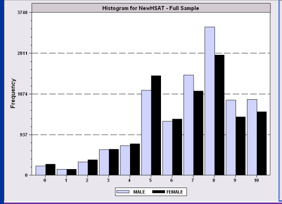
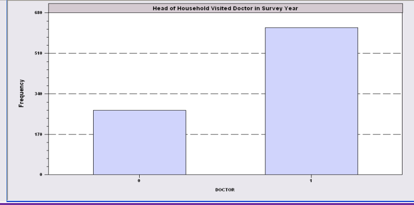
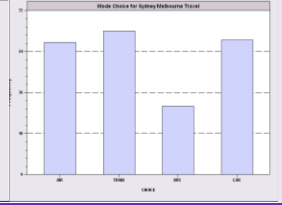
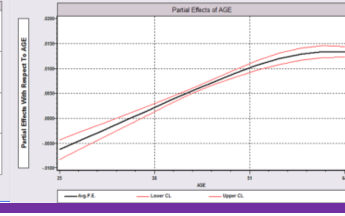
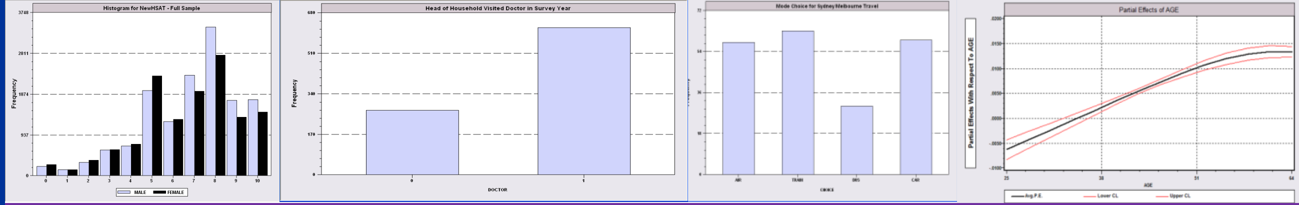
   

Discrete Choice Modeling

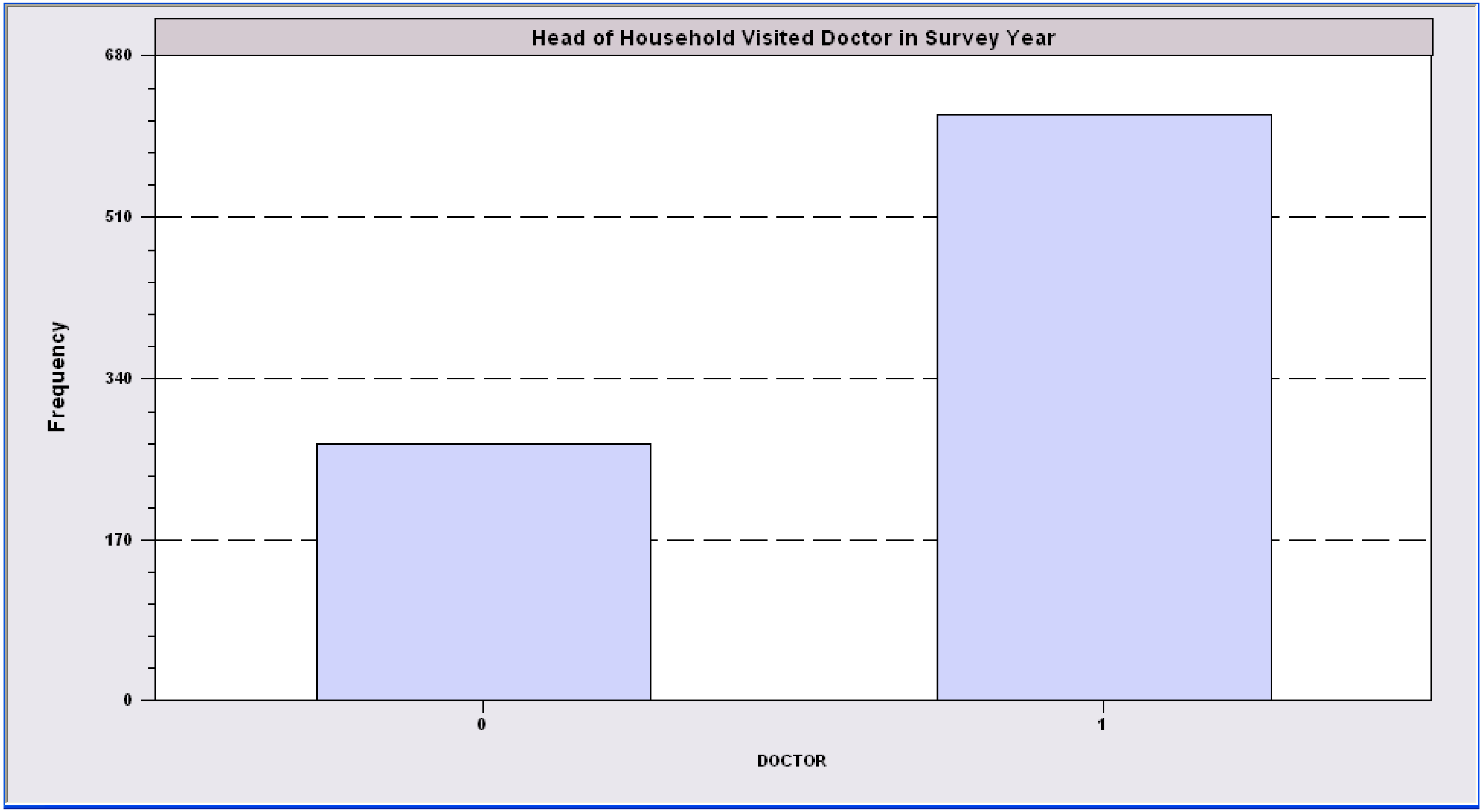
William Greene
Stern School of Business
New York University

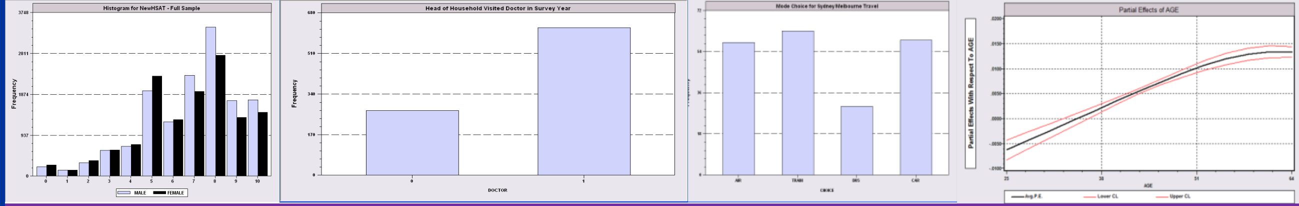


A Random Utility Approach

- Underlying Preference Scale, U*(choices)
- Revelation of Preferences:
  - U*(choices) $\leq$ 0 $\rightarrow$ Choice “0”
  - U*(choices) > 0 $\rightarrow$ Choice “1”

Binary Outcome: Visit Doctor





A Model for Binary Choice

- Yes or No decision (Buy/NotBuy, Do/NotDo)
- Example, choose to visit physician or not
- Model: Net utility of visit at least once

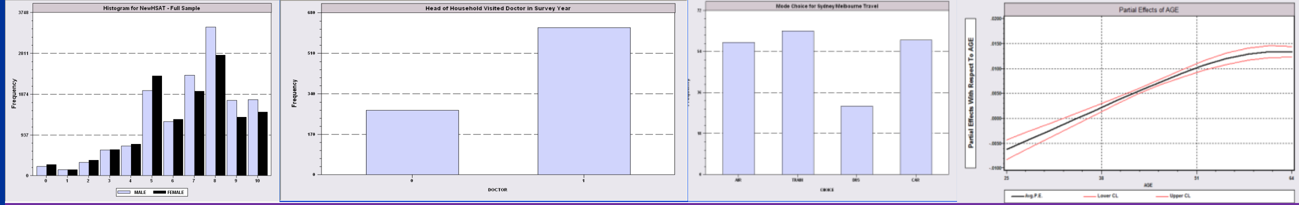
$$U_{visit} = \alpha+\beta_1 Age + \beta_2 Income + \gamma Sex + \varepsilon$$

Random Utility

Choose to visit if net utility is positive

Net utility = $U_{visit} - U_{not\ visit}$

- Data: **X** = [1,age,income,sex]
  y = 1 if choose visit, $\Leftrightarrow U_{visit} > 0$, 0 if not.



Choosing Between the Two Alternatives

Modeling the Binary Choice

$$\text{Net Utility} = U_{\text{visit}} - U_{\text{not visit}}$$

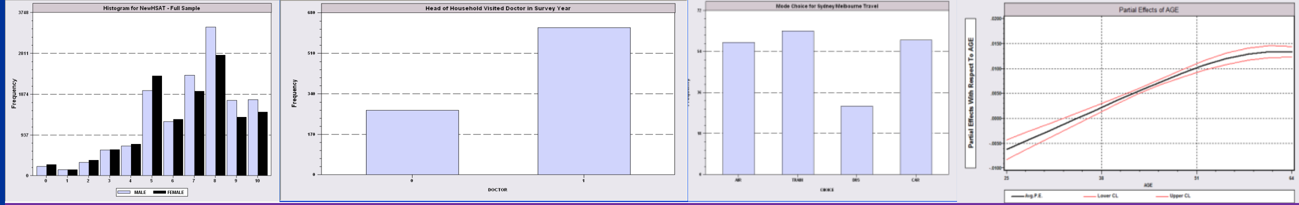
$$\text{Normalize} \quad U_{\text{not visit}} = 0$$

$$\text{Net } U_{\text{visit}} = \alpha + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_3 \text{Sex} + \varepsilon$$

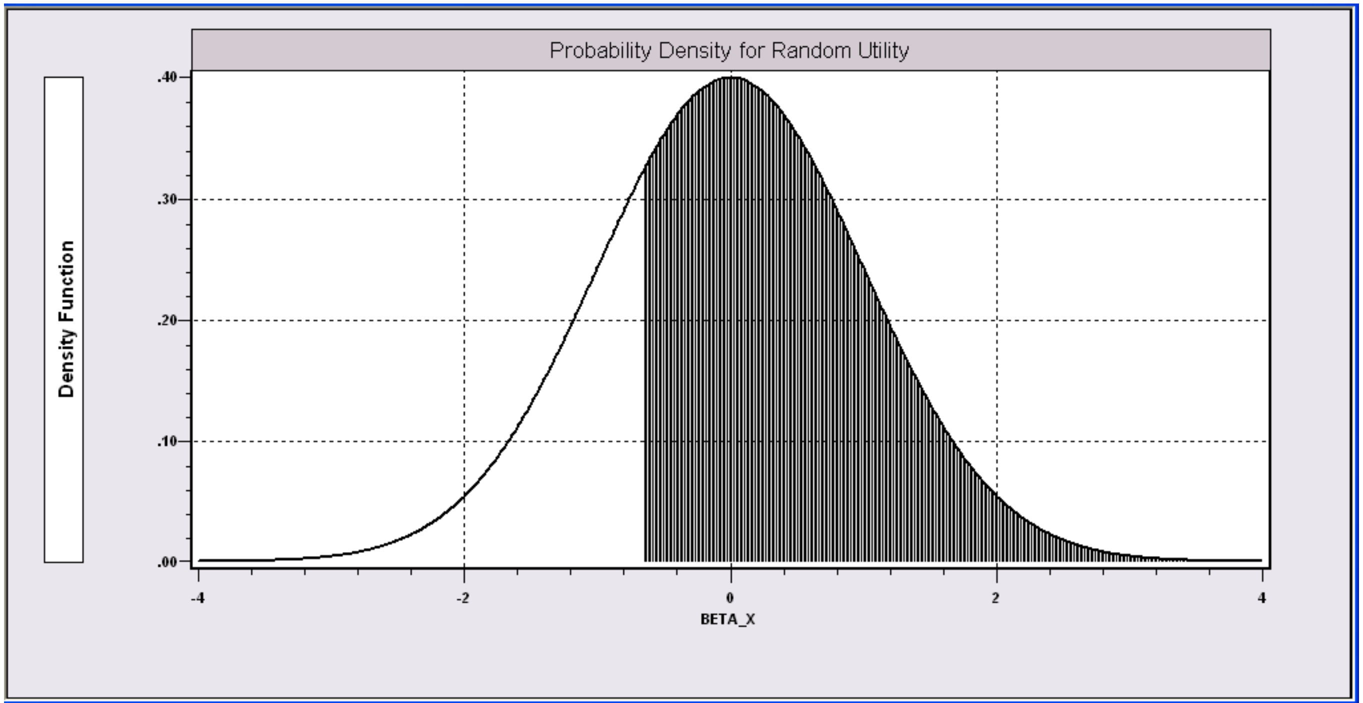
Chooses to visit: $U_{\text{visit}} > 0$

$$\alpha + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_3 \text{Sex} + \varepsilon > 0$$

$$\varepsilon > -[\alpha + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_3 \text{Sex}]$$

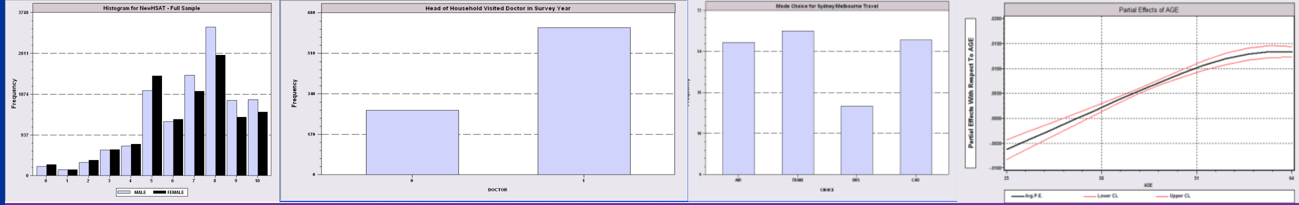


Probability Model for Choice Between Two Alternatives



Probability is governed by ε, the random part of the utility function.

$$\varepsilon > -[\alpha + \beta_1 Age + \beta_2 Income + \beta_3 Sex]$$



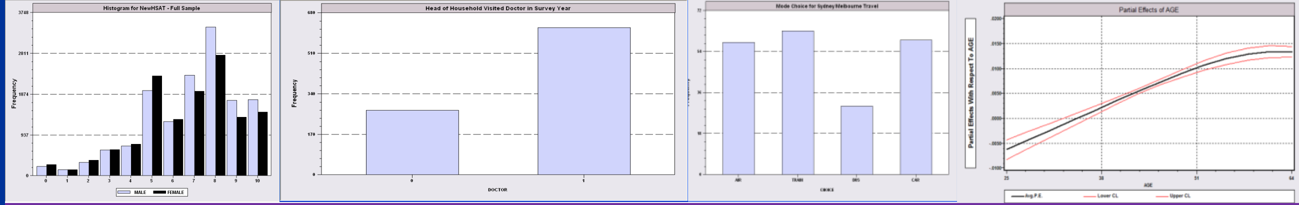
Application

27,326 Observations

- 1 to 7 years, panel
- 7,293 households observed
- We use the 1994 year, 3,337 household observations

Descriptive Statistics for 4 variables

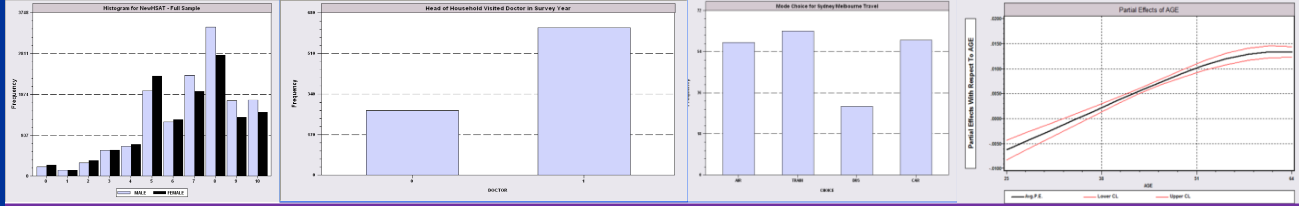
Variable	Mean	Std.Dev.	Minimum	Maximum	Cases	Missing
DOCTOR	.657980	.474456	0.0	1.0	3377	0
AGE	42.62659	11.58599	25.0	64.0	3377	0
INCOME	.444764	.216586	.034000	3.0	3377	0
FEMALE	.463429	.498735	0.0	1.0	3377	0



Binary Choice Data

Listing of current sample

Line	Observation	ID	DOCTOR	AGE	INCOME	FEMALE
1	1	1	1	54	.30500	0
2	2	1	0	55	.45101	0
3	3	1	0	56	.35000	0
4	4	2	0	44	.30500	1
5	5	2	1	45	.31828	1
6	6	2	1	46	.35000	1
7	7	2	1	48	.35305	1
8	8	3	0	58	.14340	1
9	9	3	0	60	.30000	1
10	10	3	1	61	.11000	1
11	11	3	1	62	.10000	1
12	12	4	1	29	.13000	1
13	13	5	1	27	.06500	0
14	14	5	1	28	.06000	0
15	15	5	0	31	.15500	0
16	16	6	1	25	.16000	0
17	17	6	1	26	.30000	0
18	18	6	0	27	.30000	0
19	19	6	1	28	.20000	0
20	20	6	1	31	.18000	0
21	21	7	0	26	.30000	1
22	22	7	0	27	.20000	1
23	23	7	1	30	.18000	1
24	24	8	1	64	.15000	0
25	25	9	1	30	.24000	0



An Econometric Model

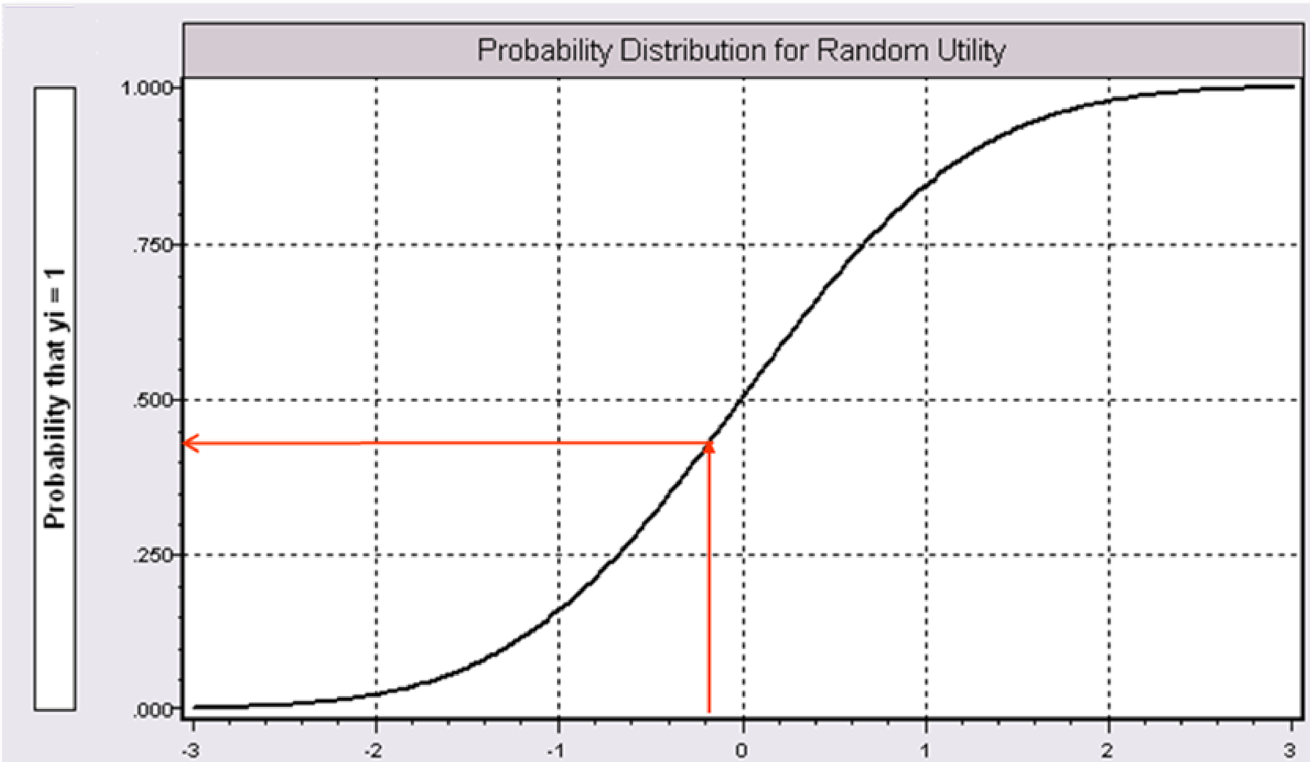
- Choose to visit iff $U_{visit} > 0$
  - $U_{visit} = \alpha + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_3 \text{Sex} + \varepsilon$
  - $U_{visit} > 0 \Leftrightarrow \varepsilon > -(\alpha + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_3 \text{Sex})$

    $\varepsilon < \alpha + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_3 \text{Sex}$

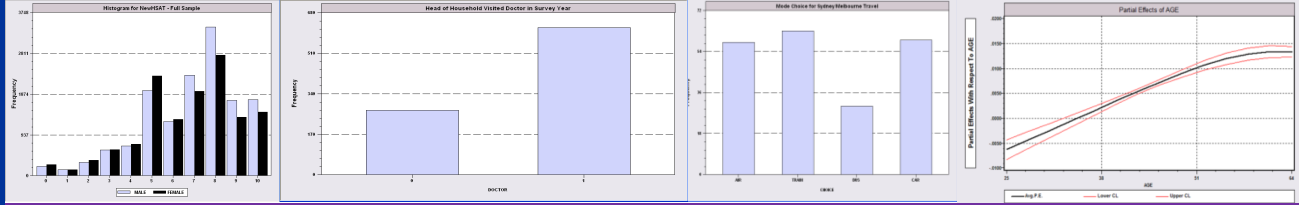
- Probability model: For any person observed by the analyst,

  $\text{Prob(visit)} = \text{Prob}[\varepsilon \leq \alpha + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_3 \text{Sex}]$

- Note the relationship between the unobserved ε and the outcome

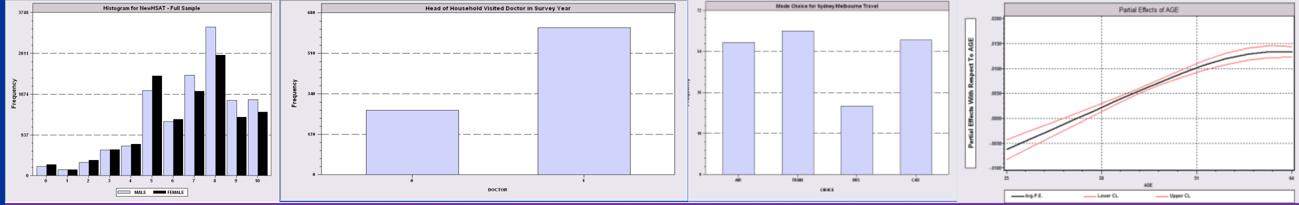


$\alpha+\beta_1 \text{Age} + \beta_2 \text{ Income} + \beta_3 \text{ Sex}$

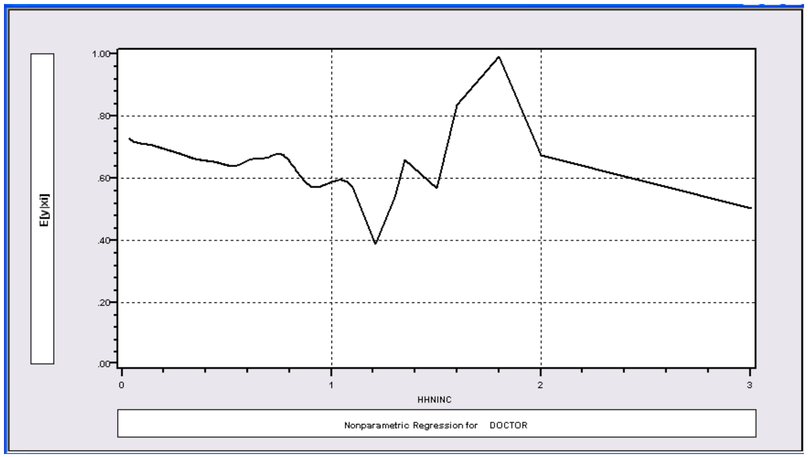


Modeling Approaches

- **Nonparametric – “relationship”**
  - Minimal Assumptions
  - Minimal Conclusions
- **Semiparametric – “index function”**
  - Stronger assumptions
  - Robust to model misspecification (heteroscedasticity)
  - Still weak conclusions
- **Parametric – “Probability function and index”**
  - Strongest assumptions – complete specification
  - Strongest conclusions
  - Possibly less robust. (Not necessarily)
- **The linear probability “model”**

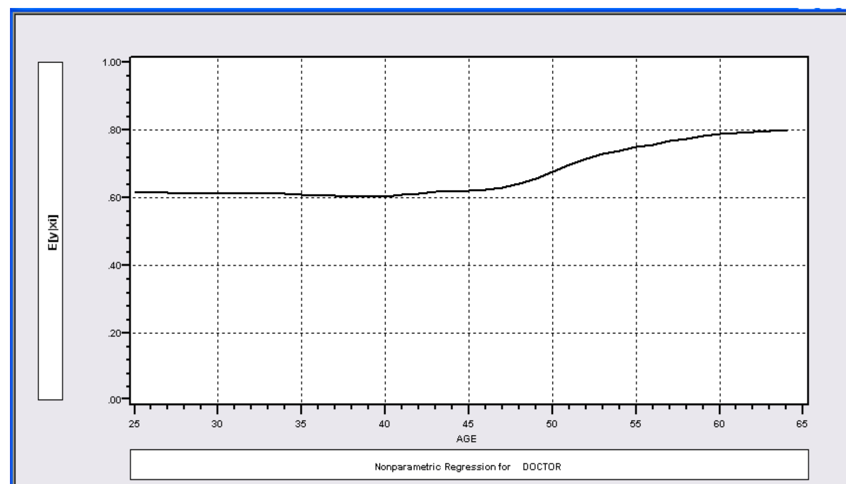


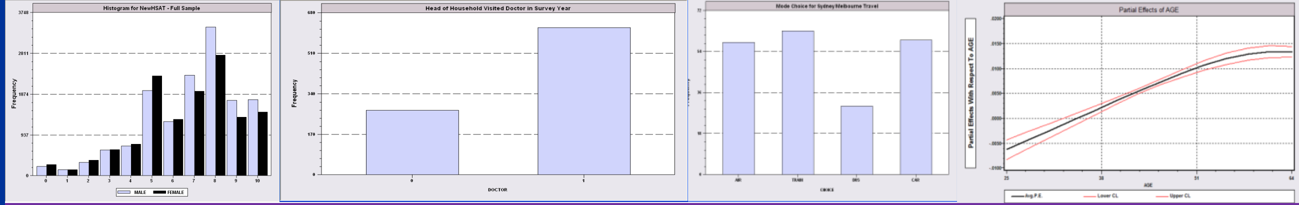
Nonparametric Regressions



P(Visit)=f(Income)

P(Visit)=f(Age)



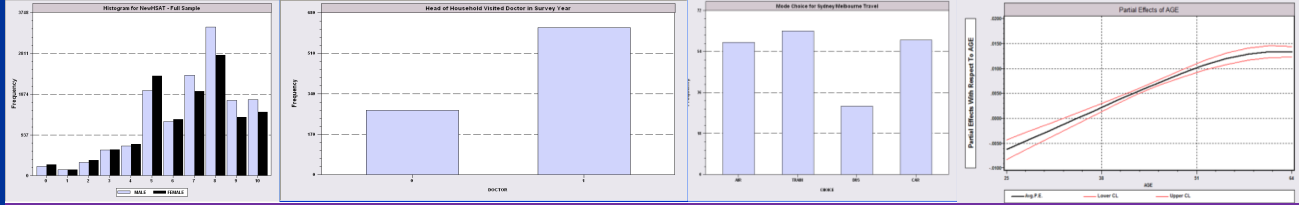


Klein and Spady Semiparametric No specific distribution assumed

```
-----------------------------------------------------------------------------
Semiparametric Binary Choice Model
Dependent variable                   DOCTOR
Log likelihood function         -2902.04013
Restricted log likelihood       -2908.96085
Chi squared [   2 d.f.]            13.84142
Significance level                   .00099
McFadden Pseudo R-squared          .0023791
Estimation based on N =    4481, K =      2
Inf.Cr.AIC  =     5808.1 AIC/N =      1.296
Logistic kernel fn. Bandwidth =      .24628
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient     Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[DOCTOR=1]
     AGE|     .03136**     .01269     2.47   .0134      .00649      .05622
  INCOME|    -.30808       .88120     -.35   .7266    -2.03519     1.41903
  FEMALE|     1.0     .....(Fixed Parameter).....
Constant|      0.0    .....(Fixed Parameter).....
--------+--------------------------------------------------------------------
```

Note necessary normalizations. Coefficients are relative to FEMALE.

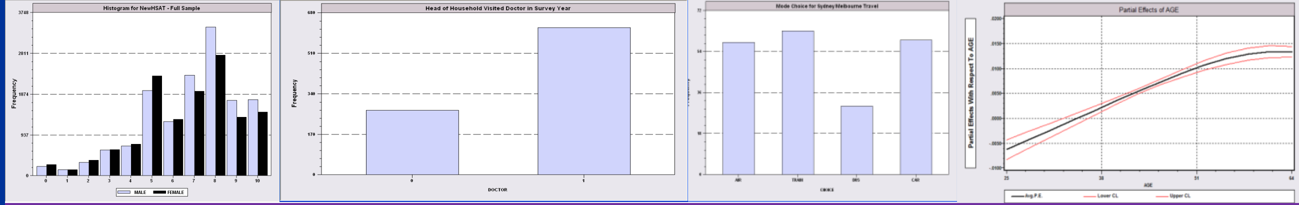
$\mathrm{Prob}(y_i = 1 \mid x_i) = G(\boldsymbol{\beta}'\mathbf{x})$ G is estimated by kernel methods



Fully Parametric

- Index Function: $U^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$
- Observation Mechanism: $y = 1[U^* > 0]$
- Distribution: $\varepsilon \sim f(\varepsilon)$; Normal, Logistic, …
- Maximum Likelihood Estimation:

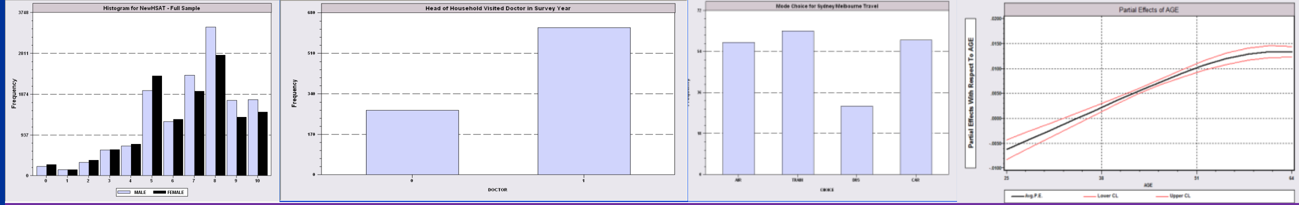
$$\text{Max}(\boldsymbol{\beta}) \log L = \Sigma_i \log \text{Prob}(Y_i = y_i | x_i)$$



Parametric: Logit Model

```
-----------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable               DOCTOR
Log likelihood function     -2097.48109
Restricted log likelihood   -2169.26982
Chi squared [   3](P= .000)   143.57744
Significance level               .00000
McFadden Pseudo R-squared      .0330935
Estimation based on N =    3377, K =   4
Inf.Cr.AIC  =    4203.0 AIC/N =    1.245
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient      Error       z     |z|>Z*        Interval
--------+--------------------------------------------------------------------
Constant|    -.42085***    .15810     -2.66   .0078     -.73072    -.11099
     AGE|     .02365***    .00328      7.21   .0000      .01722     .03008
  INCOME|    -.44198***    .16936     -2.61   .0091     -.77393    -.11003
  FEMALE|     .63825***    .07551      8.45   .0000      .49026     .78624
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

What do these mean?



Parametric Model Estimation

- How to estimate α, β_1, β_2, β_3?
  - The technique of maximum likelihood

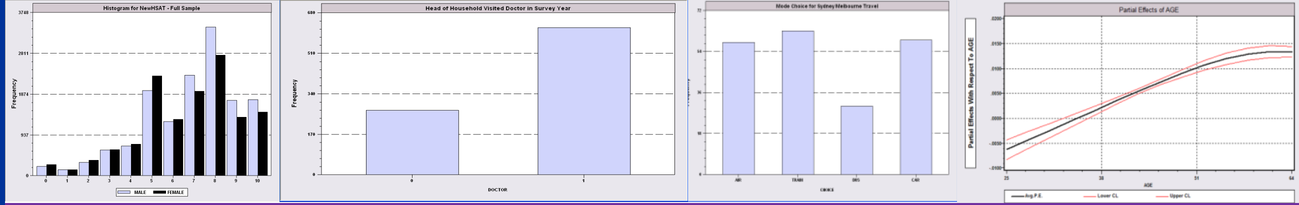
$$L = \prod_{y=0} \text{Prob}[y = 0 \,|\, \mathbf{x}] \times \prod_{y=1} \text{Prob}[y = 1 \,|\, \mathbf{x}]$$

  - Prob[y=1] =

    Prob[ε > -(α + β_1 Age + β_2 Income + β_3 Sex)]

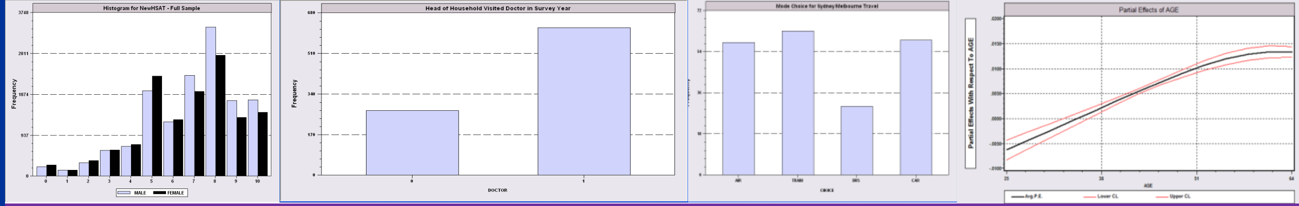
    Prob[y=0] = 1 - Prob[y=1]

- Requires a model for the probability



Completing the Model: F(Ɛ)

- The distribution
  - Normal: **PROBIT**, natural for behavior
  - Logistic: **LOGIT**, allows “thicker tails”
  - Gompertz: **EXTREME VALUE**, asymmetric
  - Others…
- Does it matter?
  - Yes, large difference in estimates
  - Not much, quantities of interest are more stable.

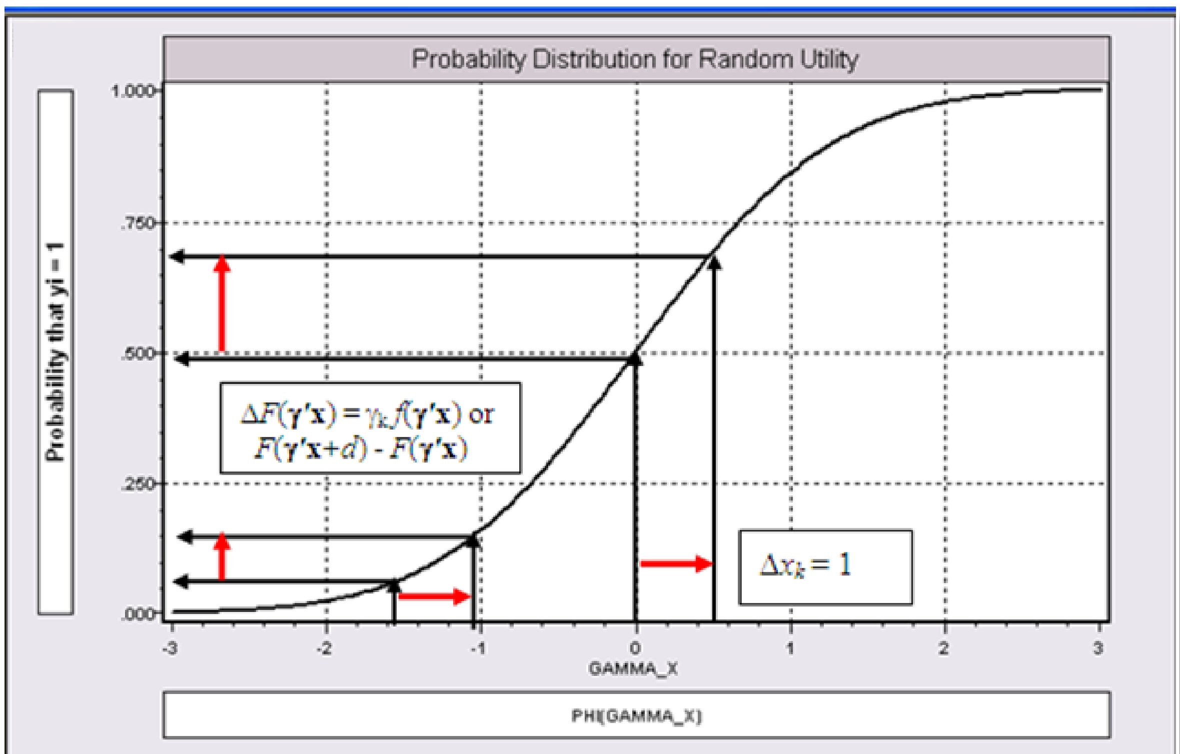


Estimated Binary Choice Models

	LOGIT		PROBIT		EXTREME VALUE	
Variable	Estimate	t-ratio	Estimate	t-ratio	Estimate	t-ratio
Constant	-0.42085	-2.662	-0.25179	-2.600	0.00960	0.078
Age	0.02365	7.205	0.01445	7.257	0.01878	7.129
Income	-0.44198	-2.610	-0.27128	-2.635	-0.32343	-2.536
Sex	0.63825	8.453	0.38685	8.472	0.52280	8.407
Log-L	-2097.48		-2097.35		-2098.17	
Log-L(0)	-2169.27		-2169.27		-2169.27	

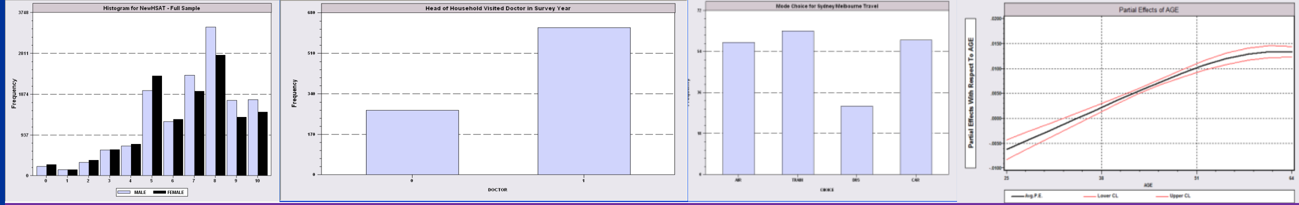
Log-L(0) = log likelihood for a model that has only a constant term.
Ignore the t ratios for now.

Effect on Predicted Probability of an Increase in Age



$$\alpha + \boxed{\beta_1 \text{ (Age+1)}} + \beta_2 \text{ (Income)} + \beta_3 \text{ Sex}$$

(β_1 is positive)

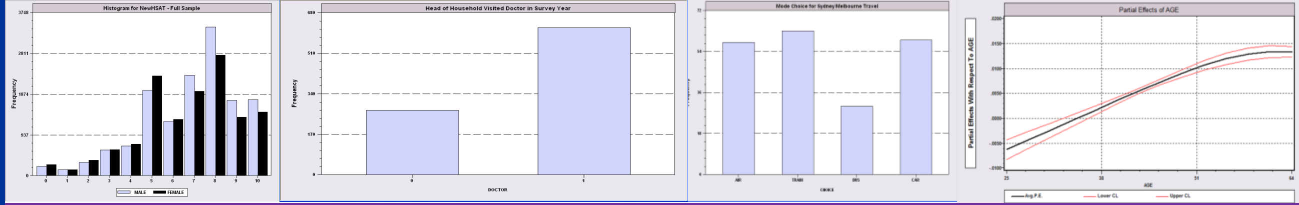


Partial Effects in Probability Models

- Prob[Outcome] = some $F(\alpha+\beta_1\text{Income}\ldots)$
- "Partial effect" = $\partial F(\alpha+\beta_1\text{Income}\ldots) / \partial\text{"}x\text{"}$ (derivative)
  - Partial effects are derivatives
  - Result varies with model

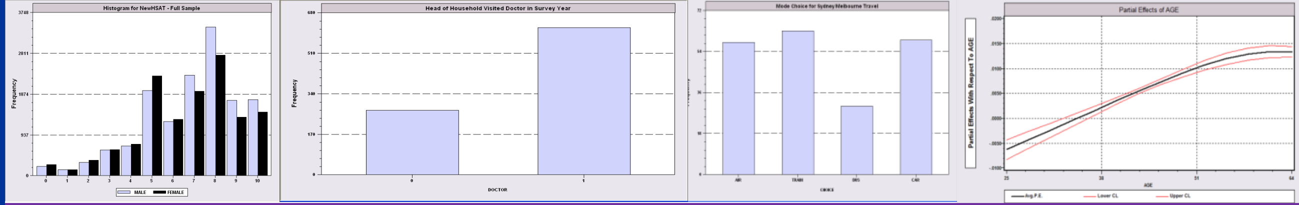
Logit: $\partial F(\alpha+\beta_1\text{Income}\ldots) / \partial\mathbf{x}$	=	Prob * (1-Prob) $\times\ \boldsymbol{\beta}$
Probit: $\partial F(\alpha+\beta_1\text{Income}\ldots) / \partial\mathbf{x}$	=	Normal density $\times\ \boldsymbol{\beta}$
Extreme Value: $\partial F(\alpha+\beta_1\text{Income}\ldots) / \partial\mathbf{x}$	=	Prob * (-log Prob) $\times\ \boldsymbol{\beta}$

  - Scaling usually erases model differences



Estimated Partial Effects

	LOGIT		PROBIT		EXTREME VALUE	
	Estimate	t ratio	Estimate	t ratio	Estimate	t ratio
Age	.00527	7.235	.00527	7.269	.00506	6.291
Income	-.09844	-2.611	-.09897	-2.636	-.09711	-2.527
Female	.14026	8.663	.13958	8.264	.13539	8.747

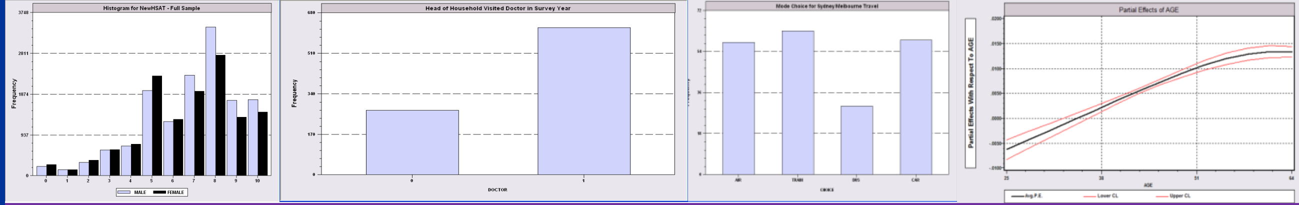


Partial Effect for a Dummy Variable

- Prob[y_i = 1|$\mathbf{x}_i$,d_i]   = F(β'$\mathbf{x}_i$+γd_i)

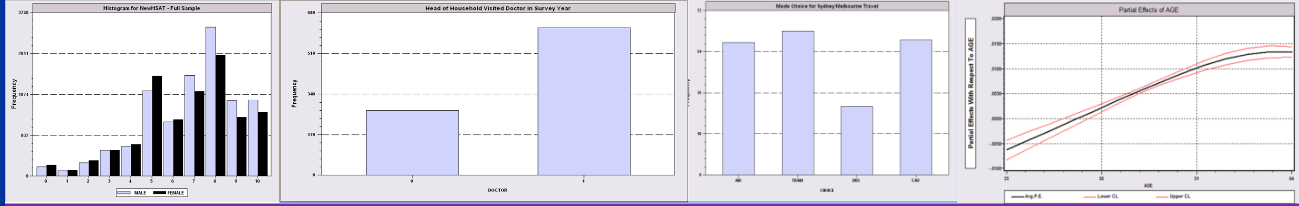
  = conditional mean
- Partial effect of d

  Prob[y_i = 1|$\mathbf{x}_i$, d_i=1]   -   Prob[y_i = 1|$\mathbf{x}_i$, d_i=0]
- Probit: $\delta(d_i) = \Phi\left(\hat{\boldsymbol{\beta}}'\overline{\mathbf{x}} + \hat{\gamma}\right) - \Phi\left(\hat{\boldsymbol{\beta}}'\overline{\mathbf{x}}\right)$



Partial Effect – Dummy Variable

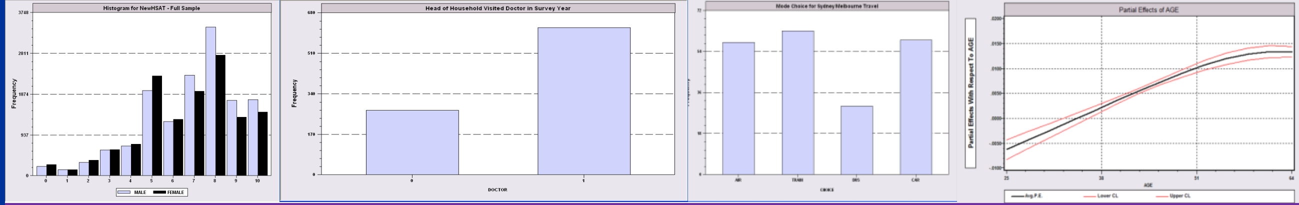
```
------------------------------------------------------------------------
Partial derivatives of E[y] = F[*]  with
respect to the vector of characteristics
They are computed at the means of the Xs
Observations used for means are All Obs.
--------+---------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]  Elasticity
--------+---------------------------------------------------------------
        |Index function for probability
Constant|    -.09186***        .03550       -2.588   .0097
     AGE|     .00527***        .00073        7.269   .0000       .33855
  INCOME|    -.09897***        .03755       -2.636   .0084      -.06632
        |Marginal effect for dummy variable is P|1 - P|0.
  FEMALE|     .13958***        .01618        8.624   .0000       .09745
--------+---------------------------------------------------------------
Note: ***, **, * = Significance at 1%, 5%, 10% level.
Elasticity for a binary variable = marginal effect/Mean.
------------------------------------------------------------------------
```



Computing Partial Effects

- Compute at the data means?
  - Simple
  - Inference is well defined.

- Average the individual effects
  - More appropriate?
  - Asymptotic standard errors are complicated.



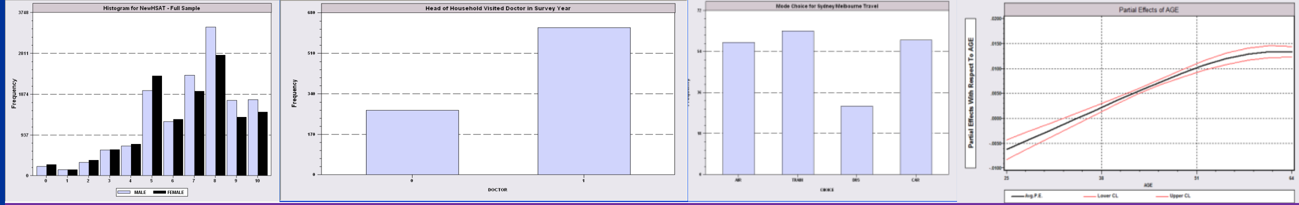
Average Partial Effects

$$\text{Probability} \;=\; P_i = F(\boldsymbol{\beta}'\mathbf{x}_i)$$

$$\text{Partial Effect} = \frac{\partial P_i}{\partial \mathbf{x}_i} = \frac{\partial F(\boldsymbol{\beta}'\mathbf{x}_i)}{\partial \mathbf{x}_i} = f(\boldsymbol{\beta}'\mathbf{x}_i) \times \boldsymbol{\beta} = \mathbf{d}_i$$

$$\text{Average Partial Effect} = \frac{1}{n}\sum\nolimits_{i=1}^{n} \mathbf{d}_i \;=\boldsymbol{\beta}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} f(\boldsymbol{\beta}'\mathbf{x}_i)\right)$$

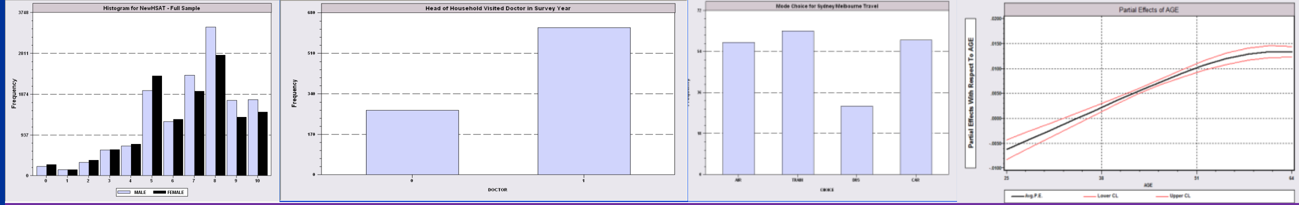
are estimates of $\boldsymbol{\delta} = E[\mathbf{d}_i]$ under certain assumptions.



Average Partial Effects vs. Partial Effects at Data Means

Variable	Mean
ME_AGE	.00511838
ME_INCOM	-.0960923
ME_FEMAL	.137915

PROBIT	
Estimate	t ratio
.00527	7.269
-.09897	-2.636
.13958	8.264



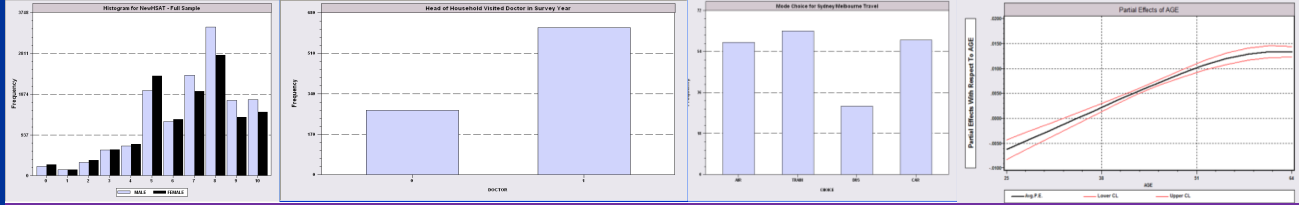
Practicalities of Nonlinearities

The software does not know that agesq = age^2.

```
PROBIT     ; Lhs=doctor
           ; Rhs=one,age,agesq,income,female
           ; Partial effects $
```

The software knows that age * age is age^2.

```
PROBIT     ; Lhs=doctor
           ; Rhs=one,age,age*age,income,female $
PARTIALS   ; Effects : age $
```



Partial Effect for Nonlinear Terms

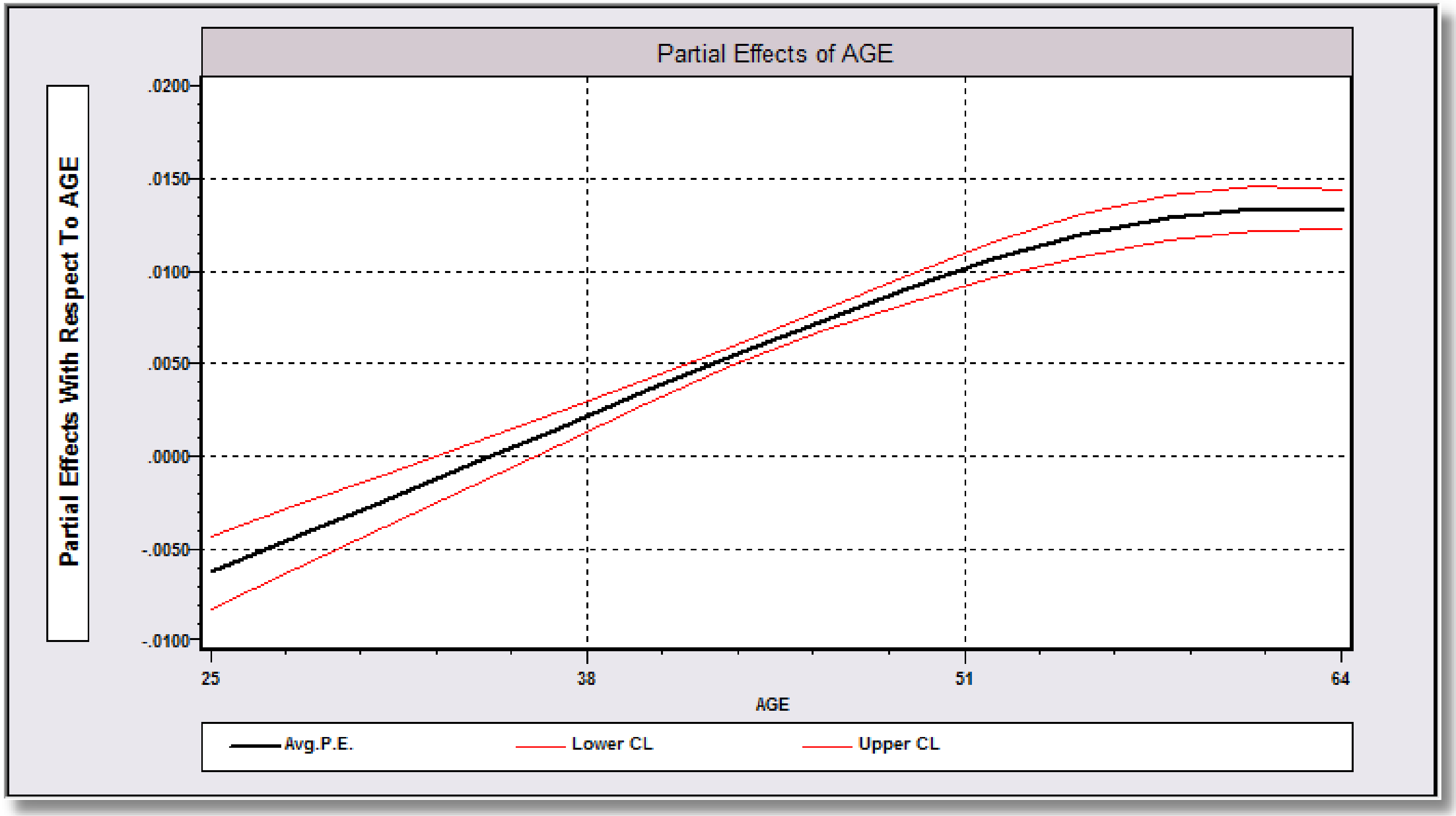
$$\text{Prob} = \Phi[\alpha + \beta_1\text{Age} + \beta_2\text{Age}^2 + \beta_3\text{Income} + \beta_4\text{Female}]$$

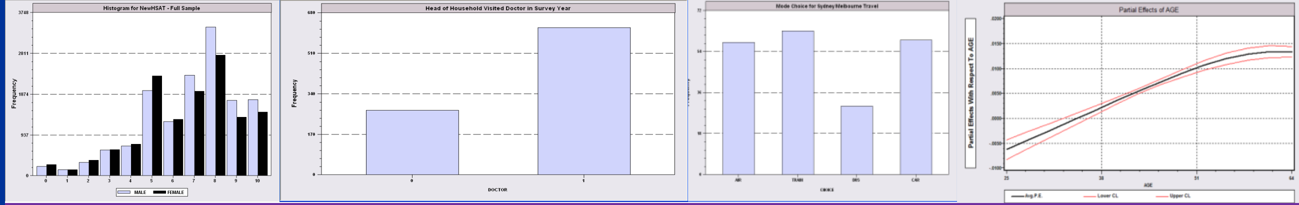
$$\frac{\partial\text{Prob}}{\partial\text{Age}} = \phi[\alpha + \beta_1\text{Age} + \beta_2\text{Age}^2 + \beta_3\text{Income} + \beta_4\text{Female}] \times (\beta_1 + 2\beta_2\text{Age})$$

$$= \begin{array}{l} \phi(1.30811 - .06487\,Age + .0091\,Age^2 - .17362\,Income + .39666\,Female) \\ \times[(-.06487 + 2(.0091)\,Age] \end{array}$$

Must be computed at specific values of Age, Income and Female

Average Partial Effect: Averaged over Sample Incomes and Genders for Specific Values of Age





The Linear Probability "Model"

$$\text{Prob}(y = 1 | \mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$$

$$E[y | \mathbf{x}] = 0 * \text{Prob}(y = 1 | \mathbf{x}) + 1\text{Prob}(y = 1 | \mathbf{x}) = \text{Prob}(y = 1 | \mathbf{x})$$

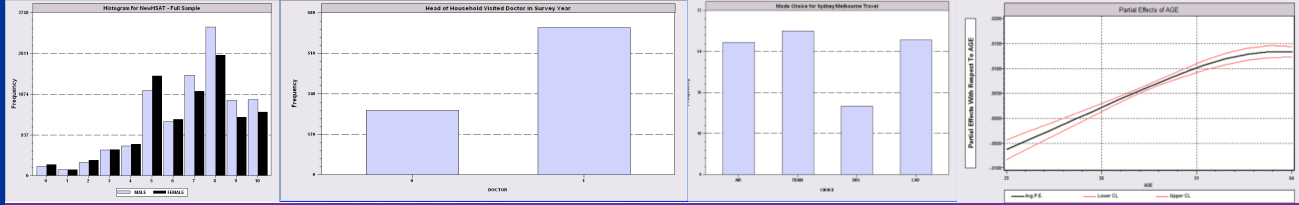
$$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$$

ROTTEN APPLES: AN INVESTIGATION OF THE
PREVALENCE AND PREDICTORS
OF TEACHER CHEATING

Brian A. Jacob
Steven D. Levitt

Working Paper 9413
http://www.nber.org/papers/w9413

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2002



The Dependent Variable equals zero for 98.9% of the observations. In the sample of 163,474 observations, the LHS variable equals 1 about 1,500 times.

Table 9: OLS Estimates of the Relationship between Cheating and Classroom Characteristics

Independent variables	Dependent variable = Indicator of classroom cheating (1)	(2)	(3)	(4)
Social promotion policy	0.0011 (0.0013)	0.0011 (0.0013)	0.0015 (0.0013)	0.0023 (0.0009)
School probation policy	0.0020 (0.0014)	0.0019 (0.0014)	0.0021 (0.0014)	0.0029 (0.0013)
Prior classroom achievement	-0.0047 (0.0005)	-0.0028 (0.0005)	-0.0016 (0.0007)	-0.0028 (0.0007)
Social promotion*classroom achievement	--	-0.0049 (0.0014)	-0.0051 (0.0014)	-0.0046 (0.0012)
School probation*classroom achievement	--	-0.0070 (0.0013)	-0.0070 (0.0013)	-0.0064 (0.0013)
Mixed grade classroom	-0.0084 (0.0007)	-0.0085 (0.0007)	-0.0089 (0.0008)	-0.0089 (0.0012)
% of students included in official reporting	0.0252 (0.0031)	0.0249 (0.0031)	0.0141 (0.0037)	0.0131 (0.0037)
School*Year Fixed Effects	No	No	No	Yes
Number of observations	163,474	163,474	163,474	163,474

Notes: The unit of observation is classroom*grade*year*subject and the sample includes years eight years (1993 to 2000), four subjects (reading comprehension and three math sections) and five grades (three to seven). The dependent variable is the cheating indicator derived using the 95th percentile cutoff. Robust standard errors clustered by school*year are shown in parenthesis. Other variables included in the regressions in column 1 and 2 include a

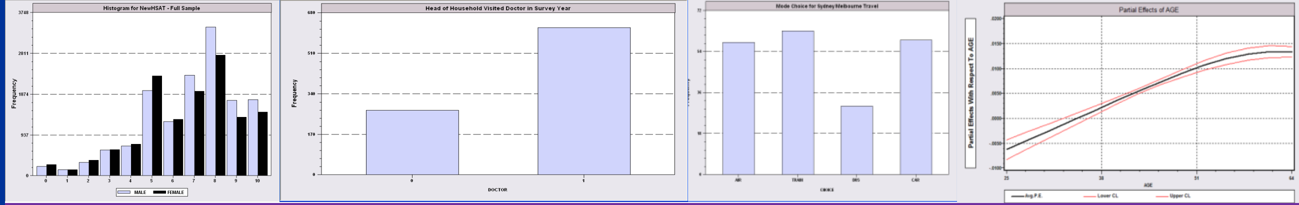


Table 10: In Cheating Classrooms, for Whom do Teachers Cheat?

Independent variables	Dependent variable = Teacher cheated for the student (1)	(2)	(3)	(4)
Prior achievement in the bottom quartile	0.011 (0.038)	--	-0.007 (0.075)	--
Prior achievement in the 2nd quartile	0.057 (0.024)	--	0.069 (0.039)	--
Prior achievement in the 3rd quartile	0.023 (0.067)	--	-0.012 (0.141)	--
Prior achievement (linear measure)	--	0.0004 (0.0003)	--	0.0005 (0.0004)
Prior achievement (linear) * High-stakes	--	-0.0007 (0.0004)	--	-0.0007 (0.0005)
Excluded from test reporting	-0.045 (0.014)	-0.048 (0.014)	-0.045 (0.021)	-0.052 (0.020)
Male	-0.009 (0.004)	-0.009 (0.004)	-0.014 (0.005)	-0.013 (0.005)
Black	0.005 (0.011)	0.006 (0.011)	0.004 (0.024)	0.001 (0.023)
Hispanic	-0.010 (0.010)	-0.008 (0.009)	0.006 (0.023)	0.004 (0.022)
Age	-0.010 (0.004)	-0.012 (0.004)	-0.015 (0.005)	-0.017 (0.005)
Sample	Full		Low-Achieving Schools	
Number of observations	39,216		23,010	

Notes: The sample includes only those classrooms that were categorized as cheating based on the 95th percentile cutoff in a particular subject and year. The dependent variable takes on the value of one if a *student's* answer string and test score pattern was suspicious at the 90th percentile level, suggesting that the teacher had cheated for that student in the particular subject and year. All models include fixed effects for classroom*year. Low achieving schools are defined as those in which fewer than 25% of students met national norms in reading in 1995. The equations are estimated using 2SLS where a student's test scores at t-2 are used to instrument for the student's t-1 achievement level. Robust standard errors are shown in parenthesis.

2SLS for a binary dependent variable.

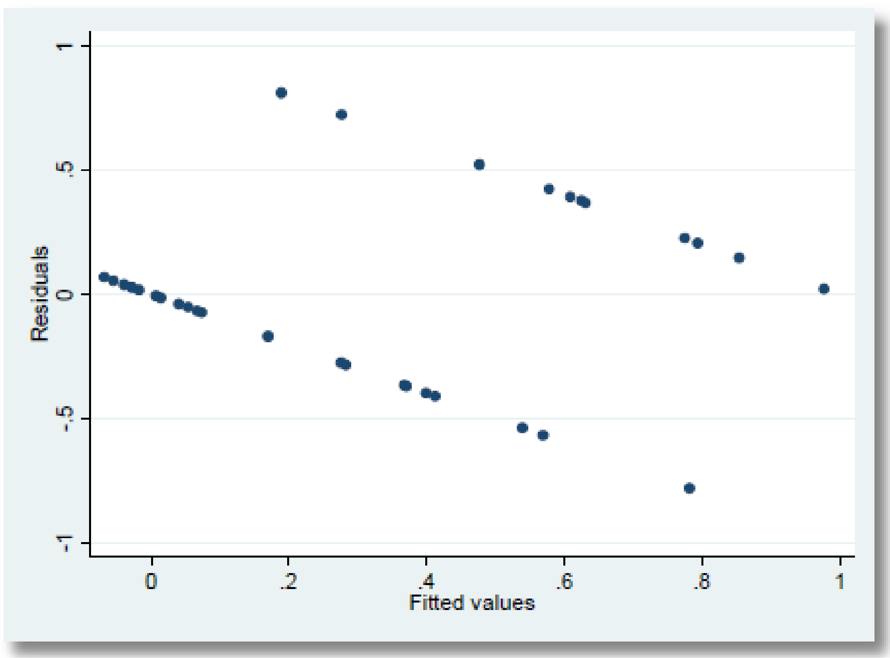
$$\text{Prob}(y=1|\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$$

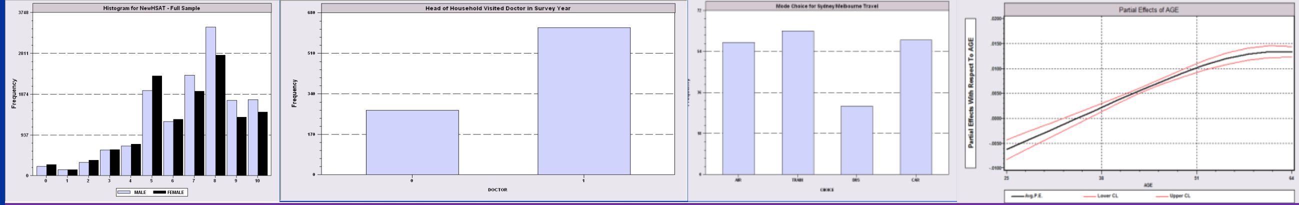
$$E[y|\mathbf{x}] = 0 * \text{Prob}(y=1|\mathbf{x}) + 1\text{Prob}(y=1|\mathbf{x}) = \text{Prob}(y=1|\mathbf{x})$$

$$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$$

$$\text{Residuals}: \; e \; = \; y \; - \hat{\boldsymbol{\beta}}'\mathbf{x} = \mathbf{1} - \hat{\boldsymbol{\beta}}'\mathbf{x} \text{ if } y = 1, \text{ or } 0 - \hat{\boldsymbol{\beta}}'\mathbf{x} \text{ if } y = 0$$

The standard errors make no sense because the stochastic properties of the "disturbance" are inconsistent with the observed variable.





$$\text{Prob}(y=1|\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$$

$$E[y|\mathbf{x}] = 0 * \text{Prob}(y=1|\mathbf{x}) + 1\text{Prob}(y=1|\mathbf{x}) = \text{Prob}(y=1|\mathbf{x})$$

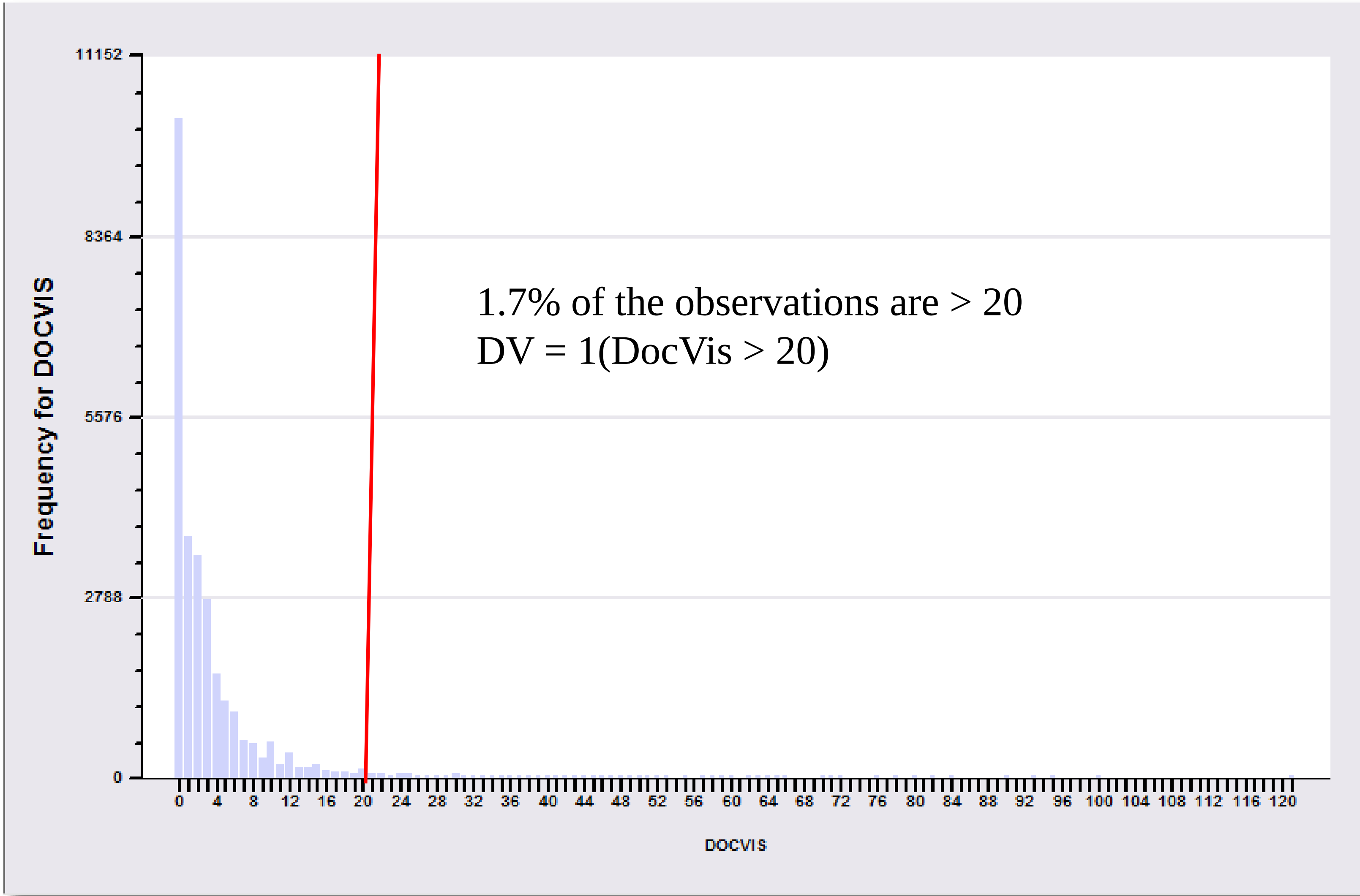
$$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$$

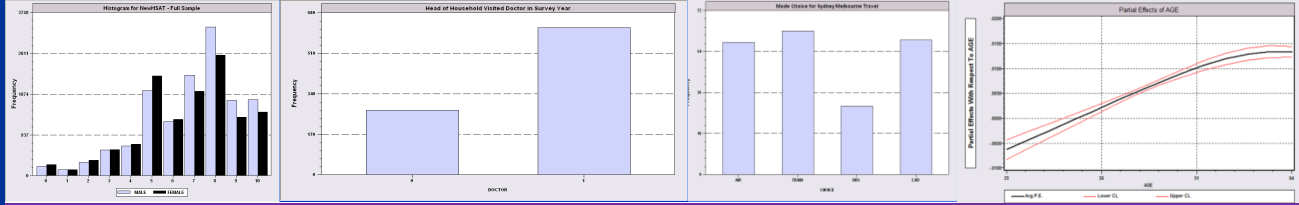
$$\text{Residuals: } e = y - \hat{\boldsymbol{\beta}}'\mathbf{x} = \mathbf{1} - \hat{\boldsymbol{\beta}}'\mathbf{x} \text{ if } y = 1, \text{ or } 0 - \hat{\boldsymbol{\beta}}'\mathbf{x} \text{ if } y = 0$$

The standard errors make no sense because the stochastic properties of the "disturbance" are inconsistent with the observed variable.

The variance of y|x equals $\text{Prob}(y=0|\mathbf{x})\text{Prob}(y=1|\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}(1-\boldsymbol{\beta}'\mathbf{x})$

The "disturbances" are heteroscedastic. Users of the LPM always seem to worry about clustering. They never seem to worry about heteroscedasticity.


1.7% of the observations are > 20
DV = 1(DocVis > 20)
Frequency for DOCVIS
11152
8364
5576
2788
0
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72 76 80 84 88 92 96 100 104 108 112 116 120
DOCVIS



Binomial Probit Model
Dependent variable DV

DV	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Index function for probability					
Constant	-2.23278***	.19860	-11.24	.0000	-2.62203	-1.84353
AGE	.01053***	.00186	5.65	.0000	.00688	.01418
EDUC	-.02047*	.01095	-1.87	.0616	-.04193	.00099
MARRIED	-.12096***	.04625	-2.61	.0089	-.21161	-.03030
PUBLIC	.29821***	.09436	3.16	.0016	.11327	.48314
HEALTHY	-.85776***	.04959	-17.30	.0000	-.95496	-.76057

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

Partial derivatives of E[y] = F[*] with
respect to the vector of characteristics
Average partial effects for sample obs.

DV	Partial Effect	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval		
AGE	.00041***	.7425D-04	5.56	.0000	.00027	.00056	
EDUC	-.00080*	.00043	-1.87	.0621	-.00164	.00004	
MARRIED	-.00504**	.00205	-2.46	.0139	-.00906	-.00103	#
PUBLIC	.00919***	.00223	4.12	.0000	.00482	.01356	#
HEALTHY	-.03140***	.00186	-16.92	.0000	-.03503	-.02776	#

Partial effect for dummy variable is E[y|x,d=1] - E[y|x,d=0]

Ordinary least squares regression ....
LHS=DV Mean = .01749
Standard deviation = .13110
Fit R-squared = .01955 R-bar squared .01937

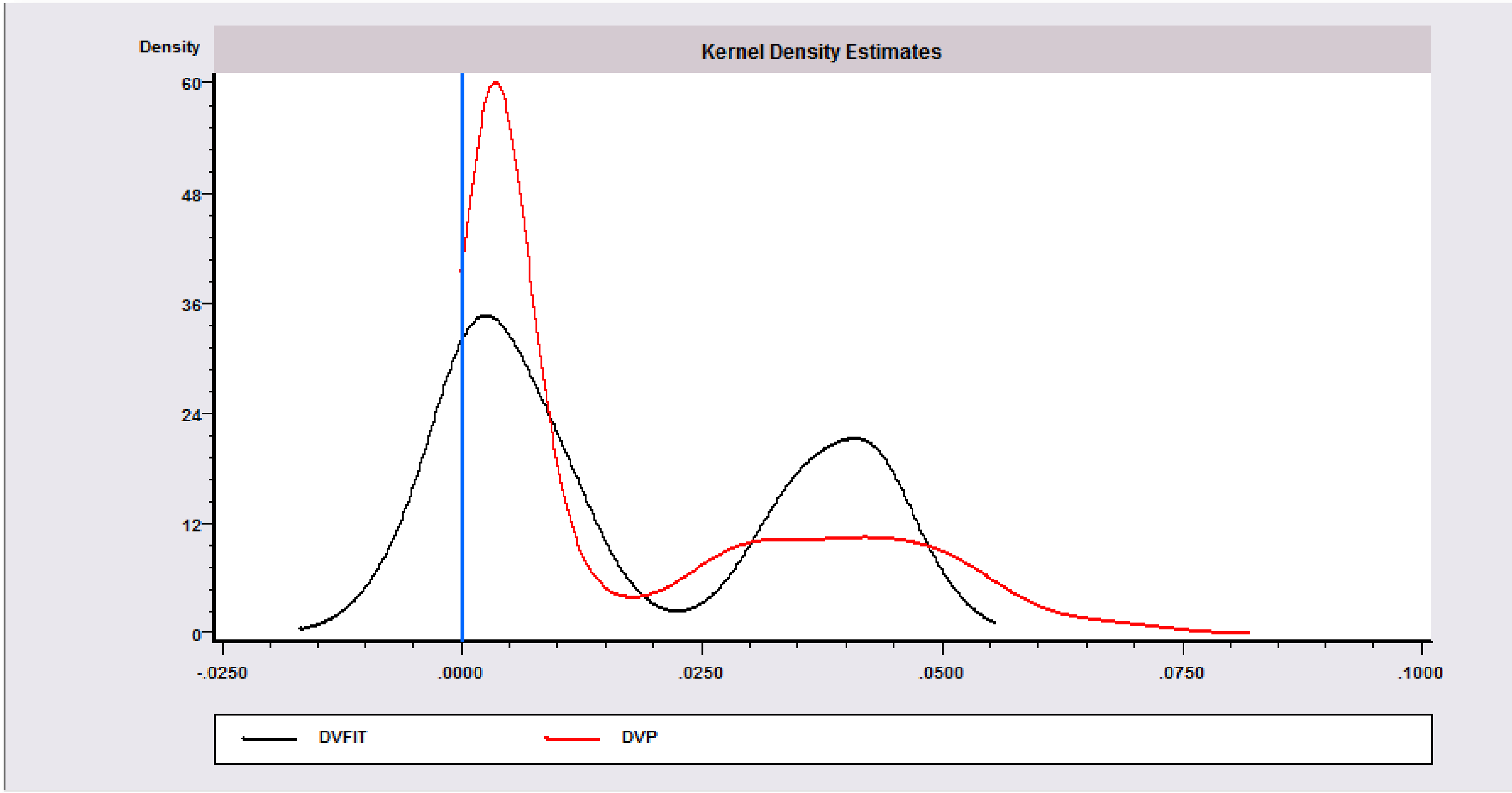
DV	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	.02278***	.00682	3.34	.0008	.00942	.03614
AGE	.00044***	.7315D-04	5.98	.0000	.00029	.00058
EDUC	-.00059	.00037	-1.62	.1060	-.00131	.00013
MARRIED	-.00520***	.00187	-2.78	.0055	-.00887	-.00153
PUBLIC	.00700***	.00263	2.66	.0077	.00185	.01215
HEALTHY	-.03261***	.00166	-19.59	.0000	-.03588	-.02935

What does OLS Estimate?

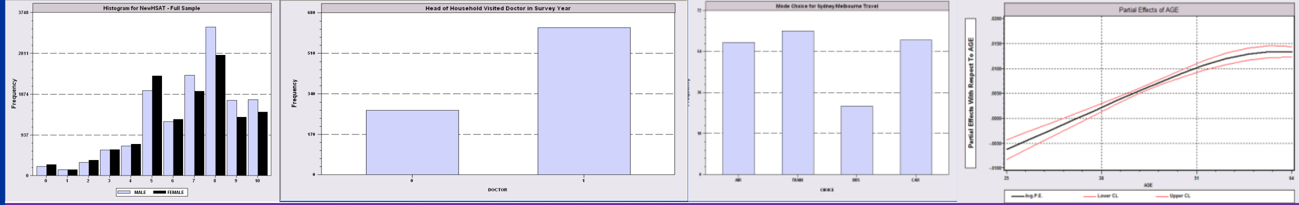
MLE

Average Partial Effects

OLS Coefficients

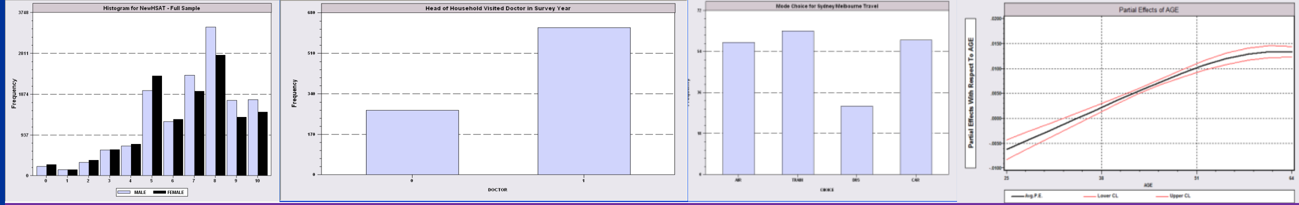


Kernel Density Estimates
Density
60
48
36
24
12
0
-.0250
.0000
.0250
.0500
.0750
.1000
DVFIT
DVP

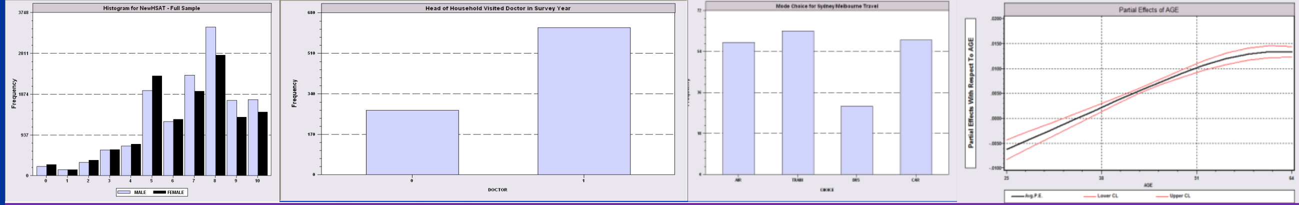


Negative Predicted Probabilities

```
|-> create;dvn=dvfit<0$
|-> cross;rhs=dvn;lhs=dv$

Cross Tabulation
                              DVN
+---------+---------------+-------+
|       DV|      0       1| Total|
+---------+---------------+-------+
|        0|  22281    4567| 26848|
|        1|    463      15|   478|
+---------+---------------+-------+
|    Total|  22744    4582| 27326|
+---------+---------------+-------+
```

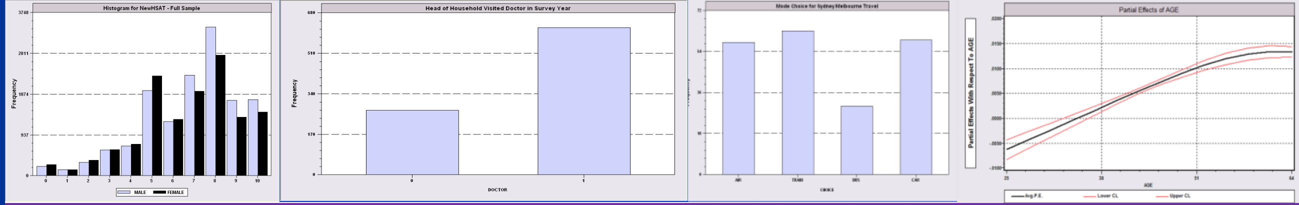


Measuring Fit



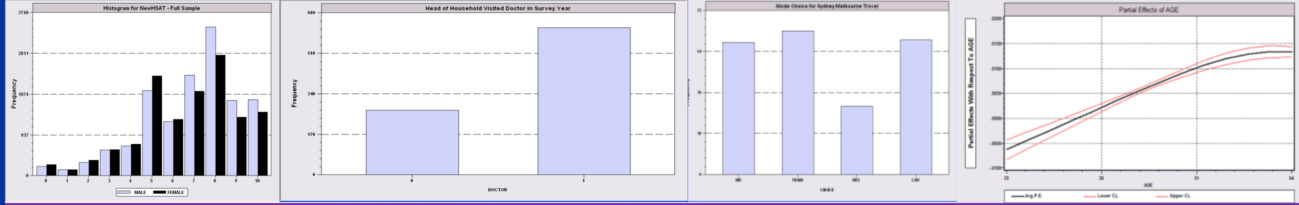
How Well Does the Model Fit?

- **There is no R squared.**
  - Least squares for linear models is computed to maximize R^2
  - There are no residuals or sums of squares in a binary choice model
  - The model is not computed to optimize the fit of the model to the data
- **How can we measure the "fit" of the model to the data?**
  - "Fit measures" computed from the log likelihood
    - **"Pseudo R squared" = 1 – logL/logL0**
    - Also called the "likelihood ratio index"
    - Others… - these do not measure fit.
  - Direct assessment of the effectiveness of the model at predicting the outcome



Log Likelihoods

- $\log L = \sum_i \log \text{density}\,(y_i|\mathbf{x}_i,\boldsymbol{\beta})$
- For probabilities
  - Density is a probability
  - Log density is < 0
  - LogL is < 0
- For other models, log density can be positive or negative.
  - For linear regression, $\log L = -N/2(1+\log 2\pi+\log(\mathbf{e}'\mathbf{e}/N)]$
  - Positive if $s^2 < .058497$



Likelihood Ratio Index

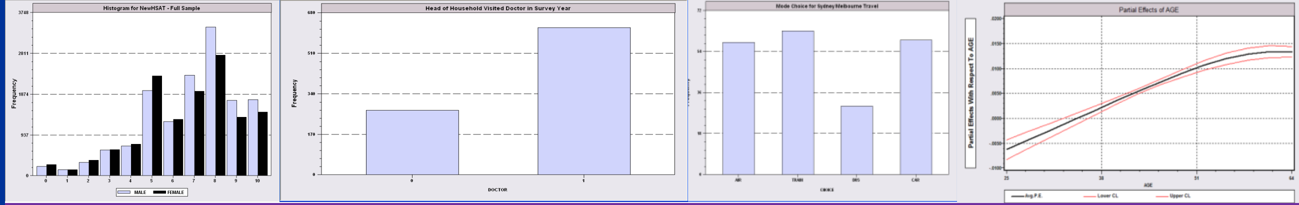
$$\log L = \sum_{i=1}^{N} \{(1-y_i)\log[1-F(\boldsymbol{\beta}'\mathbf{x}_i)] + y_i \log F(\boldsymbol{\beta}'\mathbf{x}_i)\}$$

1. Suppose the model predicted $F(\boldsymbol{\beta}'\mathbf{x}_i) = 1$ whenever y=1 and $F(\boldsymbol{\beta}'\mathbf{x}_i) = 0$ whenever y=0. Then, logL = 0. [$F(\boldsymbol{\beta}'\mathbf{x}_i)$ cannot equal 0 or 1 at any finite β.]

2. Suppose the model always predicted the same value, $\mathrm{F}(\beta_0)$

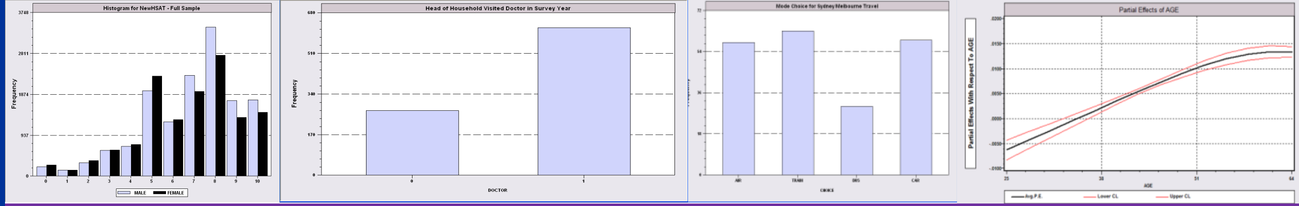
$$\begin{aligned} \mathrm{LogL}_0 &= \sum_{i=1}^{N} \{(1-y_i)\log[1-\mathrm{F}(\beta_0)] + y_i \log \mathrm{F}(\beta_0)\} \\ &= N_0 \log[1-\mathrm{F}(\beta_0)] + N_1 \log \mathrm{F}(\beta_0) \\ &< 0 \end{aligned}$$

$$\mathrm{LRI} = 1 - \frac{\log L}{\log L_0}. \text{ Since logL} > \mathrm{logL}_0 \quad 0 \leq \mathrm{LRI} < 1.$$



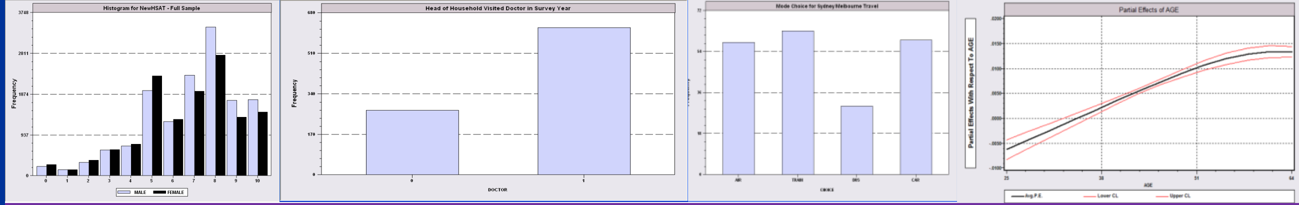
The Likelihood Ratio Index

- Bounded by 0 and 1-ε
- Rises when the model is expanded
- **Values between 0 and 1 have no meaning**
- Can be strikingly low.
- Should not be used to compare models
  - Use logL
  - Use information criteria to compare nonnested models



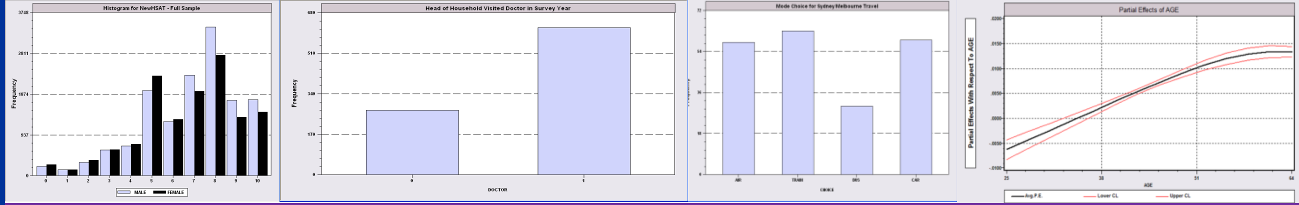
Fit Measures Based on LogL

```
----------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                  DOCTOR
Log likelihood function        -2085.92452  <-  Full model          LogL
Restricted log likelihood      -2169.26982  <-  Constant term only LogL0
Chi squared [    5 d.f.]        166.69058
Significance level                  .00000
McFadden Pseudo R-squared         .0384209  <-  1 - LogL/logL0
Estimation based on N =    3377, K =    6
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              1.23892     4183.84905       -2LogL + 2K
Fin.Smpl.AIC     1.23893     4183.87398       -2LogL + 2K + 2K(K+1)/(N-K-1)
Bayes IC         1.24981     4220.59751       -2LogL + KlnN
Hannan Quinn     1.24282     4196.98802       -2LogL + 2Kln(lnN)
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|    1.86428***       .67793        2.750   .0060
     AGE|    -.10209***       .03056       -3.341   .0008      42.6266
   AGESQ|     .00154***       .00034        4.556   .0000      1951.22
  INCOME|     .51206          .74600         .686   .4925       .44476
 AGE_INC|    -.01843          .01691       -1.090   .2756      19.0288
  FEMALE|     .65366***       .07588        8.615   .0000       .46343
--------+-------------------------------------------------------------
```



Fit Measures Based on Predictions

- Computation
  - Use the model to compute predicted probabilities
  - Use the model and a rule to compute predicted y = 0 or 1
- Fit measure compares predictions to actuals

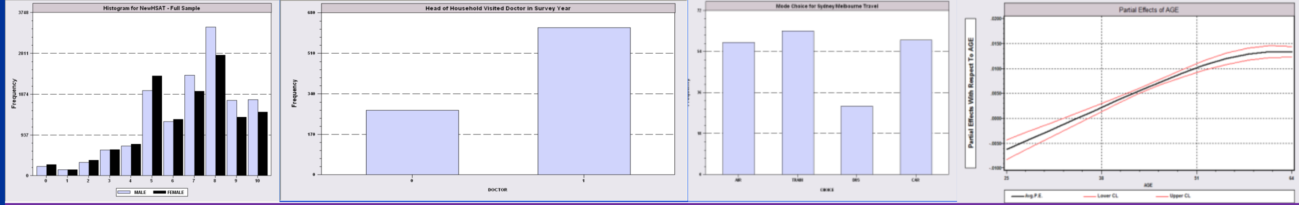


Predicting the Outcome

- Predicted probabilities

  P = F(a + b_1Age + b_2Income + b_3Female+…)

- Predicting outcomes
  - Predict y=1 if P is "large"
  - Use 0.5 for "large" (more likely than not)
  - Generally, use $\hat{y} = 1 \text{ if } \hat{P} > P^*$
- Count successes and failures



Cramer Fit Measure

$$\hat{\mathrm{F}} = \text{Predicted Probability}$$

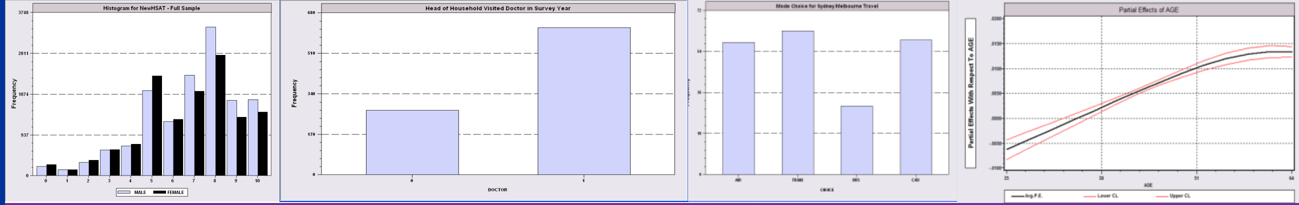
$$\hat{\lambda} = \frac{\Sigma_{i=1}^{N} y_i \hat{\mathrm{F}}}{\mathrm{N}_1} - \frac{\Sigma_{i=1}^{N} (1 - y_i) \hat{\mathrm{F}}}{\mathrm{N}_0}$$

$$\hat{\lambda} = \left(\text{Mean } \hat{\mathrm{F}} \,|\, \text{when } y = 1\right) - \left(\text{Mean } \hat{\mathrm{F}} \,|\, \text{when } y = 0\right)$$

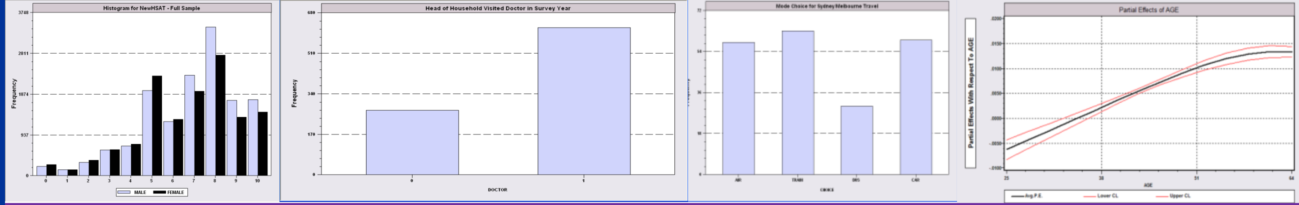
= reward for correct predictions minus

penalty for incorrect predictions

```
+----------------------------------------+
| Fit Measures Based on Model Predictions|
| Efron                      =     .04825|
| Ben Akiva and Lerman       =     .57139|
| Veall and Zimmerman        =     .08365|
| Cramer                     =     .04771|
+----------------------------------------+
```



Hypothesis Testing in Binary Choice Models



Covariance Matrix for the MLE

Log Likelihood

$$\log L = \sum_{i=1}^{N} \{(1-y_i)\log[1-F(\boldsymbol{\beta}'\mathbf{x}_i)] + y_i \log F(\boldsymbol{\beta}'\mathbf{x}_i)\}$$

We focus on the standard choices of $F(\boldsymbol{\beta}'\mathbf{x}_i)$, probit and logit.

Both distributions are symmetric; F(t)=1-F(-t). Therefore, the terms in the sums are

$$\log L_i = \log F[q_i(\boldsymbol{\beta}'\mathbf{x}_i)] \text{where } \mathrm{q}_i = 2y_i - 1$$

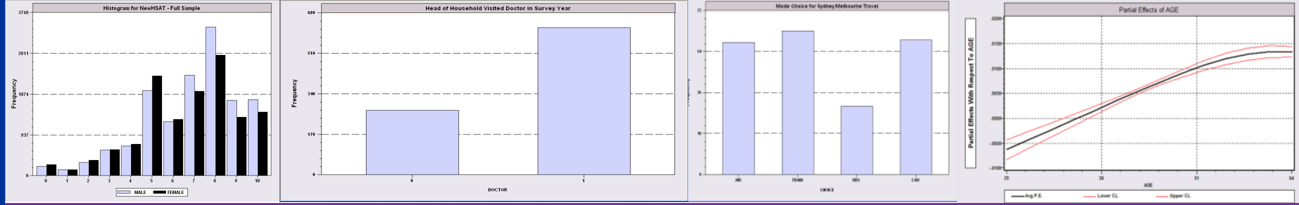
$$\frac{\partial \log L_i}{\partial \boldsymbol{\beta}} = \frac{F'[q_i(\boldsymbol{\beta}'\mathbf{x}_i)]}{F[q_i(\boldsymbol{\beta}'\mathbf{x}_i)]}(q_i\mathbf{x}_i) = \mathrm{q}_i \frac{F_i'}{F}\mathbf{x}_i = \mathbf{g}_i$$

$$\frac{\partial^2 \log L_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \left\{\frac{F''}{F} - \left(\frac{F'}{F}\right)^2\right\}(q_i\mathbf{x}_i)(q_i\mathbf{x}_i)' = \mathbf{H}_i$$

These simplify considerably. Note $\mathrm{q}_i^2 = 1$.

For the logit model, F=Λ, F′=Λ(1-Λ) and F″=Λ(1-Λ)(1-2Λ).

For the probit model, F=Φ, F′= ϕ and F″ = -$[q_i(\boldsymbol{\beta}'\mathbf{x}_i)]\phi$



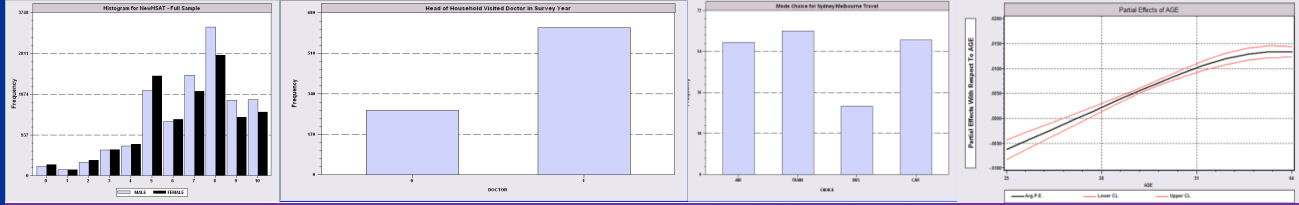
Simplifications

Logit: $g_i = y_i - \Lambda_i \qquad H_i = \Lambda_i(1-\Lambda_i) \qquad E[H_i] = \Psi_i = \Lambda_i(1-\Lambda_i)$

Probit: $g_i = \dfrac{q_i\phi_i}{\Phi_i} \qquad H_i = \dfrac{(q_i\boldsymbol{\beta}'\mathbf{x}_i)\phi_i}{\Phi_i} + \left(\dfrac{\phi_i}{\Phi_i}\right)^2, \quad E[H_i] = \Psi_i = \dfrac{\phi_i^2}{\Phi_i(1-\Phi_i)}$

Estimators: Based on H_i, $E[H_i]$ and g_i^2 all functions evaluated at $(q_i\boldsymbol{\beta}'\mathbf{x}_i)$

Actual Hessian:	$\text{Est.Asy.Var}[\hat{\beta}] = \left[\sum_{i=1}^{N} H_i \mathbf{x}_i\mathbf{x}_i'\right]^{-1}$
Expected Hessian:	$\text{Est.Asy.Var}[\hat{\beta}] = \left[\sum_{i=1}^{N} \Psi_i \mathbf{x}_i\mathbf{x}_i'\right]^{-1}$
BHHH:	$\text{Est.Asy.Var}[\hat{\beta}] = \left[\sum_{i=1}^{N} g_i^2 \mathbf{x}_i\mathbf{x}_i'\right]^{-1}$



Robust Covariance Matrix

"Robust" Covariance Matrix: $\mathbf{V} = \mathbf{A}\,\mathbf{B}\,\mathbf{A}$

$\mathbf{A}$ = negative inverse of second derivatives matrix

$$= \text{estimated E}\left[-\frac{\partial^2 \log L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right]^{-1} = \left[-\sum_{i=1}^{N}\frac{\partial^2 \log \text{Prob}_i}{\partial\hat{\boldsymbol{\beta}}\partial\hat{\boldsymbol{\beta}}'}\right]^{-1}$$

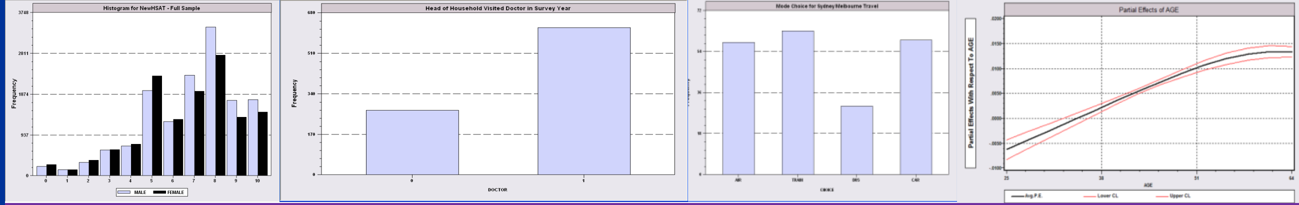
$\mathbf{B}$ = matrix sum of outer products of first derivatives

$$= \text{estimated E}\left[\frac{\partial \log L}{\partial\boldsymbol{\beta}}\frac{\partial \log L}{\partial\boldsymbol{\beta}'}\right] = \left[\sum_{i=1}^{N}\frac{\partial \log \text{Prob}_i}{\partial\hat{\boldsymbol{\beta}}}\frac{\partial \log \text{Prob}_i}{\partial\hat{\boldsymbol{\beta}}'}\right]^{-1}$$

For a logit model, $\mathbf{A} = \left[\sum_{i=1}^{N}\hat{P}_i(1-\hat{P}_i)\mathbf{x}_i\mathbf{x}_i'\right]^{-1}$

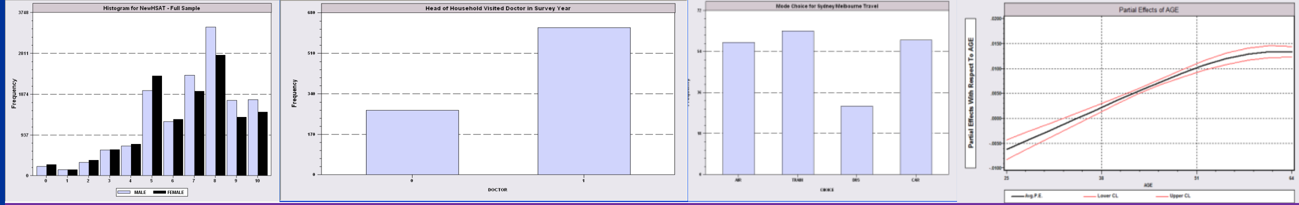
$$\mathbf{B} = \left[\sum_{i=1}^{N}(y_i-\hat{P}_i)^2\mathbf{x}_i\mathbf{x}_i'\right] = \left[\sum_{i=1}^{N}e_i^2\mathbf{x}_i\mathbf{x}_i'\right]$$

(Resembles the White estimator in the linear model case.)



Robust Covariance Matrix for Logit Model

```
--------+----------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+----------------------------------------------------------
        |Robust Standard Errors
Constant|     1.86428***        .68442        2.724   .0065
     AGE|     -.10209***        .03115       -3.278   .0010     42.6266
   AGESQ|      .00154***        .00035        4.446   .0000     1951.22
  INCOME|      .51206           .75103         .682   .4954      .44476
 AGE_INC|     -.01843           .01703       -1.082   .2792     19.0288
  FEMALE|      .65366***        .07585        8.618   .0000      .46343
--------+----------------------------------------------------------
        |Conventional Standard Errors Based on Second Derivatives
Constant|     1.86428***        .67793        2.750   .0060
     AGE|     -.10209***        .03056       -3.341   .0008     42.6266
   AGESQ|      .00154***        .00034        4.556   .0000     1951.22
  INCOME|      .51206           .74600         .686   .4925      .44476
 AGE_INC|     -.01843           .01691       -1.090   .2756     19.0288
  FEMALE|      .65366***        .07588        8.615   .0000      .46343
```

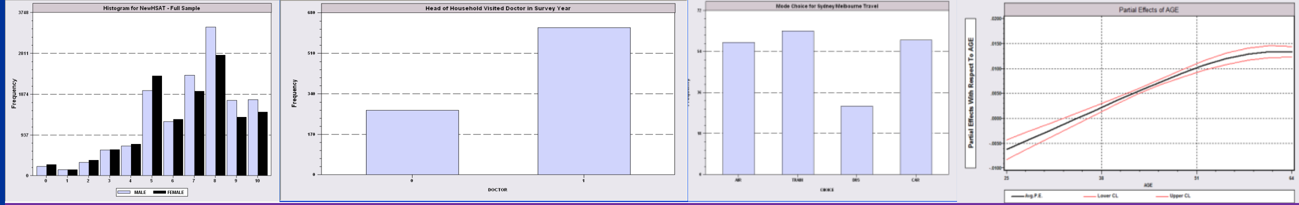


Base Model for Hypothesis Tests

```
----------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                  DOCTOR
Log likelihood function        -2085.92452
Restricted log likelihood      -2169.26982
Chi squared [    5 d.f.]         166.69058
Significance level                  .00000
McFadden Pseudo R-squared         .0384209
Estimation based on N =    3377, K =     6
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              1.23892     4183.84905
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|    1.86428***       .67793         2.750   .0060
     AGE|    -.10209***       .03056        -3.341   .0008     42.6266
   AGESQ|     .00154***       .00034         4.556   .0000     1951.22
  INCOME|     .51206          .74600          .686   .4925      .44476
 AGE_INC|    -.01843          .01691        -1.090   .2756     19.0288
  FEMALE|     .65366***       .07588         8.615   .0000      .46343
--------+-------------------------------------------------------------
```

H_0: Age is not a significant determinant of Prob(Doctor = 1)

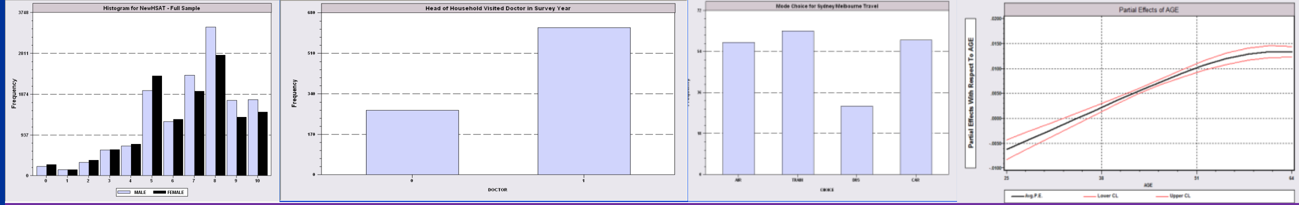
H_0: $\beta_2 = \beta_3 = \beta_5 = 0$



Likelihood Ratio Tests

- Null hypothesis restricts the parameter vector
- Alternative relaxes the restriction
- Test statistic: Chi-squared =
  2 (LogL|Unrestricted model –
  LogL|Restrictions) $\geq$ 0

Degrees of freedom = number of restrictions

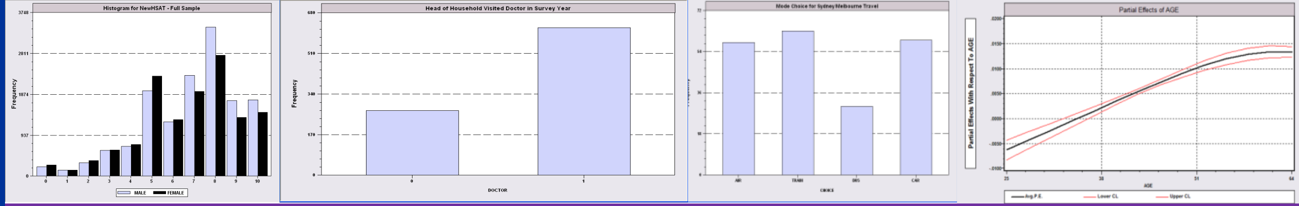


LR Test of H_0

```
UNRESTRICTED MODEL
Binary Logit Model for Binary Choice
Dependent variable                    DOCTOR
Log likelihood function          -2085.92452
Restricted log likelihood        -2169.26982
Chi squared [    5 d.f.]           166.69058
Significance level                    .00000
McFadden Pseudo R-squared           .0384209
Estimation based on N =     3377, K =    6
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC             1.23892      4183.84905
```

```
RESTRICTED MODEL
Binary Logit Model for Binary Choice
Dependent variable                    DOCTOR
Log likelihood function          -2124.06568
Restricted log likelihood        -2169.26982
Chi squared [    2 d.f.]            90.40827
Significance level                    .00000
McFadden Pseudo R-squared           .0208384
Estimation based on N =     3377, K =    3
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC             1.25974      4254.13136
```

Chi squared[3] = 2[-2085.92452 - (-2124.06568)] = 77.46456



Wald Test

- Unrestricted parameter vector is estimated
- Discrepancy: $\mathbf{q} = \mathbf{Rb} - \mathbf{m}$ (or $\mathbf{r}(\mathbf{b},\mathbf{m})$ if nonlinear) is computed
- Variance of discrepancy is estimated:

$$\text{Var}[\mathbf{q}] = \mathbf{R}\,\mathbf{V}\,\mathbf{R}'$$

- Wald Statistic is $\mathbf{q}'[\text{Var}(\mathbf{q})]^{-1}\mathbf{q} = \mathbf{q}'[\mathbf{RVR}']^{-1}\mathbf{q}$

Wald Test

Rb − m = q

R:

0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	0	1	0

b:

1.86428
-0.102093
0.00154004
0.512059
-0.018433
0.653659

\- m:

0
0
0

= q:

-0.102093
0.00154004
-0.018433

RVR’

R:

0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	0	1	0

V:

0.459584	-0.019886	0.000198178	-0.231031	0.00556395	0.00125442
-0.019886	0.000934022	-9.98629e-006	0.00516338	-0.00014392	-0.000160159
0.000198178	-9.98629e-006	1.14272e-007	-1.03054e-006	2.83849e-007	1.56007e-006
-0.231031	0.00516338	-1.03054e-006	0.556523	-0.0122686	-0.000805883
0.00556395	-0.00014392	2.83849e-007	-0.0122686	0.000285847	2.77535e-005
0.00125442	-0.000160159	1.56007e-006	-0.000805883	2.77535e-005	0.00575729

R’:

0	0	0
1	0	0
0	1	0
0	0	0
0	0	1

=

0.000934022	-9.98629e-006	-0.00014392
-9.98629e-006	1.14272e-007	2.83849e-007
-0.00014392	2.83849e-007	0.000285847

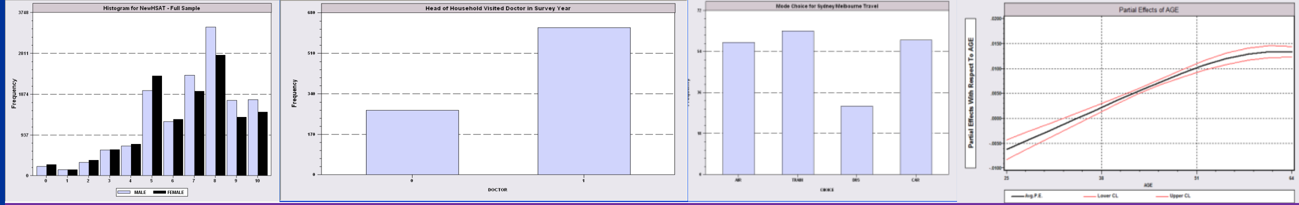
-0.102093	0.00154004	-0.018433

$\left[\begin{array}{ccc} 0.000934022 & -9.98629e\text{-}006 & -0.00014392 \\ -9.98629e\text{-}006 & 1.14272e\text{-}007 & 2.83849e\text{-}007 \\ -0.00014392 & 2.83849e\text{-}007 & 0.000285847 \end{array} \right]^{-1}$

-0.102093
0.00154004
-0.018433

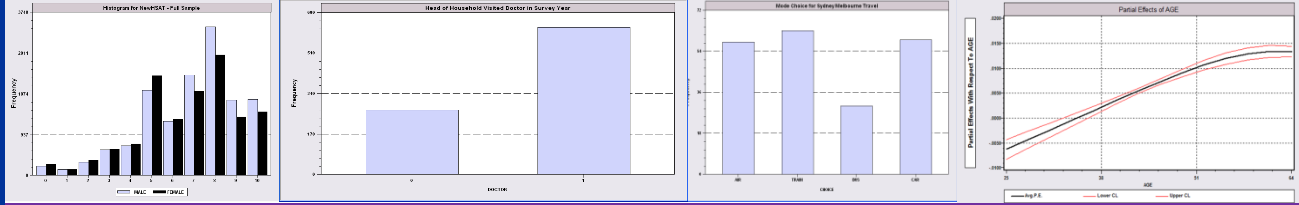
= 69.0541

Chi squared[3] = 69.0541



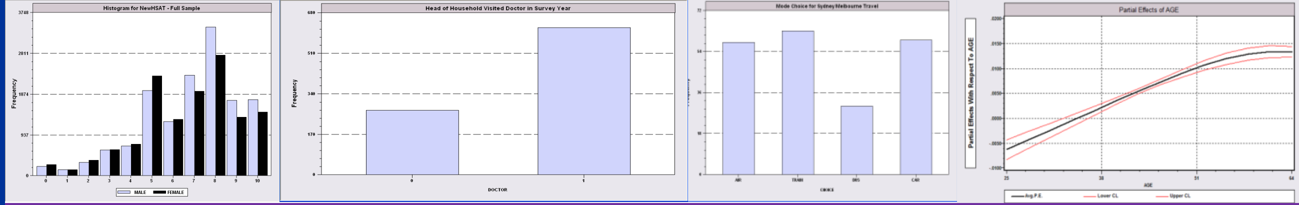
Lagrange Multiplier Test

- Restricted model is estimated
- Derivatives of unrestricted model and variances of derivatives are computed at restricted estimates
- Wald test of whether derivatives are zero tests the restrictions
- Usually hard to compute – difficult to program the derivatives and their variances.

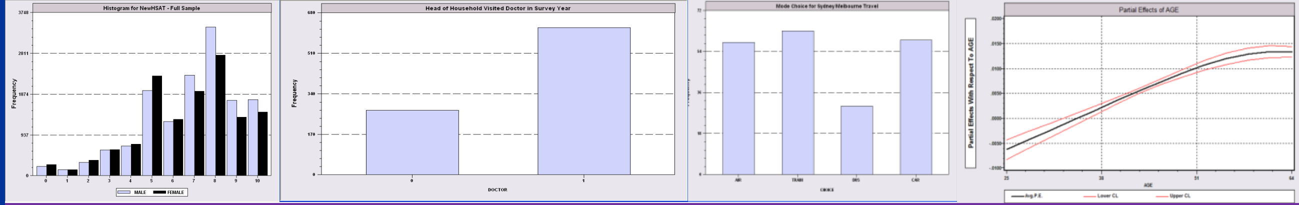


LM Test for a Logit Model

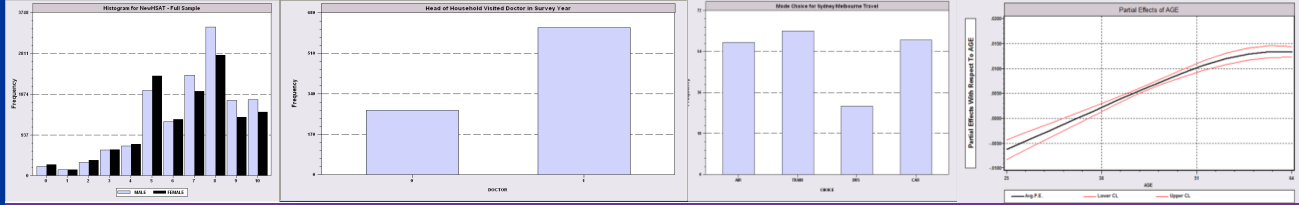
- Compute $\mathbf{b}_0$ (subject to restictions) (e.g., with zeros in appropriate positions.
- Compute $P_i(\mathbf{b}_0)$ for each observation.
- Compute $e_i(\mathbf{b}_0) = [y_i - P_i(\mathbf{b}_0)]$
- Compute $\mathbf{g}_i(\mathbf{b}_0) = \mathbf{x}_i e_i$ using full $\mathbf{x}_i$ vector
- $LM = [\Sigma_i \mathbf{g}_i(\mathbf{b}_0)]'[\Sigma_i \mathbf{g}_i(\mathbf{b}_0)\mathbf{g}_i(\mathbf{b}_0)]^{-1}[\Sigma_i \mathbf{g}_i(\mathbf{b}_0)]$




```
? Logit Model with quadratic and interaction
Namelist ; x=one,age,age*age,income,
               age*income,female $
Logit      ; if[year=1994]
           ; Lhs = doctor
           ; Rhs = x
? Constrained MLE. Force 3 coefficients to = 0
           ; cml:b(2)=0,b(3)=0,b(5)=0
           ; Prob = p$
? First derivative (scale part)
Create     ; gi= (doctor - p) ; gi2 = gi*gi $
? Second derivative (scale part)
Create     ; hi=p*(1-p)$
? LM statistic based on BHHH estimator
Matrix ;if[year=1994] ; list ; G = X'gi $
Matrix ;if[year=1994] ; List ; LM = g'*<X'[gi2]X>*g $
? LM statistic uses internal routine
Logit    ; if[year=1994] ; Lhs=doctor ; Rhs=x
         ; Start = b ; Maxit=0$
? LM statistic based on actual second derivatives
Matrix ;if[year=1994] ; List ; ML = g'*<X'[hi]X>*g $
```



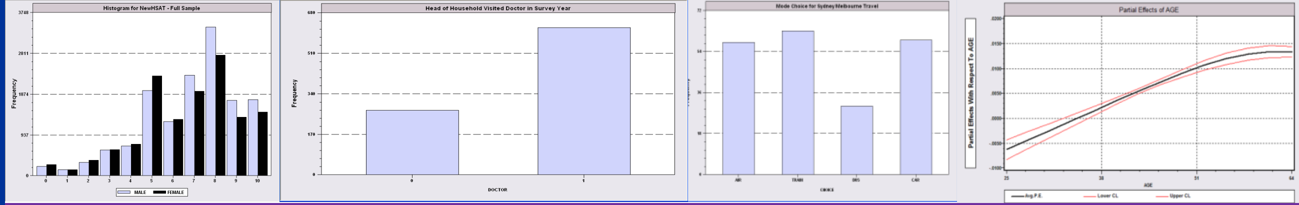

```
-----------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                    DOCTOR
Log likelihood function          -2124.06568
Restricted log likelihood        -2169.26982
Chi squared [   5](P= .000)         90.40827
Significance level                    .00000
McFadden Pseudo R-squared           .0208384
Estimation based on N =    3377, K =    3
Inf.Cr.AIC  =    4254.1 AIC/N =      1.260
Linear constraints imposed                 3
--------+--------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
  DOCTOR|  Coefficient     Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|      .52822***     .08978     5.88  .0000      .35227     .70418
     AGE|     0.0      .....(Fixed Parameter).....
 AGE*AGE|     0.0      .....(Fixed Parameter).....
  INCOME|    -.37810**       .16741    -2.26  .0239    -.70623    -.04998
        |Interaction AGE*INCOME
_ntrct02|     0.0      .....(Fixed Parameter).....
  FEMALE|      .67750***     .07483     9.05  .0000      .53084     .82416
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```




```
|-> Create   ; gi= (doctor - p) ; gi2 = gi*gi $
|-> Matrix ;if[year=1994] ; list ; G = X'gi $
      G|            1
--------+--------------
       1|   .239344E-05
       2|       2268.60
       3|       212205.
       4|   .968396E-06
       5|       849.705
       6|   .238041E-05
|-> Matrix ;if[year=1994] ; List ; ML = g'*<X'[gi2]X>*g $
     ML|            1
--------+--------------
       1|       81.4583
```

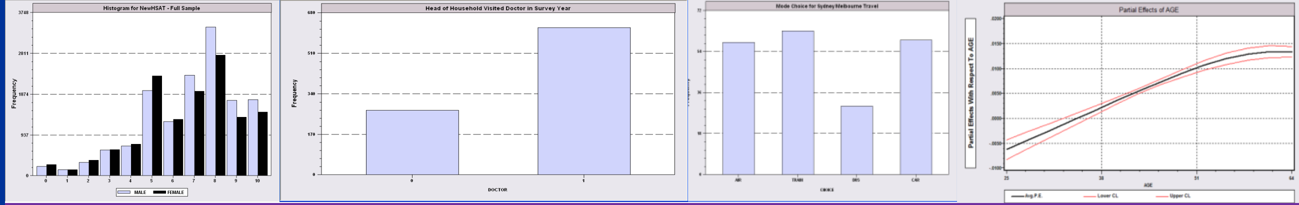
```
|-> Matrix ;if[year=1994] ; List ; ML = g'*<X'[hi]X>*g $

      ML|             1
--------+--------------
       1|       71.6745
```

```
 B_AND_G|             1              2
--------+------------------------------
       1|       .528225   .239344E-05
       2|       .000000       2268.60
       3|       .000000       212205.
       4|     -.378105    .968396E-06
       5|       .000000       849.705
       6|       .677500   .238041E-05
```

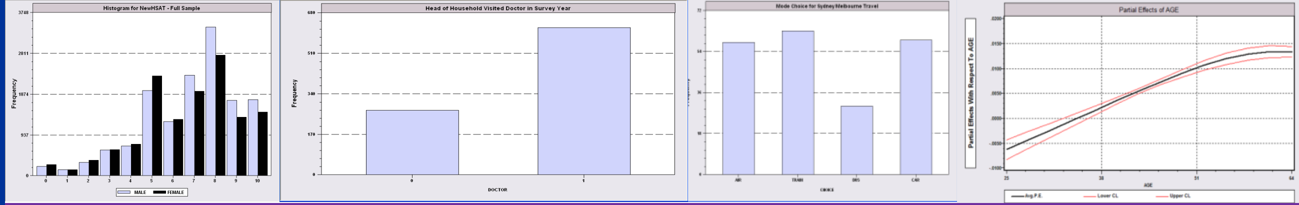


```
|-> Logit  ; if[year=1994] ; Lhs=doctor ; Rhs=x
    ; Start = b ; Maxit=0$
Maximum of      0 iterations. Exit iterations with status=1.
Maxit = 0. Computing LM statistic at starting values.
No iterations computed and no parameter update done.

-----------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                DOCTOR
LM Stat. at start values        71.67452
LM statistic kept as scalar     LMSTAT
Log likelihood function      -2124.06568
Restricted log likelihood    -2169.26982
Chi squared [   5](P= .000)     90.40827
Significance level                .00000
McFadden Pseudo R-squared       .0208384
Estimation based on N =    3377, K =   6
Inf.Cr.AIC  =    4260.1 AIC/N =    1.262
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  DOCTOR|  Coefficient      Error         z   |z|>Z*        Interval
--------+--------------------------------------------------------------------
Constant|      .52822        .66783      .79  .4290      -.78069   1.83714
     AGE|      0.0           .02967      .00 1.0000 -.58161D-01  .58161D-01
 AGE*AGE|      0.0           .00032      .00 1.0000 -.63007D-03  .63007D-03
  INCOME|     -.37810        .72928     -.52  .6041     -1.80747   1.05126
        |Interaction AGE*INCOME
_ntrct02|      0.0           .01625      .00 1.0000 -.31844D-01  .31844D-01
  FEMALE|      .67750***     .07522     9.01  .0000       .53007    .82493
--------+--------------------------------------------------------------------
```

```
|-> Matrix ;if[year=1994] ; List ; ML = g'*<X'[hi]X>*g $

      ML|             1
--------+--------------
       1|       71.6745
```



Inference About Partial Effects



Marginal Effects for Binary Choice

$$\textbf{LOGIT:}\quad \mathrm{E}[y\,|\,\overline{\mathbf{x}}] = \exp\left(\hat{\boldsymbol\beta}'\overline{\mathbf{x}}\right) \Big/ \left[1+\exp\left(\hat{\boldsymbol\beta}'\overline{\mathbf{x}}\right)\right] = \Lambda\left(\hat{\boldsymbol\beta}'\overline{\mathbf{x}}\right)$$

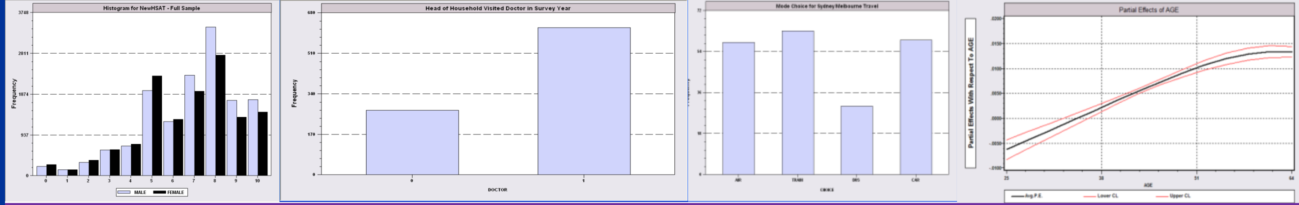
$$\hat{\boldsymbol\delta} = \frac{\partial \mathrm{E}[y\,|\,\overline{\mathbf{x}}]}{\partial\overline{\mathbf{x}}} = \left[\Lambda\left(\hat{\boldsymbol\beta}'\overline{\mathbf{x}}\right)\right]\left[1-\Lambda\left(\hat{\boldsymbol\beta}'\overline{\mathbf{x}}\right)\right]\hat{\boldsymbol\beta}$$

$$\textbf{PROBIT}\quad \mathrm{E}[y\,|\,\overline{\mathbf{x}}] = \Phi\left(\hat{\boldsymbol\beta}'\overline{\mathbf{x}}\right)$$

$$\hat{\boldsymbol\delta} = \frac{\partial \mathrm{E}[y\,|\,\overline{\mathbf{x}}]}{\partial\overline{\mathbf{x}}} = \left[\phi\left(\hat{\boldsymbol\beta}'\overline{\mathbf{x}}\right)\right]\hat{\boldsymbol\beta}$$

$$\textbf{EXTREME VALUE}\quad \mathrm{E}[y\,|\,\overline{\mathbf{x}}] = \mathrm{P}_1 = \exp\left[-\exp\left(-\hat{\boldsymbol\beta}'\overline{\mathbf{x}}\right)\right]$$

$$\hat{\boldsymbol\delta} = \frac{\partial \mathrm{E}[y\,|\,\overline{\mathbf{x}}]}{\partial\overline{\mathbf{x}}} = \mathrm{P}_1[-\log \mathrm{P}_1]\hat{\boldsymbol\beta}$$



The Delta Method

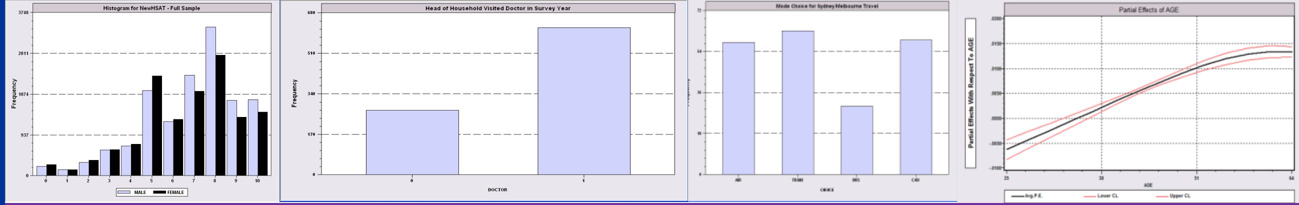
$$\hat{\delta} = f\left(\hat{\beta},\mathbf{x}\right),\quad \mathbf{G}\left(\hat{\beta},\mathbf{x}\right) = \frac{\partial f\left(\hat{\beta},\mathbf{x}\right)}{\partial\hat{\beta}'},\quad \hat{\mathbf{V}} = \text{Est.Asy.Var}\left[\hat{\beta}\right]$$

$$\textbf{Probit}\quad \mathbf{G} = \left[\phi\left(\hat{\beta}'\overline{\mathbf{x}}\right)\right] \left\{\mathbf{I} - \left(\hat{\beta}'\overline{\mathbf{x}}\right)\hat{\beta}\overline{\mathbf{x}}'\right\}$$

$$\textbf{Logit}\quad \mathbf{G} = \left[\Lambda\left(\hat{\beta}'\overline{\mathbf{x}}\right)\right]\left[1-\Lambda\left(\hat{\beta}'\overline{\mathbf{x}}\right)\right] \left\{\mathbf{I} + \left[1-2\Lambda\left(\hat{\beta}'\overline{\mathbf{x}}\right)\right]\hat{\beta}\overline{\mathbf{x}}'\right\}$$

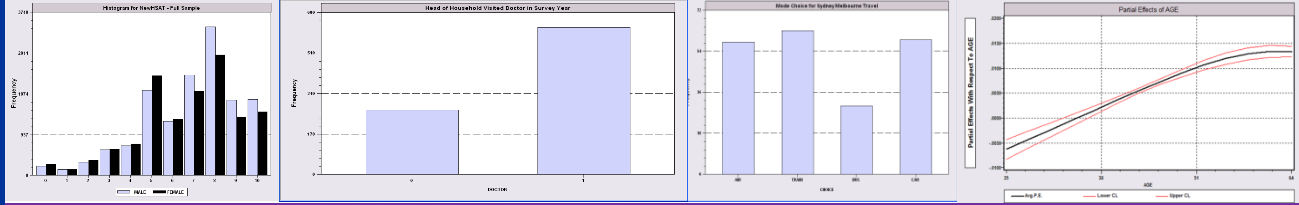
$$\textbf{ExtVlu}\quad \mathbf{G} = \left[\text{P}_1\left(\hat{\beta},\mathbf{x}\right)\right]\left[-\log \text{P}_1\left(\hat{\beta},\mathbf{x}\right)\right]\left\{\mathbf{I} - \left[1+\log \text{P}_1\left(\hat{\beta},\mathbf{x}\right)\right]\hat{\beta}\overline{\mathbf{x}}'\right\}$$

$$\text{Est.Asy.Var}\left[\hat{\delta}\right] = \left[\mathbf{G}\left(\hat{\beta},\mathbf{x}\right)\right]\hat{\mathbf{V}}\left[\mathbf{G}\left(\hat{\beta},\mathbf{x}\right)\right]$$



Computing Effects

- Compute at the data means?
  - Simple
  - Inference is well defined
- Average the individual effects
  - More appropriate?
  - Asymptotic standard errors more complicated.
- Is testing about marginal effects meaningful?
  - f(**b’x**) must be > 0; **b** is highly significant
  - How could f(**b’x**)***b** equal zero?

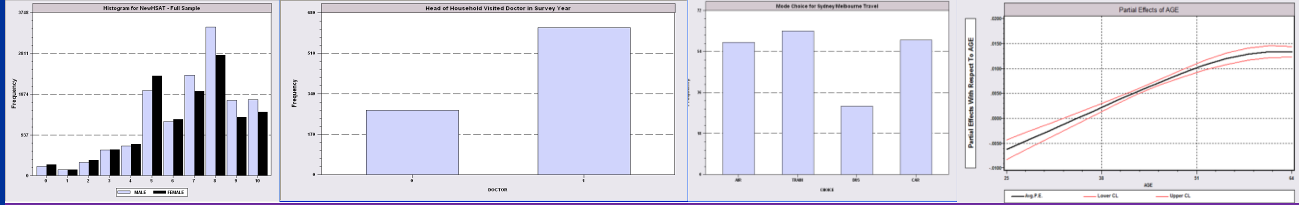


APE vs. Partial Effects at the Mean

Delta Method for Average Partial Effect

$$\text{Estimator of Var}\left[\frac{1}{N}\sum_{i=1}^{N}\text{PartialEffect}_i\right] = \bar{\mathbf{G}}\ \mathbf{Var}\left[\hat{\boldsymbol{\beta}}\right]\bar{\mathbf{G}}'$$

```
--> partials ; effects: hhninc/female ; summary $
--------------------------------------------------------------------
Partial Effects for Probit Probability Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
--------------------------------------------------------------------
                   Partial     Standard
(Delta method)     Effect       Error      |t|   95% Confidence Interval
--------------------------------------------------------------------
      HHNINC       -.05496     .03762     1.46     -.12869        .01877
   *  FEMALE        .14021     .01599     8.77      .10886        .17155
--------------------------------------------------------------------
--> partials ; effects: hhninc/female ; summary ; means$
--------------------------------------------------------------------
Partial Effects for Probit Probability Function
Partial Effects Computed at data Means
* ==> Partial Effect for a Binary Variable
--------------------------------------------------------------------
                   Partial     Standard
(Delta method)     Effect       Error      |t|   95% Confidence Interval
--------------------------------------------------------------------
      HHNINC       -.06374     .04009     1.59     -.14232        .01484
      FEMALE        .15045     .01752     8.59      .11611        .18479
--------------------------------------------------------------------
```



Method of Krinsky and Robb

Estimate **β** by Maximum Likelihood with **b**

Estimate asymptotic covariance matrix with **V**

Draw R observations **b**(r) from the normal population N[**b**,**V**]

b(r) = **b** + **C*****v**(r), **v**(r) drawn from N[0,I]
C = Cholesky matrix, **V** = **CC'**

Compute partial effects **d**(r) using **b**(r)

Compute the sample variance of **d**(r),r=1,…,R

Use the sample standard deviations of the R observations to estimate the sampling standard errors for the partial effects.

[4, 1] Cell:

	1
1	-0.251789
2	0.01445
3	-0.271285
4	0.386852

[4, 4] Cell:

	1	2	3	4
1	0.00937614	-0.000158129	-0.0043301	-0.000677015
2	-0.000158129	3.96472e-006	-1.1648e-005	-6.59447e-006
3	-0.0043301	-1.1648e-005	0.0105988	0.000122354
4	-0.000677015	-6.59447e-006	0.000122354	0.00208521

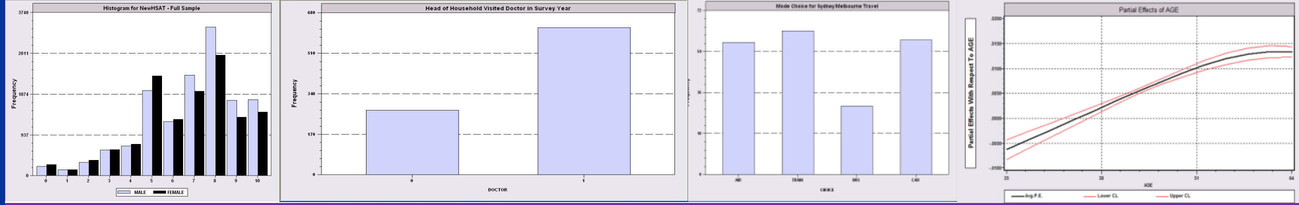
[4, 4] Cell: 0.0968304

	1	2	3	4
1	0.0968304	0	0	0
2	-0.00163305	0.00113924	0	0
3	-0.0447184	-0.074326	0.0554499	0
4	-0.00699176	-0.0158108	-0.0246251	0.0343503

Krinsky and Robb

```
-----------------------------------------------------
WALD procedure. Estimates and standard errors
for nonlinear functions and joint test of
nonlinear restrictions.
Wald Statistic               =*************
Prob. from Chi-squared[ 5] =          .00000
Krinsky-Robb method used with 1000 draws
--------+--------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er.
--------+--------------------------------------------
 Fncn(1)|      .42279***         .02273      18.599
 Fncn(2)|      .36483***         .00351     104.020
 Fncn(3)|      .00527***         .00071       7.467
 Fncn(4)|    -.09897***          .03829      -2.585
 Fncn(5)|      .14114***         .01642       8.597
--------+--------------------------------------------
Note: ***, **, * = Significance at 1%, 5%, 10% lev
-----------------------------------------------------
```

Delta Method	
PROBIT	
Estimate	**t ratio**
.00527	7.269
-.09897	-2.636
.13958	8.264



Partial Effect for Nonlinear Terms

$$\text{Prob} = \Phi[\alpha + \beta_1\text{Age} + \beta_2\text{Age}^2 + \beta_3\text{Income} + \beta_4\text{Female}]$$

$$\frac{\partial\text{Prob}}{\partial\text{Age}} = \phi[\alpha + \beta_1\text{Age} + \beta_2\text{Age}^2 + \beta_3\text{Income} + \beta_4\text{Female}] \times (\beta_1 + 2\beta_2\text{Age})$$

(1) Must be computed for a specific value of Age

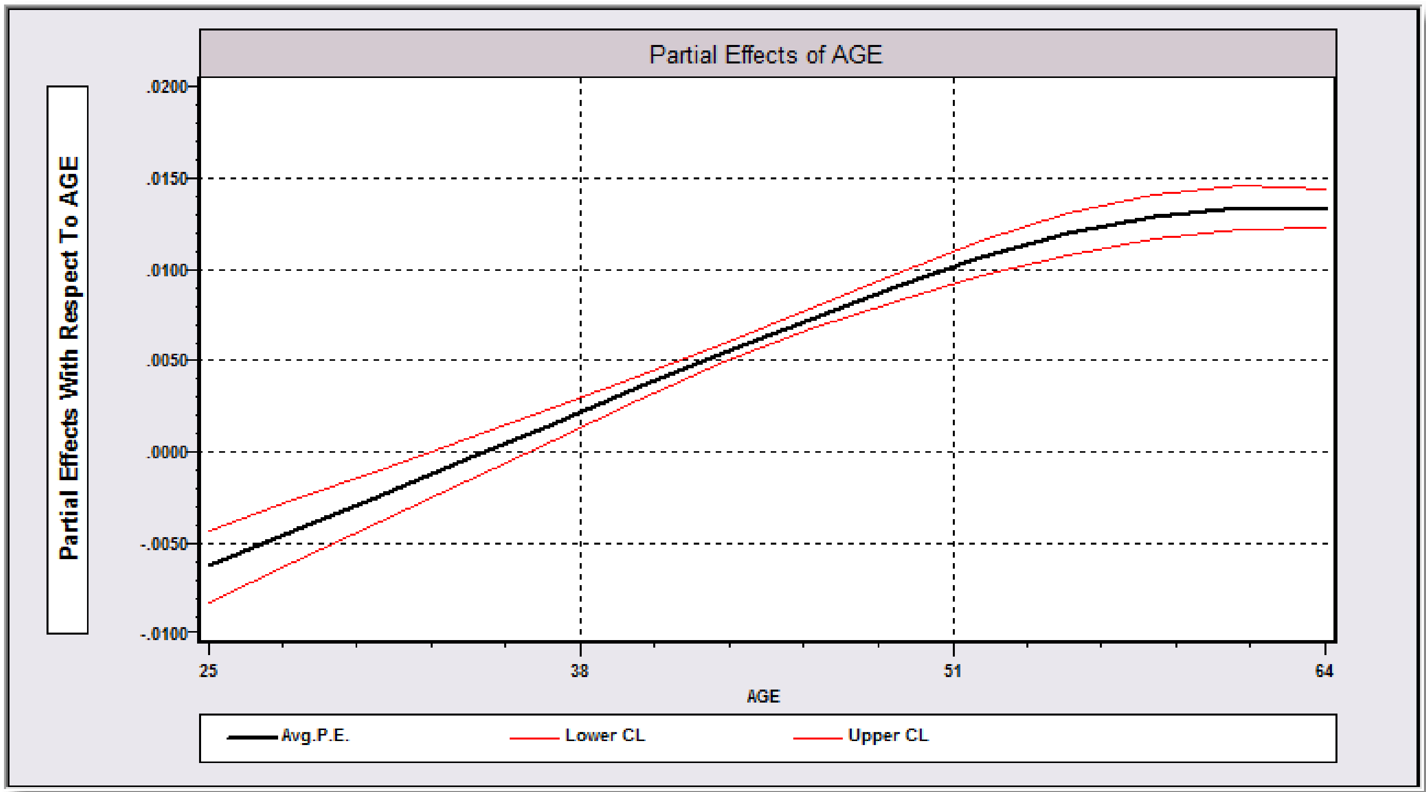
(2) Compute standard errors using delta method or Krinsky and Robb.

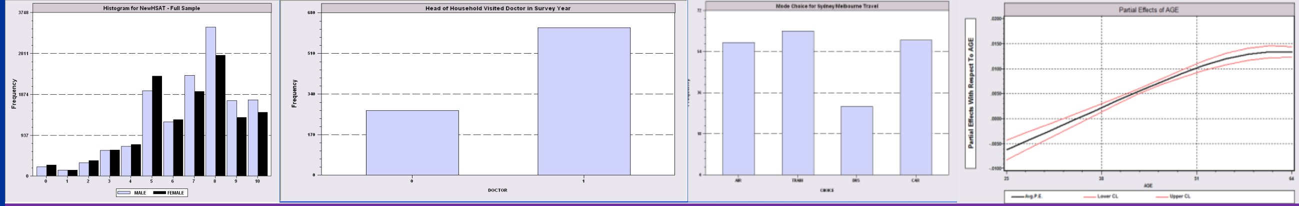
(3) Compute confidence intervals for different values of Age.

(4) Test of hypothesis that this equals zero is identical to a test that $(\beta_1 + 2\beta_2\text{Age}) = 0$. Is this an interesting hypothesis?

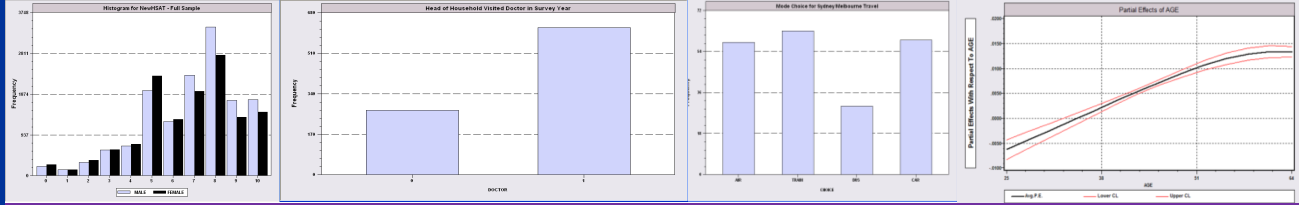
$$\frac{\partial\text{Prob}}{\partial AGE} = \begin{matrix} \phi(1.30811 - .06487\,Age + .0091\,Age^2 - .17362\,Income + .39666)\,Female) \\ \times[(-.06487 + 2(.0091)\,Age] \end{matrix}$$

Average Partial Effect: Averaged over Sample Incomes and Genders for Specific Values of Age





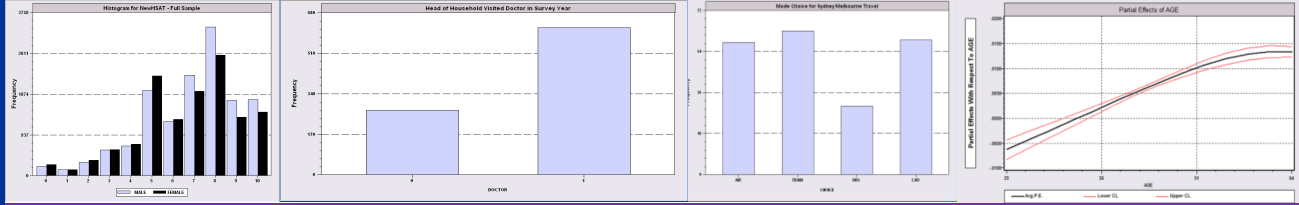
Endogeneity



Endogenous RHS Variable

- $U^* \;=\; \boldsymbol{\beta}'\mathbf{x} + \theta h + \varepsilon$
  $y \;=\; 1[U^* > 0]$

  $E[\varepsilon|h] \neq 0$ (h is endogenous)
  - Case 1: h is continuous
  - Case 2: h is binary = a treatment effect
- Approaches
  - Parametric: Maximum Likelihood
  - Semiparametric (not developed here):
    - GMM
    - Various approaches for case 2



Endogenous Continuous Variable

$$\begin{aligned} U^* &= \boldsymbol{\beta}'\mathbf{x} + \theta h + \varepsilon \\ y &= 1[U^* > 0] \\ h &= \boldsymbol{\alpha}'\mathbf{z} \qquad\quad + u \end{aligned}$$

Correlation = ρ.
This is the source of the endogeneity

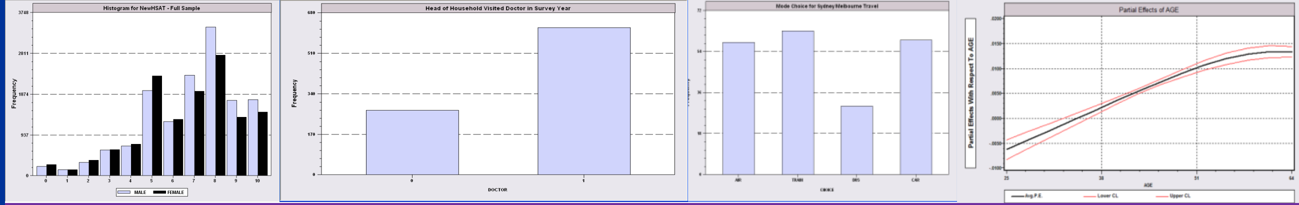
$E[\varepsilon|h] \neq 0 \Leftrightarrow \text{Cov}[u, \varepsilon] \neq 0$

Additional Assumptions:

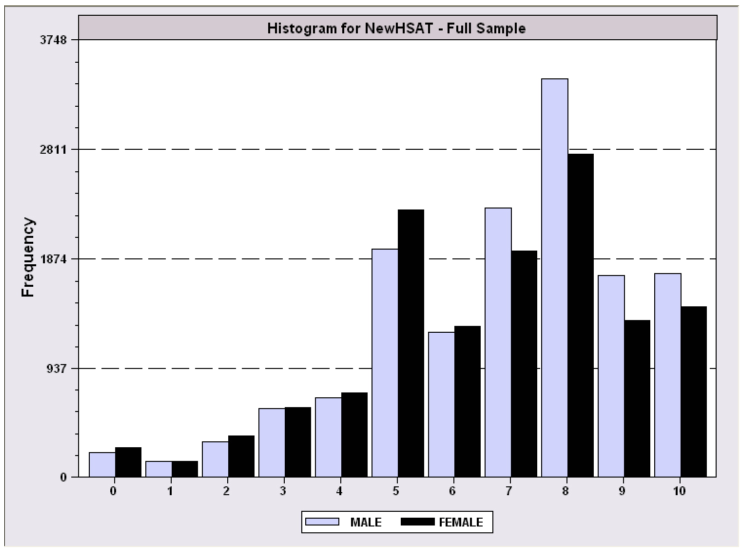
$(u,\varepsilon) \sim N[(0,0),(\sigma_u^2, \rho\sigma_u, 1)]$

z = a valid set of exogenous variables, uncorrelated with (u,ε)

This is not IV estimation. Z may be uncorrelated with X without problems.



Endogenous Income



0 = Not Healthy

1 = Healthy

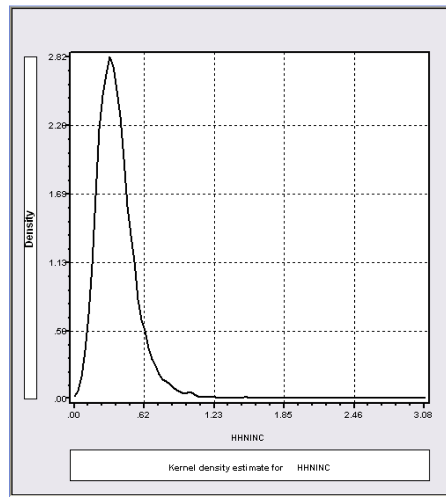
Healthy = 0 or 1

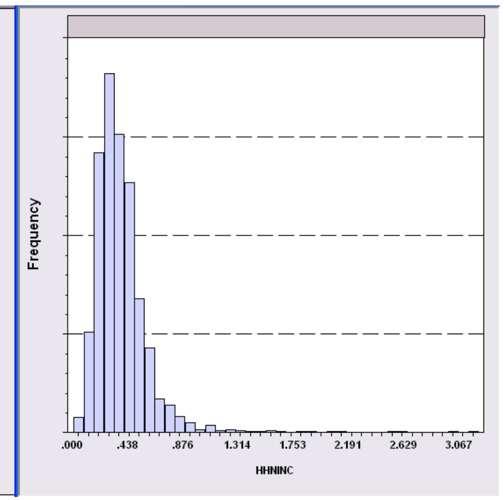
Age, Married, Kids, Gender, Income

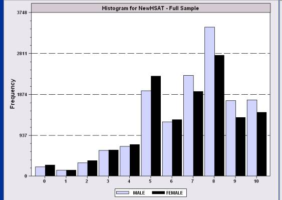
Determinants of Income (observed and unobserved) also determine health satisfaction.

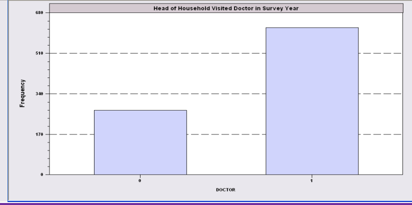
Income responds to

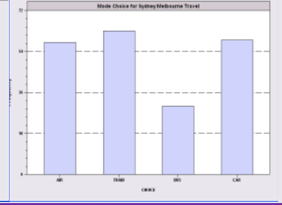
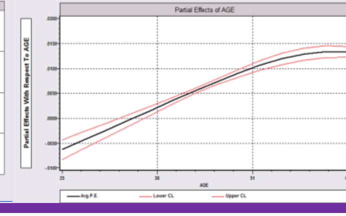
Age, Age^2, Educ, Married, Kids, Gender









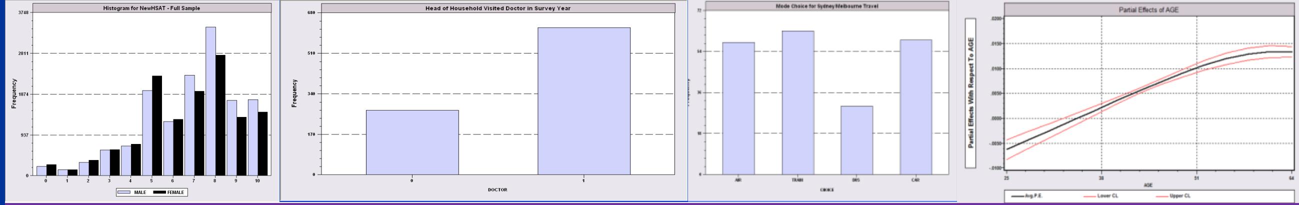
Estimation by ML (Control Function)

Probit fit of y to $\mathbf{x}$ and h will not consistently estimate $(\boldsymbol{\beta},\theta)$ because of the correlation between h and ε induced by the correlation of u and ε. Using the bivariate normality,

$$\text{Prob}(y=1\,|\,\mathbf{x},h)=\Phi\left[\frac{\boldsymbol{\beta}'\mathbf{x}+\theta h+(\rho/\sigma_u)u}{\sqrt{1-\rho^2}}\right]$$

Insert $u_i = (h_i - \boldsymbol{\alpha}'\mathbf{z})/\sigma_u$ and include f(h|$\mathbf{z}$) to form logL

$$\text{logL}=\sum_{i=1}^{N}\left\{\begin{array}{l}\log\Phi\left[(2y_i-1)\left(\dfrac{\boldsymbol{\beta}'\mathbf{x}_i+\theta h_i+\rho\left(\dfrac{h_i-\boldsymbol{\alpha}'\mathbf{z}_i}{\sigma_u}\right)}{\sqrt{1-\rho^2}}\right)\right]+\\ \log\dfrac{1}{\sigma_u}\phi\left[\left(\dfrac{h_i-\boldsymbol{\alpha}'\mathbf{z}_i}{\sigma_u}\right)\right]\end{array}\right\}$$



Two Approaches to ML

(1) **Full information ML.** Maximize the full log likelihood with respect to $(\boldsymbol{\beta},\theta,\sigma_u,\boldsymbol{\alpha},\rho)$

(The built in Stata routine IVPROBIT does this. It is not an instrumental variable estimator; it is a FIML estimator.)

Note also, this does not imply replacing h with a prediction from the regression then using probit with $\hat{h}$ instead of h.

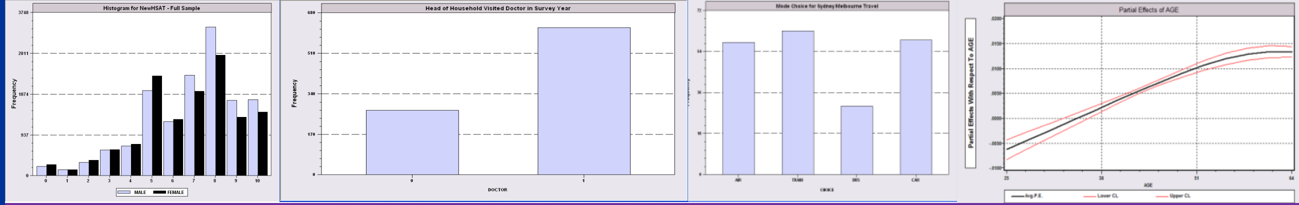
(2) **Two step limited information ML. (Control Function)**

(a) Use OLS to estimate $\boldsymbol{\alpha}$ and σ_u with **a** and s.

(b) Compute $\hat{v}_i = \hat{u}_i/s = (h_i - \mathbf{a}'\mathbf{z}_i)/s$

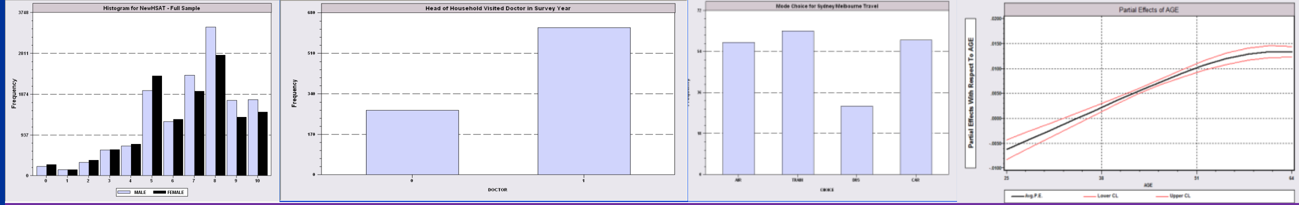
(c) $\log\hat{\Phi}\left[\dfrac{\boldsymbol{\beta}'\mathbf{x}_i + \theta h_i + \rho\hat{v}_i}{\sqrt{1-\rho^2}}\right] = \log\hat{\Phi}\left[\boldsymbol{\delta}'\mathbf{x}_i + \lambda h_i + \tau\hat{v}_i\right]$

The second step is to fit a probit model for y to $(\mathbf{x},h,\hat{v})$ then solve back for $(\boldsymbol{\beta},\theta,\rho)$ from $(\boldsymbol{\delta},\lambda,\tau)$ and from the previously estimated **a** and s. Use the delta method to compute standard errors.



FIML Estimates

```
----------------------------------------------------------------------
Probit with Endogenous RHS Variable
Dependent variable                 HEALTHY
Log likelihood function        -6464.60772
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Coefficients in Probit Equation for HEALTHY
Constant|    1.21760***       .06359      19.149   .0000
     AGE|    -.02426***       .00081     -29.864   .0000     43.5257
 MARRIED|    -.02599          .02329      -1.116   .2644      .75862
  HHKIDS|     .06932***       .01890       3.668   .0002      .40273
  FEMALE|    -.14180***       .01583      -8.959   .0000      .47877
  INCOME|     .53778***       .14473       3.716   .0002      .35208
        |Coefficients in Linear Regression for INCOME
Constant|    -.36099***       .01704     -21.180   .0000
     AGE|     .02159***       .00083      26.062   .0000     43.5257
   AGESQ|    -.00025***     .944134D-05  -26.569   .0000     2022.86
    EDUC|     .02064***       .00039      52.729   .0000     11.3206
 MARRIED|     .07783***       .00259      30.080   .0000      .75862
  HHKIDS|    -.03564***       .00232     -15.332   .0000      .40273
  FEMALE|     .00413**        .00203       2.033   .0420      .47877
        |Standard Deviation of Regression Disturbances
Sigma(w)|     .16445***       .00026     644.874   .0000
        |Correlation Between Probit and Regression Disturbances
Rho(e,w)|    -.02630          .02499      -1.052   .2926
--------+-------------------------------------------------------------
```



Partial Effects: Scaled Coefficients

Conditional Mean

$$\text{E}[y\,|\,\mathbf{x},h] = \Phi(\boldsymbol{\beta}'\mathbf{x} + \theta h)$$

$$h = \boldsymbol{\alpha}'\mathbf{z} + u = \boldsymbol{\alpha}'\mathbf{z} + \sigma_u v \text{ where } v \sim \text{N}[0,1]$$

$$\text{E}[\text{y}|\text{x},\mathbf{z},v] = \Phi[\boldsymbol{\beta}'\mathbf{x} + \theta(\boldsymbol{\alpha}'\mathbf{z} + \sigma_u v)]$$

Partial Effects. Assume $\mathbf{z} = \mathbf{x}$ (just for convenience)

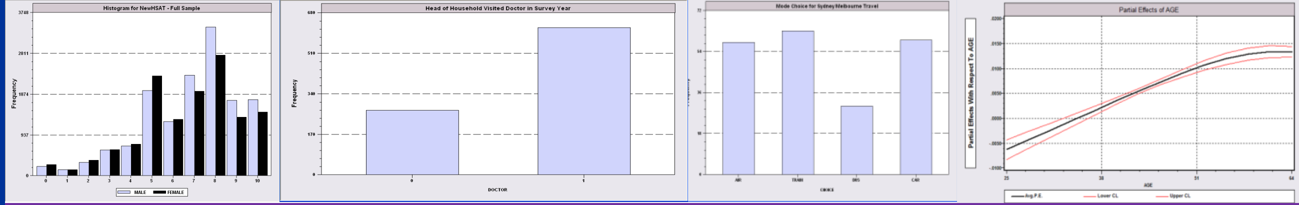
$$\frac{\partial \text{E}[\text{y}|\mathbf{x},\mathbf{z},v]}{\partial \mathbf{x}} = \phi[\boldsymbol{\beta}'\mathbf{x} + \theta(\boldsymbol{\alpha}'\mathbf{z} + \sigma_u v)](\boldsymbol{\beta} + \theta\boldsymbol{\alpha})$$

$$\frac{\partial \text{E}[\text{y}|\mathbf{x},\mathbf{z}]}{\partial \mathbf{x}} = \text{E}_v\left[\frac{\partial \text{E}[\text{y}|\mathbf{x},\mathbf{z},v]}{\partial \mathbf{x}}\right] = (\boldsymbol{\beta} + \theta\boldsymbol{\alpha})\int_{-\infty}^{\infty} \phi[\boldsymbol{\beta}'\mathbf{x} + \theta(\boldsymbol{\alpha}'\mathbf{z} + \sigma_u v)]\phi(v)dv$$

The integral does not have a closed form, but it can easily be simulated :

$$Est.\frac{\partial \text{E}[\text{y}|\mathbf{x},\mathbf{z}]}{\partial \mathbf{x}} = (\boldsymbol{\beta} + \theta\boldsymbol{\alpha})\frac{1}{R}\sum_{r=1}^{\text{R}} \phi[\boldsymbol{\beta}'\mathbf{x} + \theta(\boldsymbol{\alpha}'\mathbf{z} + \sigma_u v_r)]$$

For variables only in $\mathbf{x}$, omit $\theta\alpha_{\text{k}}$. For variables only in $\mathbf{z}$, omit β_{k}.



Partial Effects

```
+--------------------------------------+
| Listed Calculator Results            |
+--------------------------------------+
SB        =         .380919
```

$\theta = 0.53778$

Matrix - BMODEL

[7, 1] Cell: 1.2176

	1
1	1.2176
2	-0.02426
3	0
4	0
5	-0.02599
6	0.06932
7	-0.1418

Matrix - ALPHA

[7, 1] Cell: -0.36099

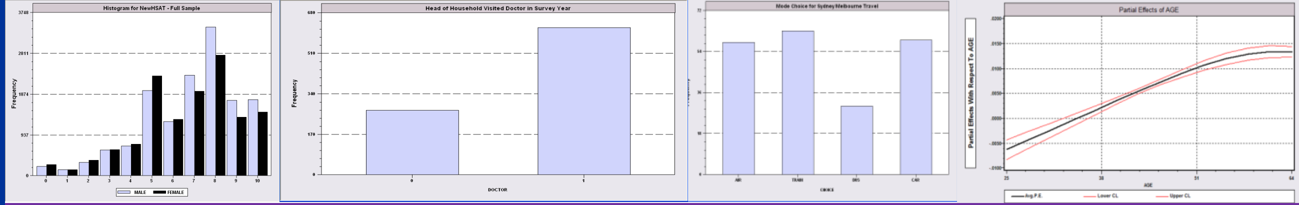
	1
1	-0.36099
2	0.02159
3	-0.00025
4	0.02064
5	0.07783
6	-0.03564
7	0.00413

Matrix - SCALEDB

[7, 1] Cell: 0.299023

	1
1	0.299023
2	-0.00369571
3	-3.92804e-005
4	0.00324299
5	0.00463535
6	0.0146532
7	-0.0407803

The scale factor is computed using the model coefficients, means of the variables and 35,000 draws from the standard normal population.



Endogenous Binary Variable

$$U^* = \boldsymbol{\beta}'\mathbf{x} + \theta h + \boxed{\varepsilon}$$
$$y = 1[U^* > 0]$$
$$h^* = \boldsymbol{\alpha}'\mathbf{z} + \boxed{u}$$
$$h = 1[h^* > 0]$$

$\Leftarrow$ **Correlation = ρ.**
This is the source of the endogeneity

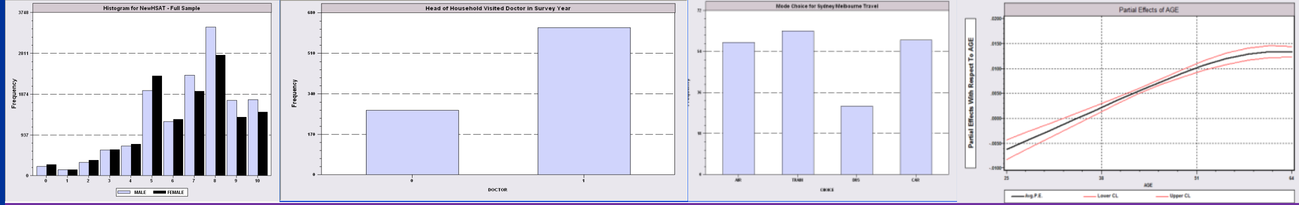
$$E[\varepsilon|h^*] \neq 0 \Leftrightarrow \text{Cov}[u, \varepsilon] \neq 0$$

Additional Assumptions:

$$(u,\varepsilon) \sim N[(0,0),(\sigma_u^2, \rho\sigma_u, 1)]$$

$\mathbf{z}$ = a valid set of exogenous variables, uncorrelated with (u,ε)

This is not IV estimation. Z may be uncorrelated with X without problems.



Endogenous Binary Variable

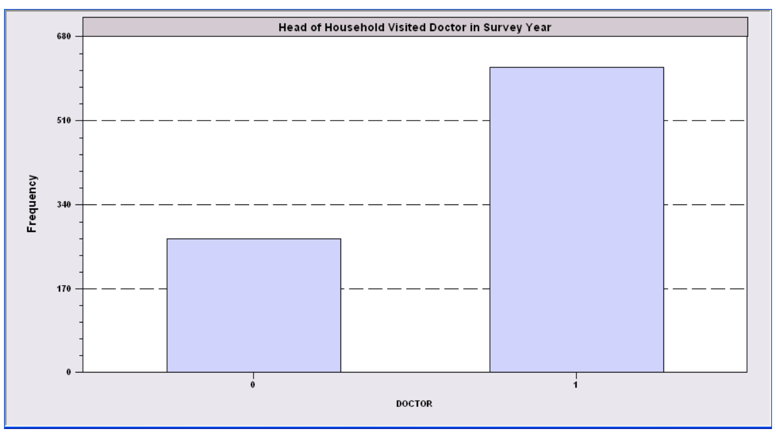
$$P(Y = y, H = h) = P(Y = y|H = h) \times P(H=h)$$

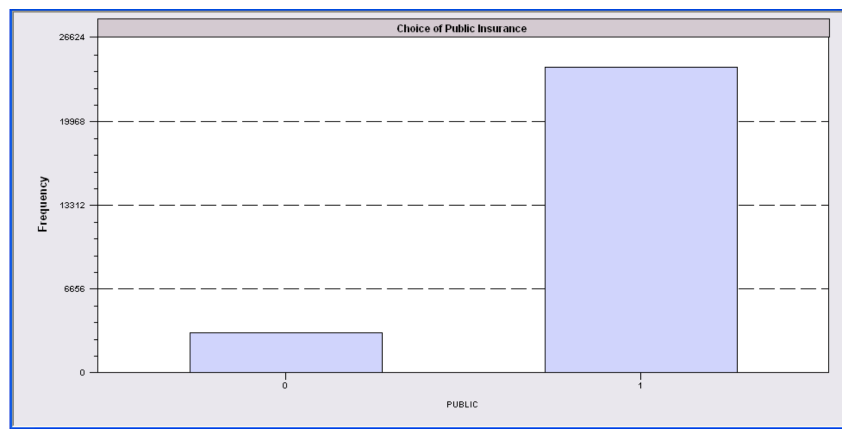
This is a simple bivariate probit model.

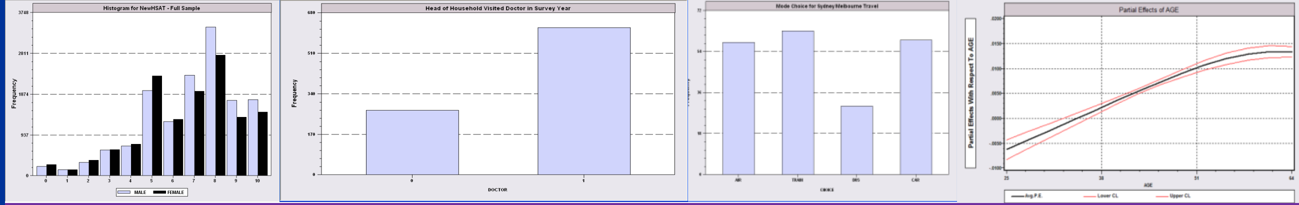
Not a simultaneous equations model - the estimator is FIML, not any kind of least squares.

Doctor = F(age,age^2,income,female,**Public**)

Public = F(age,educ,income,married,kids,female)

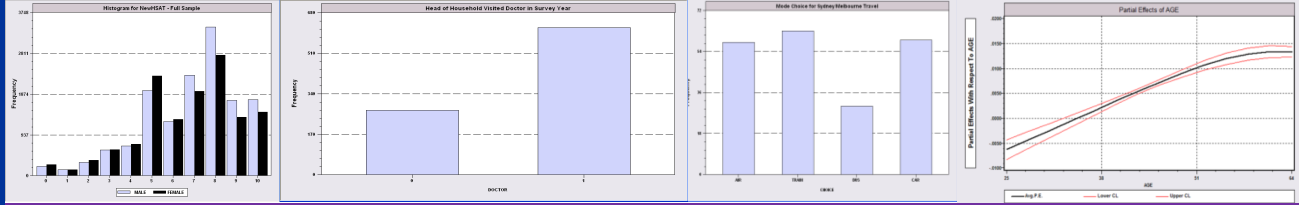






FIML Estimates

```
-----------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable                DOCPUB
Log likelihood function     -25671.43905
Estimation based on N =   27326, K =   14
--------+--------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+--------------------------------------------------------------
        |Index    equation for DOCTOR
Constant|     .59049***       .14473        4.080   .0000
     AGE|    -.05740***       .00601       -9.559   .0000      43.5257
   AGESQ|     .00082***    .681660D-04     12.100   .0000      2022.86
  INCOME|     .08883*         .05094        1.744   .0812       .35208
  FEMALE|     .34583***       .01629       21.225   .0000       .47877
  PUBLIC|     .43533***       .07357        5.917   .0000       .88571
        |Index    equation for PUBLIC
Constant|    3.55054***       .07446       47.681   .0000
     AGE|     .00067          .00115         .581   .5612      43.5257
    EDUC|    -.16839***       .00416      -40.499   .0000      11.3206
  INCOME|    -.98656***       .05171      -19.077   .0000       .35208
 MARRIED|    -.00985          .02922        -.337   .7361       .75862
  HHKIDS|    -.08095***       .02510       -3.225   .0013       .40273
  FEMALE|     .12139***       .02231        5.442   .0000       .47877
        |Disturbance correlation
RHO(1,2)|    -.17280***       .04074       -4.241   .0000
--------+--------------------------------------------------------------
```



Partial Effects

Conditional Mean

$$E[y\,|\,\mathbf{x},h] = \Phi(\boldsymbol{\beta}'\mathbf{x} + \theta h)$$

$$\begin{aligned} E[y\,|\,\mathbf{x},\mathbf{z}] &= E_h E[y\,|\,\mathbf{x},h] \\ &= \text{Prob}(h=0\,|\,\mathbf{z})E[y\,|\,\mathbf{x},h=0] + \text{Prob}(h=1\,|\,\mathbf{z})E[y\,|\,\mathbf{x},h=1] \\ &= \Phi(-\boldsymbol{\alpha}'\mathbf{z})\Phi(\boldsymbol{\beta}'\mathbf{x}) + \Phi(\boldsymbol{\alpha}'\mathbf{z})\Phi(\boldsymbol{\beta}'\mathbf{x}+\theta) \end{aligned}$$

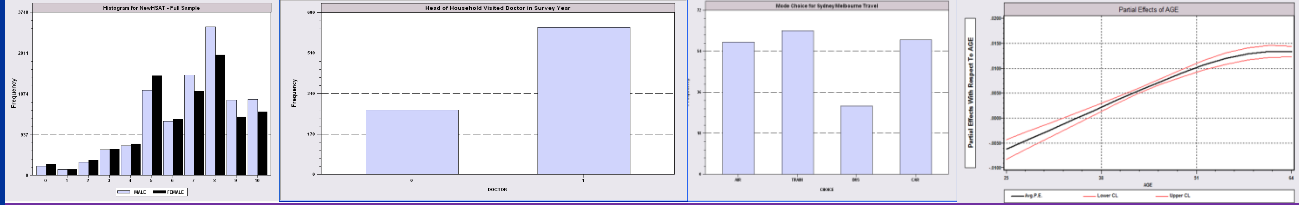
Partial Effects

Direct Effects

$$\frac{\partial E[y\,|\,\mathbf{x},\mathbf{z}]}{\partial \mathbf{x}} = \left[\Phi(-\boldsymbol{\alpha}'\mathbf{z})\phi(\boldsymbol{\beta}'\mathbf{x}) + \Phi(\boldsymbol{\alpha}'\mathbf{z})\phi(\boldsymbol{\beta}'\mathbf{x}+\theta)\right]\boldsymbol{\beta}$$

Indirect Effects

$$\begin{aligned} \frac{\partial E[y\,|\,\mathbf{x},\mathbf{z}]}{\partial \mathbf{z}} &= \left[-\phi(-\boldsymbol{\alpha}'\mathbf{z})\Phi(\boldsymbol{\beta}'\mathbf{x}) + \phi(\boldsymbol{\alpha}'\mathbf{z})\Phi(\boldsymbol{\beta}'\mathbf{x}+\theta)\right]\boldsymbol{\alpha} \\ &= \phi(\boldsymbol{\alpha}'\mathbf{z})\left[\Phi(\boldsymbol{\beta}'\mathbf{x}+\theta) - \Phi(\boldsymbol{\beta}'\mathbf{x})\right]\boldsymbol{\alpha} \end{aligned}$$



Identification Issues

- Exclusions are not needed for estimation
- Identification is, in principle, by “functional form”
- Researchers usually have a variable in the treatment equation that is not in the main probit equation “to improve identification”
- A fully simultaneous model
  - $y1 = f(\mathbf{x}1,y2)$, $y2 = f(\mathbf{x}2,y1)$
  - Not identified even with exclusion restrictions
  - (Model is “incoherent”)