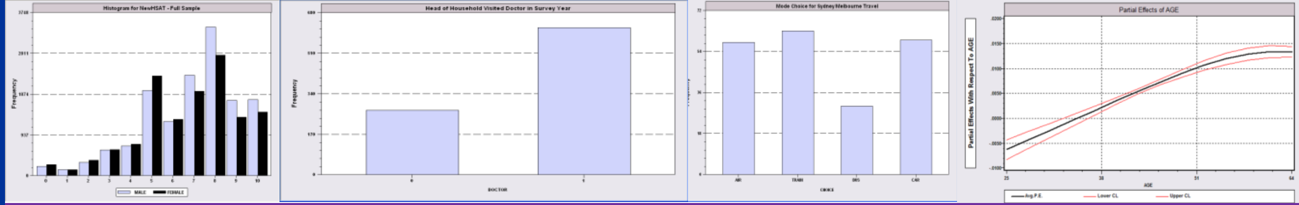
Elasticity wrt change of X in row choice on Prob[column choice]

PRICE	BRAND1	BRAND2	BRAND3	NONE
BRAND1	-.3805	.1047	.1047	.1047
BRAND2	.1492	-.4815	.1492	.1492
BRAND3	.1348	.1348	-.3261	.1348
NONE	.0000	.0000	.0000	.0000

MNL

Elasticity wrt change of X in row choice on Prob[column choice]

PRICE	BRAND1	BRAND2	BRAND3	NONE
BRAND1	-.8895	.2907	.2907	.2907
BRAND2	.3127	-1.2216	.3127	.3127
BRAND3	.3664	.3664	-.7548	.3664
NONE	.0000	.0000	.0000	.0000

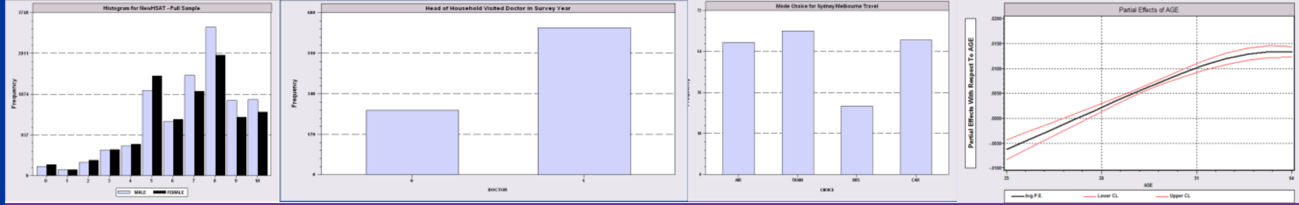


Scaling as Unobserved Heterogeneity

Removing the Scale Factor Confound in Multinomial Logit Choice Models to Obtain Better Estimates of Preference[1]

Jay Magidson
Statistical Innovations Inc.

Jeroen K. Vermunt
Tilburg University

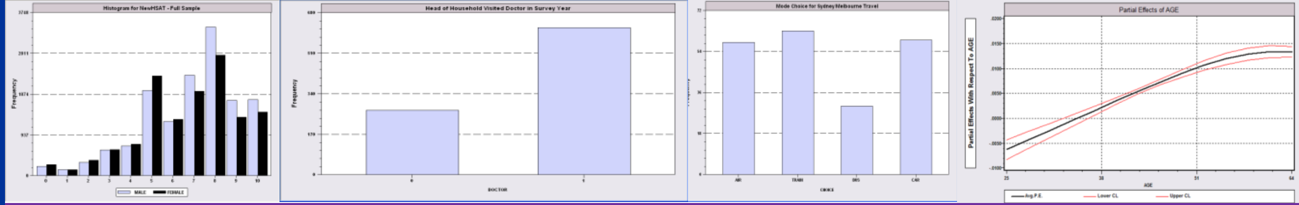


Two Way Latent Class?

	Less certain subgroup (s = 1): λ[s]=1	More certain subgroup (s = 2): λ[s]=2
Segment 1 (X*=1): $\beta_1 = (0.5, 0.0, -0.5)$	joint class (1,1)	joint class (1,2)
Segment 2 (X*=2): $\beta_1 = (-0.2, -0.8, 1.0)$	joint class (2,1)	joint class (2,2)

$$\text{Prob(Choice=j|class=c,risk=s)}=\frac{\exp[\sigma_s \boldsymbol{\beta}_c' \mathbf{x}_j]}{\sum_{alt=1}^{J} \exp[\sigma_s \boldsymbol{\beta}_c' \mathbf{x}_{alt}]}, \quad \sigma_1 = 1$$

$$\text{Prob(Choice=j)}=\sum_{c=1}^{C} \sum_{s=1}^{S} \pi(c,s) \frac{\exp[\sigma_s \boldsymbol{\beta}_c' \mathbf{x}_j]}{\sum_{alt=1}^{J} \exp[\sigma_s \boldsymbol{\beta}_c' \mathbf{x}_{alt}]}$$

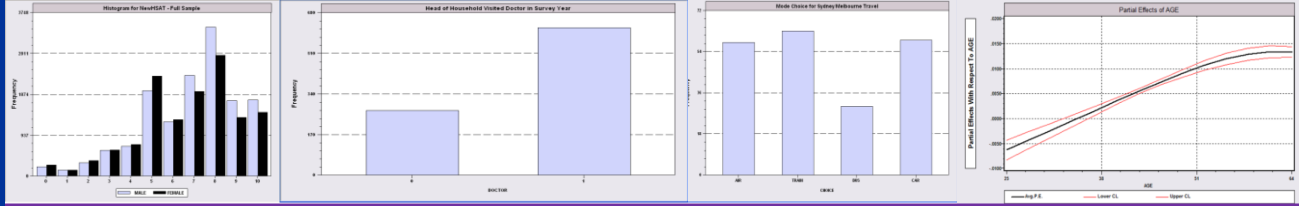


Appendix: Maximum Simulated Likelihood

$$\log L(\boldsymbol{\theta}, \boldsymbol{\Omega}) =$$

$$\sum_{i=1}^{N} \log \int_{\boldsymbol{\beta}_i} \prod_{t=1}^{T} f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\beta}_i, \boldsymbol{\theta}) h(\boldsymbol{\beta}_i \mid \mathbf{z}_i, \boldsymbol{\Omega}) d\boldsymbol{\beta}_i$$

$$\boldsymbol{\Omega} = \boldsymbol{\beta}, \boldsymbol{\Delta}, \phi_{\mathbf{1}}, \ldots, \phi_K, \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_K, \boldsymbol{\Gamma}$$



Monte Carlo Integration

(1) Integral is of the form

$$K = \int_{\text{range of } v} g(v|\text{data},\boldsymbol{\beta})\, f(v|\boldsymbol{\Omega})\, dv$$

where f(v) is the density of random variable v possibly conditioned on a set of parameters $\boldsymbol{\Omega}$ and $g(v|\text{data},\boldsymbol{\beta})$ is a function of data and parameters.

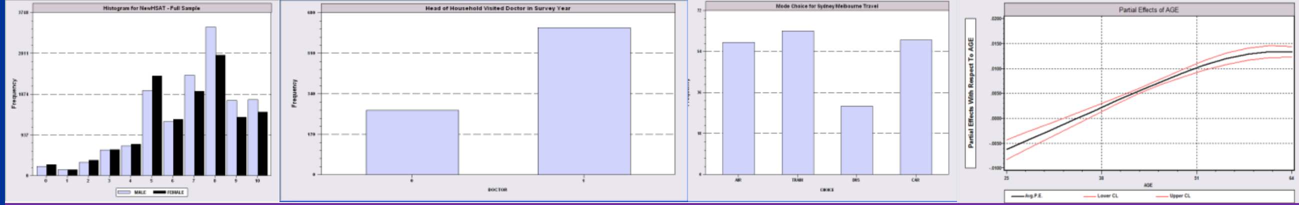
(2) By construction, $K(\boldsymbol{\Omega}) = E[g(v|\text{data},\boldsymbol{\beta})]$

(3) Strategy:

a. Sample R values from the population of v using a random number generator.

b. Compute average $\bar{K} = (1/R)\sum_{r=1}^{R} g(v_r|\text{data},\boldsymbol{\beta})$

By the law of large numbers, plim $\bar{K}$ = K.



Monte Carlo Integration

$$\frac{1}{R}\sum_{r=1}^{R} f(u_{ir}) \xrightarrow{P} \int_{u_i} f(u_i)g(u_i)du_i = E_{u_i}[f(u_i)]$$

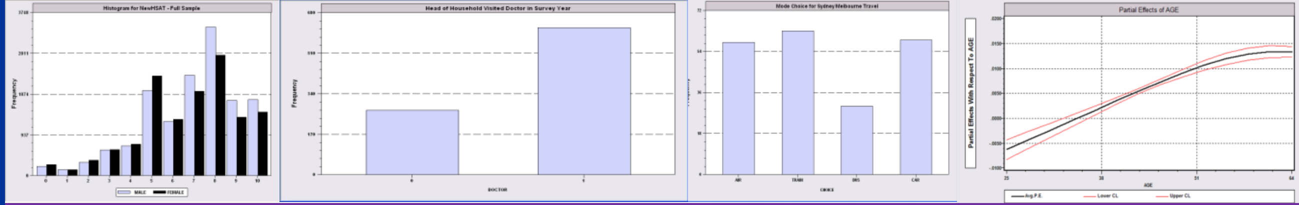
(Certain smoothness conditions must be met.)

Drawing u_{ir} by 'random sampling'

$u_{ir} = t(v_{ir}),\ v_{ir} \sim U[0,1]$

E.g., $u_{ir} = \sigma\Phi^{-1}(v_{ir}) + \mu$ for $N[\mu,\sigma^2]$

Requires many draws, typically hundreds or thousands



Example: Monte Carlo Integral

$$\int_{-\infty}^{\infty} \Phi(x_1 + .9v)\Phi(x_2 + .9v)\Phi(x_3 + .9v)\frac{\exp(-v^2/2)}{\sqrt{2\pi}}dv$$

where Φ is the standard normal CDF and

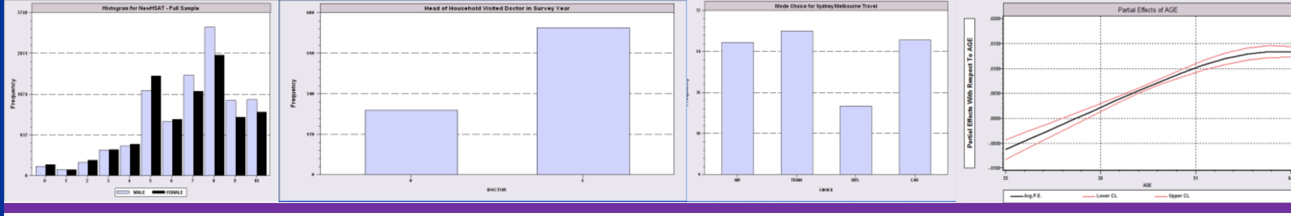
$x_1 = .5$, $x_2 = -.2$, $x_3 = .3$.

The weighting function for v is the standard normal.

Strategy: Draw R (say 1000) standard normal random draws, v_r. Compute the 1000 functions

$\Phi(x_1 + .9v_r)\Phi(x_2 + .9v_r)\Phi(x_3 + .9v_r)$ and average them.

(Based on 100, 1000, 10000, I get .28746, .28437, .27242)



Simulated Log Likelihood for a Mixed Probit Model

Random parameters probit model

$$f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\beta}_i) = \Phi[(2y_{it} - 1)\mathbf{x}_{it}'\boldsymbol{\beta}_i]$$

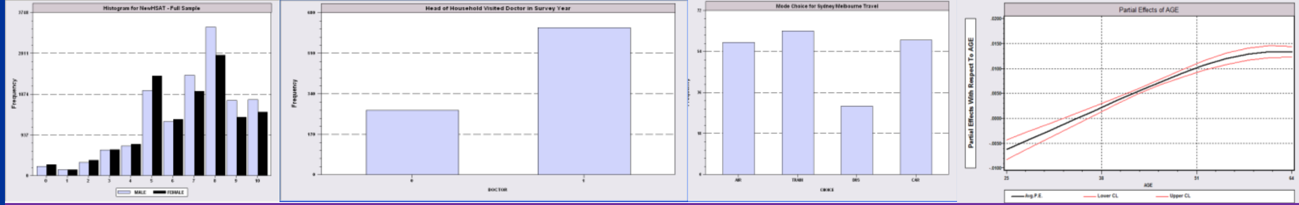
$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i$$

$$\mathbf{u}_i \sim N[\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Lambda}^2\boldsymbol{\Gamma}']$$

$$\text{LogL}(\boldsymbol{\beta},\boldsymbol{\Gamma},\boldsymbol{\Lambda}) = \sum_{i=1}^{N} \log \int_{\boldsymbol{\beta}_i} \prod_{t=1}^{T_i} \Phi[(2y_{it} - 1)\mathbf{x}_{it}'\boldsymbol{\beta}_i]\, N[\boldsymbol{\beta}, \boldsymbol{\Gamma}\boldsymbol{\Lambda}^2\boldsymbol{\Gamma}']d\boldsymbol{\beta}_i$$

$$\text{LogL}^S = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} \Phi[(2y_{it} - 1)\mathbf{x}_{it}'(\boldsymbol{\beta} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}\mathbf{v}_{ir})]$$

We now maximize this function with respect to $(\boldsymbol{\beta},\boldsymbol{\Gamma},\boldsymbol{\Lambda})$.



Generating Random Draws

Most common approach is the "inverse probability transform"

Let u = a random draw from the standard uniform (0,1).

Let x = the desired population to draw from

Assume the CDF of x is $F(x)$.

The random draw is then $x = F^{-1}(u)$.

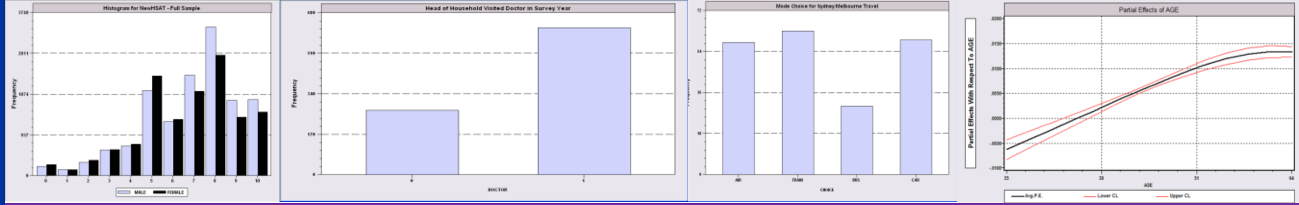
Example : exponential, θ. $f(x)=\theta\exp(-\theta x)$, $F(x)=1-\exp(-\theta x)$

Equate u to $F(x)$, $x = -(1/\theta)\log(1-u)$.

Example: Normal(μ,σ). Inverse function does not exist in closed form. There are good polynomial approximations to produce a draw from N[0,1] from a U(0,1).

Then $x = \mu+\sigma v$.

This leaves the question of how to draw the U(0,1).



Drawing Uniform Random Numbers

Computer generated random numbers are not random; they are Markov chains that look random.

The Original IBM SSP Random Number Generator for 32 bit computers.

SEED originates at some large odd number

d3 = 2147483647.0

d2 = 2147483655.0

d1=16807.0

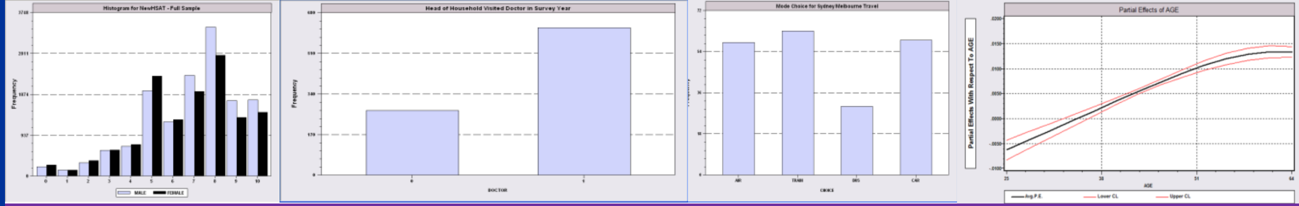
SEED=Mod(d1*SEED,d3) ! MOD(a,p) = a - INT(a/p) * p

X=SEED/d2 is a pseudo-random value between 0 and 1.

Problems:

(1) Short period. Based on 32 bits, so recycles after $2^{31}-1$ values

(2) Evidently not very close to random. (Recent tests have discredited this RNG)



Quasi-Monte Carlo Integration Based on Halton Sequences

Coverage of the unit interval is the objective,

not randomness of the set of draws.

Halton sequences --- Markov chain

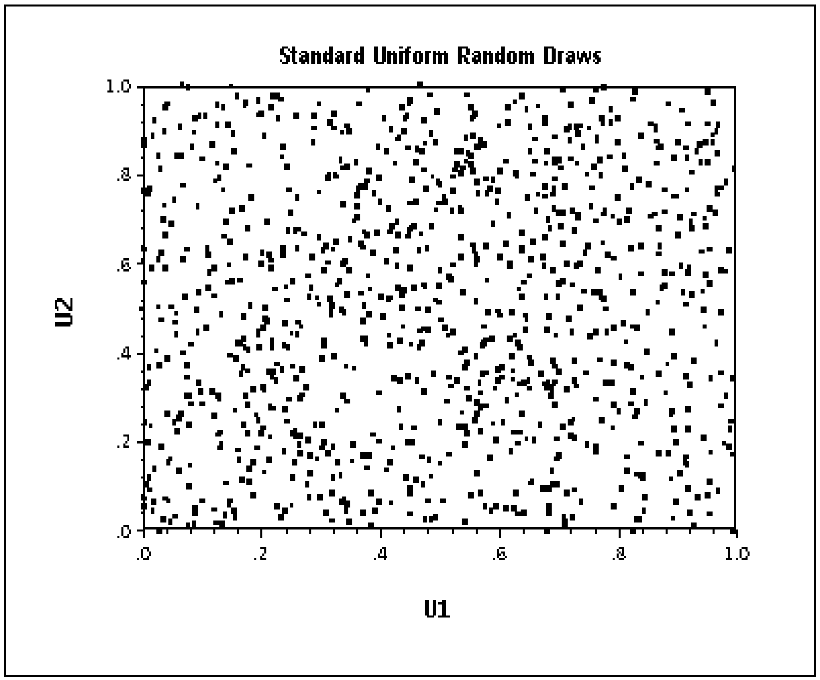
p = a prime number,

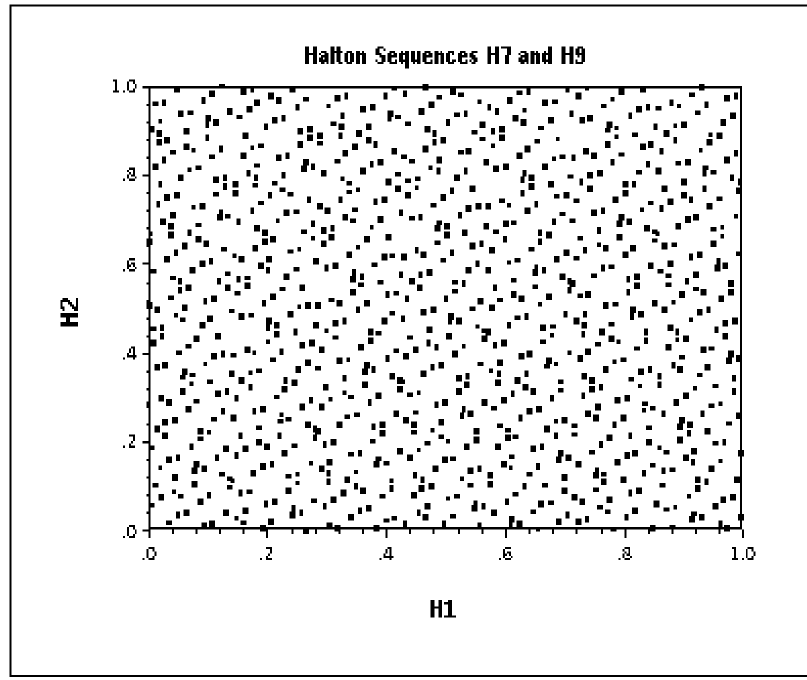
r= the sequence of integers, decomposed as $\sum_{i=0}^{I} b_i p^i$

H(r|p) $= \sum_{i=0}^{I} b_i p^{-i-1}$, r = r_1,... (e.g., 10,11,12,...)

For example, using base p=5, the integer r=37 has b_0 = 2, b_1 = 2, and b_2 = 1; ($37=1x5^2 + 2x5^1 + 2x5^0$). Then H(37|5) = $2\times5^{-1} + 2\times5^{-2} + 1\times5^{-3}$ = 0.448.

Halton Sequences vs. Random Draws





Requires far fewer draws – for one dimension, about 1/10. Accelerates estimation by a factor of 5 to 10.