ICPSR 2018

Modeling Categorical Outcomes

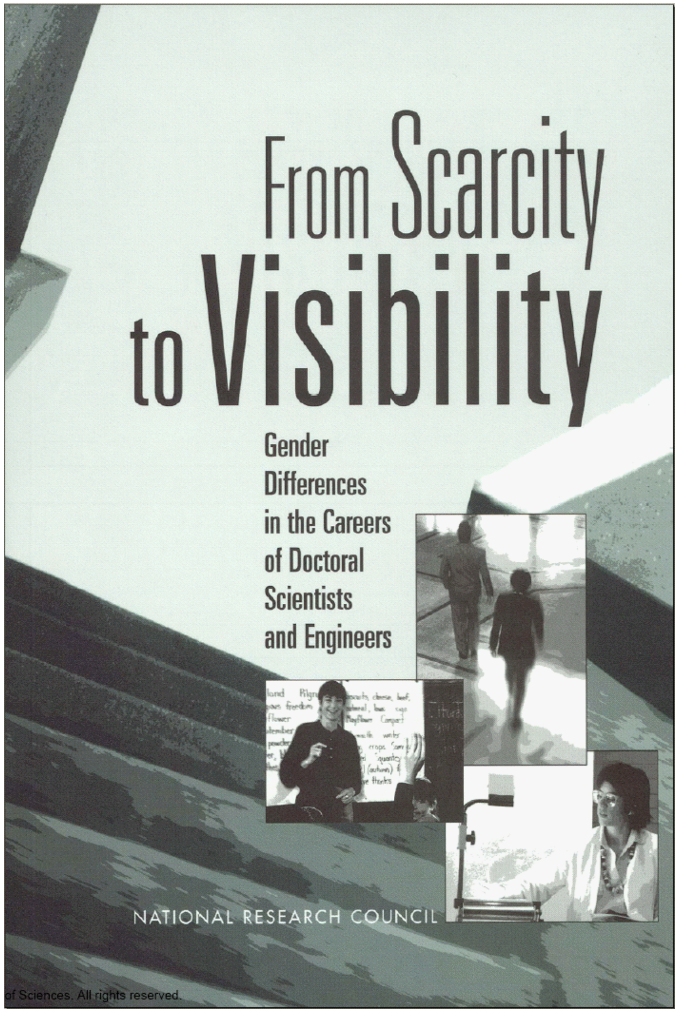
Advanced methods of interpretation

Scott Long – β1a draft

mco18-part2-b1-2018-04-02.docx

Table of contents

β1a Comparing Groups



Motivating example

Do levels of tenure differ for men and women with similar characteristics?

Are the effects of scientific productivity on receiving tenures the same for men and women?

Readings and examples

Long & Mustillo 2018. Using predictions to compare groups in regression models for binary outcomes. See references there.

mdo18-groups-.do*

Statistical and substantive issues

1. In LRM,
   - Fit separate models for each groups
   - Compare coefficients across groups.
1. <u>Substantive concern</u>: Academy panel did not find logit coefficients informative.
2. <u>Statistical problem</u>: Paul Allison's working paper said:

   Differences in the estimated coefficients <u>tell us nothing</u> about the group differences in the effect of articles on tenure.

3. Are the logit based analyses in the NAS report wrong?

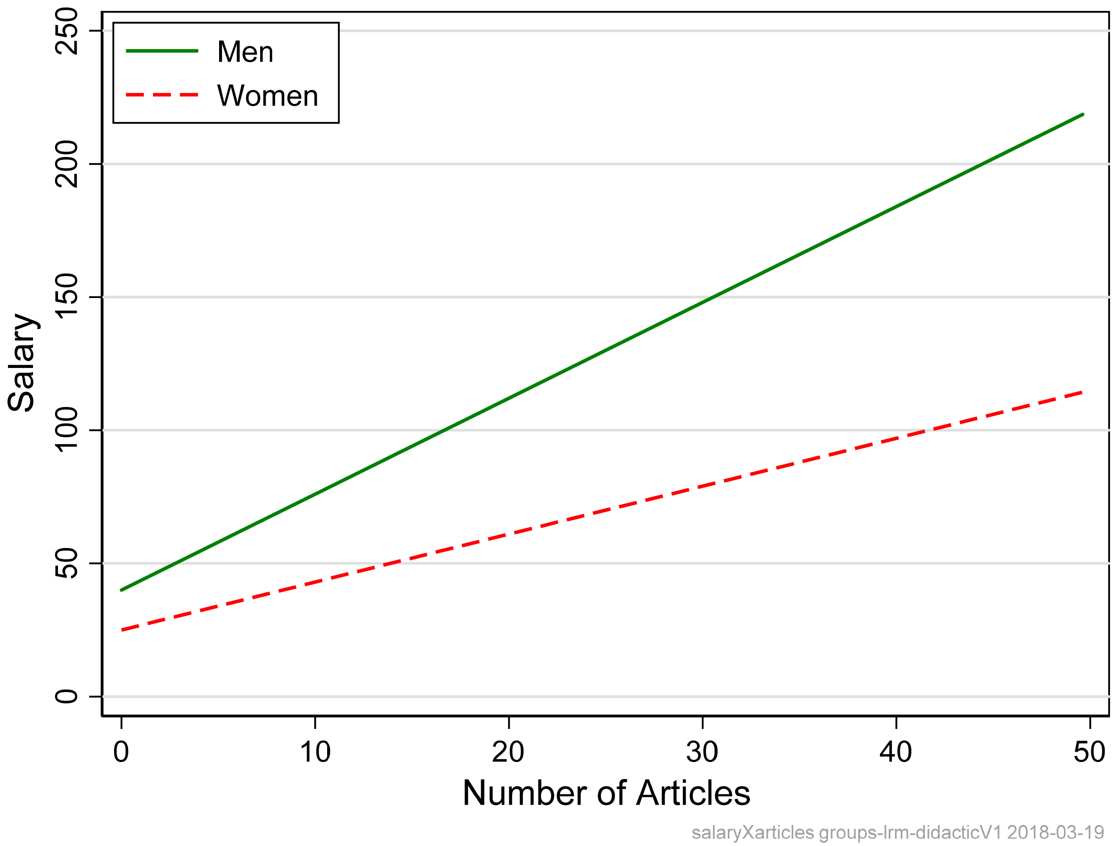
Roadmap

1. Linear regression model for group comparisons.
1. Binary regression model and scalar identification of regression coefficients
2. Group comparisons using predictions and marginal effects
3. Examples

LRM: comparing regressions

Men: $y = \alpha^m + \beta^m_{articles} articles + \beta^m_{prestige} prestige + \varepsilon$

Women: $y = \alpha^w + \beta^w_{articles} articles + \beta^w_{prestige} prestige + \varepsilon$



LRM: comparing effect of articles

1. Do men and women have the same return for articles?

$$H_0: \beta^w_{articles} = \beta^m_{articles}$$

2. Test statistic standardizes the difference

$$t = \frac{\hat{\beta}^w_{articles} - \hat{\beta}^m_{articles}}{\sqrt{Var\left(\hat{\beta}^w_{articles}\right) + Var\left(\hat{\beta}^m_{articles}\right) - 2Cov\left(\hat{\beta}^w_{articles}, \hat{\beta}^m_{articles}\right)}}$$

3. $Cov\left(\hat{\beta}^w_{articles}, \hat{\beta}^m_{articles}\right) = 0$ since the samples do not overlap
   - Might not be true with complex sampling

LRM: Group comparing entire model

1. The hypothesis that regression planes are identical is:

$$H_0: \alpha^w = \alpha^m;\ \beta^w_{articles} = \beta^m_{articles};\ \beta^w_{prestige} = \beta^m_{prestige}$$

If H_O is true, it does not imply that $R^2_w = R^2_m$.

- Indeed, I expect less explained variation for women due to discrimination.

LRM: Sources of groups differences

1. Groups can differ in
   - Coefficients
   - Unobserved heterogeneity
2. In LRM we can estimates coefficients and unobserved heterogeneity
3. In BRM we cannot estimate unobserved heterogeneity
   - This causes an identification problem for the coefficients

Binary regression model

1. Allow group differences in regression coefficients:

$$\text{Women:} \quad \Pr(y=1) = \Lambda\left(\alpha^w + \beta^w_{articles} articles + \beta^w_{prestige} prestige\right)$$

$$\text{Men:} \quad \Pr(y=1) = \Lambda\left(\alpha^m + \beta^m_{articles} articles + \beta^m_{prestige} prestige\right)$$

2. We want to tests

$$H_0: \beta^w_{articles} = \beta^m_{articles}$$

3. Unfortunately, standard tests are invalid:

   Because of an *identification problem*, the usual tests of this hypothesis tell us <u>nothing</u> about the underlying impact of articles for men and women (Allison 1999).

Scalar identification in the BRM

1. A LV approach makes this easy to see, but other approaches are possible
2. Regress latent y* on x:

$$y^* = \alpha + \beta x + \varepsilon$$

3. The mean, variance and form of ε are assumed:
   - Profit: ε is normal(0,1)
   - Logit: ε is logistic(0, $\pi^2/3$)
4. The variance cannot be estimated and is assumed
   - The values are largely a matter of convenience and tradition

Computing Pr(y) from y*

1. The observed y is linked to the latent y*

$$y = \begin{cases} 1 & \text{if } y^* > 0 \\ 0 & \text{if } y^* \le 0 \end{cases}$$

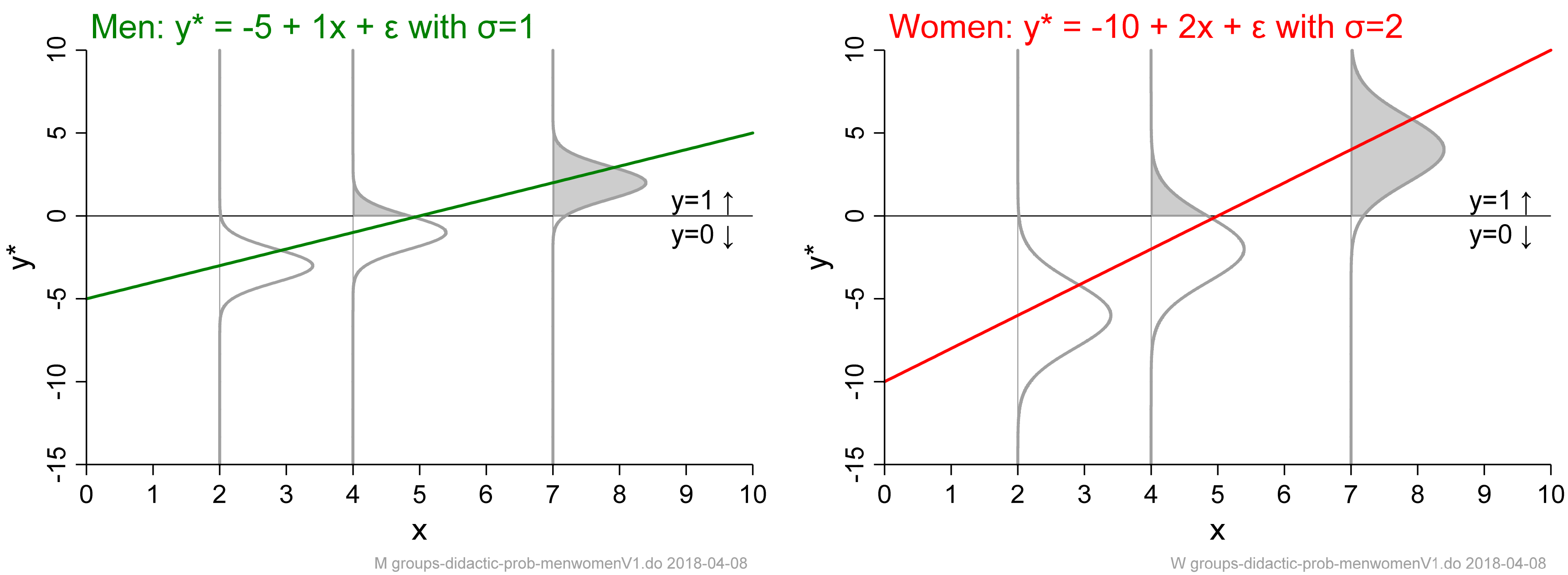
2. The probability Pr(y=1|x) depends on:
   - The error distribution
   - The regression coefficients
   - The value of x where the probability is computed

3. Formally,

$$\begin{aligned} \Pr\left(y = 1 \mid x\right) &= \Pr\left(y^* > 0 \mid x\right) \\ &= \Pr\left(\varepsilon < \left[\alpha + \beta x\right] \mid x\right) \end{aligned}$$

4. The identification problem is illustrated in this graph

Identification of Pr(y) but not β



1. β_X is <u>larger</u> for women than men.
2. ΔPr(y)/Δx is <u>identical</u> for women and men.

Identification of βs and group comparisons

1. Regress y* on articles:

$$\text{Women:} \quad y^* = \alpha^w + \beta^w_{\text{articles}}\text{articles} + \varepsilon_m$$

$$\text{Men:} \quad y^* = \alpha^m + \beta^m_{\text{articles}}\text{articles} + \varepsilon_w$$

2. I want to test if articles have the same effect for men and women:

$$\beta^w_{\text{articles}} = \beta^m_{\text{articles}}$$

3. Substantively, I expect:

$$\sigma^2_w > \sigma^2_m$$

4. Software traditionally assumes

Logit: $\text{Var}(\varepsilon) = \pi^2 / 3$

Probit: $\text{Var}(\varepsilon) = 1$

Fitting probit model for each groups

1. The ***true*** variance of ε: $Var\left(\varepsilon_g\right) = \sigma_g$

2. Software rescales ε so that: $Var\left(\tilde{\varepsilon}_g\right) = 1 = Var\left(\varepsilon_g / \sigma_g\right)$

3. For women, the model fit by software

$$\frac{y^*}{\sigma_w} = \frac{\alpha^w}{\sigma_w} + \frac{\beta^w_{\text{articles}}}{\sigma_w}\text{articles} + \frac{\varepsilon_w}{\sigma_w}$$

$$= \tilde{\alpha}^w + \tilde{\beta}^w_{\text{articles}}\text{articles} + \tilde{\varepsilon}_w, \text{ where } \tilde{\sigma}_w = 1$$

4. For men

$$\frac{y^*}{\sigma_m} = \frac{\alpha^m}{\sigma_m \sigma_m} + \frac{\beta^m_{\text{articles}}}{\sigma_m}\text{articles} + \frac{\varepsilon_m}{\sigma_m}$$

$$= \tilde{\alpha}^m + \tilde{\beta}^m_{\text{articles}}\text{articles} + \tilde{\varepsilon}_m, \text{ where } \tilde{\sigma}_m = 1$$

5. I want to test:

$$H_0^{NoTilda}: \beta^{w}_{\text{articles}} = \beta^{m}_{\text{articles}}$$

6. Standard software tests:

$$H_0^{Tilda}: \tilde{\beta}^{w}_{\text{articles}} = \tilde{\beta}^{m}_{\text{articles}}$$

7. The problem raised by Allison is:

- If $\tilde{\beta}^{w}_{\text{articles}} = \tilde{\beta}^{m}_{\text{articles}}$, then $\beta^{w}_{\text{articles}} = \beta^{m}_{\text{articles}}$ only if $\sigma_m^2 = \sigma_w^2$.
- But, we don't know if $\sigma_m^2 = \sigma_w^2$.

Allison test of the equality of true coefficients

1. *If* you are interested in comparing β_x^w and β_x^m
2. Assume for some variable z

$$\beta_z^w / \beta_z^m = 1$$

3. The ratio $\tilde{\beta}_z^w / \tilde{\beta}_z^m$ is the relative size of σ_m and σ_w since:

$$\frac{\tilde{\beta}_z^w}{\tilde{\beta}_z^m} = \frac{\cancel{\beta_z^w} / \sigma_w}{\cancel{\beta_z^m} / \sigma_m} = \frac{\sigma_m}{\sigma_w}$$

4. This provides leverage to test the underlying coefficients:

$$H_0\text{: } \beta_x^w = \beta_x^m$$

5. This works only if you can justify the assumption $\beta_z^w = \beta_z^m$
   - Different assumptions can lead to different results
   - There is no test of whether the assumption is correct

Comparing effects across groups

1. Tests of group equality of regression coefficients are an appealing way to compare groups
2. Even without problems in their application, regression coefficients are not in a natural metric
   - Are you interested in effects x on y* or in the log(p/(1-p))?
3. While odds ratios are a more natural metric, they are especially misleading in group comparisons.

Group comparisons of odds ratios

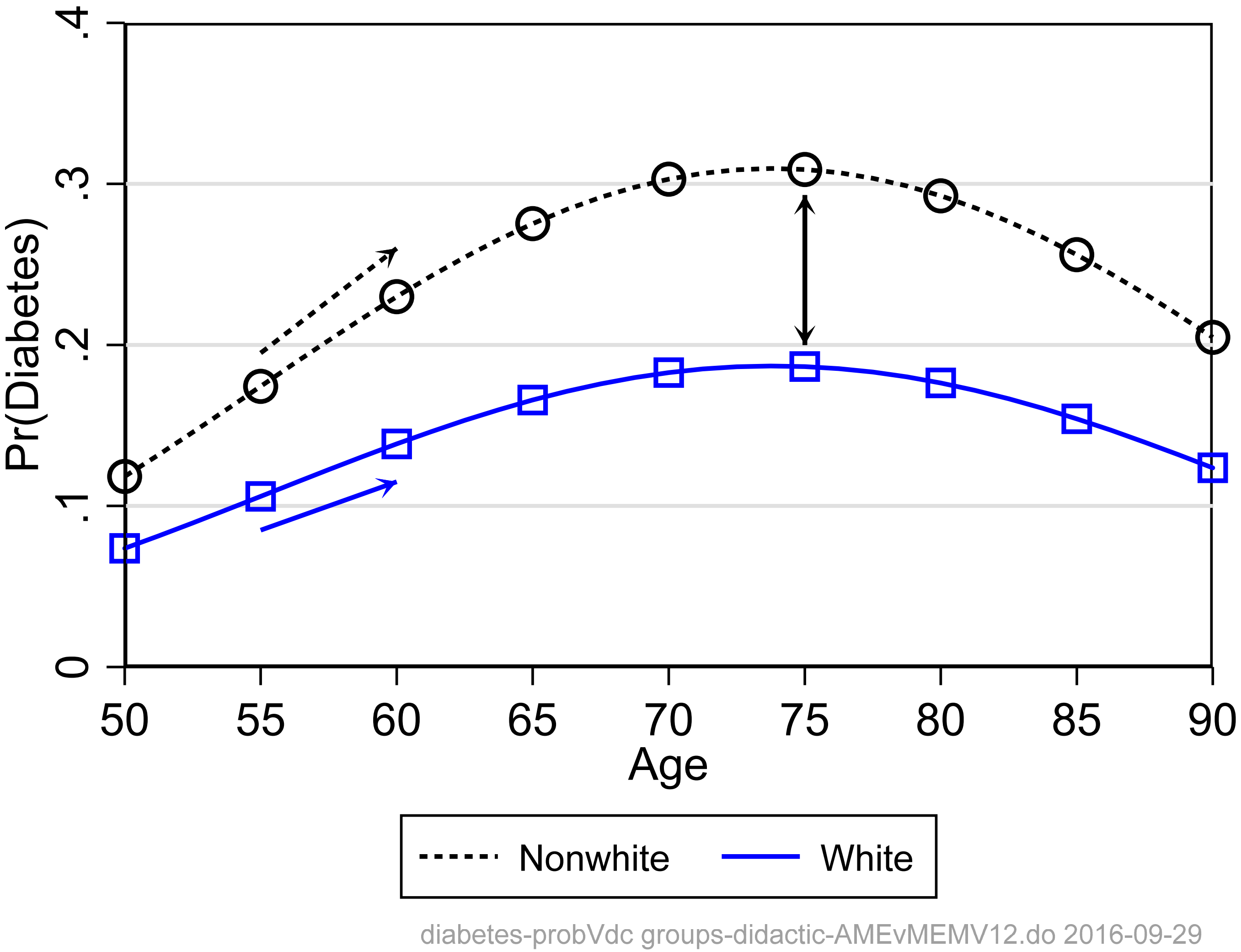
1. ORs can be similar across groupds while DCs can be very different

Quantity	Group 0	Group 1	
π(x1=1)	.026	.4	
π(x1=0)	.010	.2	
DC(x1)	.016	.2	< = Groups differ
Ω(x1=1)	.0267	.6667	
Ω(x1=0)	.0101	.2500	
OR(x1)	2.64	2.67	< = Groups are similar

Group comparisons of marginal effects

1. Comparing effects on the probability have many advantages
2. The substantive meaning is clear
3. The can deal with interactions and polynomials

Comparisons of probabilities & marginal effects



Roadmap

1. We consider probabilities and DCs for two groups using logit.
2. Methods generalize to
   - Any number of groups
   - Any model that makes predictions
   - With any measure of effect (e.g., MC, RRR)
   - Methods are useful for LRM when there are nonlinearities

Notation

$\mathbf{x}$	Vector with K regressors
$\mathbf{x}^*$	Specific values of the x's
g	Group (g=1 and g=0)
$\boldsymbol{\gamma}^g$	Vector of regression parameters for group g (<u>not</u> $\boldsymbol{\beta}^g$)
$\pi(\mathbf{x}, g)$	$Pr_g(y=1\|\mathbf{x})$ using compact notation

Fitting a model for multiple groups

1. If $Cov\left(\hat{\beta}_k^0, \hat{\beta}_k^1\right) = 0$, you can fit separate equations for each group

   $$\text{Group 0: } \pi(\mathbf{x}, g = 0) = F(\mathbf{x}'\boldsymbol{\gamma}^0)$$

   $$\text{Group 1: } \pi(\mathbf{x}, g = 1) = F(\mathbf{x}'\boldsymbol{\gamma}^1)$$

2. Or fit one equation with interactions:

   $$\pi\left(\mathbf{x}, g\right) = F\left(\left[g \times \mathbf{x}'\boldsymbol{\gamma}^1\right] + \left[(1-g) \times \mathbf{x}'\boldsymbol{\gamma}^0\right]\right)$$

   so that:

   $\pi\left(\mathbf{x}, g = 1\right) = F\left(\mathbf{x}'\boldsymbol{\gamma}^1\right)$ and $\pi\left(\mathbf{x}, g = 0\right) = F\left(\mathbf{x}'\boldsymbol{\gamma}^0\right)$

3. Joint estimations facilitates post-estimation computations and might be required with complex sampling.
4. A regressor can be eliminated for one group by constraining $\gamma_k^g = 0$.

Comparing predictions

1. A <u>group differences</u> is the DC with respect to group at given values of **x**:

$$\frac{\Delta\pi(\mathbf{x}=\mathbf{x}^*)}{\Delta g}=\pi(\mathbf{x}=\mathbf{x}^*,g{=}1)-\pi(\mathbf{x}=\mathbf{x}^*,g{=}0)$$

2. To test if the conditional probabilities are equal:

$$H_0\text{: }\frac{\Delta\pi(\mathbf{x}=\mathbf{x}^*)}{\Delta g}=0 \quad \text{or} \quad H_0\text{: }\pi(\mathbf{x}=\mathbf{x}^*,g=0)=\pi(\mathbf{x}=\mathbf{x}^*,g=1)$$

Ways to use group differences

1. Predictions at a single location
   - Do white men at 40 have the same probability as nonwhite men at 40?
2. Tables comparing groups at combinations of charcateristics
   - Racial differences in diabetes for men and women by levels of education.
3. Plots for continuous variables
   - Do nonwhites and whites differ in the probability of diabetes as they age?

Comparing marginal effects

Discrete changes

1. $DC_g(x_k)$ is the change in probability as x_k changes from *start* to *end* holding other variables at specific values:

$$\frac{\Delta\pi(\mathbf{x}=\mathbf{x}^*,g)}{\Delta x_k(start \rightarrow end)} = \pi(x_k = end, \mathbf{x}=\mathbf{x}^*, g) - \pi(x_k = start, \mathbf{x}=\mathbf{x}^*, g)$$

Comparing $DC(x_k)$ across groups

1. To test if effects are equal

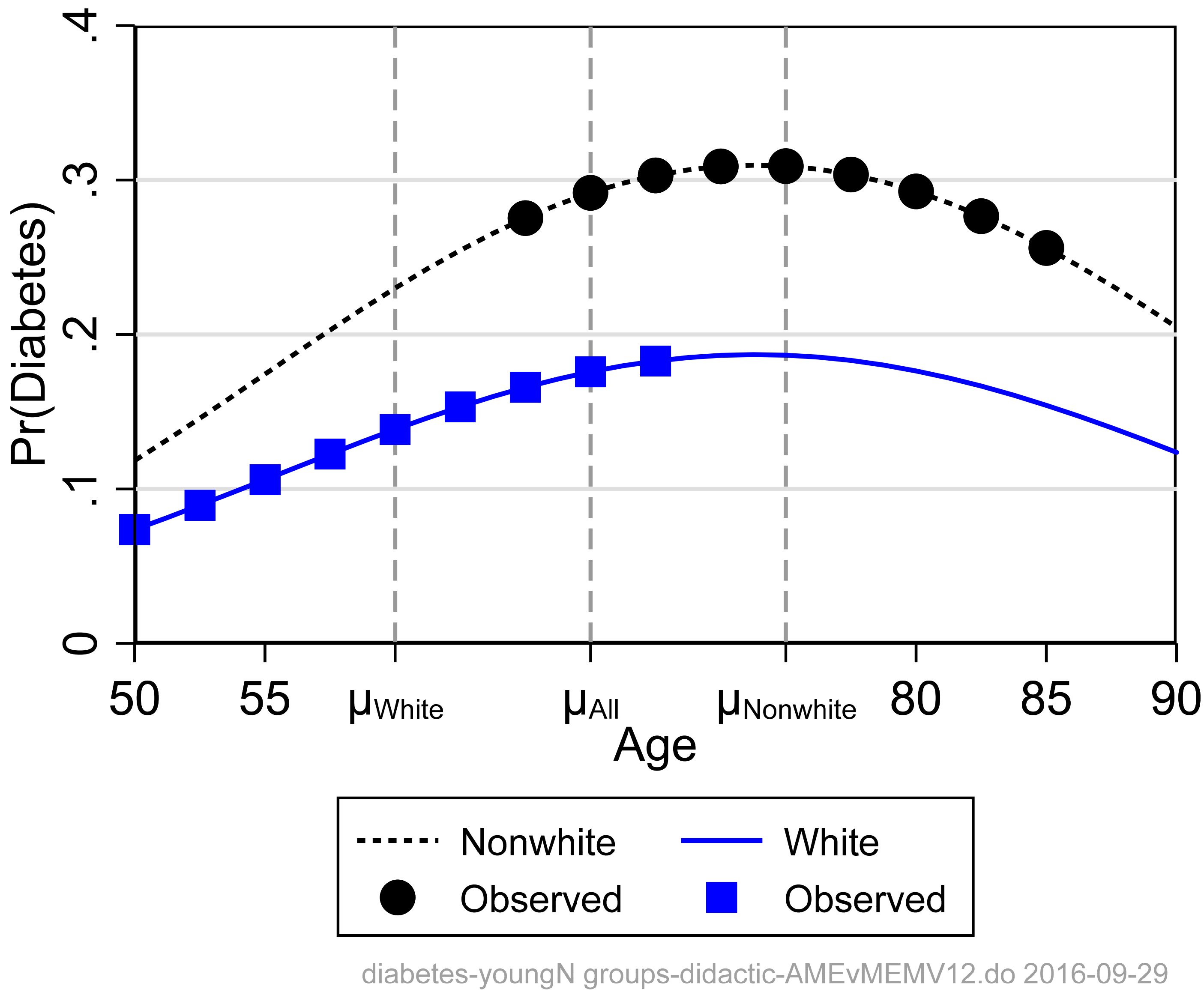
$$H_0: \frac{\Delta\pi(\mathbf{x}=\mathbf{x}^*,g=1)}{\Delta x_k(start \rightarrow end)} = \frac{\Delta\pi(\mathbf{x}=\mathbf{x}^*,g=0)}{\Delta x_k(start \rightarrow end)}$$

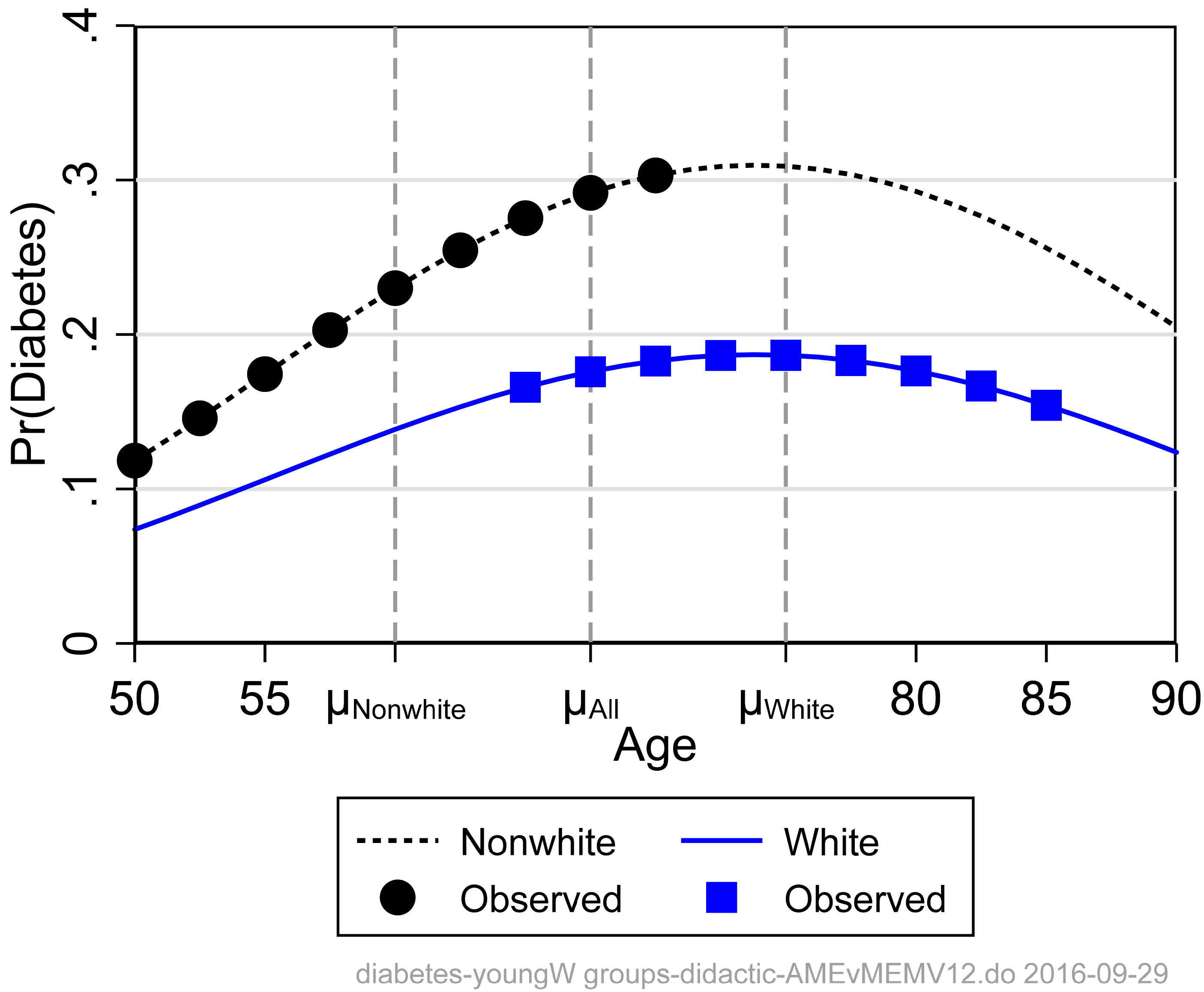
2. Equivalently, test the group difference in $DC(x_k)$

$$H_0: \frac{\Delta\pi(\mathbf{x}=\mathbf{x}^*,g=1)}{\Delta x_k(start \rightarrow end)} - \frac{\Delta\pi(\mathbf{x}=\mathbf{x}^*,g=0)}{\Delta x_k(start \rightarrow end)} = \frac{\Delta^2\pi(\mathbf{x}=\mathbf{x}^*)}{\Delta x_k(start \rightarrow end)\Delta g} = 0$$

Ways to summarize the discrete change

1. DC(x_k) depends on values of all x's where the change is estimated
   - According, how do you summarize the effect?
2. Most common summary measure
   - DCR(x_k) computes the effect at representative values such as DCM
   - ADC(x_k) averages DC($x_k | \mathbf{x}_i$) for observations in the group
3. DCRs and ADCs highlight different ways in which groups can differ.
   - DCRs reflect differences in the curve at the same values of the regressors
   - ADCs reflect differences in the curves and in the distribution of regressors.
4. The "best" measure is the one that answers your question.
   - Despite some literature suggesting DCR bad, ADC good!
5. Consider these graphs…


Pr(Diabetes)
0
.1
.2
.3
.4
50
55
μWhite
μAll
μNonwhite
80
85
90
Age
Nonwhite
White
Observed
Observed
diabetes-youngN groups-didactic-AMEvMEMV12.do 2016-09-29


Pr(Diabetes)
0
.1
.2
.3
.4
50
55
μNonwhite
μAll
μWhite
80
85
90
Age
Nonwhite
Observed
White
Observed
diabetes-youngW groups-didactic-AMEvMEMV12.do 2016-09-29

Discrete change at representative values (DCR)

1. $DCR_g(x_k \mid \mathbf{x}=\mathbf{x}^*)$

$$\frac{\Delta\pi(\mathbf{x}=\mathbf{x}^*, g)}{\Delta x_k(start \to end)} = \pi(x_k=end, \mathbf{x}=\mathbf{x}^*, g) - \pi(x_k=start, \mathbf{x}=\mathbf{x}^*, g)$$

2. To compare $DCR_g(x_k)$ across groups at the mean:

$$\frac{\Delta^2\pi(\mathbf{x}=\overline{\mathbf{x}})}{\Delta x_k(\overline{x}_k \to \overline{x}_k+s_k)\Delta g} = \frac{\Delta\pi(\mathbf{x}=\overline{\mathbf{x}}, g=1)}{\Delta x_k(\overline{x}_k \to \overline{x}_k+s_k)} - \frac{\Delta\pi(\mathbf{x}=\overline{\mathbf{x}}, g=0)}{\Delta x_k(\overline{x}_k \to \overline{x}_k+s_k)}$$

3. If x_k is binary,

$$\frac{\Delta^2\pi(\mathbf{x}=\mathbf{x}^*)}{\Delta x_k(0 \to 1)\Delta g} = \frac{\Delta\pi(\mathbf{x}=\mathbf{x}^*, g=1)}{\Delta x_k(0 \to 1)} - \frac{\Delta\pi(\mathbf{x}=\mathbf{x}^*, g=0)}{\Delta x_k(0 \to 1)}$$

4. Remember:
   - DCRs do not reflect group differences in the distribution of the regressors

Average discrete change (ADC)

1. $ADC_g(x_k)$ averages DCs at observed $\mathbf{x}_i$'s for each observation in a group.
2. For observation *i* in group *g*, the $DC_i(x_k)$

$$\frac{\Delta\pi(\mathbf{x}{=}\mathbf{x}_i, g)}{\Delta x_k(start_i \rightarrow end_i)} = \pi(x_k{=}end_i, \mathbf{x}{=}\mathbf{x}_i, g) - \pi(x_k{=}start_i, \mathbf{x}{=}\mathbf{x}_i, g)$$

3. For continuous x_k, compute ADC as x_k increases by δ from observed x_{ik}:

$$\frac{\Delta\pi(\mathbf{x}{=}\mathbf{x}_i, g)}{\Delta x_k(x_{ik} \rightarrow x_{ik}{+}\delta)} = \pi(x_k{=}x_{ik}{+}\delta, \mathbf{x}{=}\mathbf{x}_i, g) - \pi(x_k{=}x_{ik}, \mathbf{x}{=}\mathbf{x}_i, g)$$

4. Changes between fixed values are possible such age from 60 to 65 or a binary variable from 0 to 1.

$$\frac{\Delta\pi(\mathbf{x} = \mathbf{x}_i, g)}{\Delta x_k(start \rightarrow end)} = \pi(x_k{=}end, \mathbf{x}{=}\mathbf{x}_i, g) - \pi(x_k{=}start, \mathbf{x}{=}\mathbf{x}_i, g)$$

5. The $ADC(x_k|g)$ is the average of the DC for each observation in the group:

$$\text{ADC}^{g}_{x_k} = \frac{1}{N_g}\sum_{i\in g}\frac{\Delta\pi(\mathbf{x} = \mathbf{x}_i, g)}{\Delta x_k(start_i \rightarrow end_i)}$$

Example: Racial differences in health outcomes

The literatures suggests racial differences in diabetes and health that might decline with age. Obesity and physical activity might affect racial differences in diabetes.

Data

1. Health and Retirement Study (HRS) of older adults in the US. In Stata, run **`search groupsbrm`** to obtain the data.

2. Primary variables

```
goodhlth    Has good health?
diabetes    Has diabetes?
white       Is white?
age         Age
active      Is physically active?
obese       Is obese?
female      Is female?
highschool  Has high school degree?
married     Is married?
ihsincome   Inverse hyperbolic sine of income
```

Descriptive statistics

1. Groups differ on the levels of all variables.

Variable	White (N=12,427) Mean	White (N=12,427) Std Dev	Nonwhite (N=3,799) Mean	Nonwhite (N=3,799) Std Dev	Difference ΔMean	Difference p
goodhlth	0.769	—	0.565	—	0.205	<.001
diabetes	0.162	—	0.281	—	-0.120	<.001
age	66.514	10.421	64.099	9.677	2.415	<.001
active	0.303	—	0.223	—	0.080	<.001
obese	0.286	—	0.390	—	-0.104	<.001
female	0.532	—	0.575	—	-0.043	<.001
highschool	0.853	—	0.563	—	0.289	<.001
married	0.692	—	0.541	—	0.150	<.001
ihsincome	4.523	1.001	3.809	1.164	0.714	<.001

Note: ΔMean is group difference in the means.

Logit model of good health

The tests of $\beta_w = \beta_m$ are invalid as per Allison.

Logit model for good health (N =16,226).

	White			Nonwhite			H0: βw = βn	
Variable	1: βw	2: ORw	4: p	5: βn	6: ORn	8: p	9: F	10: p
Constant	-0.488	—	0.086	-1.541	—	<.001	4.71	0.034
female	0.144	1.155	0.006	-0.118	0.888	0.176	6.96	0.011
highschool	0.800	2.225	<.001	0.816	2.262	<.001	0.02	0.891
married	-0.056	0.945	0.341	-0.169	0.844	0.124	0.70	0.406
ihsincome	0.556	1.744	<.001	0.583	1.792	<.001	0.16	0.695
age	-0.018	0.982	<.001	-0.008	0.992	0.072	3.47	0.068
obese	-0.573	0.564	<.001	-0.361	0.697	0.002	3.10	0.084

Note: OR is the odds ratio; tests of H0: βw = βn are for didactic purposes.

Logit models of diabetes

The tests of $\beta_w = \beta_m$ are invalid as per Allison.

Logit model for diabetes (N =16,226).

	White			Nonwhite			H0: βw = βn	
Variable	1: βw	2: ORw	4: p	5: βn	6: ORn	8: p	9: F	10: p
Constant	-9.627	—	<.001	-11.169	—	<.001	0.40	0.529
female	-0.400	0.670	<.001	-0.079	0.924	0.427	7.27	0.009
highschool	-0.253	0.776	0.001	-0.142	0.867	0.160	0.92	0.342
married	0.070	1.073	0.333	0.064	1.066	0.548	0.00	0.960
ihsincome	-0.189	0.828	<.001	-0.131	0.877	<.001	1.61	0.210
age	0.243	1.274	<.001	0.299	1.348	<.001	0.65	0.422
agesq	-0.002	0.998	<.001	-0.002	0.998	<.001	0.78	0.382
active	-4.048	0.017	0.302	-2.557	0.078	0.678	0.04	0.849
activeXage	0.115	1.122	0.331	0.052	1.053	0.783	0.07	0.791
activeXagesq	-.0009	0.999	0.310	-.0003	1.000	0.835	0.11	0.737
obese	1.163	3.199	<.001	0.740	2.095	<.001	9.35	0.003

Note: OR is the odds ratio; tests of H0: βw = βn are for didactic purposes.

Comparing marginal effects

Comparing $ADC_g(x_k)$ for diabetes

Panel A: Average discrete change for logit model for diabetes (N=16,226).

Variable	White 1: ADC	2: p	Nonwhite 3: ADC	4: p	Difference 5: ADC	6: p
female	-0.051	<0.001	-0.015	0.431	-0.036	0.089
highschool	-0.033	0.001	-0.027	0.160	-0.006	0.763
married	0.009	0.330	0.012	0.546	-0.003	0.875
ihsincome	-0.022	<0.001	-0.024	<0.001	0.002	0.779
age	0.011	<0.001	0.027	<0.001	-0.016	0.002
active	-0.053	<0.001	-0.084	<0.001	0.031	0.110
obese	0.169	<0.001	0.146	<0.001	0.022	0.408

Note: The effect of age is for a five-year change.

On average being female significantly decreases the probability of diabetes by .051 (p<.001) for white respondents, with a decrease to .015 (p<.001) for nonwhites. The effects of gender differ by .036, which is significant at the .10 level but not the .05 level.

Comparing $DCM_g(x_k)$ for diabetes

Panel B: DCM for logit model for diabetes (N=16,226).

Variable	White 1: DCM	2: p	Nonwhite 3: ADC	4: p	Difference 5: ADC	6: p
female	-0.057	<0.001	-0.016	0.431	-0.041	0.070
highschool	-0.038	0.001	-0.029	0.162	-0.008	0.716
married	0.010	0.328	0.013	0.545	-0.003	0.895
ihsincome	-0.025	<0.001	-0.026	<0.001	0.001	0.916
age	0.014	<0.001	0.023	<0.001	-0.009	0.169
active	-0.049	<0.001	-0.083	0.006	0.034	0.319
obese	0.190	<0.001	0.158	<0.001	0.032	0.264

Note: The effect of age is for a five-year change.

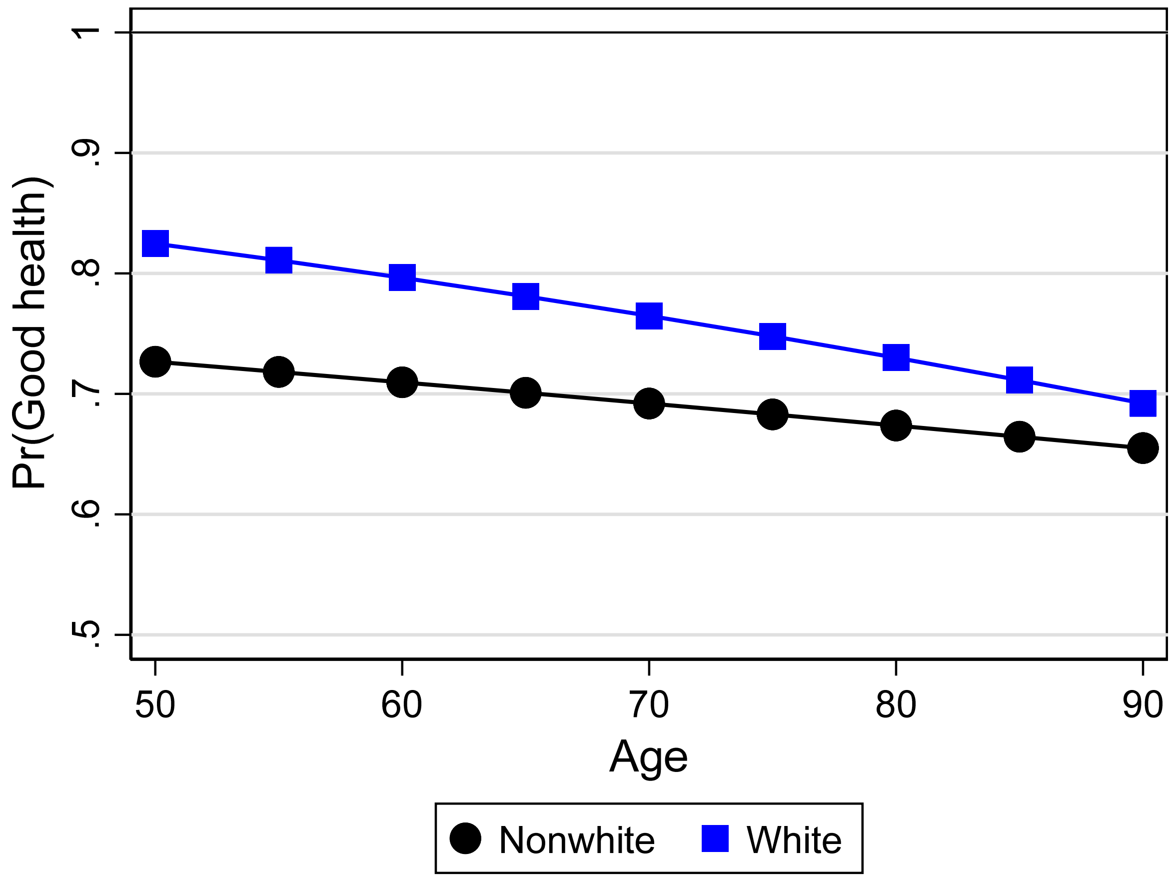
1. ADC and DCM do not always lead to the same conclusions as illustrated by the effect of a five-year increase in age.

   While the ADC of age is significantly larger for nonwhites than whites (p=.002), the effect of age at the mean not differ for whites and nonwhites (p=.169).

2. The different conclusions reflects the different age distributions for whites and nonwhites.

Plots: Race, age and good health

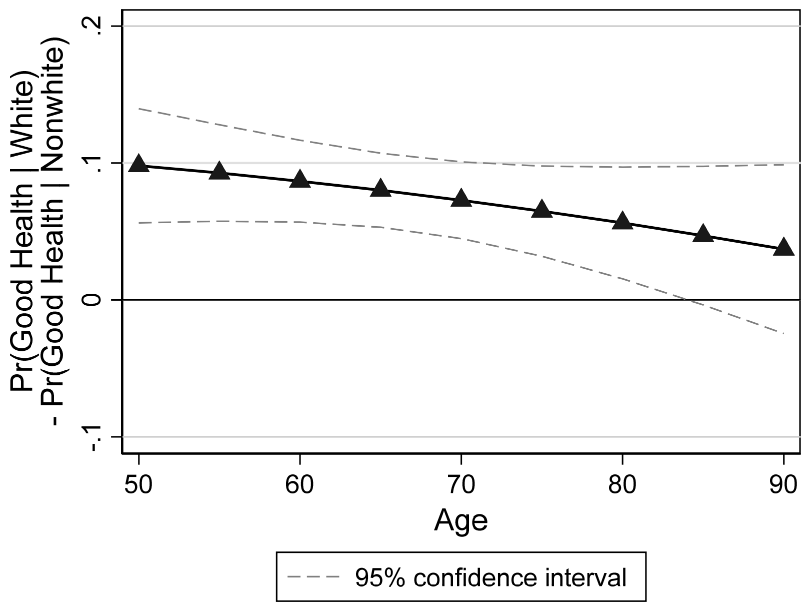
1. Do racial disparities in health change with age?

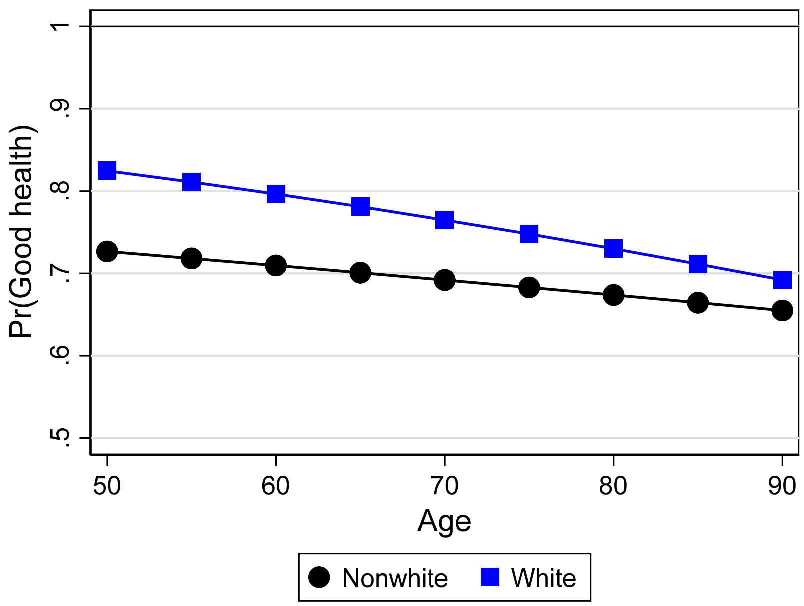


Whites have a higher probability of good health that decreases from .10 at age 50 to less than .04 at 90.

Are race differences significant?

1. We can plot the race difference conditional on age.
   - When the CI falls below 0, the difference is nonsignificant/
2. Racial differences by age show that differences are significant till age 85.





Whites have a higher probability of good health that decreases from .10 at age 50 to less than .04 at 90. Differences are significant at the .05 level until age 85.

Differences in probabilities and differences in effects

1. Since curves are nearly linear, the effects of age can be summarized by changes in the probability of good health as age increase from 50 to 90.

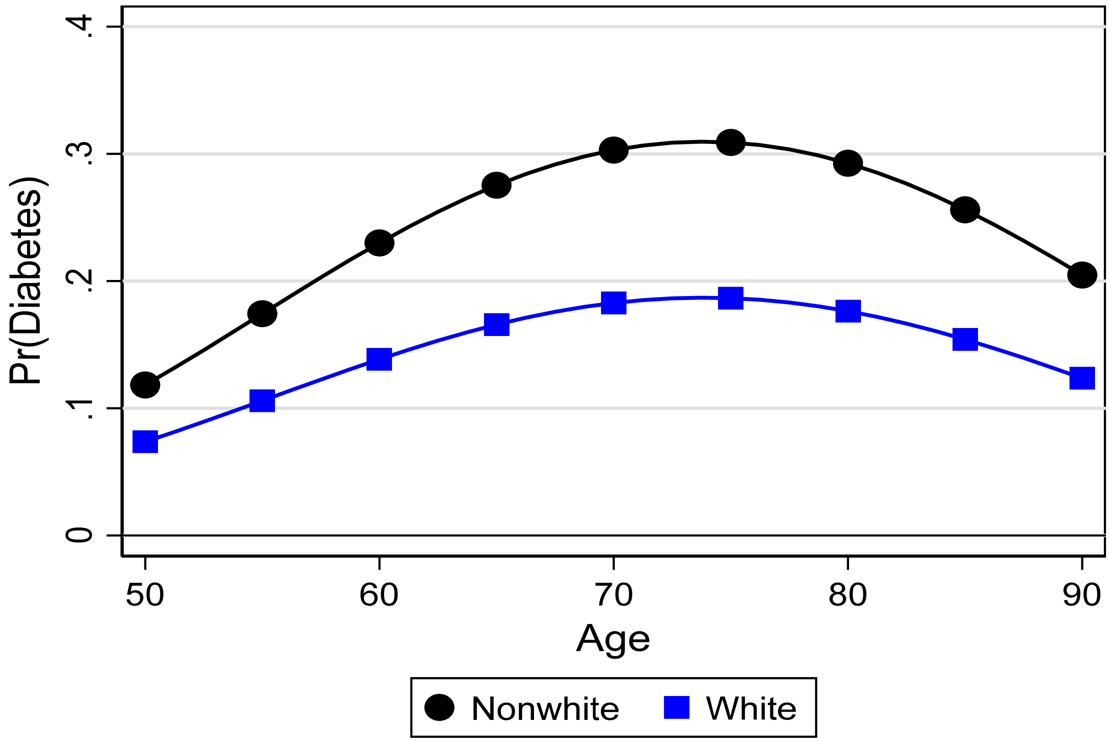
```
Change over range of age
                          |   lincom    pvalue        ll        ul
--------------------------+----------------------------------------
        #1 DCage if white |   -0.133     0.000    -0.176    -0.089
     #2 DCage if nonwhite |   -0.072     0.077    -0.151     0.008
 #3 DCage race difference |   -0.061     0.174    -0.150     0.028
```

While the effect of age on good health is larger for whites than non-whites, the difference is not significant (p=.17).

2. Levels of health differ significantly, but the effects of age do not.
3. We compute these effects by computing differences and second differences in the predicted probabilities.
4. The figures for good health show that we do not need the figures!

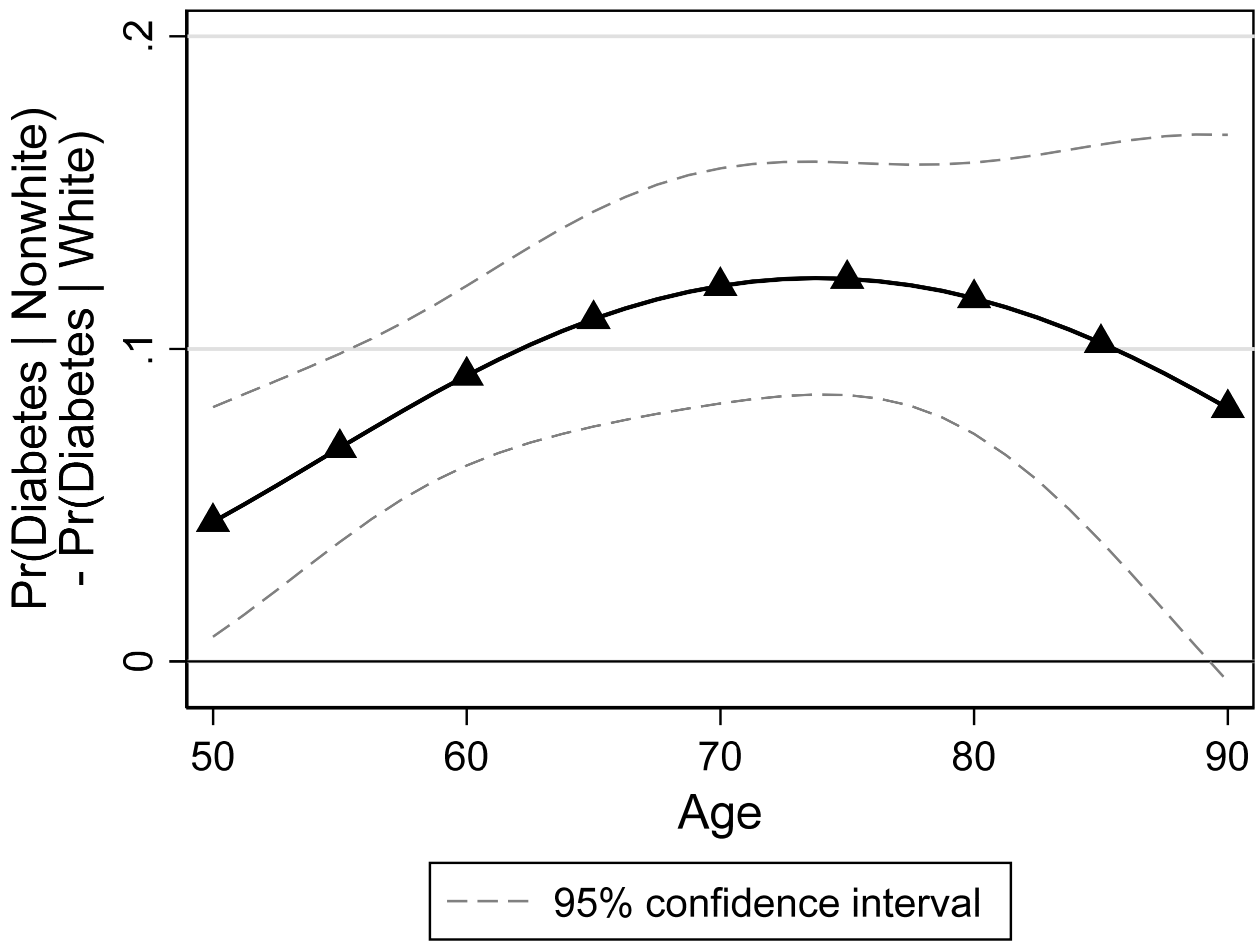
Plots: Race, age and diabetes

1. The effect of age on diabetes is more complex.



For both groups the probability of diabetes increases from 50 to 75 before declining. While whites have a smaller probability of diabetes at all ages, the difference is smallest at 50 where it is about .04, then increases to a maximum of .12 at 75 before decreasing to .08 at 90.

2. To determine significance, we plot racial differences with CIs and determine that the differences are significant until about 88.

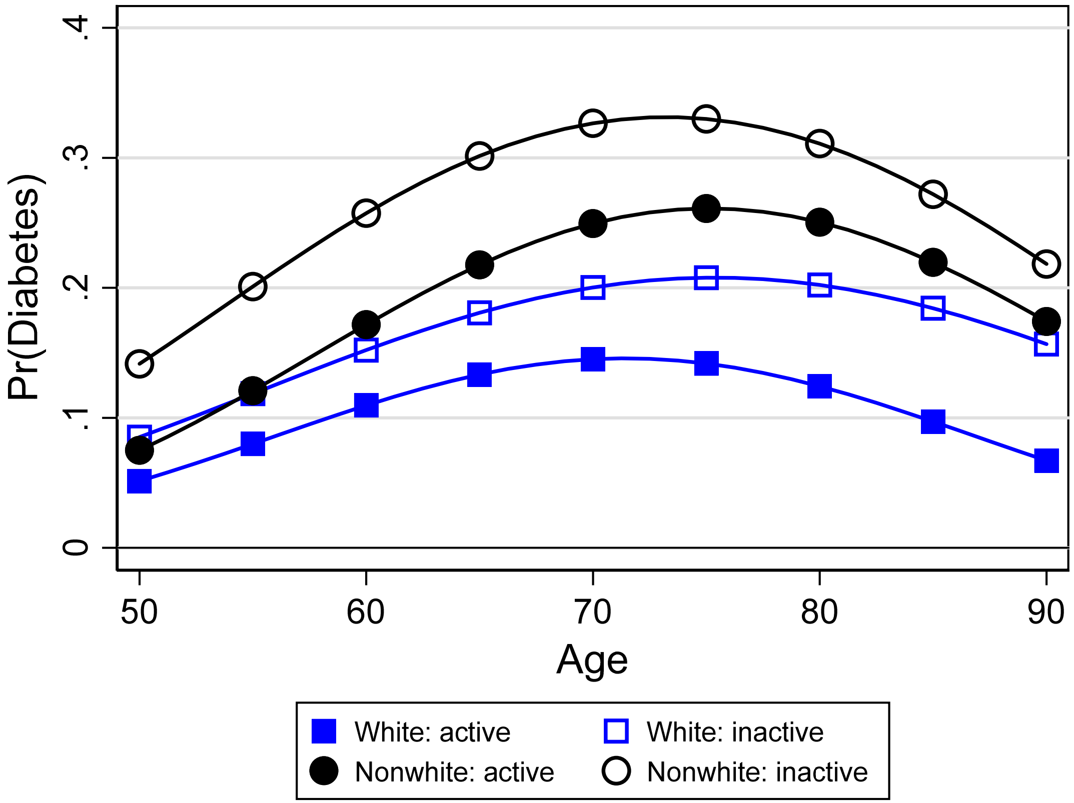


Plots: Race, age and diabetes by activity

1. Are the benefits of activity different for nonwhites and whites?
2. Our model includes age and activity in complex ways:

   `c.age c.age#c.age i.active i.active#c.age i.active#c.age#c.ages`

3. Graphs of probabilities are too complex to understand easily



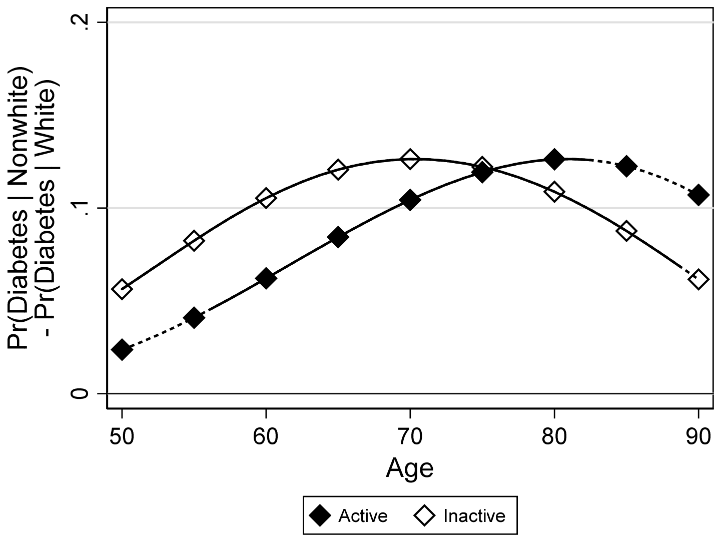
Racial differences with age and activity

1. To simplify, I compute comparisons that emphasize different aspects of our research question.
   - DC_g(race|activity, **x**)
   - DC_g(activity|race, **x**)
2. DC_g(race|activity, **x**): racial differences given activity over age
   - Comparing solid symbols show race differences if active
   - Comparing hollow symbols show race differences if inactive
3. DC_g(activity|race, **x**): activity differences by race over age
   - Comparing solid and open squares for activity differences if white
   - Comparing solid and open squares for activity differences if nonwhite

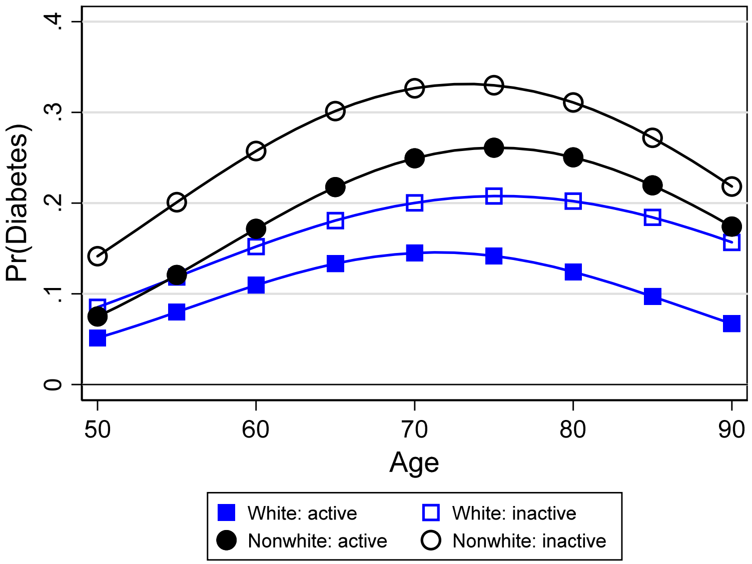
Avoiding CIs

4. With multiple lines, CIs can get confusing
5. I use a dashed DC line to indicate a difference is NOT significant

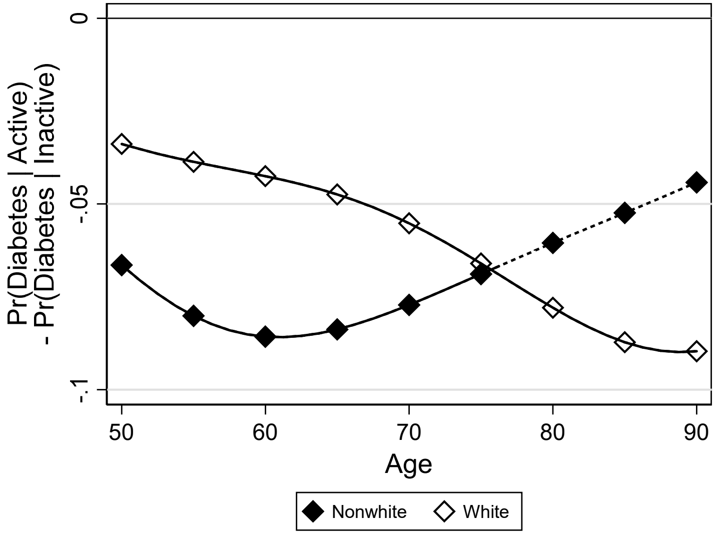
Overview



Pr(Diabetes | Nonwhite)
- Pr(Diabetes | White)
0
.1
.2
50
60
70
80
90
Age
Active
Inactive

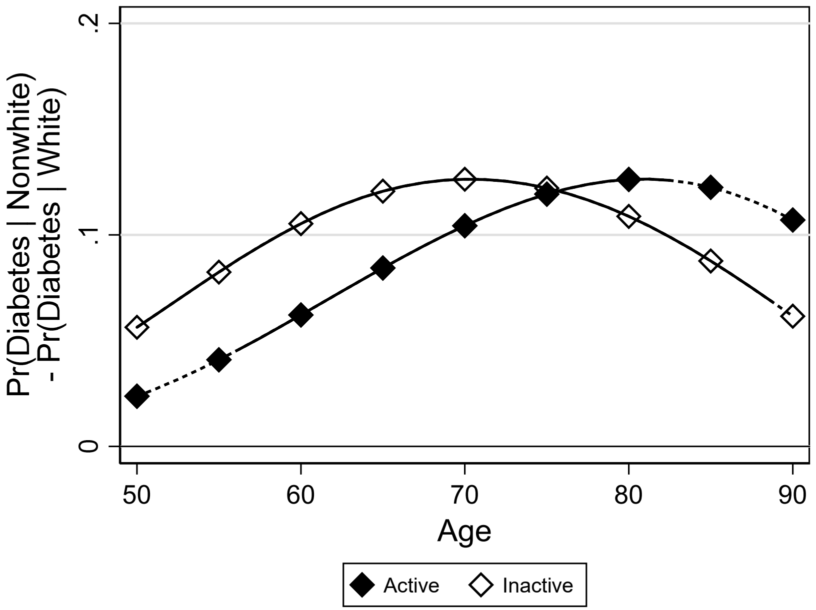


Pr(Diabetes)
0
.1
.2
.3
.4
50
60
70
80
90
Age
White: active
White: inactive
Nonwhite: active
Nonwhite: inactive



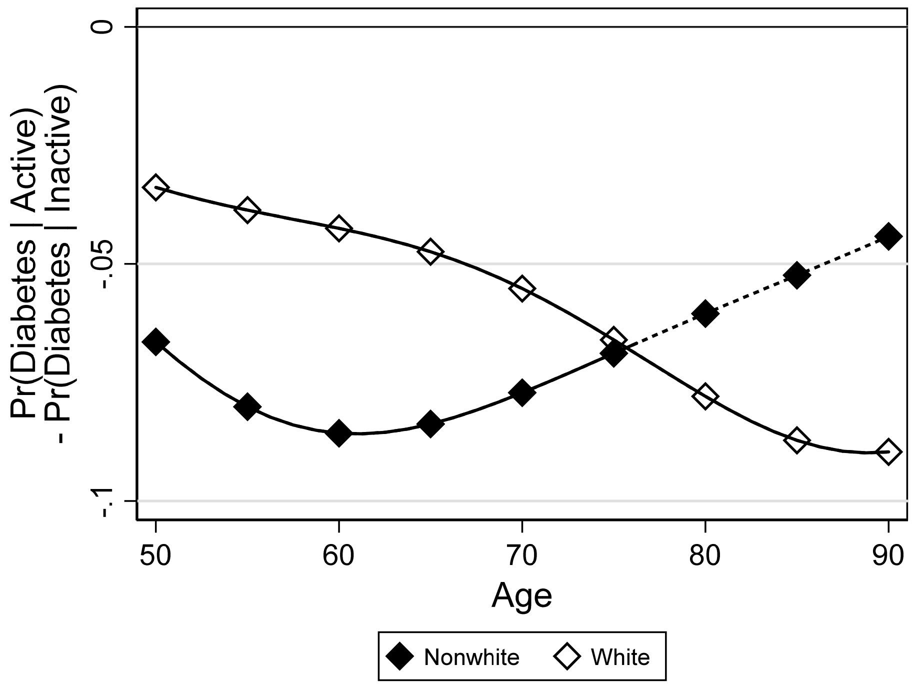
Pr(Diabetes | Active)
- Pr(Diabetes | Inactive)
0
-.05
-.1
50
60
70
80
90
Age
Nonwhite
White

Race differences by level of activity and age



While the benefits of being white occur for those who are active and those who are not, the strength and timing of the benefits differ by level of activity. For those who are not active, the advantages for whites increase from age 50 to 70 before decreasing thereafter. Differences are significant at all ages except 90. For those who are active, the same pattern occurs, but the effects are weaker at younger ages than they are for those who are inactive. The differences increase from age 50 to 80, becoming statistically significant at age 57. At age 80 the differences begin to decrease and are no longer significant.

Activity differences by race and age



The benefits of activity occur differently for whites and nonwhites. For whites the protective effect of activity is smaller (i.e., less negative) at younger ages and increases in magnitude until age 90. For nonwhites the effect gets stronger from age 50 to 60 before decreasing till age 90; after age 76 the effects are not significant. Tests of racial differences in the effect of activity are significant at the .10 level between ages 55 and 61, reach significance at the .05 level at age 58 where the difference reaches its maximum of .044, and are not significant at other ages.

Tables: diabetes, gender, and obesity by race

1. Do racial differences vary by gender and obesity?
2. This is the source data we need to unpack

	Women		Men		Effect of obesity		
	1: Obese	2: Not	3: Obese	4: Not	5: Women	6: Men	7: Diff
1: White	0.278	0.107	0.365	0.152	0.170*	0.212*	-0.042*
2: Nonwhite	0.389	0.233	0.408	0.248	0.156*	0.161*	-0.005
3: Difference	-0.112*	-0.126*	-0.044	-0.096*	-0.014	-0.052†	-0.038*

Note: Other variables held at their means.* = p<.01; † = p<.10 for two-tailed test.

3. This complex table is explained on the next page

Unpacking a complicated table

G-gender O-obesity R-race	Girls Women 1:Obese	2:Thin	Boys Men 3:Obese	4:Thin	Effect of obesity 5:Girls	6:Boys	7:Diff
1: Wht	π(wgo)	π(wgt)	π(wbo)	π(wbt)	Δπ/ΔO|wg	Δπ/ΔO|wb	Δπ/ΔOG|w
2: Non	π(ngo)	π(ngt)	π(wbo)	π(nbt)	Δπ/ΔO|ng	Δπ/ΔO|nb	Δπ/ΔOG|n
3: Dif	Δπ/ΔR|go	Δπ/ΔR|gt	Δπ/ΔR|bo	Δπ/ΔR|bo	Δπ/ΔOR|g	Δπ/ΔOR|b	Δπ/ΔOGR

Key:

R=race	w=white	n=nonwhite
G=gender	b=boy	g=girl
O=obesity	o=obese	t=thin
π(RGO):	Prob give R, G and O	
Δπ/ΔR|GO:	Race difference given G and O	
Δπ/ΔO|RG:	DC(obese) given R and G	
Δπ/ΔOR|G	Race difference in DC(obese) given G	
Δπ/ΔOG|R	Gender difference in DC(obese) given R	
Δπ/ΔOGR	Race diff in gender diff in obesity difference	

4. To simplify $\Delta\pi(female = p, obese = q, \mathbf{x}=\bar{\mathbf{x}})$ is written as $\Delta\pi(female = p, obese = q)$.

Race differences in diabetes given gender and obesity

1. The last row show estimates of:

$$\frac{\Delta\pi(female = p, obese = q)}{\Delta white}$$

2. We find:

The probabilities of whites being diagnosed with diabetes are smaller than those for nonwhites for all combinations of obesity and gender. The largest racial differences, shown in row 3, is -.126 for women who are not obese and the smallest difference is a nonsignificant -.044 for obese men.

3. To test if racial differences are equal for men and women given obesity, we can compute a second difference (not shown in table):

$$\frac{\Delta\pi(obese = q)}{\Delta female\, \Delta white} = \frac{\Delta\pi(female = 1, obese = q)}{\Delta white} - \frac{\Delta\pi(female = 0, obese = q)}{\Delta white}$$

The effect of race for obese men and women differ by -.068 = (-.112 - -.044) which is significant at the .02 level. For those who are not obese, the gender difference is smaller and not significant (p=.13).

Effects of obesity on diabetes

1. To formalize these findings, we estimate DC(obese|gender,race):

$$\frac{\Delta\pi(female=p, white=r)}{\Delta obese}$$

Obesity significantly increases the probability of diabetes by about .16 for all groups except white men where the effect is .21.

2. To test racial differences DC(obese) the last row, columns 5-6:

$$\frac{\Delta\pi(female=p)}{\Delta obese\,\Delta white} = \frac{\Delta\pi(female=p, white=1)}{\Delta obese} - \frac{\Delta\pi(female=p, white=0)}{\Delta obese}$$

Racial differences in the effects of obesity are small and not significant for women, but larger and marginally significant for men (p=.09).

3. To gender dofferences DC(obese) using rows 1 and 2 of column 7:

$$\frac{\Delta\pi(white=r)}{\Delta obese\,\Delta female} = \frac{\Delta\pi(female=1, white=r)}{\Delta obese} - \frac{\Delta\pi(female=0, white=r)}{\Delta obese}$$

The effect of obesity for whites is .04 larger (p<.001) for men than women, but the gender difference is small and nonsignificant for nonwhite respondents.

Alternatives when computing effects

1. Instead of global means, effects could be computed at local means
2. To reflect group differences in the distribution of regressors, compute ADC(obesity) by race and gender.
3. The most effective approach for comparing effects depends on the specific questions motivating the research.

Comparing regression coefficients for diabetes

1. Suppose you still want to compare the logit coefficients.
2. Consider alternative tests of the hypothesis that the regression coefficients for gender and obesity are equal for whites and nonwhites.
3. The of coefficients for diabetes has from standard Wald tests of $H_0\colon \gamma_k^{\mathrm{W}} = \gamma_k^{\mathrm{N}}$.

   If these tests were appropriate, which they are not due to the scalar identification of the coefficients, we would conclude that the protective effects of being female are significantly larger for whites than nonwhites (p<.01) and that the health costs of obesity are significantly greater for whites than nonwhites (p<.01).

4. This contradicts our conclusions using discrete changes.

   The ADC for obesity is .022 larger for whites than nonwhites, but the difference is not significant (p=.41); similarly the DCM is .032 larger for whites (p=.26). The ADC and DCM for being female are about .04 less negative for nonwhites than whites, but the differences are not significant (p=.09; p=.07).

Applying Allison's test

Testing the equality of regression coefficients in models for diabetes

Coefficients constrained to be equal

	Full model		M1: ihsincome		M2: age, active†		M3: obese	
Variable	Δβ	p	Δβ	p	Δβ	p	Δβ	p
female	-0.321	0.009	-0.199	0.164	-0.363	0.013	-0.176	0.120
obese	0.423	0.003	0.069	0.810	0.530	0.064	0.000‡	1.000‡

Note: Results for the full model are from table 3. Tests from models with coefficients constrained to be equal estimated with a location scale model.
† All coefficients involving age and/or active are constrained to be equal.
‡ Coefficient for obese are constrained to be equal.

Three equality assumptions we can impose

- M1: coefficients for *ihsincome* are constrained to be equal;
- M2: coefficients that include *age* or *active* are constrained
- M3: coefficients for *obese* are constrained.
- Models with other constrains did not converge.

Racial differences in regression coefficients for *female*

- M1: $p=.164$
- M2: $p=.013$
- M3: $p=.120$

Racial differences in regression coefficients for *obese*

- M1: $p=.810$
- M2: $p=.064$

The models are empirically indistinguishable

- The predictions from M1 and M3 are identical
- The predictions for M3 have correlations of about .999

Code for comparing groups

Joint fitting models using factor syntax

```
logit goodhlth ibn.white ///
                    ibn.white#(i.female … i.obese), nocon
```

- `ibn` means indicator variable where you want dummies for both values
- `ibn.white` : estimates intercepts for both groups
- `ibn.white#(i.female … i.obese)` : intercations with regressors
- `nocon` : do not include intercept

Example of output:

```
              |
     goodhlth |      Coef.
--------------+------------------------------------
        white |
    Nonwhite  |   -1.541056   : Nonwhite intercept
       White  |   -.4880763   : White intercept
              |
 white#female |
     Nonwhite#|
      Female  |   -.1183793   : Nonwhite β for female
```

```
White#Female  |   .1437669  : White β for female
::
```

- **`white`** is the variable with values **`Nonwhite`** and **`White`**
- **`white#female`** is the interaction of **`white`** with **`female`**
- **`Nonwhite#Female`** interacts female with being nonwhite and is 0 for those who are white
- **`White#Female`** interacts female with being white and is 0 for those who are nonwhites

Computing marginal effects

Comparing ADC for indicator variables

1. **over(white)** averages over the groups defined by **white**

```
. est restore fullmodel
. mtable, dydx(female) over(white) post brief dec(4)

Expression: Pr(diabetes), predict()

           |  d Pr(y)
 ----------+---------
  Nonwhite |  -0.0150
     White |  -0.0506
```

2. **mlincom** creates a table with 1st and 2nd differences

```
. qui mlincom (2), rowname(ADC female: White) add clear
. qui mlincom (1), rowname(ADC female: Nonwhite) add
. mlincom (2-1), rowname(ADC female: Difference) twidth(25) add

                          |   lincom    pvalue        ll        ul
--------------------------+----------------------------------------
ADC female                |
                    White |   -0.051     0.000    -0.066    -0.036
                 Nonwhite |   -0.015     0.431    -0.053     0.023
               Difference |   -0.036     0.089    -0.077     0.006
```

Comparing ADC from continuous variables

1. **`at(age=gen(age))`** computes predictions at observed age
2. **`at(age=gen(age+5))`** computes predications at 5 plus observed age

```
. est restore fullmodel
. mtable, at(age=gen(age)) at(age=gen(age+5)) over(white) post

                |    Pr(y)
 ---------------+----------
    0.white#c.1 |   0.2811
    1.white#c.1 |   0.1615
    0.white#c.2 |   0.3080
    1.white#c.2 |   0.1729
```

3. Compute the effects

```
. qui mlincom (4-2), rowname(ADC age+5: White) clear
. qui mlincom (3-1), rowname(ADC age+5: Nonwhite) add
. mlincom (4-2) - (3-1), ///
>         rowname(ADC age+5: Difference) twidth(25) add

                         |   lincom    pvalue        ll        ul
-------------------------+----------------------------------------
ADC age+5                |
                   White |    0.011     0.000     0.008     0.015
                Nonwhite |    0.027     0.000     0.018     0.036
              Difference |   -0.016     0.002    -0.025    -0.006
```

Compare DCM for female

```
. est restore fullmodel
. mtable, dydx(female) post at((omeans) _all white=(0 1))

Expression: Pr(diabetes), predict()

           |    white   d Pr(y)
 ----------+-------------------
         1 |        0   -0.0160
         2 |        1   -0.0573

Specified values of covariates

           |        1.             1.          1.                          1.
           |    female       highsc~l     married   ihsinc~e       age     active
 ----------+----------------------------------------------------------------------
   Current |     .5392           .805       .6667      4.405     66.11        .29
::
. qui mlincom (2),   rowname(DCM female: White)       add clear
. qui mlincom (1),   rowname(DCM female: Nonwhite)    add
. mlincom (2-1), rowname(DCM female: Difference)  add
```

Compare DCM for age

1. Get start and end values

```
. svy: mean age // compute start and end for age

             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
         age |   66.11455   .1723443       65.7693     66.4598

. global ageST = _b[age] // mean
. global ageEN = _b[age] + 5 // mean + 5
```

2. Predictions where `(omeans) _all` is the overall means

```
. mtable, at((omeans) _all age=($ageST $ageEN) white=(0 1)) post dec(4)

Expression: Pr(diabetes), predict()

           |    white        age      Pr(y)
 ----------+-----------------------------
         1 |        0      66.11     0.2832
         2 |        0      71.11     0.3063
         3 |        1      66.11     0.1706
         4 |        1      71.11     0.1849

Specified values of covariates
::
```

3. First and second differences:

```
. qui mlincom (4-3),     rowname(DCM age+5: White)      add
. qui mlincom (2-1),     rowname(DCM age+5: Nonwhite)   add
. mlincom (4-3) - (2-1), rowname(DCM age+5: Difference) add

             |   lincom    pvalue        ll        ul
-------------+----------------------------------------
DCM age+5    |
       White |    0.014     0.000     0.009     0.019
    Nonwhite |    0.023     0.000     0.011     0.035
  Difference |   -0.009     0.169    -0.021     0.004
```

Plots

Plotting predictions

```
mgen, at(white=1 age=(50(5)90)) atmeans stub(AprW) ///
    predlabel(White) replace  // probabilities for whites

mgen, at(white=0 age=(50(5)90)) atmeans stub(AprN) ///
    predlabel(Nonwhite) replace   // probabilities for nonwhites

// plot probabilities
twoway ///
    (connected  AprWpr1  AprWage,  $LINwht)  ///
    (connected  AprNpr1  AprWage,  $LINnon)  ///
    (scatter    AprWpr1  AprWage,  $LINwht)  ///
    (scatter    AprNpr1  AprWage,  $LINnon), ///
    ytitle(Pr(Good health)) xtitle(Age) ///
    xlab(50(10)90) ylab(.5(.1)1, gmin gmax grid) ///
    yline(1, lcol(black) lwid(*.6)) legend(rows(1) order(4 3)) ///
    scale(1.2)
```

Plot group differences

```
mgen, dydx(white) at(age=(50(5)90)) atmeans stub(Adc) ///
    predlabel(DC) replace // DCrace
label var Adcll1 "95% confidence interval"
label var Adcul1 "95% confidence interval"


* to reverse directory of DC
gen RAdcd_pr1  = -1 * Adcd_pr1
gen RAdcll1    = -1 * Adcll1
gen RAdcul1    = -1 * Adcul1
label var Adcll1 "95% confidence interval"

twoway  ///
    (connected  Adcd_pr1      Adcage,  $LINdc)  /// line
    (connected  Adcll1 Adcul1 Adcage,  $LINci)  /// ci
    (scatter    Adcd_pr1      Adcage,  $LINdc), /// line
    ytitle("Pr(Good Health | White)" "- Pr(Good Health | Nonwhite)") ///
        xtitle(Age) xlab(50(10)90) ylab(-.1(.1).2, grid) ///
        yline(-.1 .2, lcol(black*.2) lwid(*.7)) ///
        legend(pos(6) ring(1) cols(1) order(3) symxsize(7)) ///
        scale(1.2) ///
        yline(0, lcol(black) lwid(*.6))
```

Computing change over range of age

```
. mtable, at(age=(50 90) white=(0 1)) atmeans post

Expression: Pr(goodhlth), predict()

          |    white        age      Pr(y)
----------+-----------------------------
        1 |        0         50      0.727
        2 |        0         90      0.655
        3 |        1         50      0.825
        4 |        1         90      0.692
::
. qui mlincom (4-3),   rowname(#1 DCage if white) clear
. qui mlincom (2-1),   rowname(#2 DCage if nonwhite) add
. mlincom (4-3)-(2-1), rowname(#3 DCage race difference) add
```

Tables

1. To examine if racial differences vary by gender and obesity:

	Women		Men		Effect of obesity		
	1: Obese	2: Not	3: Obese	4: Not	5: Women	6: Men	7: Diff
1: White	0.278	0.107	0.365	0.152	0.170*	0.212*	-0.042*
2: Nonwhite	0.389	0.233	0.408	0.248	0.156*	0.161*	-0.005
3: Difference	-0.112*	-0.126*	-0.044	-0.096*	-0.014	-0.052†	-0.038*

Note: Other variables held at their means.*=p<.01; †=p<.10 for two-tailed test.

- Probabilities are in rows 1-2, columns 1-4
- Row 3 is racial differences in probabilities given obesity and gender
- Columns 5-6 are DCs of obesity given gender
- Column 7 compare DCs of obesity for men and women

2. To simplify $\Delta\pi(female = p, obese = q, \mathbf{x}=\overline{\mathbf{x}})$ is written as $\Delta\pi(female = p, obese = q)$.

Code: Pr(diabetes | female, obese, race)

```
. //  #3.1 Pr(diabetes | female, obese, race)
. //  Table 4 rows 1-2, columns 1-4

. qui est restore fullmodel

. //  Pr(diabetes | female, obese, race)

. mtable, at(white=(0 1) female=(0 1) obese=(0 1)) atmeans stats(est
pvalue)

Expression: Pr(diabetes), predict()

           |     white    female     obese      Pr(y)         p
 ----------+----------------------------------------------------
         1 |         0         0         0      0.248     0.000
::
         8 |         1         1         1      0.278     0.000

Specified values of covariates

           |           1.         1.                             1.
           |     highsc~l    married   ihsinc~e       age    active
 ----------+--------------------------------------------------------
   Current |         .805       .667       4.41      66.1       .29
```

Code: DC(race | female, obese)

```
. //  #3.2 Race dif by gender and obesity: DC(race | female, obese)
. //  Table 4 row 3, columns 1-4

. mtable, dydx(white) at(female=(0 1) obese=(0 1)) ///
>     atmeans stats(est pvalue) post

Expression: Pr(diabetes), predict()

          |   female     obese   d Pr(y)         p
 ---------+----------------------------------------
        1 |        0         0    -0.096     0.000
::
        4 |        1         1    -0.112     0.000

Specified values of covariates
::
```

Code: DC(race | female, obese) by gender or obesity

```
. //  #3.2a Compare DC(race | female, obese) across gender and obesity
. //  Table 4 row 3, columns 5-6 are from rows 3 and 4 below

. qui mlincom 4-2, rowname(obese    men - obese    women)
. qui mlincom 3-1, rowname(nonobese men - nonobese women) add
. qui mlincom 4-3, rowname(women  obese - women nonobese) add
. qui mlincom 2-1, rowname(men    obese - men   nonobese) add
. mlincom, twidth(32) title(Second differences in DCobese and DCfemale)

Second differences in DCobese and DCfemale

                                 |   lincom    pvalue        ll        ul
---------------------------------+----------------------------------------
         obese men - obese women |   -0.068     0.019    -0.124    -0.012
   nonobese men - nonobese women |   -0.030     0.126    -0.070     0.009
    women obese - women nonobese |    0.014     0.603    -0.041     0.069
        men obese - men nonobese |    0.052     0.086    -0.008     0.111
```

Code: DC(obese | female, race) by gender or race

```
. //  #3.3 DC(obesity | race gender)
. //  Table 4 rows 1-2, columns 5-6

. qui est restore fullmodel
. mtable, dydx(obese) at(white=(0 1) female=(0 1)) ///
>     atmeans stats(est pvalue) post

Expression: Pr(diabetes), predict()

          |    white    female   d Pr(y)          p
----------+--------------------------------------
        1 |        0         0     0.161      0.000
        2 |        0         1     0.156      0.000
        3 |        1         0     0.212      0.000
        4 |        1         1     0.170      0.000

Specified values of covariates
::
```

```
. //  3.3a Additional tests: requires Stata 14.2+ due to bug in 14.1

. //  DC of obesity are equal across race and gender
. test [1.obese]1bn._at = [1.obese]2._at = [1.obese]3._at = [1.obese]4._at

Adjusted Wald test

 ( 1)  [1.obese]1bn._at - [1.obese]2._at = 0
 ( 2)  [1.obese]1bn._at - [1.obese]3._at = 0
 ( 3)  [1.obese]1bn._at - [1.obese]4._at = 0

       F(  3,    54) =   12.42
            Prob > F =    0.0000

. //  Compare DC of obesity for nonwhite men and white women
. test [1.obese]1bn._at = [1.obese]4._at, accum

Adjusted Wald test

 ( 1)  [1.obese]1bn._at - [1.obese]2._at = 0
 ( 2)  [1.obese]1bn._at - [1.obese]3._at = 0
 ( 3)  [1.obese]1bn._at - [1.obese]4._at = 0
 ( 4)  [1.obese]1bn._at - [1.obese]4._at = 0
       Constraint 4 dropped

       F(  3,    54) =   12.42
            Prob > F =    0.0000
```

```
. //  #3.4 Second and third differences to compare effects
. //  Table 4 rows 1-3, column 7
.
. qui mlincom 2-4, rowname(D2race|fem) clear
. qui mlincom 1-3, rowname(D2race|mal) add
. qui mlincom 4-3, rowname(D2gndr|wht) add
. qui mlincom 2-1, rowname(D2gndr|non) add
. mlincom (4-3)-(2-1), rowname(D3gndrXrace) add

             |   lincom    pvalue        ll        ul
-------------+----------------------------------------
  D2race|fem |   -0.014     0.603    -0.069     0.041
  D2race|mal |   -0.052     0.086    -0.111     0.008
  D2gndr|wht |   -0.042     0.000    -0.056    -0.028
  D2gndr|non |   -0.005     0.428    -0.016     0.007
 D3gndrXrace |   -0.038     0.000    -0.057    -0.019
```

Conclusions

1. I recommend comparing groups using predictions and marginal effect on predictions in of outcomes measured in a natural metric.
2. This approach
   - Identification is not a concern
   - The method can be applied to any regression model that makes predictions
     : it is useful in any nonlinear model including LRM
   - Any number of groups can be included
3. Deciding how to make comparisons requires careful consideration based on a substantive understanding of the process being modeled and the questions being asked.

Bibliography

Allison, Paul D. 1999. Comparing logit and probit coefficients across groups. *Sociological Methods and Research* 28:186-208.

Chow, G.C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28:591-605.

Long, J.S. 2009 (2005). Comparing groups using predicted outcomes. (www.indiana.edu/~jslsoc/research_groupdif.htm)

Long and Mustillo. 2018. Using predictions to compare groups in regression models for binary outcomes

TODO Oaxaca Decomposition

1. See in Resources the rough draft
2. Search Groups for computer code and didactic figures
3. Citation to Jann article

Example: productivity, position, and tenure

Variable	Mean	StdDev	Minimum	Maximum	Label
tenure	0.12	0.33	0.00	1.00	Is tenured?
female	0.38	0.48	0.00	1.00	Scientist is female?
year	3.86	2.30	1.00	10.00	Years in rank.
yearsq	20.17	22.15	1.00	100.00	Years in rank squared.
select	5.00	1.41	1.00	7.00	Selectivity of bachelor's
articles	7.05	6.58	0.00	73.00	Total number of articles.
prestige	2.65	0.78	0.65	4.80	Prestige of department.
presthi	0.05	0.21	0.00	1.00	Prestige is 4 or higher?

N = 2797 (person-years)

Model for women

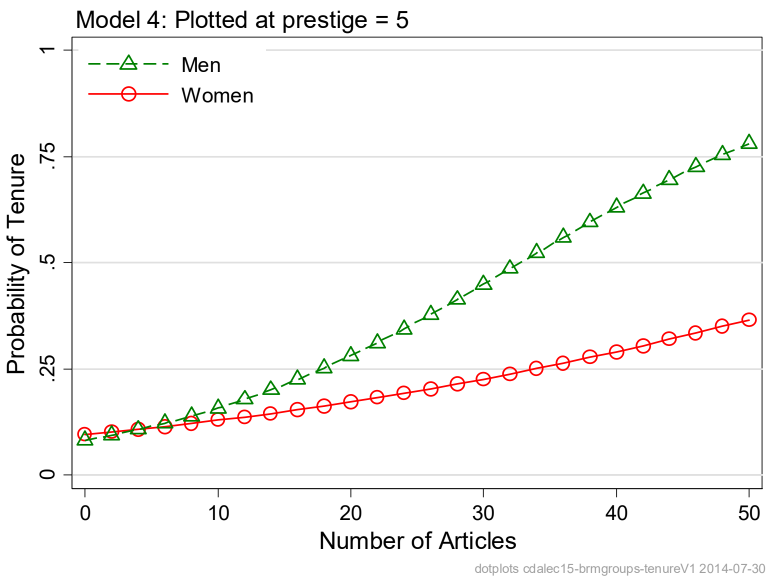
logit (N= 2797): Factor Change in Odds

WOMEN	b	z	P>|z|	e^b	e^bStdX
constant	-5.84198	-6.747	0.000	0.0029	0.0589
year	1.40777	5.472	0.000	4.0868	30.1273
yearsq	-0.09559	-4.364	0.000	0.9088	0.1857
select	0.05513	0.769	0.442	1.0567	1.1534
articles	0.03395	2.693	0.007	1.0345	1.2181
prestige	-0.37079	-2.376	0.017	0.6902	0.6013

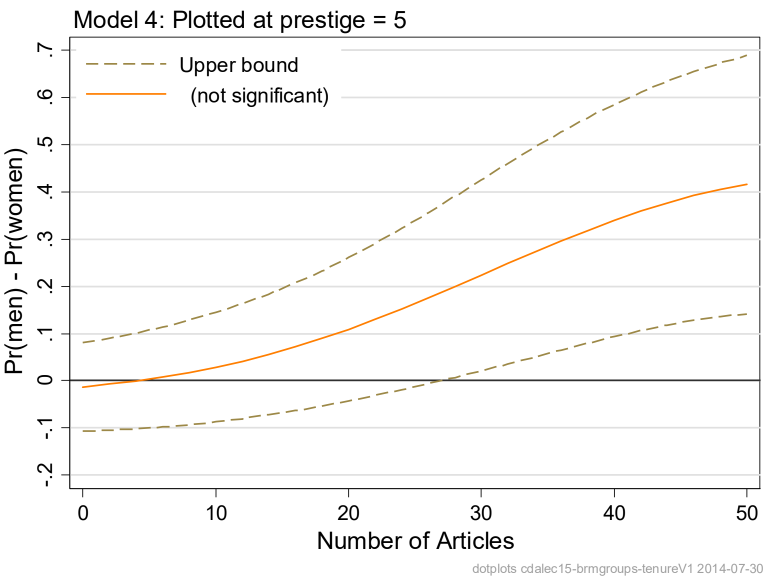
MEN	b	z	P>|z|	e^b	e^bStdX
constant	-7.68016	-11.271	0.000	0.0005	0.0241
year	1.90885	8.915	0.000	6.7454	130.9789
yearsq	-0.14322	-7.699	0.000	0.8666	0.0622
select	0.21577	3.513	0.000	1.2408	1.7711
articles	0.07369	6.367	0.000	1.0765	1.5299
prestige	-0.43119	-3.963	0.000	0.6497	0.5418

Articles with prestige = 5 (elite) other variables at means

A. Probabilities by gender

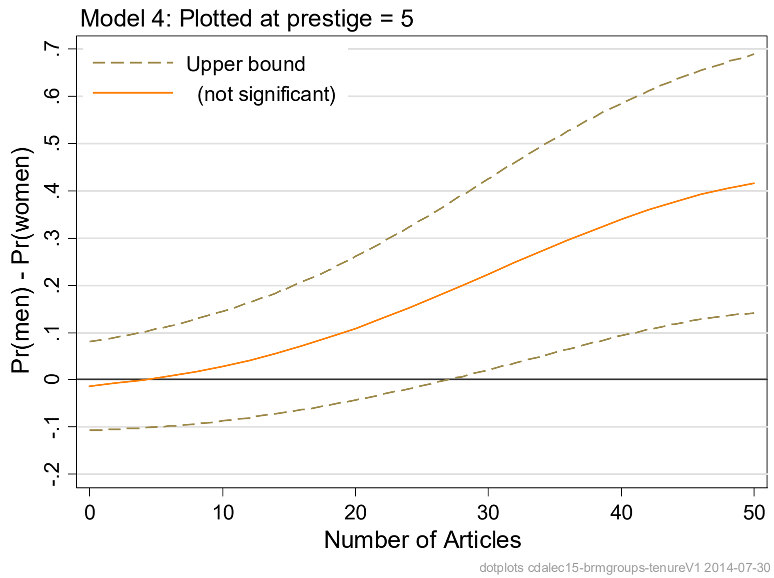


B. Group differences with CI

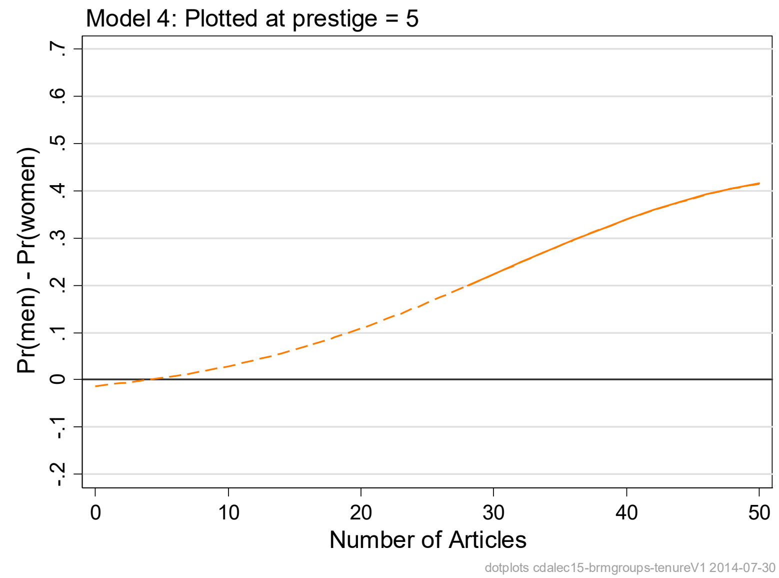


Converting CI to a broken line

B. Group difference with CI

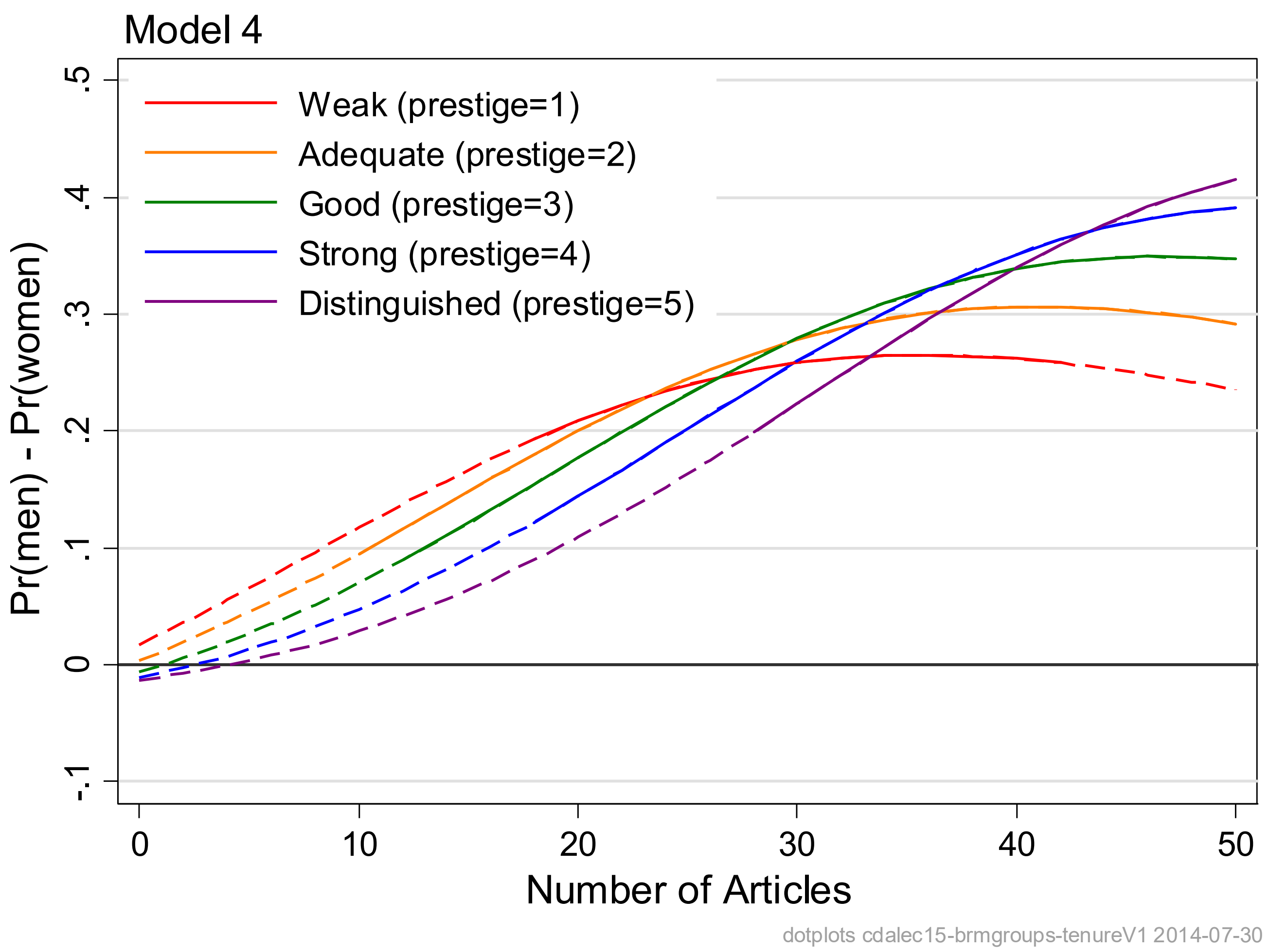


C. Group difference with broken line

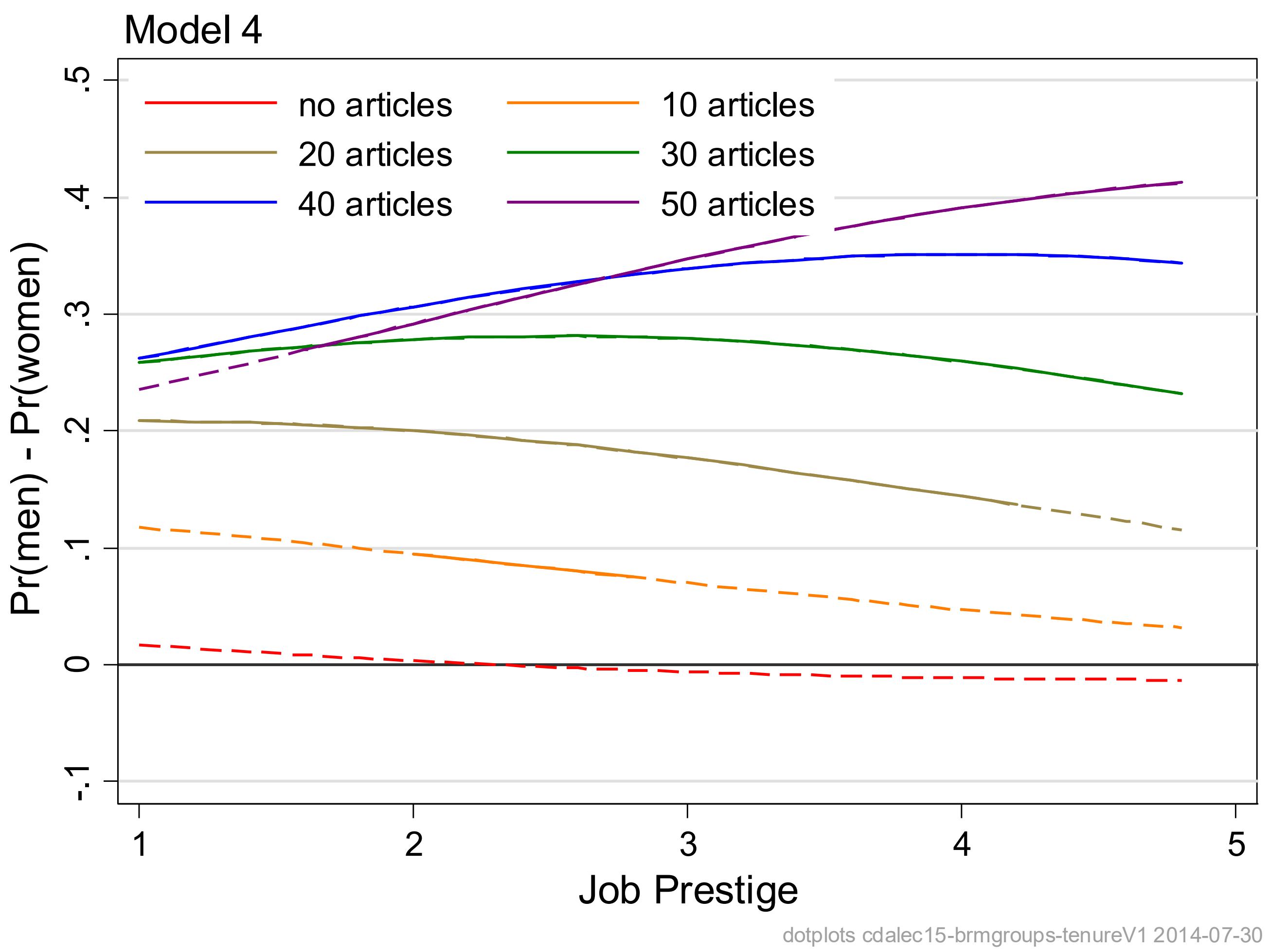


- Do this for each level of prestige
- Then combine the group differences in a single graph

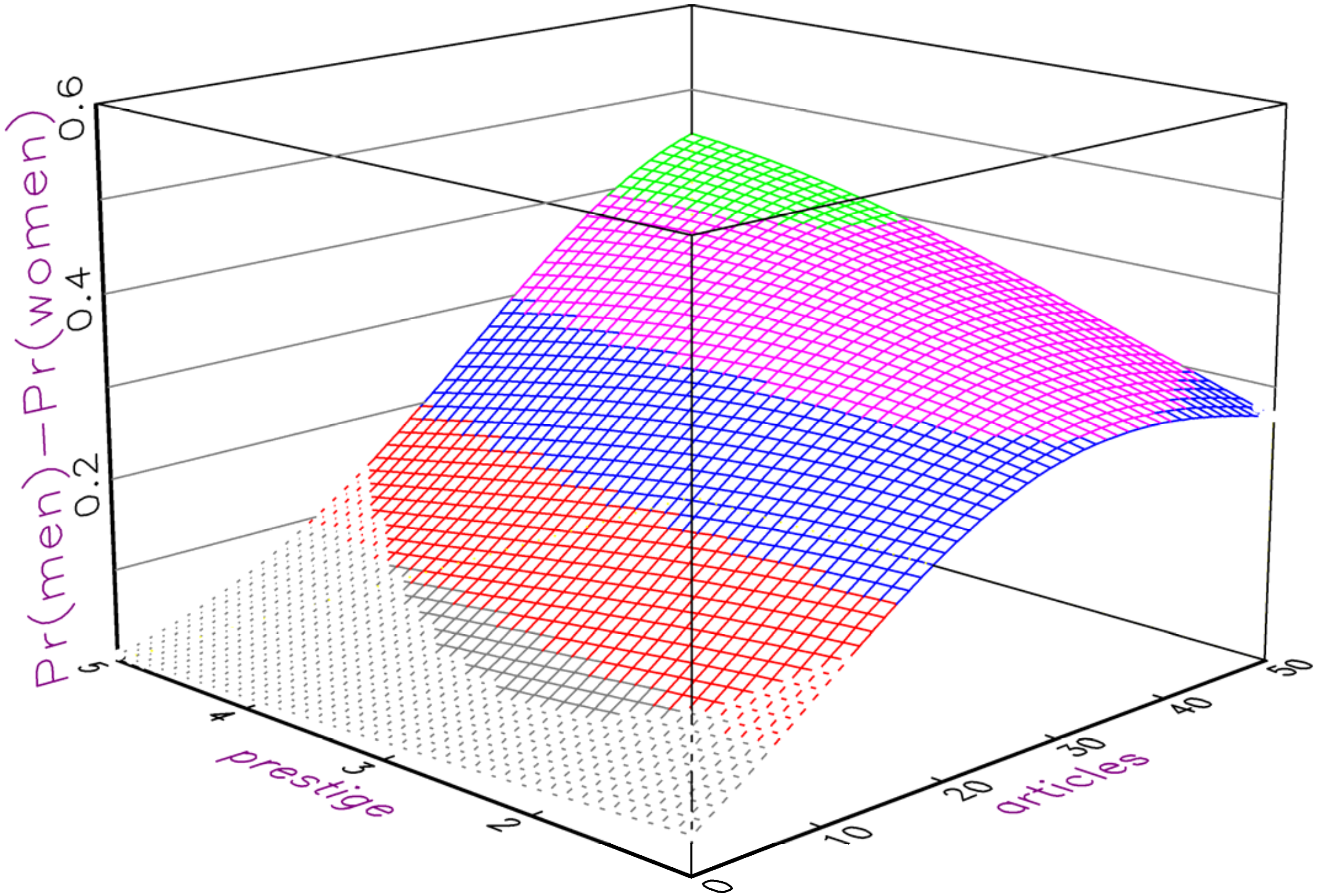
Group differences by articles & prestige



Group differences by prestige & articles



Group differences by prestige and articles



β1a Comparing Effects Across Equations

Readings and examples

Trent Mize, Long Doan, and Scott Long. 2018. A General Framework for Comparing Marginal Effects Across Models. Working paper.

See references in Mize et al.

mdo18-cme-.do*: do-files take several minutes to complete.

Overview: comparing effects across models

1. Statistical issues
   - Jointly fit multiple models
   - Jointly estimate effects with standard errors and covariances
   - Test equality of effects across models
2. Practically issues
   - Software must be able to jointly fit models and estimate effects

Examples

1. Can the relationship between two variables be explained by adding controls?
2. Does a different operationalization change the effect of interest?
3. Is a treatment more effective helpling for depression or impulse control?
4. Do conclusion from a Pew survey match those from the GSS?
5. Does the effect of age depend on whether the model is ordinal or nominal?
6. Are effects of age greater in cubic model than quadratic model?

Review of Marginal Effects

1. Consider any regression model

$$\eta(\mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta})$$

2. Examples include:
   - For BLM, $\Pr(\mathbf{x}) = \Lambda(\mathbf{x}\boldsymbol{\beta})$
   - For LRM, $y = \mathbf{x}\boldsymbol{\beta} + \varepsilon$
   - MNLM, OLM, NBRM, or nearly any other model
3. The marginal effect is the change in η for a change in x_k holding other variables at specific values.

Discrete change at representative values (DCR)

1. Change in η as x_k changes from *start* to *end* holding other variables at **x***:

$$\frac{\Delta\eta(\mathbf{x}=\mathbf{x}^*)}{\Delta x_k(start \rightarrow end)} = \eta(x_k=end, \mathbf{x}=\mathbf{x}^*) - \eta(x_k=start, \mathbf{x}=\mathbf{x}^*)$$

$$\frac{\Delta\eta(\mathbf{x}=\mathbf{x}^*)}{\Delta x_k(\overline{x}_k \rightarrow \overline{x}_k+s_k)} = \eta(x_k=\overline{x}_k+s_k, \mathbf{x}=\mathbf{x}^*) - \eta(x_k=\overline{x}_k, \mathbf{x}=\mathbf{x}^*)$$

$$\frac{\Delta\eta(\mathbf{x}=\mathbf{x}^*)}{\Delta x_k(0 \rightarrow 1)} = \eta(x_k=1, \mathbf{x}=\mathbf{x}^*) - \eta(x_k=0, \mathbf{x}=\mathbf{x}^*)$$

Average discrete change (ADC)

1. $DC_i(x_k)$ for each observation i:

$$\frac{\Delta\eta(\mathbf{x}{=}\mathbf{x}_i)}{\Delta x_k(start_i \rightarrow end_i)} = \eta(x_k{=}end_i, \mathbf{x}{=}\mathbf{x}_i) - \eta(x_k{=}start_i, \mathbf{x}{=}\mathbf{x}_i)$$

$$\frac{\Delta\eta(\mathbf{x}{=}\mathbf{x}_i)}{\Delta x_k(x_{ik} \rightarrow x_{ik}{+}\delta)} = \eta(x_k{=}x_{ik}{+}\delta, \mathbf{x}{=}\mathbf{x}_i) - \eta(x_k{=}x_{ik}, \mathbf{x}{=}\mathbf{x}_i)$$

$$\frac{\Delta\eta(\mathbf{x} = \mathbf{x}_i)}{\Delta x_k(start \rightarrow end)} = \eta(x_k{=}end, \mathbf{x}{=}\mathbf{x}_i) - \eta(x_k{=}start, \mathbf{x}{=}\mathbf{x}_i)$$

2. $ADC(x_k)$ is the average discrete change

$$\mathrm{ADC}_{x_k} = \frac{1}{N}\sum_i \frac{\Delta\eta(\mathbf{x}{=}\mathbf{x}_i)}{\Delta x_k(start_i \rightarrow end_i)})$$

3. Average can over:
   - entire sample
   - subsamples (e.g, men and women)

Testing the equality of effects

1. Models:
   - M1: logit of satisfaction on education plus controls
   - M2: logit of satisfaction on income plus controls
2. Jointly fit M1 and M2 and estimate effects (ADC, DCR, RRR,…)
   - Δ_1 = ADC(education) from M1
   - Δ_2 = ADC(income) from M2
3. Wald test statistic for Ho: $\Delta_1 = \Delta_2$

$$z = \frac{\widehat{\Delta}_1 - \widehat{\Delta}_2}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 - 2\hat{\sigma}_{1,2}}}$$

4. $\hat{\sigma}_{1,2}$ usually requires joint fitting of models and estimation of effects
5. Standard error computed with delta method, bootstrapping, or simulation.
6. Joint estimation is completed using SUEST

Joint fitting with seemingly unrelated estimation (SUEST)

Adapted from Wessie (1999)

Unstacked data of N observations

$$\begin{pmatrix} depress & sick & roles & id \\ 0 & 0 & 5 & 1 \\ 1 & 1 & 1 & 2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 8 & N \end{pmatrix}$$

1. Fit two models
   - Model 1: **`regress depress roles`**
   - Model 2: **`regress sick roles`**
2. We can estimate and test individual effects but not compare them.

Stacked data of 2N observations

$$\begin{pmatrix} y & rolesm1 & rolesm2 & ismodel1 & id \\ depress=0 & 5 & 0 & 1 & 1 \\ depress=1 & 1 & 0 & 1 & 2 \\ \vdots & \vdots & \vdots & \vdots & \\ depress=1 & 8 & 0 & 1 & N \\ sick=0 & 0 & 5 & 0 & 1 \\ sick=1 & 0 & 1 & 0 & 2 \\ \vdots & \vdots & \vdots & \vdots & \\ sick=1 & 0 & 8 & 0 & N \end{pmatrix}$$

y	First N rows are *depress* from unstacked data; last N rows are *sick*
rolesm1	First N rows are *roles* from unstacked dataset; last N fixed to 0.
rolesm2	First N rows are 0; last N contain *roles* from unstacked dataset
ismodel1	1 if observation for Model 1; 2 if observation for Model 2

Fitting models with stacked data

1. Fit models separately

```
logit y rolesm1 if ismodel1==1
logit y rolesm2 if ismodel1==2
```

2. To fit model jointly

```
logit y ismodel2 rolesm1 rolesm2, vce(cluster id)
```

- *ismodel1* is 1 if the observation is needed to fit model 1 and 0 if the observation is used to fit model 2.
- *id* is used to adjust for clustering caused by having two observations for each participant.
- **`vce(cluster id)`** adjust for the clustered data

3. Example of results of using stacked data follows…

Estimates from Models 1 and 2 fit separately and jointly

	Model 1	Model 2	Model 3
Outcome:	*depress*	*sick*	Both
roles	−0.2006 (0.0252)	−0.2608 (0.0365)	
rolesm1			−0.2006 (0.0257)
rolesm2			−0.2608 (0.0371)
ismodel1			−0.8115 (0.2311)
Constant	−0.0401 (0.1478)	−0.8516 (0.2065)	−0.0401 (0.1503)
N	4,307	4,307	8,614

SUEST and generalized SEM

1. Stacking data only works if models are the same type (e.g., two logits)
2. SUEST is a statistical method that is a more general way to "stack" results to combines estimates across models to obtain correct SE's
3. After joint estimation, marginal effects can be jointly estimated across models
4. SEM programs, like gsem, can also be used to jointly fit models
   - **`suest`** is faster and more general, but postestimation is messy
   - **`gsem`** is much slower but eaiser to use for postestimation

Examples and Applications

1. Compare effects across nested models
   - Does the effect of a x_k changes as controls are added to the model?
2. Compare effects of different operationalizations of a variable.
   - Does the way sexual orientation is measured change its effects?
3. Compares effects of the same predictor on different outcomes.
   - Does gender have the effect on diabetes as arthritis?
4. Group comparisons is a special case of this method.
5. Compare effects of x_k in OLM to the effect of x_k in MNLM
6. Does the effect of age in a model with age-squared differ from the effect in a model that adds age-cubed?

Mediation of effects using nested logit models

1. People with college degrees tend to be happier than those without.
2. Does the effect of a college degree change after accounting for earnings, occupational prestige, and other variables?

Data and variables from GSS 2000-2016

```
Variable    Unique       Mean   Min   Max   Label
----------------------------------------------------------------------------
vhappy           2   .3087023     0     1   Very Happy with Life

college          2   .3390842     0     1   R Has College Degree

wages         2936    17.4244  .1115278    70   Wages
occprest        62   45.31217        16    86   Occupational Prestige Score

married          2   .4950087      0      1   Currently Married
parent           2   .6885851      0      1   Parent
woman            2   .4930556      0      1   Woman R
conserv          2    .335612      0      1   Conservative R
reltrad          7   3.650282      1      7   Religous Tradition
year             9   2008.175   2000   2016   gss year for this respondent
age             71   42.94477     18     88   age of respondent
----------------------------------------------------------------------------
```

Models

M1: vhappy on college

M2: vhappy on college + controls

M3: vhappy on college + controls + wages

M4: vhappy on college + controls + wages + occprest

Results: ADC(college) on happiness

Panel A: ADC(college)

	Model 1: College	Model 2: + Controls	Model 3: + Wages	Model 4: + Prestige
ADC(college)	0.072***	0.060***	0.036***	0.019
	(0.010)	(0.011)	(0.011)	(0.012)

Panel B: Cross Model Differences

	Model 1:	Model 2:	Model 3:	Model 4:
Model 1		0.012**	0.036***	0.053***
		(0.004)	(0.006)	(0.007)
Model 2	0.012**		0.024***	0.041***
	(0.004)		(0.004)	(0.006)
Model 3	0.036***	0.024***		0.017***
	(0.006)	(0.004)		(0.004)
Model 4	0.053***	0.041***	0.017***	
	(0.007)	(0.006)	(0.004)	

Notes: Standard errors in parentheses. *=p≤.05, **=p≤.01, ***=p≤.001 for a two-tailed test.

Model 1 with only college as predictors

On average the probability of being happy is .072 (p<.001) higher for those with a college degree.

Model 2 adds control variables

Adding the controls decreases the effect of college by .012 (p<.001).

Model 3 examines if effect of college is mediated by hourly wages.

After controlling for wages, the effect of college continues to be significant (Δ=.036, p<.001), but is reduced by .024.

Model 4 adds wages as a mediator.

On average going to college increase the probability of a respondent reporting being very happy by .019 which is no long34 significant.

Overall, we can conclude

The relationship between a college degree and happiness is totally moderated by wages and prestige.

Fit models separately and estimates effects of college

```
local controls ///
    i.married i.parent i.woman i.conserv i.reltrad i.year c.age##c.age

logit vhappy i.college  // M1
mchange college, stat(est se p) dec(7)

logit vhappy i.college `controls'  // M2 + controls
mchange college, stat(est se p)

logit vhappy i.college c.wages `controls'  // M3 + wages
mchange college wages , amount(sd) stat(est se p)

logit vhappy i.college c.wages c.occprest `controls'  // M4 + occprest
mchange college wages occprest, amount(sd) stat(est se p)
```

Joint fitting of models

Clones to trick gsem into using the same DV in all models

```
clonevar  vhappyM1 = vhappy
lab var   vhappyM1 "M1 vhapppy college only"
clonevar  vhappyM2 = vhappy
lab var   vhappyM2 "M2 vhapppy college + controls"
clonevar  vhappyM3 = vhappy
lab var   vhappyM3 "M3 vhapppy college + controls + wages"
clonevar  vhappyM4 = vhappy
lab var   vhappyM4 "M4 vhapppy college + controls + wages + occprest"
//  NOTE: vce(robust) needed for correct covariances across models
```

Simultaneously fit models with gsem

```
gsem (vhappyM1 <- i.college, logit) ///
     (vhappyM2 <- i.college `controls', logit) ///
     (vhappyM3 <- i.college c.wages `controls', logit) ///
     (vhappyM4 <- i.college c.wages c.occprest `controls', logit) ///
     , vce(robust)

est store gsemmodel
```

Jointly estimate ADC(college) and post results

```
. margins, dydx(college) post

Average marginal effects                          Number of obs     =      9,216

dy/dx w.r.t. : 1.college
1._predict   : Predicted mean (M1 vhapppy college only), predict(mu
               outcome(vhappyM1))
2._predict   : Predicted mean (M2 vhapppy college + controls), predict(mu
               outcome(vhappyM2))
3._predict   : Predicted mean (M3 vhapppy college + controls + wages),
               predict(mu outcome(vhappyM3))
4._predict   : Predicted mean (M4 vhapppy college + controls + wages +
               occprest), predict(mu outcome(vhappyM4))

             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0.college    |  (base outcome)
-------------+----------------------------------------------------------------
1.college    |
    _predict |
          1  |    .071806   .0103344     6.95   0.000     .0515509    .0920611
          2  |   .0599197   .0105299     5.69   0.000     .0392813     .080558
          3  |   .0359563   .0111027     3.24   0.001     .0141954    .0577172
          4  |   .0191807   .0117766     1.63   0.103     -.003901    .0422625
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Compute ADC(college) and second differences across models

```
qui mlincom 1,   rowname(ADCcollege: M1)      stat(est se p) clear
qui mlincom 2,   rowname(ADCcollege: M2)      stat(est se p) add
qui mlincom 1-2, rowname(ADCcollege: M1 - M2) stat(est se p) add
qui mlincom 1,   rowname(ADCcollege: M1)      stat(est se p) add
qui mlincom 3,   rowname(ADCcollege: M3)      stat(est se p) add
qui mlincom 1-3, rowname(ADCcollege: M1 - M3) stat(est se p) add
qui mlincom 1,   rowname(ADCcollege: M1)      stat(est se p) add
qui mlincom 4,   rowname(ADCcollege: M4)      stat(est se p) add
qui mlincom 1-4, rowname(ADCcollege: M1 - M4) stat(est se p) add
qui mlincom 2,   rowname(ADCcollege: M2)      stat(est se p) add
qui mlincom 3,   rowname(ADCcollege: M3)      stat(est se p) add
qui mlincom 2-3, rowname(ADCcollege: M2 - M3) stat(est se p) add
qui mlincom 2,   rowname(ADCcollege: M2)      stat(est se p) add
qui mlincom 4,   rowname(ADCcollege: M4)      stat(est se p) add
qui mlincom 2-4, rowname(ADCcollege: M2 - M4) stat(est se p) add
qui mlincom 3,   rowname(ADCcollege: M3)      stat(est se p) add
qui mlincom 4,   rowname(ADCcollege: M4)      stat(est se p) add
mlincom 3-4, rowname(ADCcollege: M3 - M4)  stat(est se p) add twidth(15)
```

ADC(wages+SD) for models 3 and 4

```
est restore gsemmodel

qui sum wages
local sd = r(sd)
margins, ///
    predict(outcome(vhappyM3)) predict(outcome(vhappyM4)) ///
    at(wages=gen(wages)) at(wages=gen(wages + `sd')) post

qui mlincom 3,   rowname(ADCwages: M3)      stat(est se p) clear
qui mlincom 4,   rowname(ADCwages: M4)      stat(est se p) add
    mlincom 3-4, rowname(ADCwages: M3 - M4) stat(est se p) add twidth(15)
```

Effects of alternative regressors

1. Sexual orientation can be measured in different ways.
   - How someone *identifies with an orientation*
   - How someone *behaves*
2. Does classifying sexual orientation based on their *identity* leads to different conclusions than classifying based on *behavior*?

Data and variables: GSS data from 2008 – 2016

1. Outcome is whether a person views same-sex relationships as wrong (=1) or not wrong (=0).
2. Sexual orientation *identity* and sexual orientation *behavior* have three categories
   - heterosexual
   - bisexual
   - gay/lesbian.

Logit model of same-sex is wrong on orientation + controls

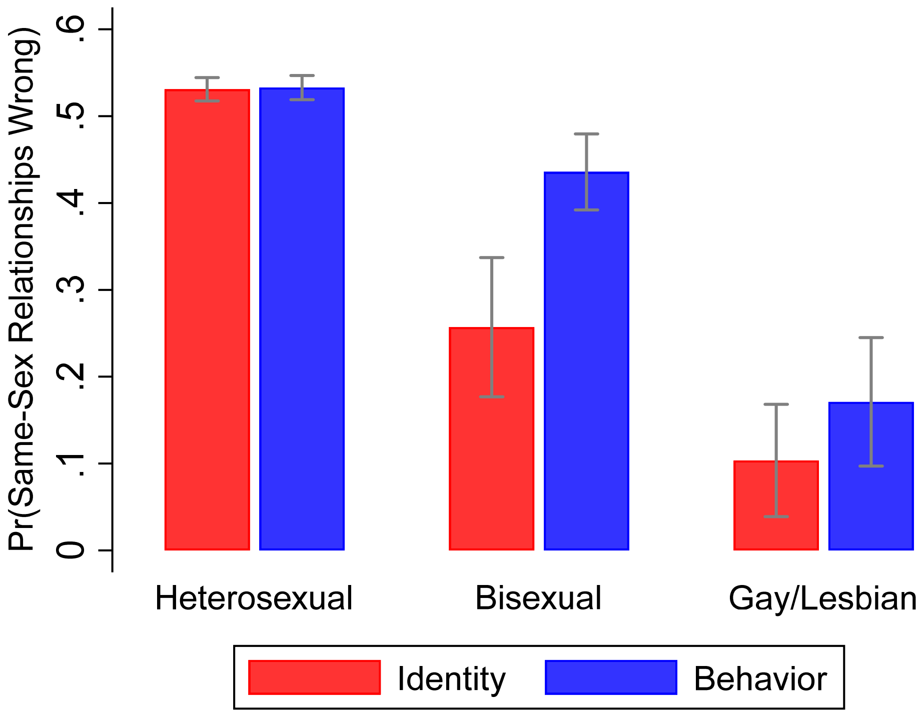
M1: sexual identity

```
iswrong <- sexident woman college age race year
```

M2: sexual behavior

```
iswrong <- sexbehav woman college age race year
```

Pr(iswrong) by sexual orientation measured two ways



Views of same-sex relationships are nearly identical across models for those who identify as heterosexual, fairly close for those who identify as gay or lesbian, but the probability of viewing same-sex relationships as wrong is substantially lower for those who are classified as bisexual based their sexual identity compared to past sexual behavior.

Pr(iswrong) by sexual orientation measured two ways

Here are the numbers behind the prior graph

Orientation	Model 1: Sexual Identity	Model 2: Sexual Behavior	Cross Model Difference
Heterosexual	0.531	0.533	0.002 (0.002)
Bisexual	0.257	0.436	0.179*** (0.041)
Gay/Lesbian	0.104	0.171	0.067 (0.040)

The probability of bisexuals saying same-sex relationships are wrong are significantly larger when orientation is measured by behavior than by identity (p<.001). The probabilities for heterosexuals and gay/lesbians do not differ signifcantly.

ADC(sexual orientation) comparing two orientations

	Model 1: Sexual Identity	Model 2: Sexual Behavior	Cross Model Difference
Heterosexual vs Bisexual	-0.274***	-0.097***	0.177***
	(0.041)	(0.023)	(0.043)
Heterosexual vs Gay/Lesbian	-0.428***	-0.362***	0.066
	(0.034)	(0.038)	(0.041)
Bisexual vs Gay/Lesbian	-0.265**	-0.153*	-0.111
	(0.044)	(0.052)	(0.059)

*=p=.05, **=p=.01, ***=p=.001 for a two-tailed test

The contrast between heterosexual and gay/lesbian and between bisexual and gay/lesbian do not differ significantly across the two models. However, there is a significantly larger difference between heterosexual and bisexual individuals when using the sexual identity measure compared to , compared sexual behavior measure. That is, for this contrast the different ways of measuring sexual orientation change the effect.

Is this clear?

Version 1 confuses me!

The contrast between heterosexual and gay/lesbian and between bisexual and gay/lesbian do not differ significantly across the two models. However, there is a significantly larger difference between heterosexual and bisexual individuals when using the sexual identity measure compared to , compared sexual behavior measure. That is, for this contrast the different ways of measuring sexual orientation change the effect.

Version 2: is this better?

The effect of being heterosexual compared to being gay or lesbian does not differ when sexual orientation is measured by behavor or identity. However, the effect of being heterosexual compared to bisexual is significantly larger (p<.001) when orientation is measured by sexual identity.

Code for example

Jointly fit models

```
clonevar samesexBEHAV = samesexB // 1=is wrong; 0=not wrong
clonevar samesexIDENT = samesexB

gsem (samesexBEHAV <- i.sexbehav i.woman i.college c.age /// model 2
                      i.race i.year, logit) ///
     (samesexIDENT <- i.sexident i.woman i.college c.age /// model 1
                      i.race i.year, logit), ///
     vce(robust)

est store gsemmodel
```

Estimate and plot probabilities by orientation

```
// model using behavior
est restore gsemmodel
margins, predict(outcome(samesexBEHAV)) at(sexbehav=(1 2 3)) post
est store probbehav
// model using identity
est restore gsemmodel
margins, predict(outcome(samesexIDENT)) at(sexident=(1 2 3)) post
est store probident

local labopt "labsize(*1.1) glwidth(*1.4) glcol(black*.7)"

coefplot (probident, color(red*.8)) (probbehav, color(black)), ///
    vertical recast(bar) barw(0.3) ///
    ciopts(recast(rcap) color(gs8)) citop ///
    legend(order(1 "Identity" 3 "Behavior")) ///
    xlab(1 "Heterosexual" 2 "Bisexual" 3 "Gay/Lesbian", noticks) ///
    ytitle("Pr(Same-Sex Relationships Wrong)",size(*.85)) ///
    ylab(0(0.1).6, `labopt') ///
    scale(1.3) /// larger text
    xscale(noline) plotregion(style(none)) // turn off x axis line
```

Predicted probabilities and differences for each model

```
est restore gsemmodel
margins, at(sexident=(1 2 3)) at(sexbehav=(1 2 3)) post
qui {
mlincom 7,   rowname("Hetero: Ident") stat(est se p) clear
mlincom 4,   rowname("Hetero: Behav") stat(est se p) add
mlincom 4-7, rowname("Hetero: Difference") stat(est se p) add
mlincom 8,   rowname("Bisexual: Ident") stat(est se p) add
mlincom 5,   rowname("Bisexual: Behav") stat(est se p) add
mlincom 5-8, rowname("Bisexual: Difference") stat(est se p) add
mlincom 9,   rowname("Gay: Ident") stat(est se p) add
mlincom 6,   rowname("Gay: Behav") stat(est se p) add
mlincom 6-9, rowname("Gay: Difference") stat(est se p) add
}
mlincom, twidth(20) title("Differences in Probabilities Across Models")
```

Pairwise comparison AMEs

```
est restore gsemmodel
margins, dydx(sexident sexbehav) post
qui {
mlincom 3,   rowname("Het vs Bi: Ident") stat(est se p) clear
mlincom 1,   rowname("Het vs Bi: Behav") stat(est se p) add
mlincom 1-3, rowname("Het vs Bi: Difference") stat(est se p) add
mlincom 4,   rowname("Het vs Gay: Ident") stat(est se p) add
mlincom 2,   rowname("Het vs Gay: Behav") stat(est se p) add
mlincom 2-4, rowname("Het vs Gay: Difference") stat(est se p) add
mlincom 1-2, rowname("Bi vs Gay: Ident") stat(est se p) add
mlincom 3-4, rowname("Bi vs Gay: Behav") stat(est se p) add
mlincom (3-4)-(1-2), rowname("Bi vs Gay: Difference") stat(est se p) add
}
mlincom, twidth(20) ///
    title("AMEs for Sexual Orientation Within and Across Models")
```

Effects of x_k on different outcomes

1. When comparing effects on different outcome the way variables are measured must be comparable.
   - Can you compare being ADC(married) to ADC(income)?
2. Two outcomes:
   - Days of poor mental health in last month
   - Days of poor physical health in last month

Data from the 2002, 2006, 2010, and 2014 General Social Survey

Variable	Unique	Mean	Min	Max	Label
mntlhlth	27	3.480245	0	30	days of poor mental health...
physhlth	28	2.535559	0	30	days of poor physical heal...
woman	2	.5132359	0	1	Woman R
married	2	.4849862	0	1	Currently Married
age	69	42.46602	18	88	age of respondent
faminc	98	39.06672	.2365	144.5027	Family Income in 1000s
race	3	1.327143	1	3	race of respondent
college	2	.3231924	0	1	R Has College Degree
parent	2	.6851047	0	1	Parent
reltrad	7	3.62663	1	7	Religious Tradition

Models: negative binomial regression

Model 1: Mental health

```
mntlhlth <- woman married age faminc race college parent year
```

Model 2: Physical health

```
physhlth <- woman married age faminc race college parent year
```

Effects in count models

1. We use the ADC on the expected # of number of days of poor health

$$\Delta E(y|\mathbf{x}) / \Delta x_k = \Delta \exp(\mathbf{x}'\boldsymbol{\beta}) / \Delta x_k$$

ADC on Days Reporting Poor Health

	Poor Mental Health Days	Poor Physical Health Days	Cross Model Difference
Woman	0.993***	0.773***	0.220
Married	-1.010***	-0.159	-0.851**
Parent	0.274	-0.269	0.543*
Age	-0.462***	0.491***	-0.953***
Family Income	-0.445***	-0.381***	-0.064
College	-0.879***	-0.542**	-0.337
Black vs White	-1.016***	-0.531*	-0.485
Other vs White	-0.438	0.145	-0.583
Other vs Black	0.578	0.676	-0.098

*=p= .05, **=p=.01, ***=p=.001 for a two-tailed test.

Results

Woman are expected to report about 0.99 more days of poor physical health per month than men and 0.77 more days of poor physical health. While the effect is 0.22 larger for mental health, the difference is not statistically significant.

Being married has a significantly larger effect on mental health than physical health. On average being married significantly reduces the days of poor mental health by 1.01 days, but the effect of marriage on physical health is nonsignificant. The effect on mental health is significantly larger (Δ= -.851).

The effect of age diverge across the two outcomes, with aging associated with fewer poor mental health days but more poor physical health days ($p<.001$).

Code

Fit models

```
gsem (mntlhlth <- i.woman i.married c.age faminc ///
                  i.race i.college i.parent i.year, nbreg) ///
     (physhlth <- i.woman i.married c.age faminc ///
                  i.race i.college i.parent i.year, nbreg), vce(robust)
est store gsemmodel
```

Effects for binary regressors

```
mlincom, clear
foreach var in woman married college parent {
    est restore gsemmodel
    margins, dydx(`var') post
    qui mlincom 1,   rowname(`var':MentHlth) stat(est se p) add
    qui mlincom 2,   rowname(`var':PhysHlth) stat(est se p) add
    qui mlincom 1-2, rowname(`var':Difference) stat(est se p) add
}
```

Effect of continuous regressors + SD

```
foreach var in age faminc {
    est restore gsemmodel
    qui sum `var'
    margins, at(`var'=gen(`var')) at(`var'=gen(`var' + `r(sd)')) post
    qui mlincom 2-1, rowname(`var':MentHlth) stat(est se p) add
    qui mlincom 4-3, rowname(`var':PhysHlth) stat(est se p) add
    qui mlincom (2-1)-(4-3), rowname(`var':Difference) stat(est se p) add
}
```

Effects for multicategory regressors

```
est restore gsemmodel
margins i.race, post
qui {
  mlincom 2-1, rowname(blackVwhite:MentHlth) stat(est se p) add
  mlincom 5-4, rowname(blackVwhite:PhysHlth) stat(est se p) add
  mlincom (2-1)-(5-4), rowname(blackVwhite:Difference) stat(est se p) add

  mlincom 3-1, rowname(otherVwhite:MentHlth) stat(est se p) add
  mlincom 6-4, rowname(otherVwhite:PhysHlth) stat(est se p) add
  mlincom (3-1)-(6-4), rowname(otherVwhite:Difference) stat(est se p) add

  mlincom 3-2, rowname(otherVblack:MentHlth) stat(est se p) add
  mlincom 6-5, rowname(otherVblack:PhysHlth) stat(est se p) add
  mlincom (3-2)-(6-5), rowname(otherVblack:Difference) stat(est se p) add
}
mlincom, twidth(20) ///
  title(AMEs for mental and physical health and cross model differences)
```

Software Implementation

1. To compare effects across models you need covariances across models.
2. Then you must jointly estimate effects across models
3. Stata's suest, gsem, and margins do this very generally
   - You may have to work hard to accomplish the same thing in other packages
4. In Stata, **`suest`** is more general but post-estimation is nasty and tedious.
5. **`gsem`** is *slow* but the syntax is much simple.

Conclusions

1. Many of the example can be approached in other ways.
   - Our approach is not a critique of that work work
2. Rather we provide a general and flexible approach fully supported by mainline software
3. Can you think of other applications?
   - Assessing alternative model specifications for nonlinearity?
   - Comparing predictions for ZIP and NBRM

β1a Nominal outcomes

Readings and examples

Long & Freese: Chapter 8 sections on MNLM

- o See references in that chapter

mdo18-nrm-.do*

Overview

1. What does it mean for an outcome to be nominal or ordinal?
2. MNLM as a set of binary logits and the IIA assumption.
3. Interpretation with probabilities, marginal effects, and odds ratios.
4. Related models.

Level of measurement

1. S.S. Stevens (1946) introduced the terms nominal and ordinal:
   - o Nominal scales have no ordering implied by the values
   - o Ordinal scales have values indicating rank ordering on one attribute.
2. Debated and critiqued when proposed, his taxonomy is firmly established.

The bias-efficiency trade-off

1. The true model is:

```
regress y x1 x2 x3
```

2. Excluding variables leads to bias (if variables are correlated)

```
regress y x1
```

3. Including extra variables leads to inefficiency (larger standard errors)

```
regress y x1 x2 x3 x4
```

Bias and inefficiency when assuming wrong level of measurement

		True Level			
		Nominal	**Ordinal**	**Interval**	**Ratio**
Assumed	**N**	**OK**	Inefficient	Inefficient	Inefficient
Level	**O**	*Biased*	**OK**	Inefficient	Inefficient
for	**I**	*Biased*	*Biased*	**OK**	Inefficient
Analysis	**R**	*Biased*	*Biased*	*Biased*	**OK**

1. Using MNLM with an ordinal outcome is a useful way to assess ordinality.
2. MNLM can even be used for interval outcomes to explore nonlinearities.

Binary logit with new notation

1. Linear in the logit for outcome A versus B:

$$\ln\left[\frac{\Pr(y=A\mid \mathbf{x})}{\Pr(y=B\mid \mathbf{x})}\right]=\ln\Omega(\mathbf{x})=\beta_{0,A|B}+\beta_{1,A|B}x_1+\beta_{2,A|B}x_2+\beta_{3,A|B}x_3$$

2. Multiplicative in the odds:

$$\Omega(\mathbf{x},x_2)=\exp\left[\ln\Omega(\mathbf{x},x_2)\right]$$

$$=e^{\beta_{0,A|B}}e^{\beta_{1,A|B}x_1}e^{\beta_{2,A|B}x_2}e^{\beta_{3,A|B}x_3}$$

3. Odds ratio:

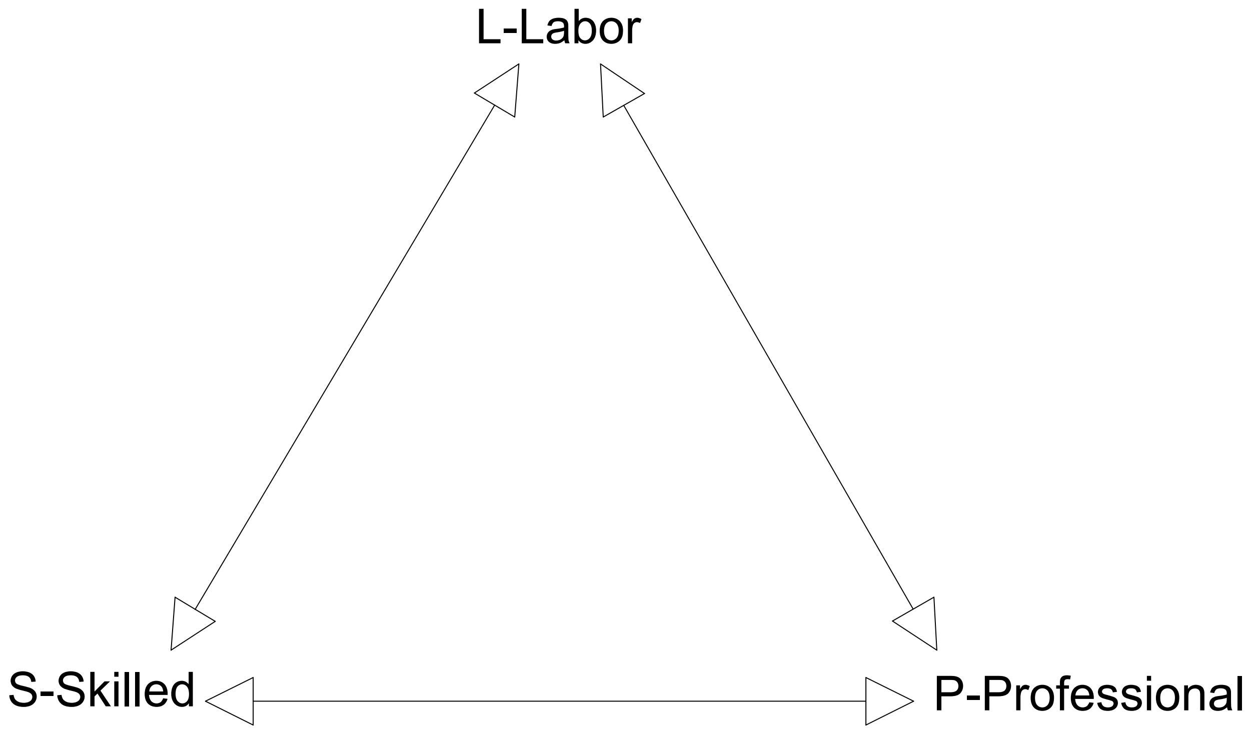
$$\frac{\Omega(\mathbf{x},x_2+1)}{\Omega(\mathbf{x},x_2)}=\frac{e^{\beta_{0,A|B}}e^{\beta_{1,A|B}x_1}e^{\beta_{2,A|B}(x_2+1)}e^{\beta_{3,A|B}x_3}}{e^{\beta_{0,A|B}}e^{\beta_{1,A|B}x_1}e^{\beta_{2,A|B}x_2}e^{\beta_{3,A|B}x_3}}=\exp\left(\beta_{2,A|B}\right)$$

Introduction to the MNLM

1. MNLM is equivalent to a set binary logits for all outcome pairs
   - With three outcomes: A vs B, A vs C, B vs C
   - Each comparison is called a contrast
2. Interpretation is complicated by large the number of parameters:
   - With 5 outcomes, there are 10 binary logits
   - With 10 outcomes, there are 45 binary logits
3. We start with a simple model.

MNLM with three outcomes

Categories L, S, and P with N_L, N_S, and N_P observations.



MNLM is a set of BLM that are simultaneously estimated.

L vs S

$$\ln\left[\frac{\Pr\left(L \mid Ed\right)}{\Pr\left(S \mid Ed\right)}\right] = \beta_{0,L|S} + \beta_{1,L|S}Ed$$

S vs P

$$\ln\left[\frac{\Pr\left(S \mid Educ\right)}{\Pr\left(P \mid Educ\right)}\right] = \beta_{0,S|P} + \beta_{1,S|P}Ed$$

L vs P

$$\ln\left[\frac{\Pr\left(L \mid Educ\right)}{\Pr\left(P \mid Educ\right)}\right] = \beta_{0,L|P} + \beta_{1,L|P}Ed$$

Redundancy among the BLMs

1. Logit for L vs P

$$\ln\left[\frac{\Pr(L \mid Ed)}{\Pr(P \mid Ed)}\right] = \ln\Pr(L \mid Ed) - \ln\Pr(P \mid Ed) + 0$$

2. Adding $0 = \ln\Pr(S \mid Ed) - \ln\Pr(S \mid Ed)$:

$$\ln\left[\frac{\Pr(L \mid Ed)}{\Pr(P \mid Ed)}\right] = \ln\Pr(L \mid Ed) - \ln\Pr(P \mid Ed) + \left[\ln\Pr(S \mid Ed) - \ln\Pr(S \mid Ed)\right]$$

$$= \left[\ln\Pr(L \mid Ed) - \ln\Pr(S \mid Ed)\right] + \left[\ln\Pr(S \mid Ed) - \ln\Pr(P \mid Ed)\right]$$

$$= \ln\left[\frac{\Pr(L \mid Ed)}{\Pr(S \mid Ed)}\right] + \ln\left[\frac{\Pr(S \mid Ed)}{\Pr(P \mid Ed)}\right]$$

Then

$$\left(\beta_{0,L|S} + \beta_{1,L|S}Ed\right) + \left(\beta_{0,S|P} + \beta_{1,S|P}Ed\right) = \left(\beta_{0,L|P} + \beta_{1,L|P}Ed\right)$$

A minimal set of coefficients

1. Since

$$\left(\beta_{0,L|S} + \beta_{1,L|S} Ed\right) + \left(\beta_{0,S|P} + \beta_{1,S|P} Ed\right) = \left(\beta_{0,L|P} + \beta_{1,L|P} Ed\right)$$

2. From any two coefficients you can compute the third

$$\beta_{L|P} = \beta_{L|S} + \beta_{S|P}$$

$$\beta_{L|S} = \beta_{L|P} - \beta_{S|P}$$

$$\beta_{S|P} = \beta_{L|P} - \beta_{L|S}$$

3. With J outcomes, J-1 comparisons allow you to determine all contrasts
   - Each set of J-1 comparisons is called a <u>minimal set</u>

Three alternative minimal sets of coefficients in 3 category MNLM

1. With base category 1:

Education	b	z	P>|z|	e^b	e^bStdX
2Skilled vs 1Labor	0.1711	2.900	0.004	1.187	1.655
3Prof vs 1Labor	0.7433	8.773	0.000	2.103	8.936

2. With base category 2:

Education	b	z	P>|z|	e^b	e^bStdX
1Labor vs 2Skilled	-0.1711	-2.900	0.004	0.843	0.604
3Prof vs 2Skilled	0.5722	7.651	0.000	1.772	5.398

3. With base category 3:

Education	b	z	P>|z|	e^b	e^bStdX
1Labor vs 3Prof	-0.7433	-8.773	0.000	0.476	0.112
2Skilled vs 3Prof	-0.5722	-7.651	0.000	0.564	0.185

4. Coefficients are linked

(1L|3P) = (1L|2S) + (2S|3P)

-0.74332 = -0.17109 + -0.57223

Comparing MNLM to a set of BLMs

1. MNLM enforces the logical relationship among the parameters.
2. They do not hold exactly if you fit separate BLMs.
3. Consider three outcomes:

   `occlsp` = 1 if labor, 2 if skilled, and 3 if professional; for MNLM

   `labrskil` = 1 if labor, 0 if skilled, else missing; for BLM

   `profskil` = 1 if professional, 0 if skilled, else missing; for BLM

   `labrprof` = 1 if labor, 0 if professional, else missing; for BLM
4. The predictor is years of education.

Comparing BLM and MNLM estimates

Three binary logits

```
. logit labrskil ed

  Odds of: 1Labor vs 0Skilled (N=225)

    labrskil |       b          z      P>|z|     e^b    e^bStdX      SDofX
-------------+------------------------------------------------------------
          ed |  -0.18398    -2.989    0.003   0.8320    0.6485     2.3536

. logit profskil ed

  Odds of: 1Prof vs 0Skilled (N=237)

    profskil |       b          z      P>|z|     e^b    e^bStdX      SDofX
-------------+------------------------------------------------------------
          ed |   0.56026     7.186    0.000   1.7511    4.9101     2.8403

. logit labrprof ed

  Odds of: 1Labor vs 0Prof (N=212)

    labrprof |       b          z      P>|z|     e^b    e^bStdX      SDofX
-------------+------------------------------------------------------------
          ed |  -0.69037    -7.115    0.000   0.5014    0.1065     3.2443
```

One MNLM

```
. mlogit occlsp ed

mlogit (N=337): Factor Change in the Odds of occlsp

Variable: ed (sd=2.9464271)

Odds comparing      |
Alternative 1       |
to Alternative 2    |       b          z      P>|z|
--------------------+------------------------------
1Labor vs 2Skilled|   -0.1711    -2.900    0.004
1Labor vs 3Prof     |   -0.7433    -8.773    0.000
3Prof  vs 2Skilled|    0.5722     7.651    0.000
---------------------------------------------------
```

Compared to the BLM results

```
                    |       b          z      P>|z|
--------------------+------------------------------
          labrskil |   -0.1840    -2.989    0.003
          labrprof |   -0.6904    -7.115    0.000
          profskil |    0.5603     7.186    0.000
---------------------------------------------------
```

Independence of irrelevant alternatives (IIA)

1. MNLM assumes that when a person chooses between choice J and choice K, the decision is not affect by other choices that are available.
   - This is the IIA assumption
   - It is why BLM can be used to estimate MNLM parameters
2. The assumption is fundamental to the model
3. Is it a weakness or a strength?
   - Opionions vary
4. Can you test the assumption?
   - Not really
5. What do you do?
   - Have choices that are distinct, nonreduncant alternatives
6. Details follow

Why IIA would fail: McFadden's buses

1. A person has two choices: $\Pr(\text{car}) = 1/2$ and $\Pr(\text{red bus}) = 1/2$

2. The odds of taking a car versus a red bus:

$$\frac{\Pr(\text{car})}{\Pr(\text{red bus})} = \frac{1/2}{1/2} = 1$$

3. A new bus company opens with identical service using blue buses.

4. IIA requires:

$$\Pr(\text{car}) = 1/3; \quad \Pr(\text{red bus}) = 1/3; \quad \Pr(\text{blue bus}) = 1/3$$

5. This maintains the odds of one

$$\frac{\Pr(\text{car})}{\Pr(\text{red bus})} = 1 = \frac{1/3}{1/3}$$

6. Substantively, we would expect a violation of IIA:

$$\Pr(\text{car}) = 1/2; \quad \Pr(\text{red bus}) = 1/4; \quad \Pr(\text{blue bus}) = 1/4$$

Does IIA make sense?

1. IIA requires that if a new choice becomes available, probabilities for prior choices adjust precisely to retain the original odds among choices
2. McFadden suggested that IIA implies that MNLM should only be used when:

   Outcome categories can plausibly be assumed to be distinct and weighed independently in the eyes of each decision maker

3. Amemiya suggested that the MNLM works well when:

   The alternatives are dissimilar

4. Specifying a model with distinct outcomes that are not substitutes for one another is reasonable, albeit ambiguous, advice.

Formal tests of IIA

1. Reviewers sometimes demand an IIA test.
   - Simulations by Cheng and Long (SMR) and other studies found that formal tests do not work well
2. Hausman and McFadden proposed a Hausman-type test of IIA
   - This compares two estimates of the same parameters.

     One estimate is consistent and efficient if the H_0 is true

     The second is consistent but inefficient
   - Cheng and Long (2006) find this tests has very poor statistical properties
3. The Small and Hsiao (1985) LR type test
   - The Small-Hsiao test works well sometimes
   - Fails completely others
   - But you can't tell when this will happen (Cheng and Long)
4. We found no test that works well in all cases.

Estimating MNLM with five outcomes

Descriptive statistics

```
occ     Occupation
white   Race: 1=white 0=nonwhite
ed      Years of education
exper   Years of work experience

-> tabulation of occ

 Occupation |      Freq.     Percent        Cum.
------------+-----------------------------------
     Menial |         31        9.20        9.20
    BlueCol |         69       20.47       29.67
      Craft |         84       24.93       54.60
   WhiteCol |         41       12.17       66.77
       Prof |        112       33.23      100.00
------------+-----------------------------------
      Total |        337      100.00

. sum white ed exper

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       white |       337    .9169139    .2764227          0          1
          ed |       337    13.09496    2.946427          3         20
       exper |       337    20.50148    13.95936          2         66
```

Output from mlogit using base(1)

1. In **mlogit** option **base(#)** sets the base or reference category.

2. Estimates for each category compared to base category

```
. mlogit occ i.white ed exper, base(1) nolog

Multinomial logistic regression                 Number of obs     =        337
                                                LR chi2(12)       =     166.09
                                                Prob > chi2       =     0.0000
Log likelihood = -426.80048                     Pseudo R2         =     0.1629

         occ |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Menial       |  (base outcome)
-------------+----------------------------------------------------------------
BlueCol      |
     1.white |   1.236504   .7244352     1.71   0.088    -.1833631    2.656371
          ed |  -.0994247   .1022812    -0.97   0.331    -.2998922    .1010428
       exper |   .0047212   .0173984     0.27   0.786    -.0293789    .0388214
       _cons |   .7412336    1.51954     0.49   0.626     -2.23701    3.719477
-------------+----------------------------------------------------------------
Craft        |
     1.white |   .4723436   .6043097     0.78   0.434    -.7120817    1.656769
          ed |   .0938154    .097555     0.96   0.336    -.0973888    .2850197
       exper |   .0276838   .0166737     1.66   0.097     -.004996    .0603636
       _cons |  -1.091353   1.450218    -0.75   0.452    -3.933728    1.751022
-------------+----------------------------------------------------------------
```

```
WhiteCol     |
     1.white |   1.571385   .9027216     1.74   0.082    -.1979166    3.340687
          ed |   .3531577   .1172786     3.01   0.003     .1232959    .5830194
       exper |   .0345959   .0188294     1.84   0.066    -.002309     .0715007
       _cons |  -6.238608   1.899094    -3.29   0.001    -9.960764   -2.516453
-------------+----------------------------------------------------------------
Prof         |
     1.white |   1.774306   .7550543     2.35   0.019     .2944273    3.254186
          ed |   .7788519   .1146293     6.79   0.000     .5541826    1.003521
       exper |   .0356509    .018037     1.98   0.048      .000299    .0710028
       _cons |  -11.51833   1.849356    -6.23   0.000      -15.143   -7.893659
------------------------------------------------------------------------------
```

3. A table for a paper might look like this...

Logit Coefficients					
Comparison		*Constant*	*WHITE*	*ED*	*EXP*
$B \mid M$	β	0.741	1.237	-0.099	0.0047
	z	0.49	1.71	-0.97	0.27
$C \mid M$	β	-1.091	0.472	0.094	0.0277
	z	-0.75	0.78	0.96	1.66
$W \mid M$	β	-6.239	1.571	0.353	0.0346
	z	-3.29	1.74	3.01	1.84
$P \mid M$	β	-11.518	1.774	0.779	0.0357
	z	-6.23	2.35	6.79	1.98

4. This table corresponds to these equations

$$\ln\Omega_{B|M}(\mathbf{x}_i) = \beta_{0,B|M} + \beta_{1,B|M}WHITE + \beta_{2,B|M}ED + \beta_{3,B|M}EXP$$
$$\ln\Omega_{C|M}(\mathbf{x}_i) = \beta_{0,C|M} + \beta_{1,C|M}WHITE + \beta_{2,C|M}ED + \beta_{3,C|M}EXP$$
$$\ln\Omega_{W|M}(\mathbf{x}_i) = \beta_{0,W|M} + \beta_{1,W|M}WHITE + \beta_{2,W|M}ED + \beta_{3,W|M}EXP$$
$$\ln\Omega_{P|M}(\mathbf{x}_i) = \beta_{0,P|M} + \beta_{1,P|M}WHITE + \beta_{2,P|M}ED + \beta_{3,P|M}EXP$$

...

Should you rely on a minimal set?

These are minimal sets of coefficients for education

Base BlueCol: 0 significant coefficients

			e^b	P>|z|
WhiteCol	vs	BlueCol	1.3978	0.720
Prof	vs	BlueCol	1.7122	0.501
Craft	vs	BlueCol	0.4657	0.227
Menial	vs	BlueCol	0.2904	0.088

Base WhiteCol: 0 significant coefficients

			e^b	P>|z|
Prof	vs	WhiteCol	1.2250	0.815
BlueCol	vs	WhiteCol	0.7154	0.720
Craft	vs	WhiteCol	0.3332	0.179
Menial	vs	WhiteCol	0.2078	0.082

Base Craft: 1 significant coefficient

			e^b	P>|z|
BlueCol	vs	Craft	2.1472	0.227
WhiteCol	vs	Craft	3.0013	0.179
Prof	vs	Craft	3.6765	0.044
Menial	vs	Craft	0.6235	0.434

Base Menial: 1 significant coefficient

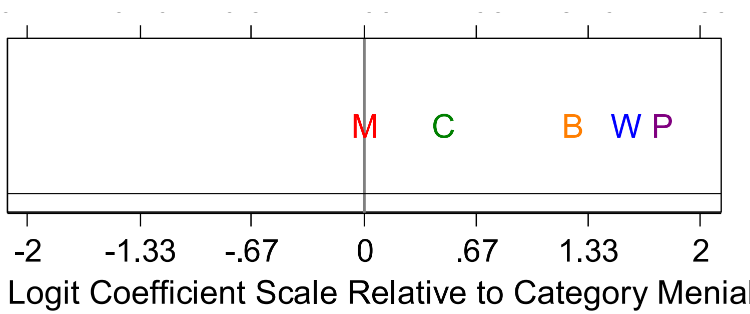
	e^b	P>|z|
Craft vs Menial	1.6037	0.434
BlueCol vs Menial	3.4436	0.088
WhiteCol vs Menial	4.8133	0.082
Prof vs Menial	5.8962	0.019

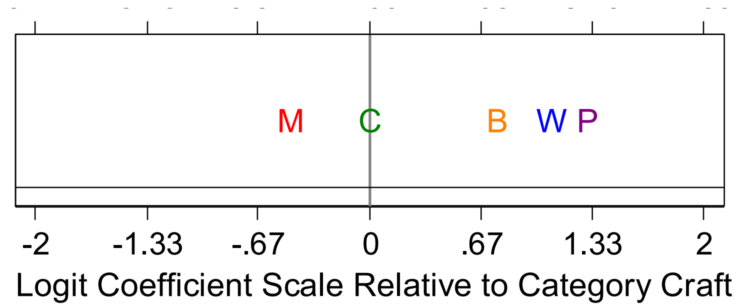
Base Prof: 2 significant coefficients

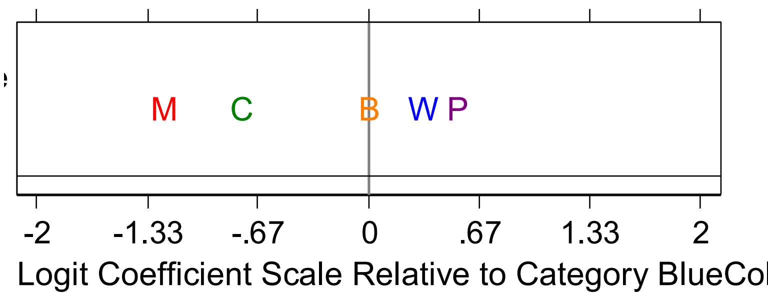
	e^b	P>|z|
WhiteCol vs Prof	0.8163	0.815
BlueCol vs Prof	0.5840	0.501
Craft vs Prof	0.2720	0.044
Menial vs Prof	0.1696	0.019

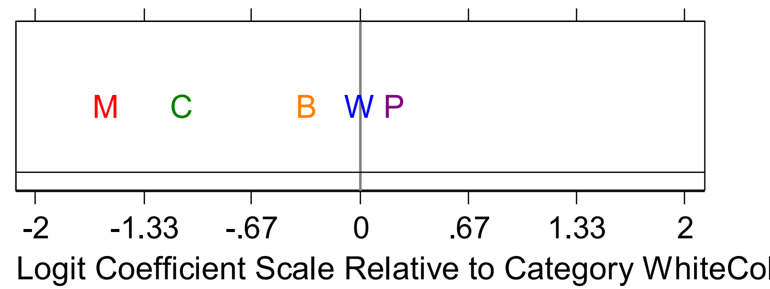
1. You should not use a minimal set to evaluate the overall significance of x_k
2. This is why minimal sets can be misleading

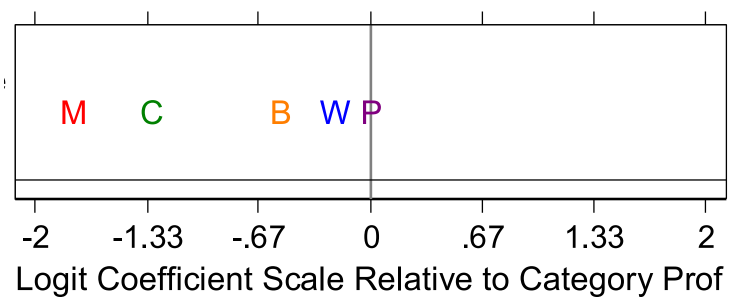
Graphs with different base categories











Looking at all ORs is overwhelming....

```
mlogit (N=337): Factor change in the odds of occ

Variable: 1.white (sd=0.276)
------------------------------------------------------------------------
                          |        b         z     P>|z|      e^b e^bStdX
--------------------------+---------------------------------------------
Menial       vs BlueCol   | -1.2365    -1.707     0.088    0.290    0.710
Menial       vs Craft     | -0.4723    -0.782     0.434    0.624    0.878
Menial       vs WhiteCol  | -1.5714    -1.741     0.082    0.208    0.648
Menial       vs Prof      | -1.7743    -2.350     0.019    0.170    0.612
BlueCol      vs Menial    |  1.2365     1.707     0.088    3.444    1.407
BlueCol      vs Craft     |  0.7642     1.208     0.227    2.147    1.235
BlueCol      vs WhiteCol  | -0.3349    -0.359     0.720    0.715    0.912
BlueCol      vs Prof      | -0.5378    -0.673     0.501    0.584    0.862
Craft        vs Menial    |  0.4723     0.782     0.434    1.604    1.139
Craft        vs BlueCol   | -0.7642    -1.208     0.227    0.466    0.810
Craft        vs WhiteCol  | -1.0990    -1.343     0.179    0.333    0.738
Craft        vs Prof      | -1.3020    -2.011     0.044    0.272    0.698
WhiteCol     vs Menial    |  1.5714     1.741     0.082    4.813    1.544
WhiteCol     vs BlueCol   |  0.3349     0.359     0.720    1.398    1.097
WhiteCol     vs Craft     |  1.0990     1.343     0.179    3.001    1.355
WhiteCol     vs Prof      | -0.2029    -0.233     0.815    0.816    0.945
Prof         vs Menial    |  1.7743     2.350     0.019    5.896    1.633
Prof         vs BlueCol   |  0.5378     0.673     0.501    1.712    1.160
Prof         vs Craft     |  1.3020     2.011     0.044    3.677    1.433
Prof         vs WhiteCol  |  0.2029     0.233     0.815    1.225    1.058
------------------------------------------------------------------------
```

Variable: ed (sd=2.946)

		b	z	P>|z|	e^b	e^bStdX
Menial	vs BlueCol	0.0994	0.972	0.331	1.105	1.340
Menial	vs Craft	-0.0938	-0.962	0.336	0.910	0.758
Menial	vs WhiteCol	-0.3532	-3.011	0.003	0.702	0.353
Menial	vs Prof	-0.7789	-6.795	0.000	0.459	0.101
BlueCol	vs Menial	-0.0994	-0.972	0.331	0.905	0.746
BlueCol	vs Craft	-0.1932	-2.494	0.013	0.824	0.566
BlueCol	vs WhiteCol	-0.4526	-4.425	0.000	0.636	0.264
BlueCol	vs Prof	-0.8783	-8.735	0.000	0.415	0.075
Craft	vs Menial	0.0938	0.962	0.336	1.098	1.318
Craft	vs BlueCol	0.1932	2.494	0.013	1.213	1.767
Craft	vs WhiteCol	-0.2593	-2.773	0.006	0.772	0.466
Craft	vs Prof	-0.6850	-7.671	0.000	0.504	0.133
WhiteCol	vs Menial	0.3532	3.011	0.003	1.424	2.831
WhiteCol	vs BlueCol	0.4526	4.425	0.000	1.572	3.794
WhiteCol	vs Craft	0.2593	2.773	0.006	1.296	2.147
WhiteCol	vs Prof	-0.4257	-4.616	0.000	0.653	0.285
Prof	vs Menial	0.7789	6.795	0.000	2.179	9.923
Prof	vs BlueCol	0.8783	8.735	0.000	2.407	13.300
Prof	vs Craft	0.6850	7.671	0.000	1.984	7.526
Prof	vs WhiteCol	0.4257	4.616	0.000	1.531	3.505

```
Variable: exper (sd=13.959)
------------------------------------------------------------------------
                          |        b         z     P>|z|      e^b e^bStdX
--------------------------+---------------------------------------------
Menial       vs BlueCol   | -0.0047    -0.271     0.786    0.995    0.936
Menial       vs Craft     | -0.0277    -1.660     0.097    0.973    0.679
Menial       vs WhiteCol  | -0.0346    -1.837     0.066    0.966    0.617
Menial       vs Prof      | -0.0357    -1.977     0.048    0.965    0.608
BlueCol      vs Menial    |  0.0047     0.271     0.786    1.005    1.068
BlueCol      vs Craft     | -0.0230    -1.829     0.067    0.977    0.726
BlueCol      vs WhiteCol  | -0.0299    -1.954     0.051    0.971    0.659
BlueCol      vs Prof      | -0.0309    -2.147     0.032    0.970    0.649
Craft        vs Menial    |  0.0277     1.660     0.097    1.028    1.472
Craft        vs BlueCol   |  0.0230     1.829     0.067    1.023    1.378
Craft        vs WhiteCol  | -0.0069    -0.495     0.621    0.993    0.908
Craft        vs Prof      | -0.0080    -0.627     0.531    0.992    0.895
WhiteCol     vs Menial    |  0.0346     1.837     0.066    1.035    1.621
WhiteCol     vs BlueCol   |  0.0299     1.954     0.051    1.030    1.517
WhiteCol     vs Craft     |  0.0069     0.495     0.621    1.007    1.101
WhiteCol     vs Prof      | -0.0011    -0.07   3    0.941    0.999    0.985
Prof         vs Menial    |  0.0357     1.977     0.048    1.036    1.645
Prof         vs BlueCol   |  0.0309     2.147     0.032    1.031    1.540
Prof         vs Craft     |  0.0080     0.627     0.531    1.008    1.118
Prof         vs WhiteCol  |  0.0011     0.073     0.941    1.001    1.015
------------------------------------------------------------------------
```

Our task is to make sense out of all of these numbers

Roadmap

1. MNLM as a probability model.
2. Fitting the model with MLE.
3. Omnibus test of a regressor.
4. Tests if categories can be combined.
5. Interpretation
   - Probabilities
   - Marginal effects
   - Odds ratios – with reservations

MNLM as a Probability Model

1. We can solve the J-1 equations for $\log\left[\Pr\left(y_i = m \mid \mathbf{x}_i\right) / \Pr\left(y_i = B \mid \mathbf{x}_i\right)\right]$:

$$\Pr\left(y_i = m \mid \mathbf{x}_i\right) = \frac{\exp\left(\mathbf{x}_i \boldsymbol{\beta}_{m|J}\right)}{\sum_{j=1}^{J} \exp\left(\mathbf{x}_i \boldsymbol{\beta}_{j|J}\right)}$$

2. You get the same values regardless of the base J you use.

ML estimation

1. $\Pr(y=m|x_i)$ is the probability of observing outcome m for person i:
   - If person i choses outcome *j*, then $p_i = \Pr(y=j|x_i)$
2. If observations are independent, the likelihood equation is:

$$L\left(\boldsymbol{\beta}_{2|J}, \ldots, \boldsymbol{\beta}_{J|J} \mid \mathbf{y}, \mathbf{X}\right) = \prod_{i=1}^{N} p_i$$

3. Solving for the parameters works well even with small samples.

Test that a variable has no effect

1. The hypothesis that x_k has no effect involves J-1 coefficients:

$$H_0: \beta_{k,B|M} = \beta_{k,C|M} = \beta_{k,W|M} = \beta_{k,P|M} = 0$$

2. This is not equivalent to combined tests of individual coefficients

$$H_0: \beta_{k,B|M} = 0 \quad H_0: \beta_{k,C|M} = 0 \quad H_0: \beta_{k,W|M} = 0 \quad H_0: \beta_{k,P|M} = 0$$

3. As with BRM, testing βs does not replace tests of marginal effects

Wald test that all coefficients for x_k are 0:

$W_k = \hat{\boldsymbol{\beta}}_k' Var\left(\hat{\boldsymbol{\beta}}_k\right)^{-1} \hat{\boldsymbol{\beta}}_k$ where $W_k \sim \chi^2_{J-1}$ if H_0 is true

LR test that all coefficients for x_k are 0: s

1. Estimate G^2_{Full} for full model, and $G^2_{\text{Restricted}}$ for $M_{\text{Restricted}}$ by excluding x_k
2. $M_{\text{Restricted}}$ has $J-1$ fewer parameters. The LR test is

$G^2_{\text{Restricted\_Full}} = G^2_{\text{Full}} - G^2_{\text{Restricted}} \sim \chi^2_{J-1}$ if H_0 is true.

Wald tests using mlogtest

```
. mlogtest, wald

Wald tests for independent variables (N=337)

  Ho: All coefficients associated with given variable(s) are 0

                 |       chi2     df    P>chi2
-----------------+-------------------------
         1.white |      8.149      4     0.086
              ed |     84.968      4     0.000
           exper |      7.995      4     0.092
```

The effect of race is not significant at the .05 level (G2=8.15, df=4).

The effect of education is significant at the .01 level.

The effect of experience is significant at the .10 level but not at the .05 level.

Wald test of regression coefficients using test

```
. test 1.white // test will work with svy estimates too
. * 1.white is the variable created from i.white!

 ( 1)   [Menial]o.white = 0          ← This is β white,M|M
 ( 2)   [BlueCol]white = 0
 ( 3)   [Craft]white = 0
 ( 4)   [WhiteCol]white = 0
 ( 5)   [Prof]white = 0
        Constraint 1 dropped

           chi2(  4) =    8.15
         Prob > chi2 =    0.0863

. test ed
  <snip>

. test exper
  <snip?
```

LR test using mlogtest

```
. estimates restore full
(results full are active now)

. mlogtest, lr

Likelihood-ratio tests for independent variables (N=337)

                 |      chi2    df   P>chi2
-----------------+-------------------------
         1.white |     8.095     4    0.088
              ed |   156.937     4    0.000
           exper |     8.561     4    0.073
```

LR tests using lrtest

```
. quietly mlogit occ white ed exper, base(1) // full model
. estimates store full

. quietly mlogit occ ed exper, base(1) // restricted model
. estimates store dropwhite

. lrtest full dropwhite

Likelihood-ratio test                          LR chi2(4)  =      8.10
(Assumption: dropwhite nested in full)         Prob > chi2 =    0.0881
```

Test if outcomes can be combined

1. If all $\beta_{k,P|W}$ are 0, then P and W are indistinguishable.

$$H_0: \beta_{1,P|W} = \beta_{2,P|W} = \beta_{3,P|W} = 0$$

2. Why test if outcomes can be combined?
3. Substantively
   - You want to know if professional and white collar jobs are differentiated by the predictors in the model
4. To simplify the model
   - You have a five point scale of party and want to combine

     Strong Democrat and Democrat

     Strong Republican and Republican
5. Decisions to combine categories must be made sequentially as now shown.

Tests for combining can lead to inconsistencies

1. Tests of hypothesis are not algebraic statements.
1. A test for combining M and B is not significant (p=.251)

   M can be combined with B ***M = B***

2. A test for combining M and C is not significant (p=.337)

   M can be combined with C ***M = C***

3. If ***M = B*** and ***M = C***, then algebraically ***B = C***
4. Yet a test for combining B and C is significant (p=.003)

   B cannot be combined with C ***B ≠ C***

5. If you decide to combine categories based on testing
   - Combine only two categories for a set of tests
   - Fit the model with the new set of outcomes
   - Test for indistinguishability in the new model

Wald tests for combining outcomes

```
. mlogtest, combine

Wald tests for combining alternatives (N=337)

Ho: All coefficients except intercepts associated with a given pair
    of alternatives are 0 (i.e., alternatives can be combined)

                   |        chi2     df     P>chi2
-------------------+--------------------------------
  Menial & BlueCol |       3.994      3     0.262
    Menial & Craft |       3.203      3     0.361
 Menial & WhiteCol |      11.951      3     0.008
     Menial & Prof |      48.190      3     0.000
   BlueCol & Craft |       8.441      3     0.038
BlueCol & WhiteCol |      20.055      3     0.000
    BlueCol & Prof |      76.393      3     0.000
  Craft & WhiteCol |       8.892      3     0.031
      Craft & Prof |      60.583      3     0.000
   WhiteCol & Prof |      22.203      3     0.000
```

Or use **`mlogtest, lrcomb`** for LR tests

Specification searches

1. Tests for combining categories and tests that all coefficients for a variable are zero can be used in a specification search.
2. Think substantively about changes to your model.
   - Do not to <u>over-fit</u> your data
   - Examine individual coefficients before revising your model
3. In models obtained from tests using the same data, significance levels are invalid.
   - Consider randomly dividing the sample into an <u>exploration sample</u> and a <u>verification sample</u>

Overview of interpretation

1. The MNLM has many parameters.
2. Do not:
   - Look only at a minimal set of parameters
   - Discuss only the stars and signs of coefficients
3. Interpretation should include
   - Predicted probabilities
   - Marginal effects
   - Odds ratios
4. Only some methods are shown in this lecture
   - Methods from the ordinal lecture that do not depend on the ordinality can be used for nominal outcomes

Example: Attitudes toward working mothers

Outcome

Working mothers can have a warm relationship with their children?

1. Responses in variable **warm** are

```
                     |      Freq.     Percent
---------------------+---------------------------
1 Strongly Disagree  |        297       12.95
2 Disagree           |        723       31.53
3 Agree              |        856       37.33
4 Strongly agree     |        417       18.19
---------------------+---------------------------
               Total |      2,293      100.00
```

2. Remember:

   Agree: support for working mothers

   Disagree: lack of support for working moms

Regressors

```
. codebook yr89 male white age prst, compact

Variable     Obs Unique      Mean  Min  Max  Label
-----------------------------------------------------------------------------
yr89        2293      2  .3986044    0    1  Survey year: 1=1989 0=1977
male        2293      2  .4648932    0    1  Gender: 1=male 0=female
white       2293      2  .8765809    0    1  Race: 1=white 0=not white
age         2293     72  44.93546   18   89  Age in years
prst        2293     58  39.58526   12   82  Occupational prestige
-----------------------------------------------------------------------------

. tab edcat

   Years of |
  education |
  in groups |      Freq.     Percent        Cum.
------------+-----------------------------------
   0-11 yrs |        670       29.22       29.22
     12 yrs |        782       34.10       63.32
  13-15 yrs |        449       19.58       82.90
  16-20 yrs |        392       17.10      100.00
------------+-----------------------------------
      Total |      2,293      100.00
```

Fitting the model

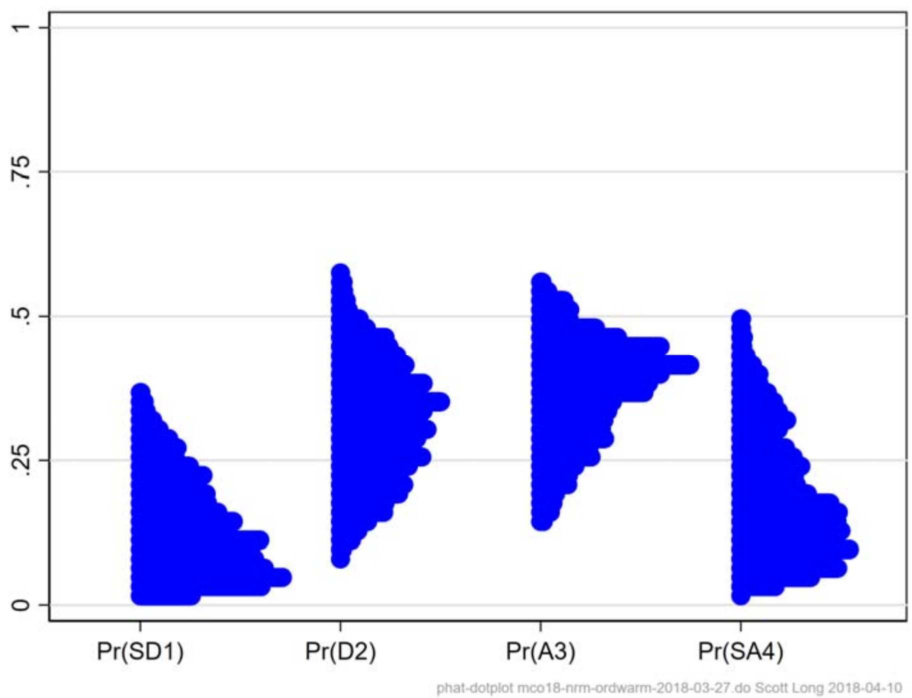
1. Treating **warm** as <u>nominal</u> and fit the model

```
mlogit warm i.yr89 i.male i.white age i.edcat prst, base(1)
```

Predictions

I start by looking for anomalies and outliers in the predictions and find none.

```
predict SDpr1 Dpr2 Apr3 SApr4
```



Average marginal effects -- we need a plot!

1. ADC(x_k) is the counterpart to β_ks in LRM
2. It produces too much information to easily absorb

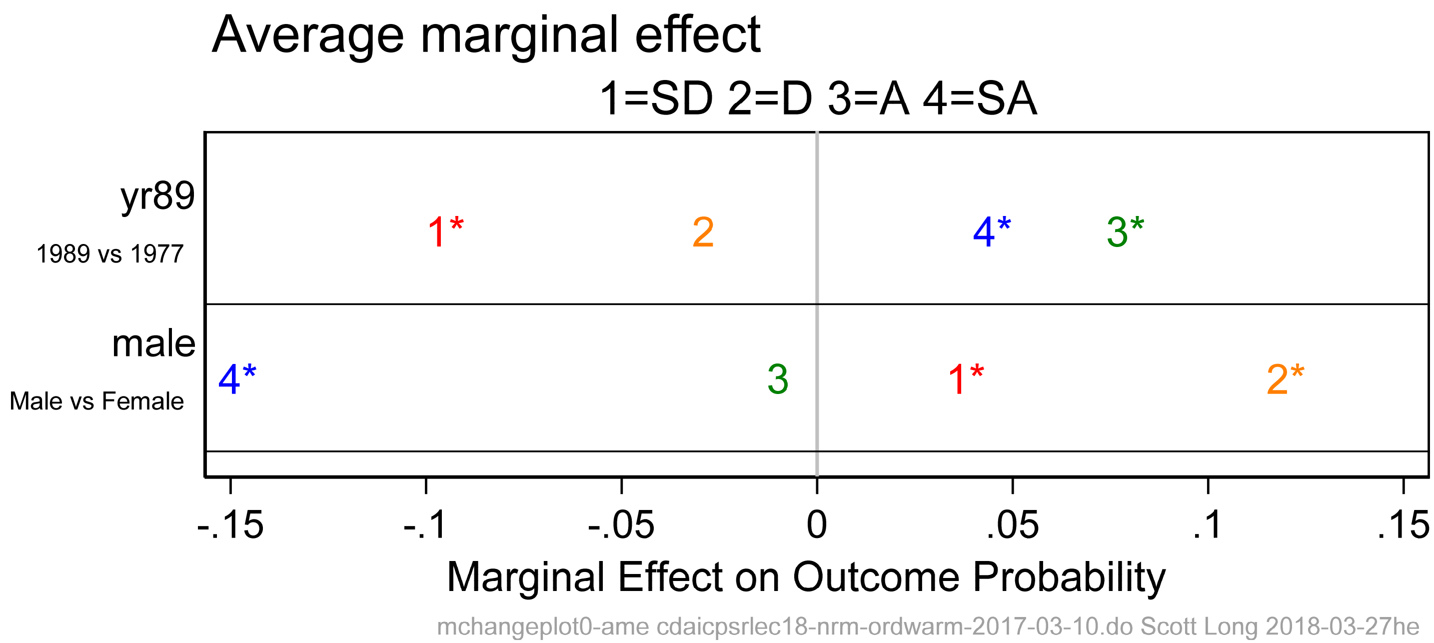
```
. mchange, amount(one sd) // no marginal changes

mlogit: Changes in Pr(y) | Number of obs = 2293

Expression: Pr(warm), predict(outcome())

                        |       1 SD         2 D         3 A        4 SA
------------------------+------------------------------------------------
yr89                    |
           1989 vs 1977 |     -0.095      -0.029       0.079       0.045
                p-value |      0.000       0.136       0.000       0.006
male                    |
         Male vs Female |      0.038       0.120      -0.010      -0.148
                p-value |      0.006       0.000       0.622       0.000
white                   |
      White vs NonWhite |      0.040       0.018      -0.006      -0.052
                p-value |      0.036       0.529       0.838       0.038
::
```

On average being male decrease the probability of strong support for working mothers by .15 (p<.001).

	1 SD	2 D	3 A	4 SA
age				
+1	0.002	0.004	-0.003	-0.003
p-value	0.000	0.000	0.000	0.000
+SD	0.037	0.059	-0.054	-0.042
p-value	0.000	0.000	0.000	0.000
edcat				
12 yrs vs 0-11 yrs	-0.012	-0.004	0.028	-0.012
p-value	0.535	0.880	0.305	0.580
13-15 yrs vs 0-11 yrs	-0.056	-0.042	0.053	0.045
p-value	0.013	0.171	0.104	0.084
16-20 yrs vs 0-11 yrs	-0.090	-0.015	0.050	0.055
p-value	0.000	0.694	0.198	0.093
13-15 yrs vs 12 yrs	-0.044	-0.038	0.025	0.057
p-value	0.029	0.167	0.380	0.011
16-20 yrs vs 12 yrs	-0.078	-0.011	0.022	0.066
p-value	0.000	0.745	0.510	0.017
16-20 yrs vs 13-15 yrs	-0.034	0.028	-0.003	0.009
p-value	0.096	0.415	0.934	0.746
prst				
+1	0.000	-0.002	0.001	0.001
p-value	0.818	0.011	0.193	0.220
+SD	0.002	-0.029	0.016	0.012
p-value	0.841	0.009	0.214	0.240

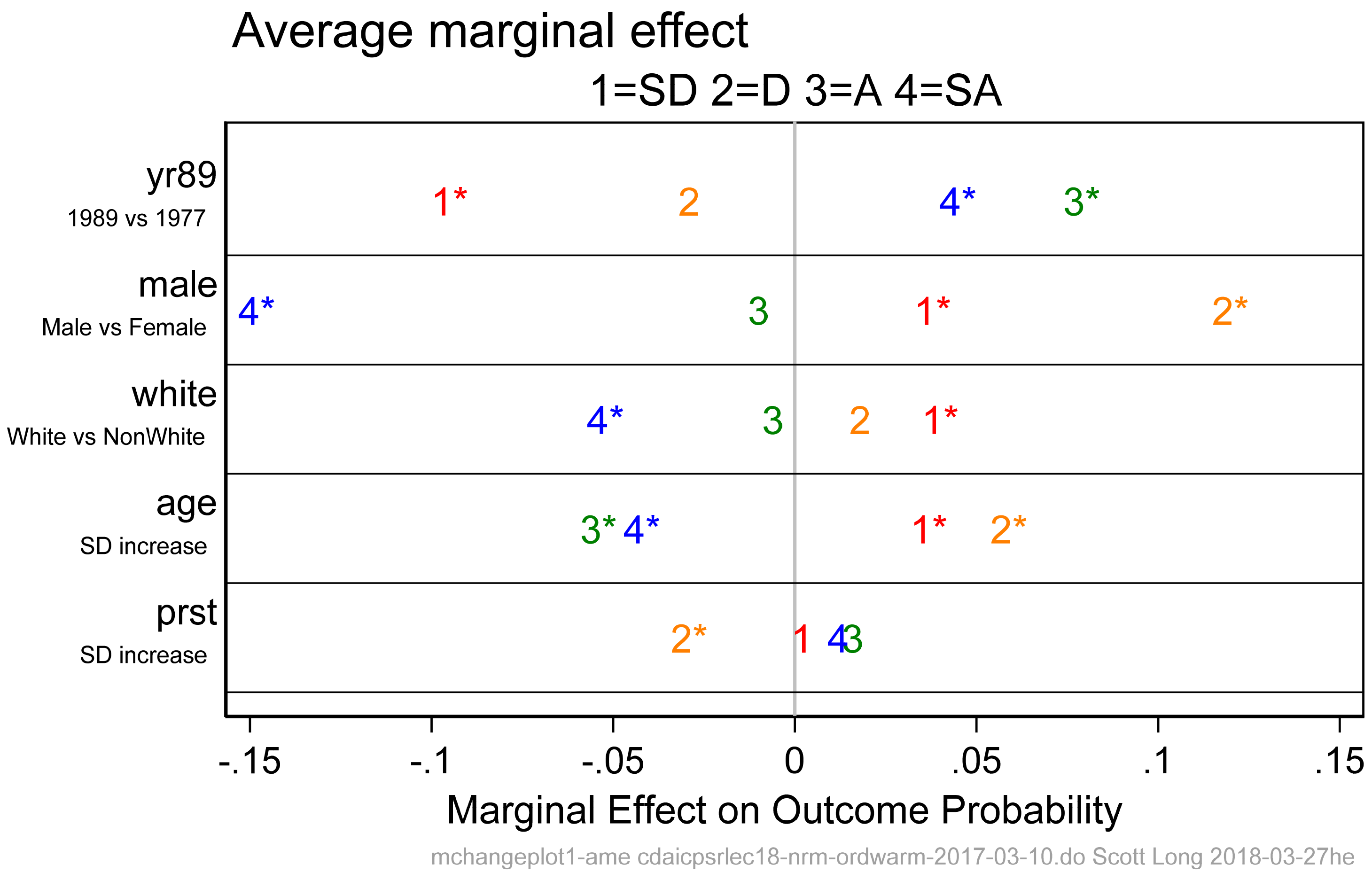
Let’s plot the effects.

Average marginal effects: two critical variables



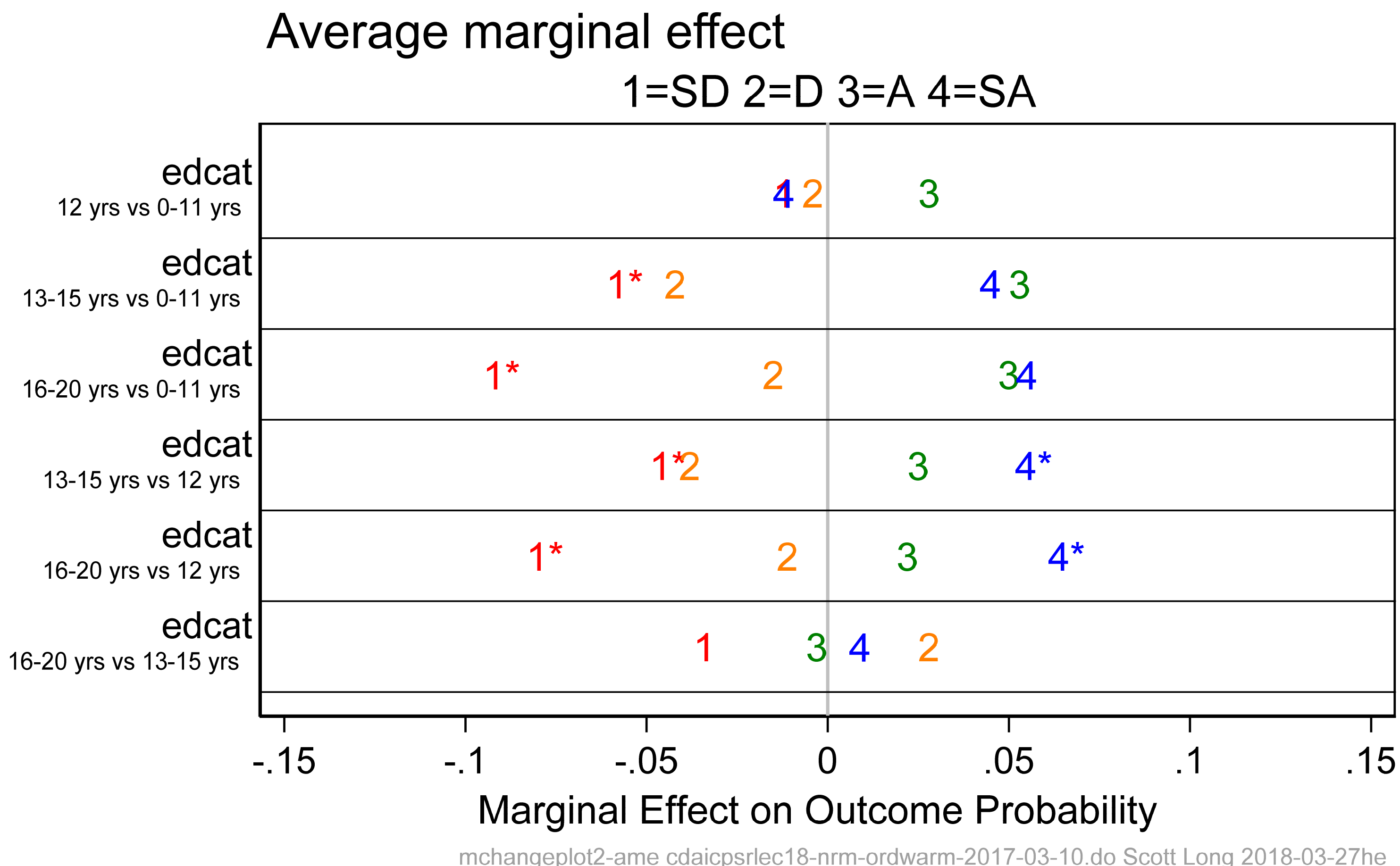
Men are more likely to have negative attitudes toward working women than are women. On average men are .15 less likely (p<.01) to strongly support working women and .12 more likely (p<.01) to be negative. From 1977 to 1989 attitudes have become more positive, with significant increases of .05 in strong support and .08 in support, with a decrease of .10 in the most negative attitudes.

Average marginal effects of continuous and binary regressors



1. Why are the colors ordered red – orange – green – blue used?
2. Is this consistent with warm being ordinal?

Average marginal effects of education



Test of βs

p=.743

p=.012

p=.002

p=.008

p=.001

p=.412

1. Compare ME's to test all βs are 0.
2. Effects vary by outcome and contrasts.

Code for mchangeplot

1. After running **mchange**, create a quick graph

```
mchangeplot
```

2. Create a customized graph:

```
mchangeplot yr89 male white age prst, /// regressors to plot
    min(-.15) max(.15) gap(.05) /// tick marks on axis
    mcol(red orange green blue) /// colors of symbols
    title(Average marginal effect, position(11)) ///
    subtitle(1=SD 2=D 3=A 4=SA) ///
    sig(.05) /// add * if sig at .05 level
    aspect(.5) leftmargin(10) // tweak shape of graph
```

Tests of regressors

1. We can test that the 4 coefficients for each regressor are simultaneously 0.
   - 2.edcat, 3.edcat, and 4.edcat are contrasts with 1.edcat

```
. mlogtest, wald

  Ho: All coefficients associated with given variable(s) are 0

                 |       chi2     df   P>chi2
-----------------+-------------------------
         1.yr89 |     54.503      3    0.000
         1.male |    100.836      3    0.000
        1.white |      7.638      3    0.054
            age |     86.556      3    0.000
        2.edcat |      1.241      3    0.743      <= 12    vs 0-11
        3.edcat |     10.994      3    0.012      <= 13-15 vs 0-11
        4.edcat |     15.119      3    0.002      <= 16-20 vs 0-11
           prst |      6.901      3    0.075
```

The effect of gender is significant (X2=100.8, df=3, p<.001).

Prestige does not significantly affect attitudes toward working women.

2. I ignore these and focus on marginal effects.

Tests for combining categories

1. To simplify analysis, I consider dichotomizing into positive vs negative attitudes.

```
. mlogtest, combine

Wald tests for combining alternatives (N=2293)

 Ho: All coefficients except intercepts associated with a given pair
     of alternatives are 0 (i.e., alternatives can be combined)

                 |       chi2     df   P>chi2
-----------------+-------------------------
      1_SD & 2_D |     38.245      8    0.000
      1_SD & 3_A |    134.132      8    0.000
     1_SD & 4_SA |    185.858      8    0.000
       2_D & 3_A |     95.727      8    0.000
      2_D & 4_SA |    172.166      8    0.000
      3_A & 4_SA |     53.660      8    0.000
```

2. The evidence does not support combining categories.

Example: Political orientation

1. 1992 American National Election Study
2. I was given this data as an example for ordinal logit.

Outcome: party affiliation

```
. use partyid4, clear
(partyid4.dta | 1992 American National Election Study | 2014-03-12)

. tab party, miss

   Party ID |      Freq.     Percent        Cum.
------------+-----------------------------------
D    StrDem |        266       19.25       19.25
d       Dem |        427       30.90       50.14
i     Indep |        151       10.93       61.07
r       Rep |        369       26.70       87.77
R    StrRep |        169       12.23      100.00
------------+-----------------------------------
      Total |      1,382      100.00
```

Regressors

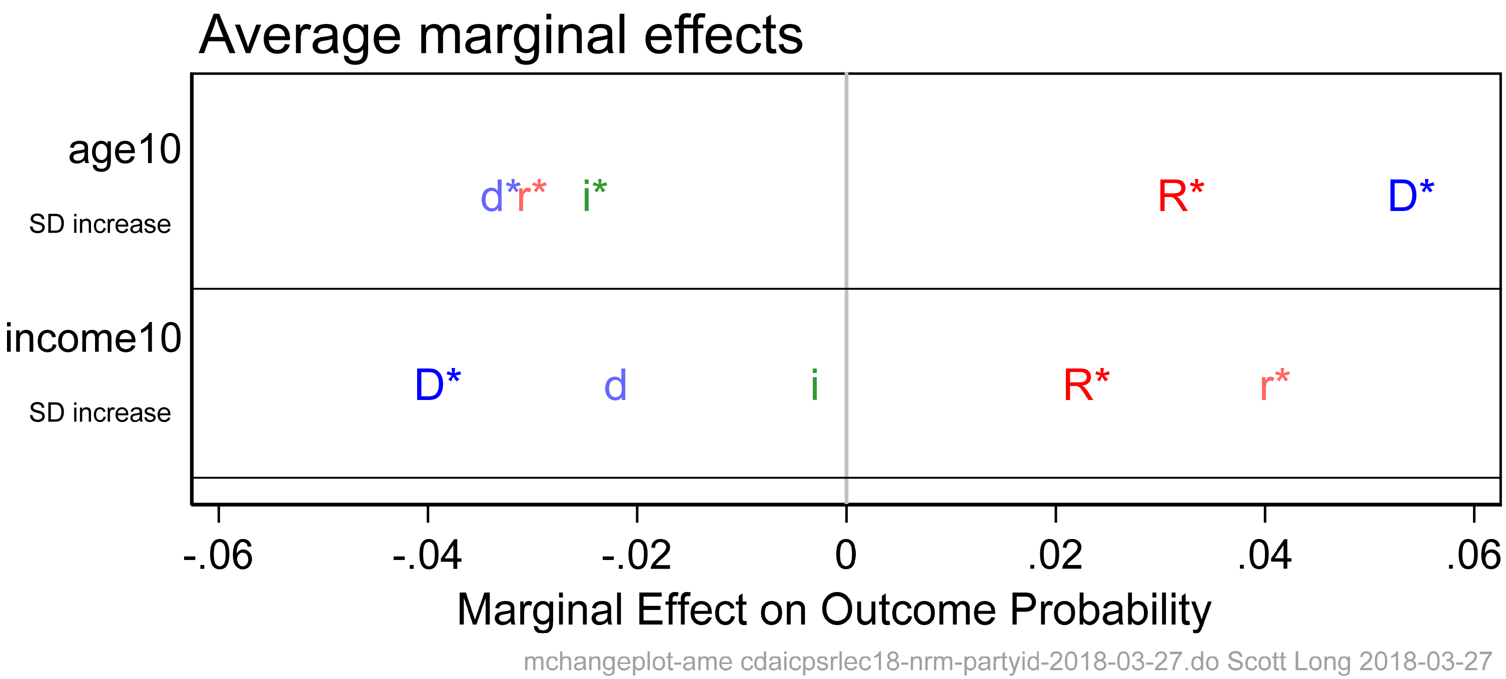
```
party   Party ID
age     Age
income  Income in $1,000s
black   Black?
female  Female?
educ    Level of education versus "not hs grad"

. sum party age income black female i.educ

   Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      party |       1382    2.817656    1.342787          1          5
        age |       1382    45.94645    16.78311         18         91
     income |       1382    37.45767    27.78148        1.5     131.25
      black |       1382    .1374819      .34448          0          1
     female |       1382    .4934877    .5001386          0          1
       educ |
   hs only  |       1382    .5803184    .4936854          0          1
   college  |       1382    .2590449    .4382689          0          1
```

"10" versions of variables divide age and income by 10.

Fit MNLM and test regressors

```
. mlogit party age10 income10 i.black i.female i.educ
::
. mlogtest, lr set(educ_set=1.highschool 1.college)

LR tests for independent variables (N=1382)

  Ho: All coefficients associated with given variable(s) are 0

                 |       chi2    df   P>chi2
-----------------+-------------------------
           age10 |     45.165     4    0.000
        income10 |     24.361     4    0.000
         1.black |    126.467     4    0.000
        1.female |      9.143     4    0.058
    1.highschool |      5.567     4    0.234
       1.college |     21.582     4    0.000
        educ_set |     26.881     8    0.001

   educ_set contains: 1.highschool 1.college
```

Average marginal effects (AME)



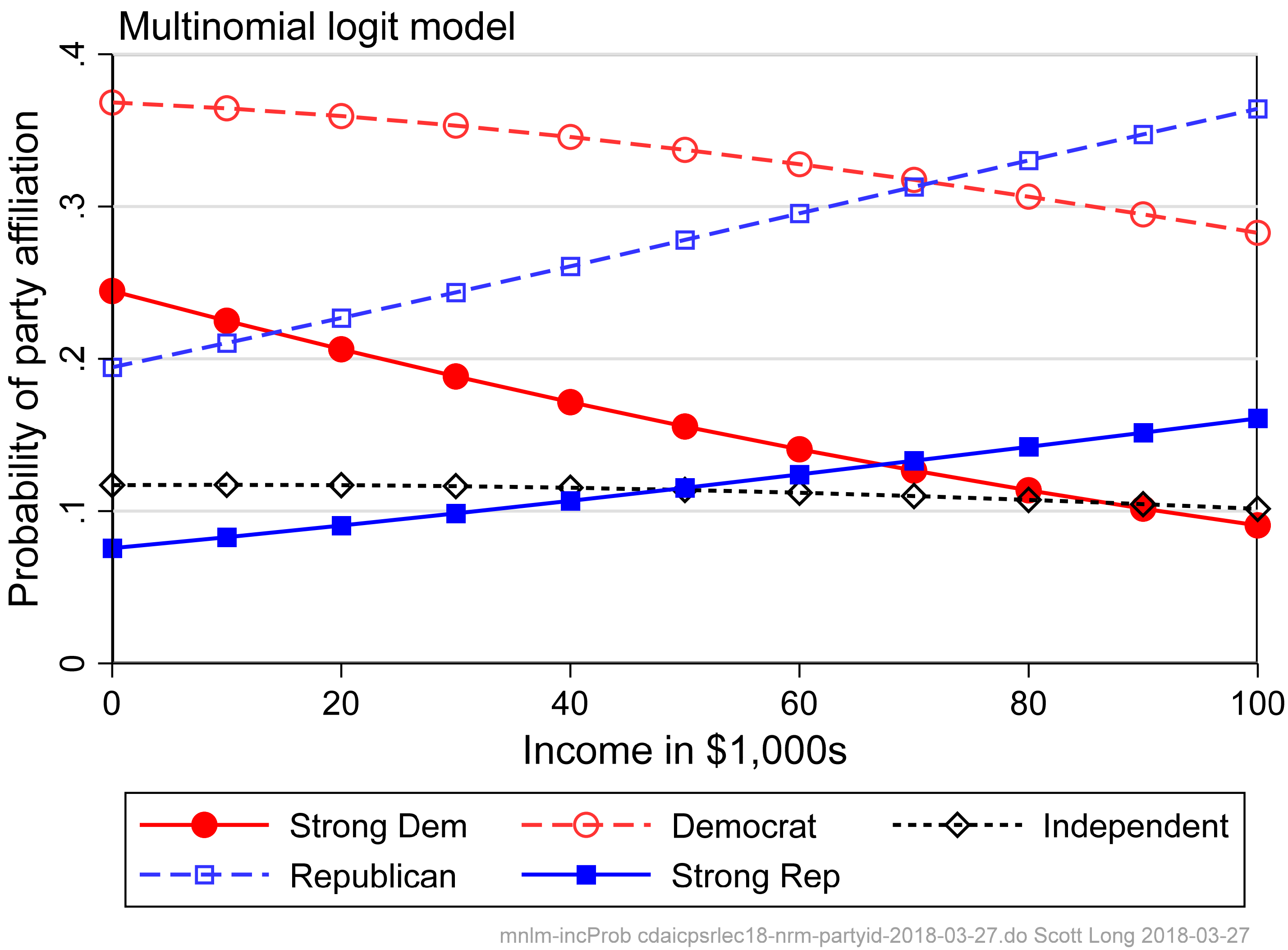
Age increases the probability of the extreme affiliations and decreases the probability of less extreme affiliations.

Income increase Republican affiliations, while decreasing Democratic affiliations.

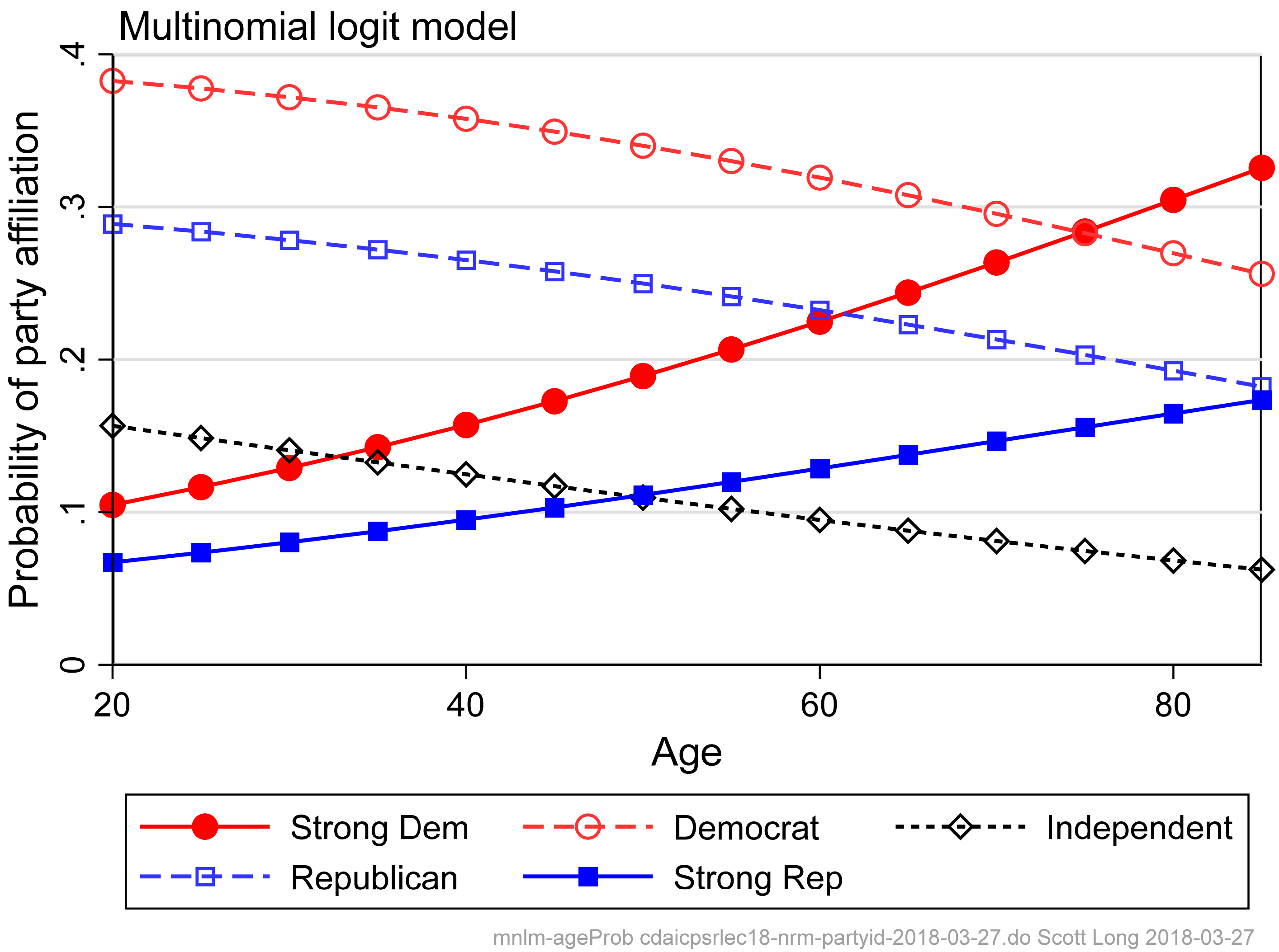
Plots of probabilities

1. For continuous regressors, plots can be useful.
2. One variable changes while others are constant.
   - Unless variables are linked, like age and age-squared
3. I create two plots:
   - Age changes while holding other variables at their means
   - Income changes holding other variables at their means
4. The probability lines are nearly straight so that the AMEs or DCMs would be good summaries.
   - The only way to know if a plot is useful is to create it

Probabilities by income



Probabilities by age



Commands for plots

Generate variables to plot

```
. mgen, atmeans at(age10=(2(.5)8.5)) stub(MNLMage)

Predictions from: margins, atmeans at(age10=(2(.5)8.5)) predict(outcome(5))

Variable        Obs Unique      Mean       Min       Max  Label
----------------------------------------------------------------------------
MNLMagepr1       14     14  .2046258  .1047786  .3256354  pr(y=SD) from margins
MNLMagell1       14     14  .1675598  .0768825  .2487619  95% lower limit
MNLMageul1       14     14  .2416917  .1326748  .4025089  95% upper limit
MNLMageage10     14     14      5.25         2       8.5  Age in decades
MNLMageCpr1      14     14  .2046258  .1047786  .3256354  pr(y<=SD)
MNLMagepr2       14     14   .329003  .2561781  .3826132  pr(y=D) from margins
MNLMagell2       14     14  .2874055  .1949605  .3331204  95% lower limit
MNLMageul2       14     14  .3706006  .3173957   .435444  95% upper limit
MNLMageCpr2      14     14  .5336288  .4873918  .5818136  pr(y<=D)
<snip>

Specified values of covariates

                         1.          1.          2.          3.
 income10        black      female        educ        educ
------------------------------------------------------
 3.745767     .1374819    .4934877    .5803184    .2590449

. mgen, atmeans at(income10=(0(1)10)) stub(MNLMinc)
::
```

Customize variables and options for plots

```
gen PLTincome = MNLMincincome10*10
label var PLTincome "Income in $1,000s"
gen PLTage = MNLMageage10*10
label var PLTage "Age"
local yaxis_p "ytitle(Probability of party affiliation)"
local yaxis_p "`yaxis_p' ylab(0(.1).4, grid) ylin(0 .4, lcol(gs14))"
local yaxis_c ///
  "ytitle(Cumulative probability) ylab(0(.2)1, grid) ylin(0 1, lcol(gs14))"
local titleopt "position(11) size(medium)"

* line options for probabilities
local line5_opts "msym(O Oh dh sh s)"
local line5_opts "`line5_opts' lwid(*1.3 *1.3 *1.3 *1.3 *1.3)"
local line5_opts "`line5_opts' lpat(solid dash shortdash dash solid)"
local line5_opts "`line5_opts' mcol(red red*.8 black blue*.8 blue)"
local line5_opts "`line5_opts' lcol(red red*.8 black blue*.8 blue)"

label var MNLMagepr1 "Strong Dem"
label var MNLMagepr2 "Democrat"
label var MNLMagepr3 "Independent"
label var MNLMagepr4 "Republican"
label var MNLMagepr5 "Strong Rep"

label var MNLMincpr1 "Strong Dem"
label var MNLMincpr2 "Democrat"
label var MNLMincpr3 "Independent"
label var MNLMincpr4 "Republican"
label var MNLMincpr5 "Strong Rep"
```

Create plots

```
* probability by age
graph twoway (connected MNLMagepr1 MNLMagepr2 MNLMagepr3 ///
    MNLMagepr4 MNLMagepr5 PLTage, `line5_opts'), ///
    title("`title'", `titleopt') `yaxis_p' `xaxis_age' ///
    legend(rows(2))

* probability by income
graph twoway (connected MNLMincpr1 MNLMincpr2 MNLMincpr3 ///
    MNLMincpr4 MNLMincpr5 PLTincome, `line5_opts'), ///
    title("`title'", `titleopt') `yaxis_p' `xaxis_inc' ///
    legend(rows(2))
```

See Long and Freese for lots of details!

Test IIA with a Hausman test

1. If IIA holds, the coefficient from binary logit:

```
. gen party15 = 1 if party==1
. replace party15 = 0 if party==5

. logit party15 age, nolog

     party15 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0102276    .005633     1.82   0.069    -.0008129    .0212681
```

2. Should be close to those from the MNLM:

```
. mlogit party age, base(5)

       party |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
StrDem       |
         age |   .0104532   .0056788     1.84   0.066    -.0006771    .0215834
::
-------------+----------------------------------------------------------------
StrRep       |  (base outcome)
------------------------------------------------------------------------------
```

3. IIA tests formally assess coefficients from alternative specifications

Hausman test

```
. mlogtest, hausman

Hausman tests of IIA assumption (N=1382)

  Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

                 |       chi2    df   P>chi2
-----------------+-------------------------
          StrDem |      4.622    20    1.000
             Dem |      0.919    21    1.000
           Indep |     -2.244    19        .
             Rep |      3.030    21    1.000
          StrRep |     -0.580    21        .

  Note: A significant test is evidence against Ho.
```

1. The Hausman test often produces negative chi-squares.
2. In simulations, its size properties are very bad.

Small-Hsiao

1. Randomly divides the samples and compares estimates (roughly speaking)
   - Random numbers are used to divide the sample

Small-Hsiao: Seed 124386

1. It appears that IIA is supported

```
. mlogtest, smhsiao

Small-Hsiao tests of IIA assumption (N=1382)

  Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

                 | lnL(full)  lnL(omit)       chi2    df   P>chi2
-----------------+-----------------------------------------------
          StrDem |  -696.753   -690.654     12.198    21    0.934
             Dem |  -565.571   -557.488     16.166    21    0.760
           Indep |  -764.563   -758.290     12.547    21    0.924
             Rep |  -621.562   -615.492     12.140    21    0.936
          StrRep |  -761.598   -752.804     17.587    21    0.675

  Note: A significant test is evidence against Ho.
```

Seed 254331

1. It appears IIA is not supported

```
. mlogtest, smhsiao

Small-Hsiao tests of IIA assumption (N=1382)

  Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

                 | lnL(full)  lnL(omit)       chi2     df   P>chi2
-----------------+------------------------------------------------
          StrDem |  -727.367   -692.048     70.639     21    0.000
             Dem |  -610.636   -573.268     74.736     21    0.000
           Indep |  -783.456   -747.654     71.604     21    0.000
             Rep |  -650.962   -615.434     71.057     21    0.000
          StrRep |  -751.887   -740.193     23.388     21    0.324

  Note: A significant test is evidence against Ho.
```

* Odds ratio plots

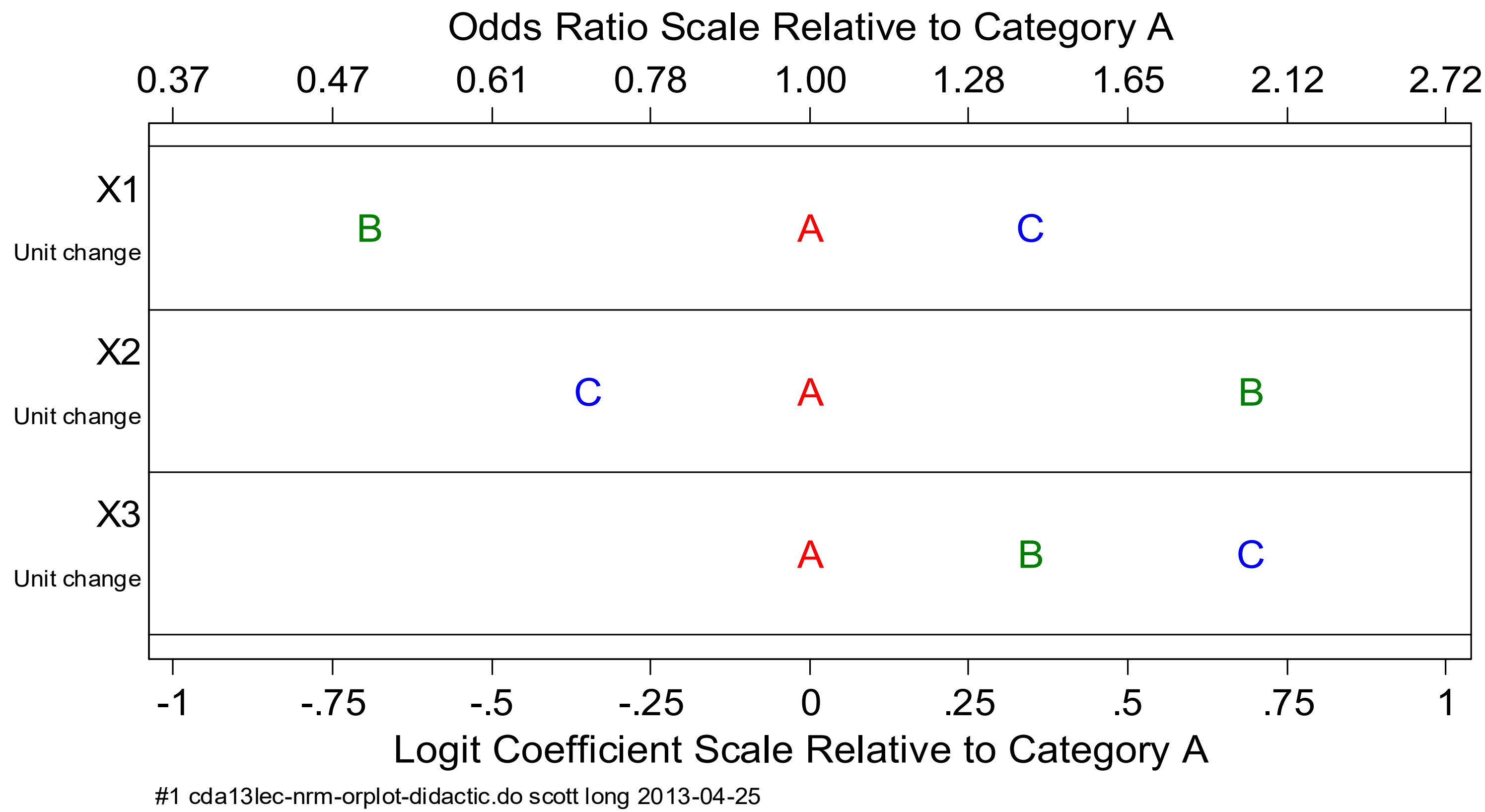
The additive links among parameters

1. Because of IIA:

$$\ln\left[\frac{\Pr(L \mid Ed)}{\Pr(P \mid Ed)}\right] = \ln\left[\frac{\Pr(L \mid Ed)}{\Pr(S \mid Ed)}\right] + \ln\left[\frac{\Pr(S \mid Ed)}{\Pr(P \mid Ed)}\right]$$

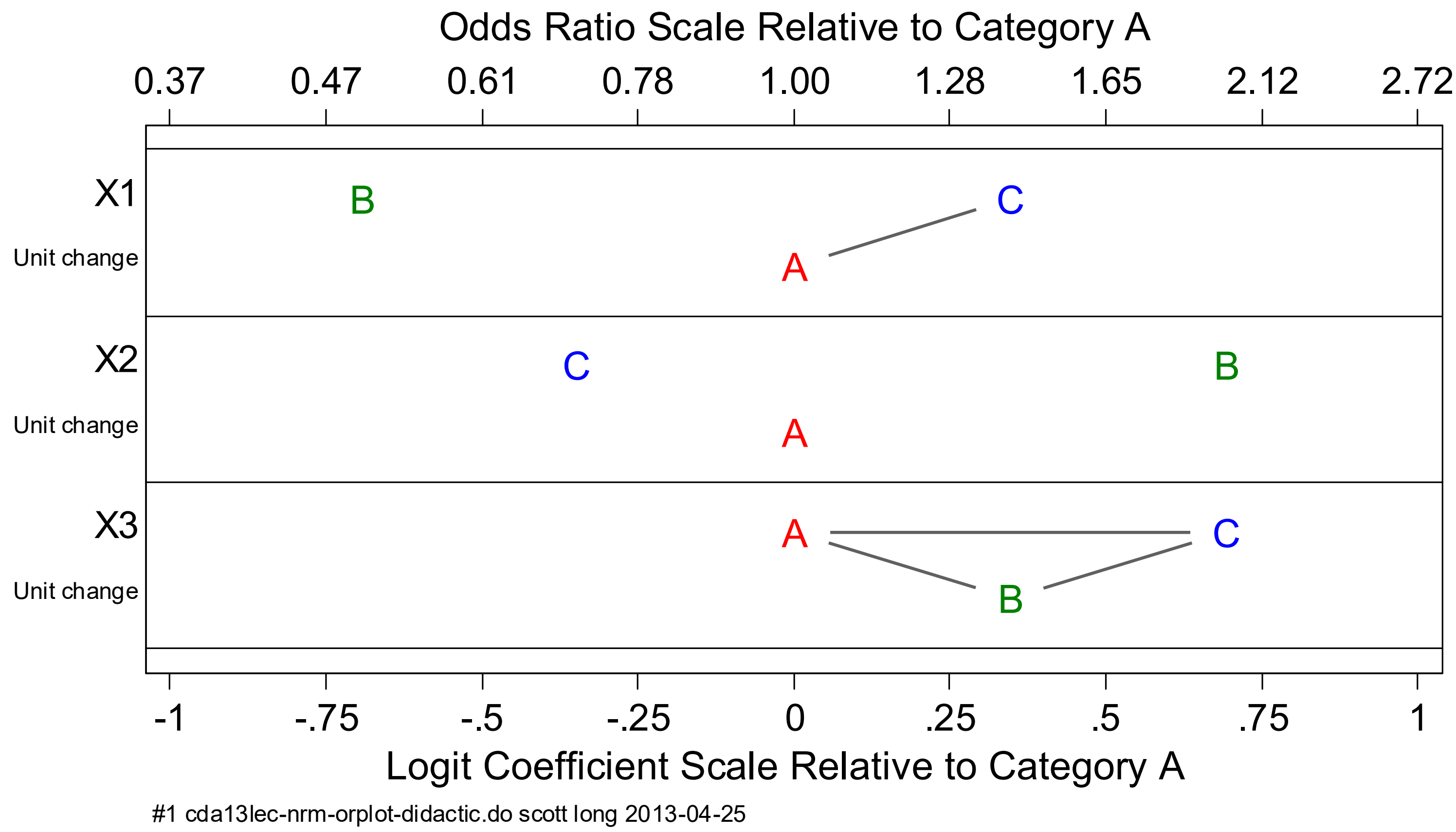
2. So that $\beta_{L \mid P} = \beta_{L \mid S} + \beta_{S \mid P}$
3. Knowing the minimal set we can determine all other parameters.
4. We can plot them just like stops on a subway line.

Odds ratio plot



1. Distance: B -> A = .693
2. Distance: A -> C = .347
3. Distance: B -> C = 1.040 = .693 + .347

A line indicates a non-significant coefficient



MNLM for occupation

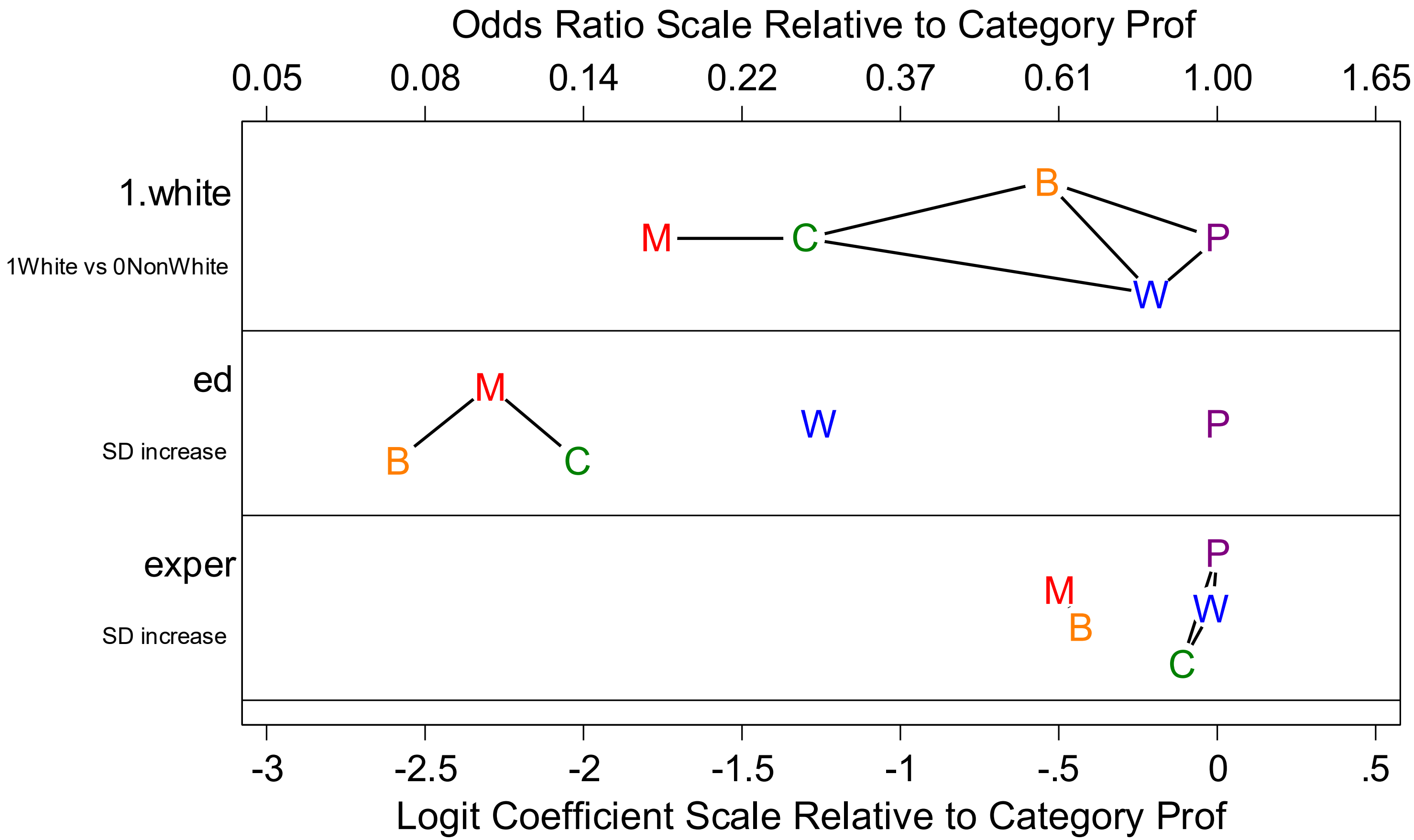
Outcomes

```
M == menial job
B == blue collar job
C == craft job
W == white collar job
P == professional job
```

Regressors

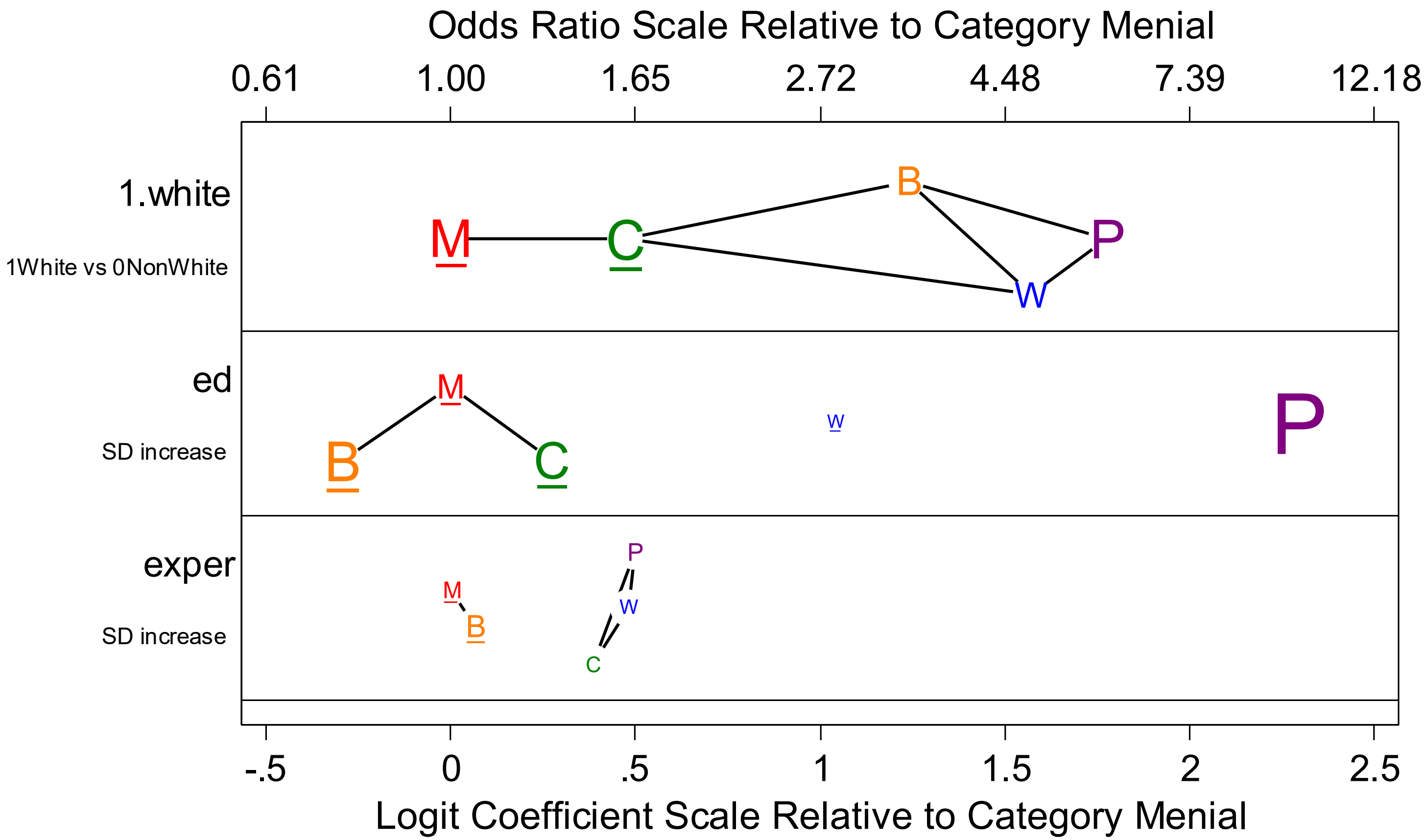
```
white == white = 1, vs nonwhite
ed    == years of education
exper == years of work experience
```

Odds ratio plot



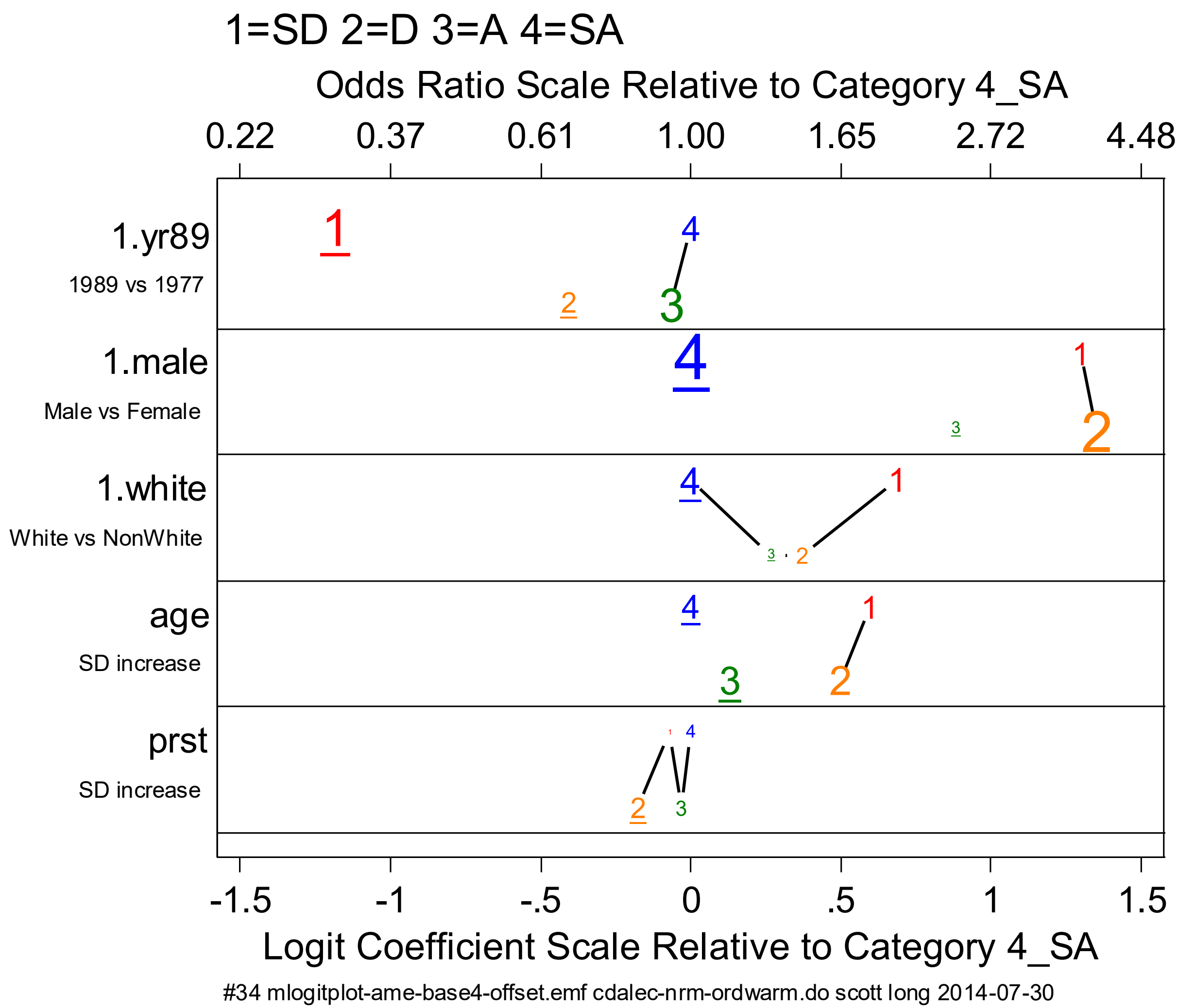
#45 mlogitplot-sig-baseP5-offset.emf cdalec-nrm-nomocc.do scott long 2014-07-30

Size of letter proportional to size of AME



#46 ormchangeplot-ame-baseP5-offset.emf cdalec-nrm-nomocc.do scott long 2014-07-30

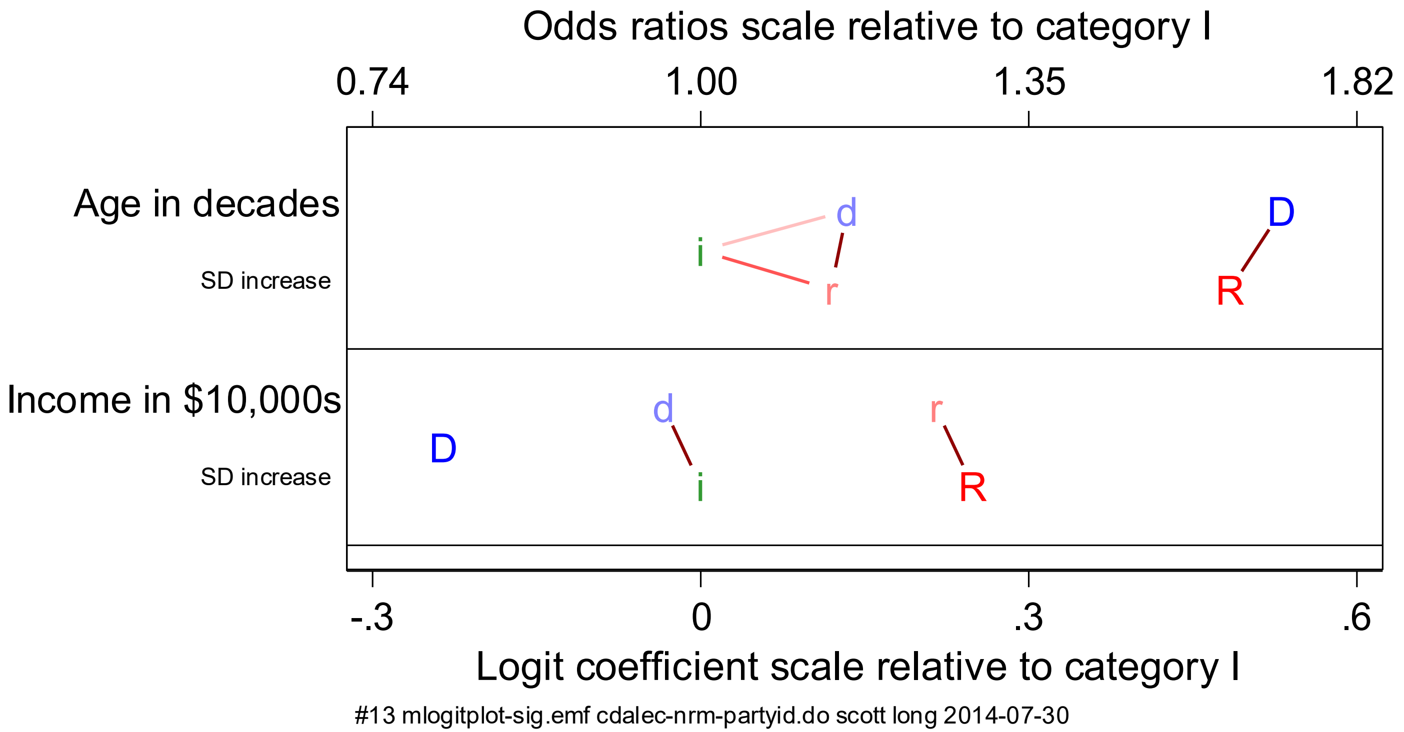
MNLM for attitudes toward working women



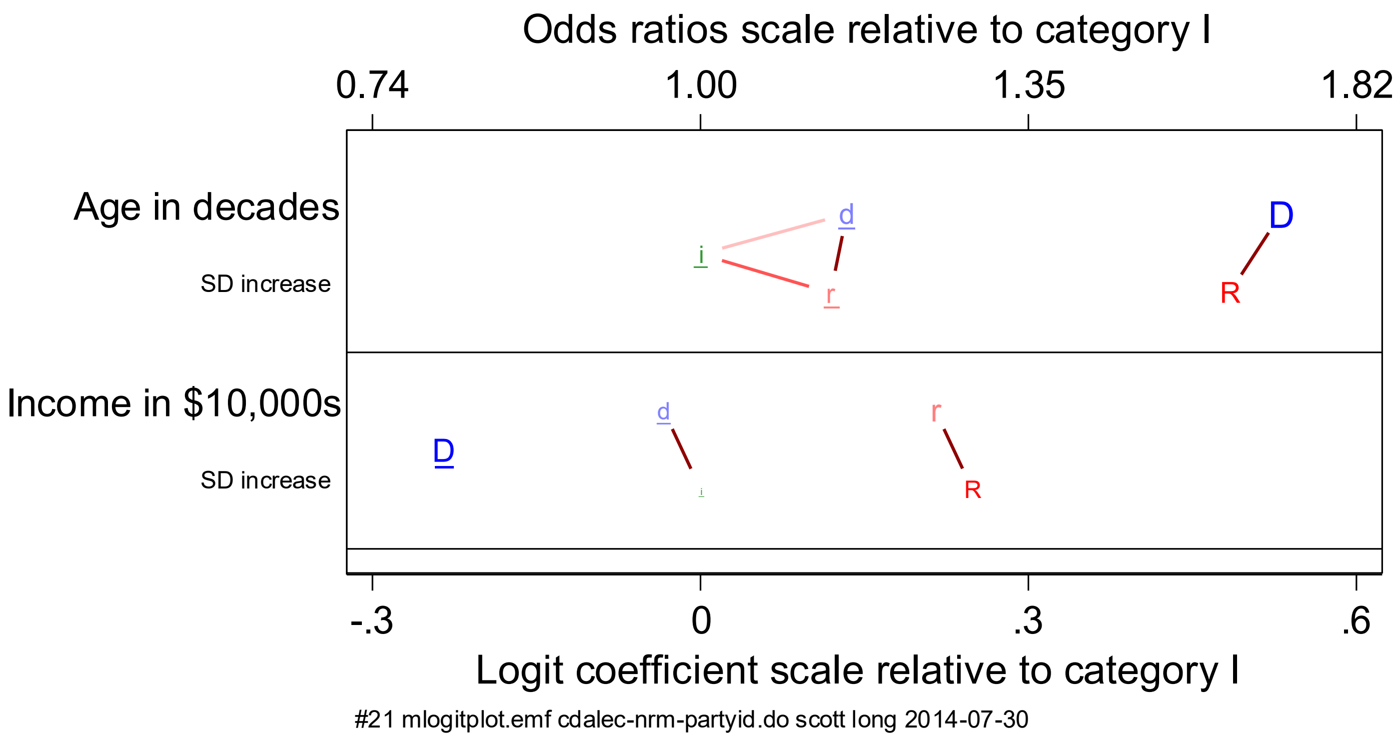
Is this consistent with warm being ordinal?

MNLM for political party

Odds ratios for age and income



Odds ratios for age and income: with marginal effects



Review of nominal LHS

1. I find MNLM to be a very useful model that is easy to interpret with practice
2. The MNLM is a set of BLMs for all pairs of outcomes.
3. IIA is a restrictive assumption that does not have an adequate test.
   - If outcomes are reasonably distinct, MNLM works well.
   - No alternative model is available.
4. MNLM can be overwhelmingly complex if you try to absorb all of the coefficients individually.
5. Plots of coefficients make it easy to uncover patterns.
   - Use plots to find patterns to explore further
   - DC plots are easy to explain
   - OR plots are less effective in papers due to complexitiy in explaining them
6. Tables of predictions are also useful (see next lecture)

β1a Ordinal outcomes

Readings and examples

Long & Freese: Chapter 7

Long, JS 2014, Regression models for nominal and ordinal outcomes *pp 173-203* in H Best C Wolf *Regression Analysis and Causal Inference*

mdo18-orm-.do*

Overview

1. What does ordinal mean? What is an ordinal regression model?
2. Derive ORM as a latent variable model
3. Apply methods of interpretation from the BRM and MNLM.
4. Explore restrictions of the ORM
   - The parallel regression/proportional odds assumption
   - Comparing OLM and MNLM
5. Alternative models for ordinal outcomes.

What does ordinal mean?

1. Are attitudes toward the following statement ordinal?

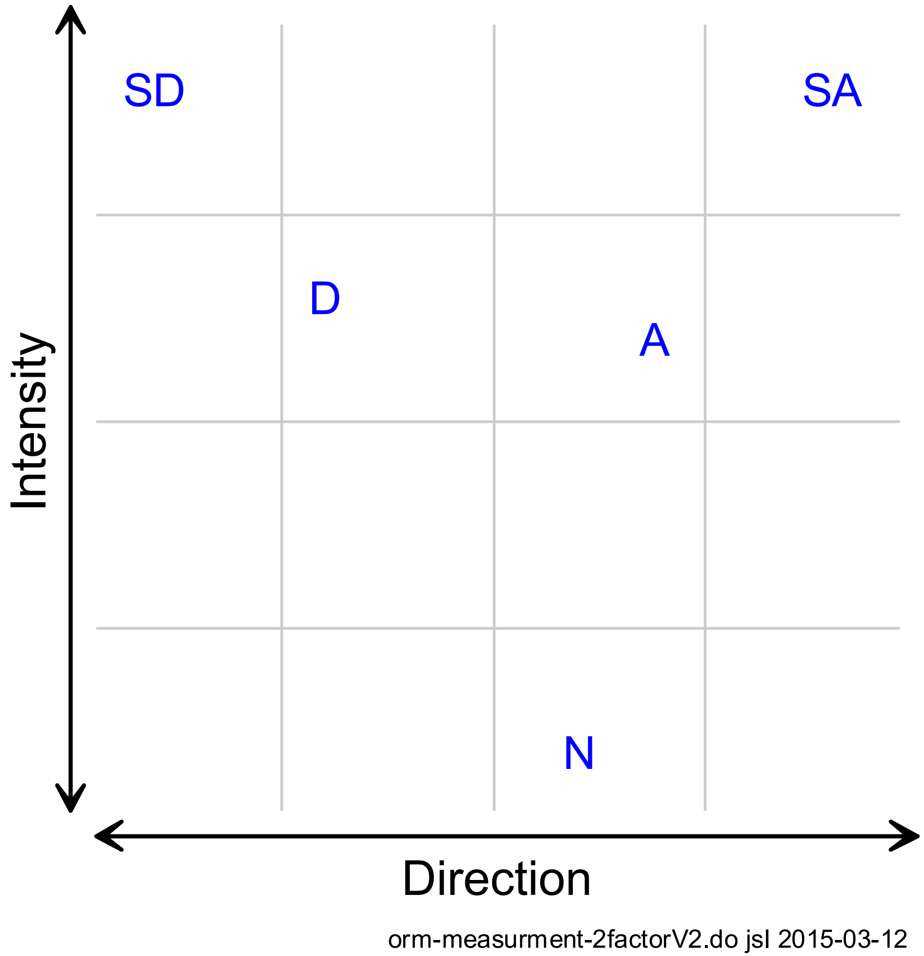
   A working mother can establish just as warm and secure of a relationship with her child as a mother who does not work.

2. An ordinal variable is ordered on a single dimension with unknown distances between categories

3. If values can be ordered does not mean they should be ordered

Variables can be ordered on multiple dimensions

- Occupations can be ordered on both status and income
- Likert scales cand reflect direction and intensity

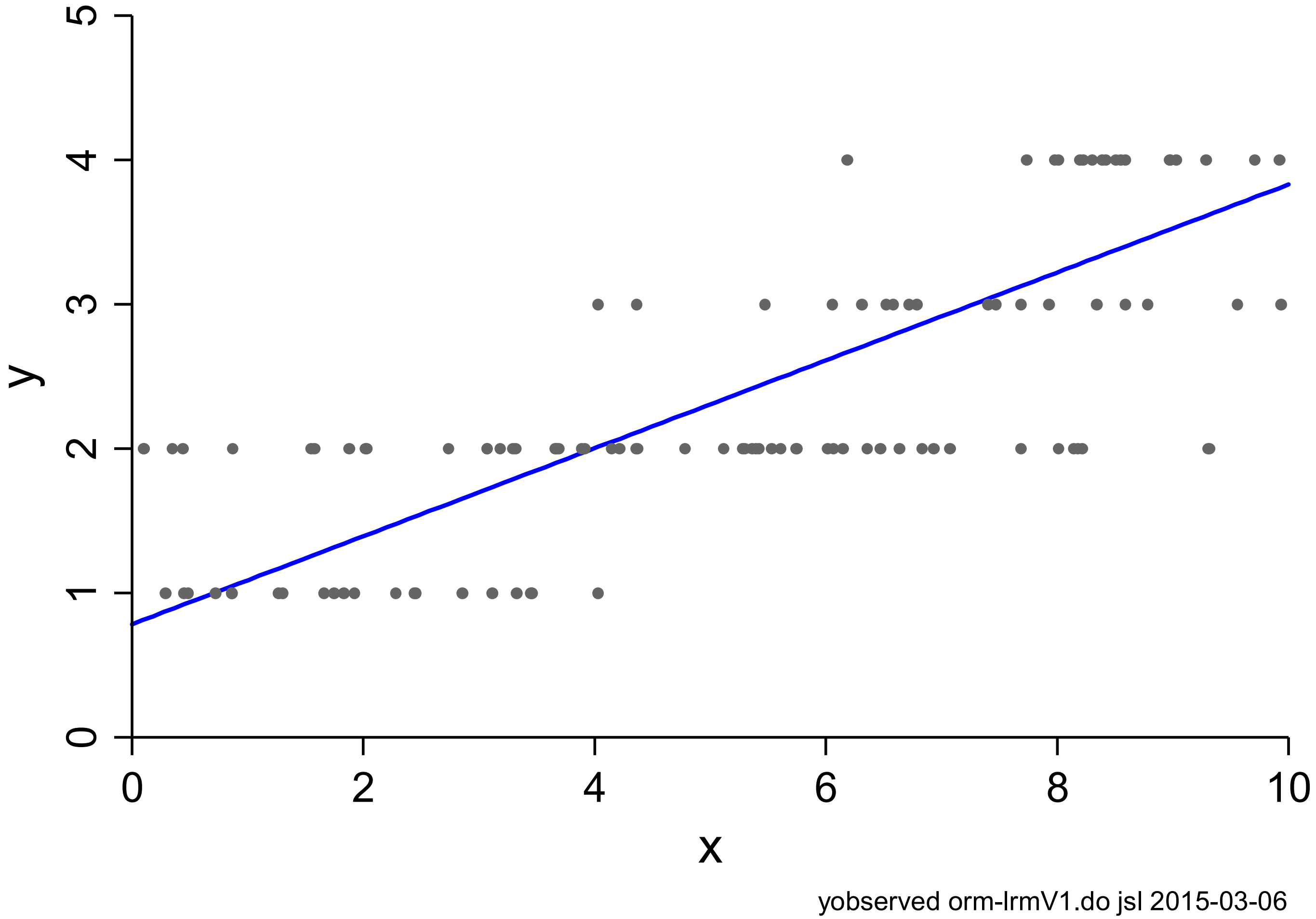


Education affects direction,

while gender affects intensity.

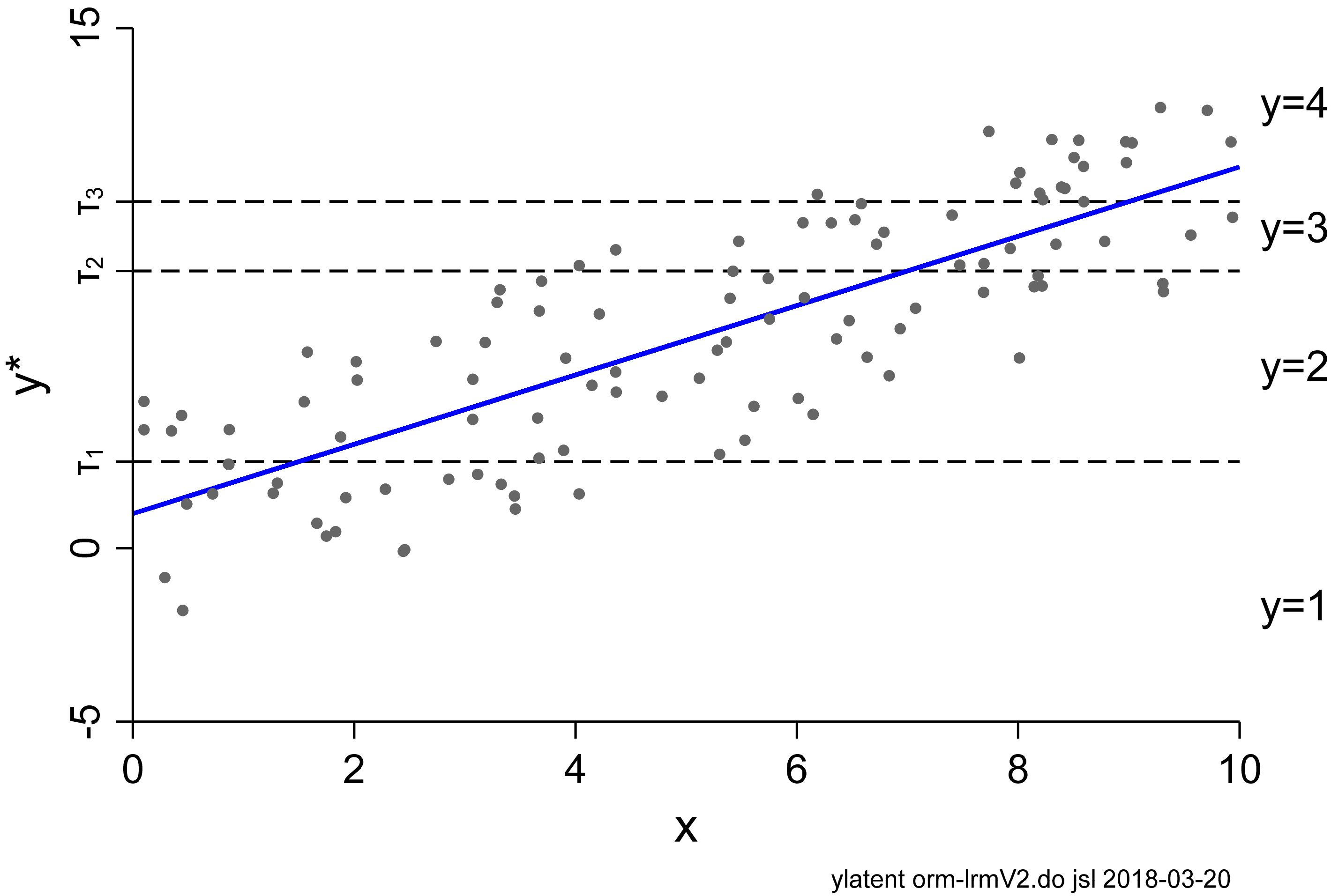
Ordinal is not interval

Treating an ordinal variable as interval can produce misleading results.



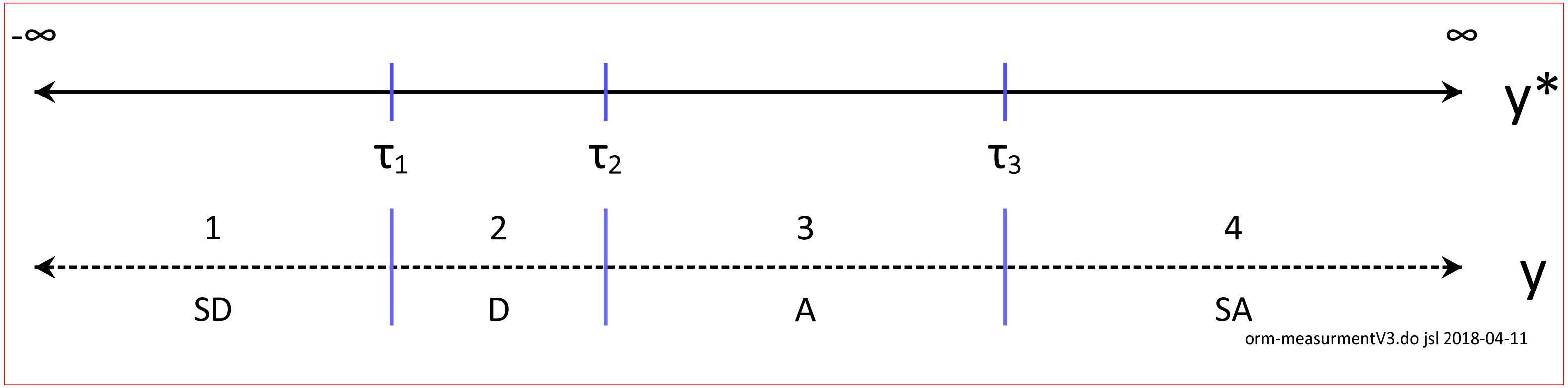
A latent variable model for ordinal outcomes

Assume $y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon$ where $\varepsilon \sim N(0,1)$ for probit and $\varepsilon \sim \lambda(0, \pi^2/3)$ for logit



The measurement model

A working mother can establish just as warm and secure of a relationship with her child as a mother who does not work.



1. The propensity y* and the observed responses are linked by:

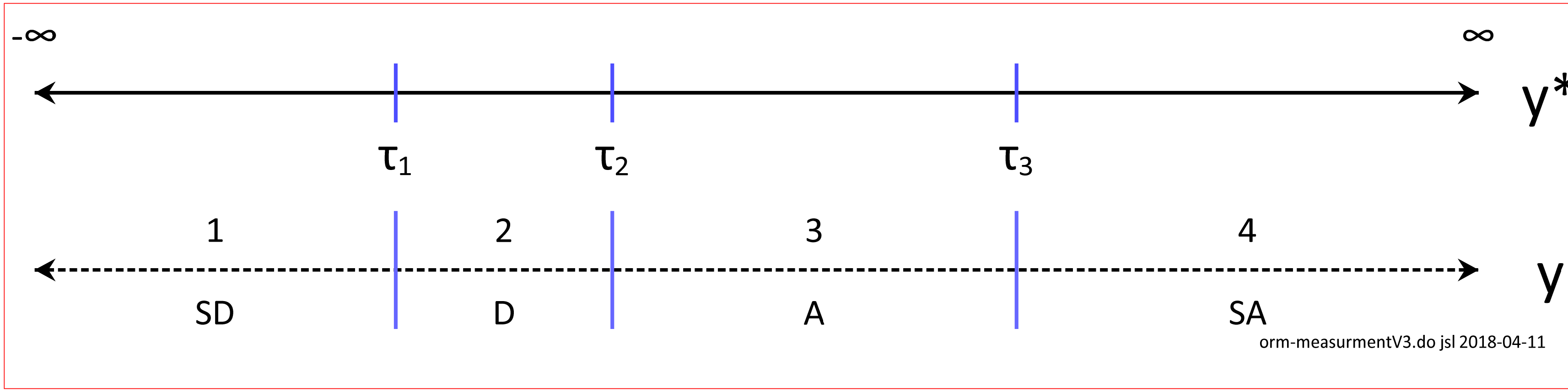
$$y_i = \begin{cases} 1 \Rightarrow \text{SD-Strongly Disagree} & \text{if } \tau_0 = -\infty \le y_i^* < \tau_1 \\ 2 \Rightarrow \text{D-Disagree} & \text{if } \tau_1 \le y_i^* < \tau_2 \\ 3 \Rightarrow \text{A-Agree} & \text{if } \tau_2 \le y_i^* < \tau_3 \\ 4 \Rightarrow \text{SA-Strongly Agree} & \text{if } \tau_3 \le y_i^* < \tau_4 = \infty \end{cases}$$

2. In general:

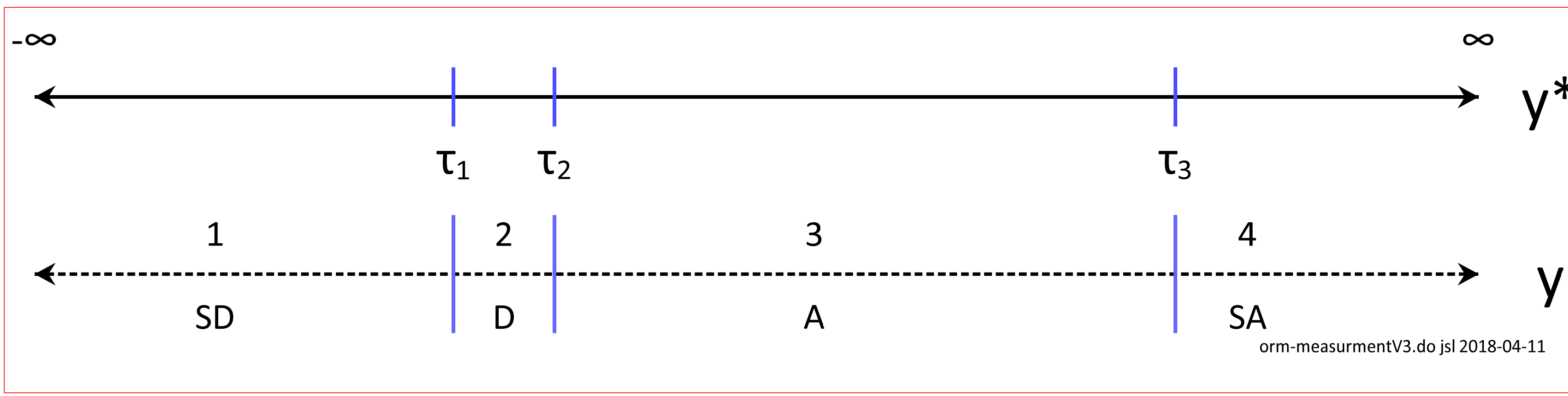
$$y_i = q \quad \text{if } \tau_{q-1} \le y_i^* < \tau_q \quad \text{for } q = 1 \text{ to } J$$

We don't know the thresholds

How can you determine if these are the correct thresholds

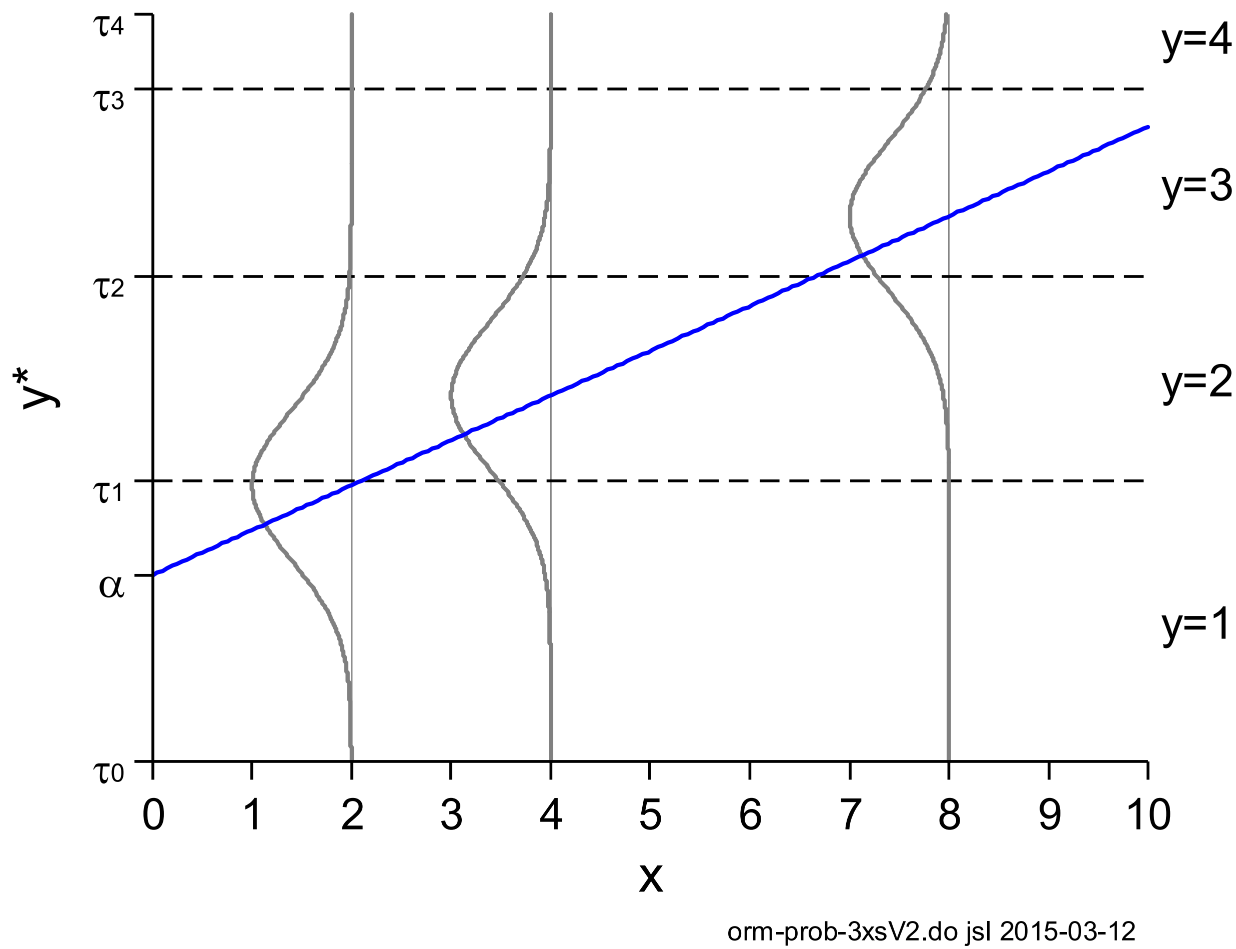


Or these are correct?



Regression linking y to y*

ML fits this model based on observed y and x's



ML estimation

1. The probability that the observed outcome *q* was observed for case *i*:

$$p_i = \Pr\left(y_i = q \mid \mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\tau}\right) = F\left(\tau_q - \mathbf{x}_i\boldsymbol{\beta}\right) - F\left(\tau_{q-1} - \mathbf{x}_i\boldsymbol{\beta}\right)$$

2. With independent observations:

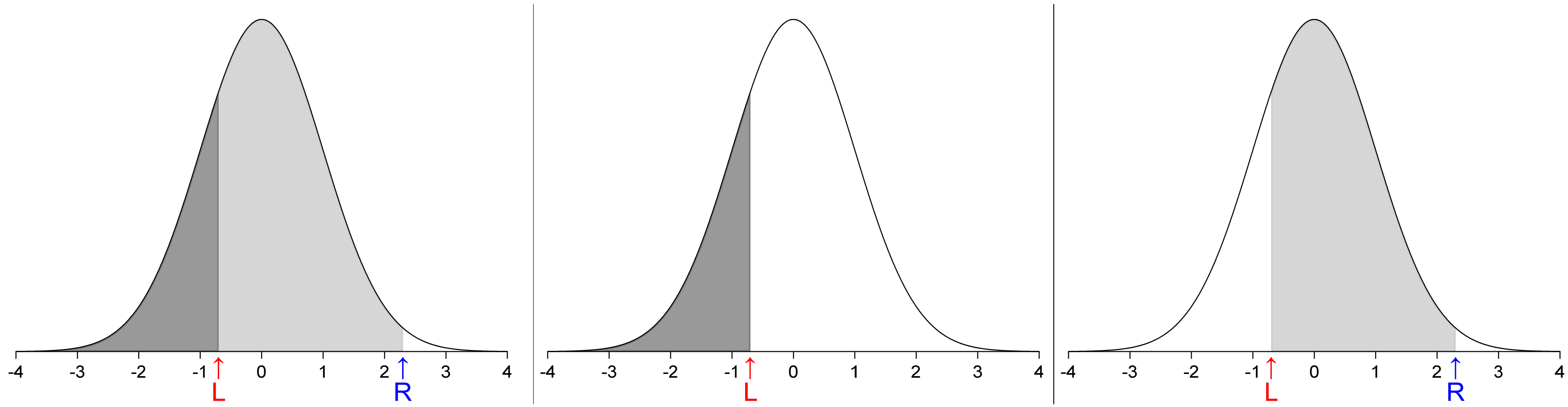
$$L\left(\boldsymbol{\beta}, \boldsymbol{\tau} \mid \mathbf{y}, \mathbf{X}\right) = \prod_{i=1}^{N} p_i$$

Software issues

1. You must know which parameterization is used for identification.
2. Different methods of maximization produce *slightly* different test statistics.
3. ORM takes longer to converge than some models.
4. Small N's in a category can lead to failure to converge.
   - You can merge adjacent categories and only lose efficiency

Computing Pr(y=k | X)

Pr(y=m | x) is the area between two thresholds



Step 1. Area ≤ R

Step 2. Area ≤ L

Step 3. Area between L & R

Computing Pr(y=2|x)

1. If *y*=2 then *y**is between τ_1 and τ_2.

$$\Pr\left(y=2\,|\,X\right)=\Pr\left(\tau_1 \;\le\; y^* < \;\tau_2\,|\,X\right)$$

2. Expanding *y**:

$$\Pr\left(y=2\,|\,X\right)=\Pr\left(\tau_1 \;\le\; \alpha+\beta x+\varepsilon \;<\; \tau_2\,|\,X\right)$$

3. Isolating ϵ:

$$\Pr\left(y=2\,|\,X\right)=\Pr\left(\tau_1-\alpha-\beta X \;\le\; \varepsilon \;<\; \tau_2-\alpha-\beta X\,|\,X\right)$$

4. Generalizing, the probability is a difference of CDFs:

$$\Pr\left(y=m\,|\,\mathbf{x}\right)=F\left(\tau_m-\mathbf{x}\boldsymbol{\beta}\right)-F\left(\tau_{m-1}-\mathbf{x}\boldsymbol{\beta}\right)$$

5. CDF(-∞) = 0 and CDF(∞) = 1 for computing the first and last categories.

Identification: you can't estimate all thresholds and intercepts

1. Suppose that true parameters are α and the τ_ms.
2. Create imposter parameters by subtracting an unknown δ:

$$\alpha^* = \alpha - \delta \qquad \text{and} \qquad \tau_q^* = \tau_q - \delta$$

3. Probabilities are unaffected since δ – δ = 0:

$$\begin{aligned}\Pr(y = q \mid x) &= F(\tau_q - \alpha - \beta x + [\delta - \delta]) - F(\tau_{q-1} - \alpha - \beta x + [\delta - \delta]) \\ &= F([\tau_q - \delta] - [\alpha - \delta] - \beta x) - F([\tau_{q-1} - \delta] - [\alpha - \delta] - \beta x) \\ &= F(\tau_q^* - \alpha^* - \beta x) - F(\tau_{q-1}^* - \alpha^* - \beta x)\end{aligned}$$

4. Identifying assumptions must be made such as:

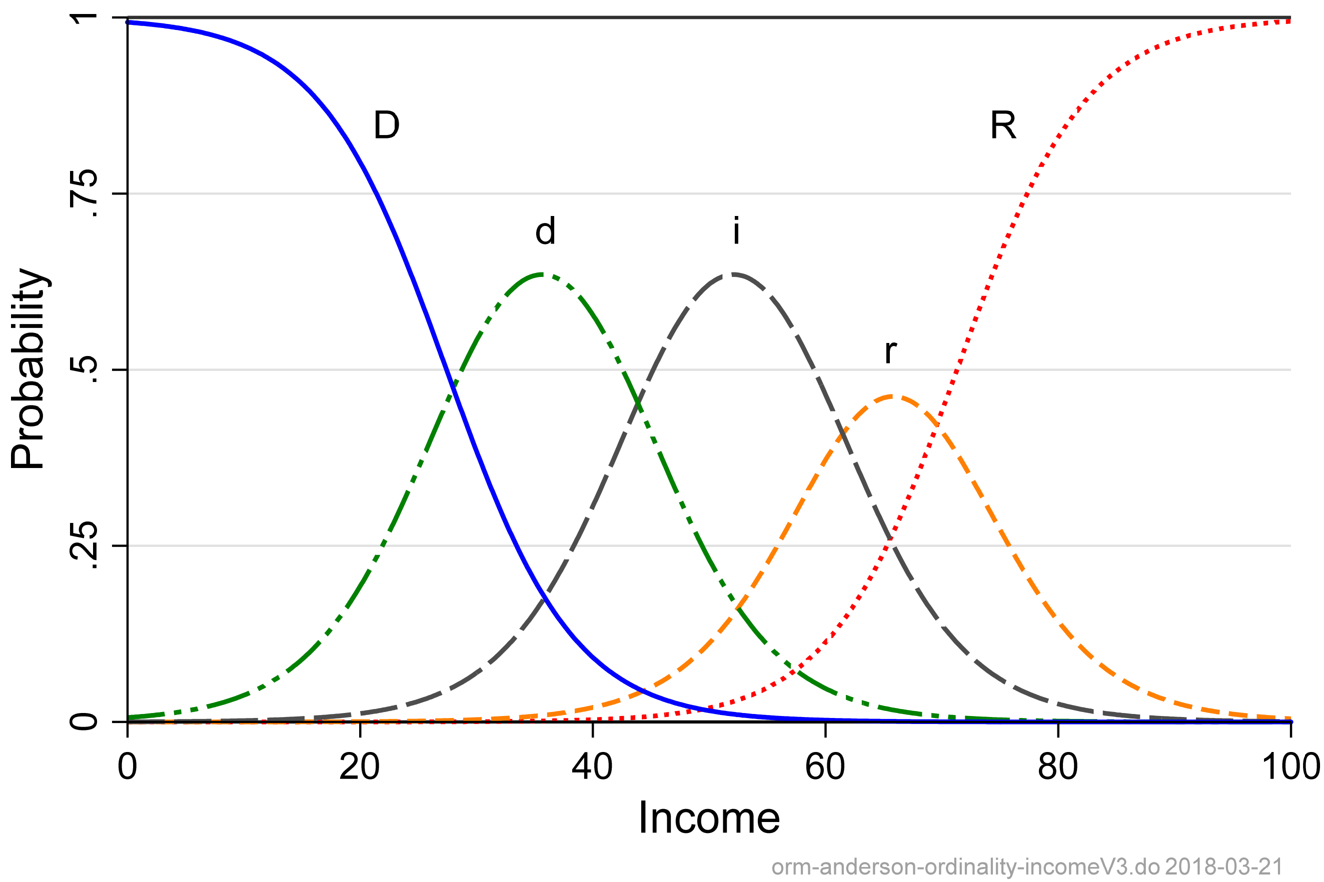
   Alternative 1: $\tau_1 = 0$ forces δ to equal τ_1

   Alternative 2: $\alpha = 0$ forces δ to equals α

5. These assumptions lead to different parameterizations that do not affect the βs or their significance or the probabilities.

What is an ordinal regression model? (Anderson 1980)

1. Inherent in the model is this pattern of between regressors and outcomes.
2. Your predictions *must* conform to this pattern



Example: Attitudes toward working mothers

A working mother can establish just as warm and secure a relationship with her child as a mother who does not work.

- Agreeing supports working women

	Freq.	Percent
1 Strong Disagree	297	12.95
2 Disagree	723	31.53
3 Agree	856	37.33
4 Strong Agree	417	18.19
Total	2,293	100.00

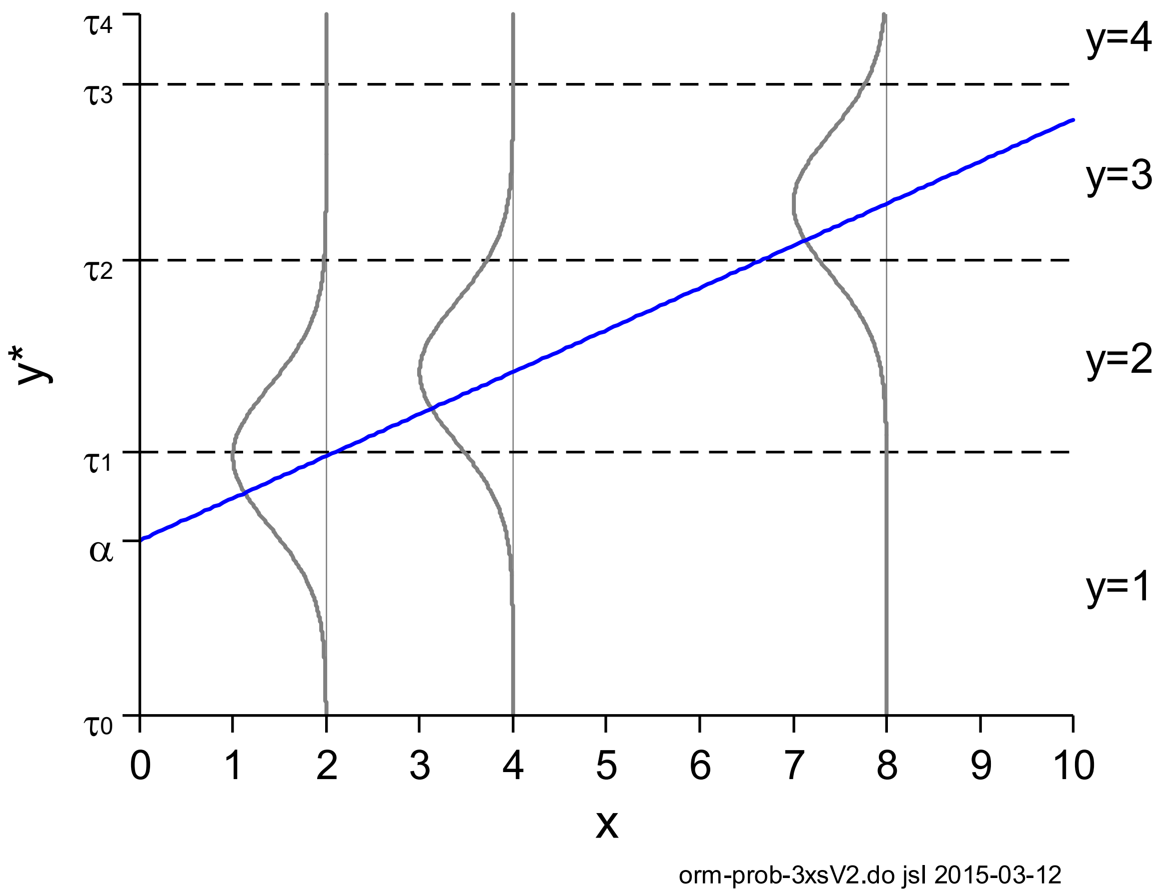
Variable	Mean	Min	Max	Label
yr89	.3986044	0	1	Survey year: 1=1989 0=1977
male	.4648932	0	1	Gender: 1=male 0=female
white	.8765809	0	1	Race: 1=white 0=not white
age	44.93546	18	89	Age in years
ed	12.21805	0	20	Years of education
prst	39.58526	12	82	Occupational prestige

Ordinal logit

```
. ologit warm i.yr89 i.male i.white age ed prst

Ordered logistic regression                     Number of obs     =       2293
                                                LR chi2(6)        =     301.72
                                                Prob > chi2       =     0.0000
Log likelihood = -2844.9123                     Pseudo R2         =     0.0504

        warm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yr89 |
       1989  |   .5239025   .0798989     6.56   0.000     .3673036    .6805014
        male |
       Male  |  -.7332997   .0784827    -9.34   0.000     -.887123   -.5794765
       white |
      White  |  -.3911595   .1183808    -3.30   0.001    -.6231816   -.1591373
         age |  -.0216655   .0024683    -8.78   0.000    -.0265032   -.0168278
          ed |   .0671728    .015975     4.20   0.000     .0358624    .0984831
        prst |   .0060727   .0032929     1.84   0.065    -.0003813    .0125267
-------------+----------------------------------------------------------------
       /cut1 |  -2.465362   .2389128                     -2.933622   -1.997102
       /cut2 |   -.630904   .2333156                     -1.088194   -.1736138
       /cut3 |   1.261854    .234018                      .8031871    1.720521
------------------------------------------------------------------------------

. estimates store olm // to restore estimates after margins, post
```

Predicted probabilities

$$\Pr\left(y=q\mid\mathbf{x}\right)=F\left(\hat{\tau}_q-\mathbf{x}\hat{\boldsymbol{\beta}}\right)-F\left(\hat{\tau}_{q-1}-\mathbf{x}\hat{\boldsymbol{\beta}}\right)$$

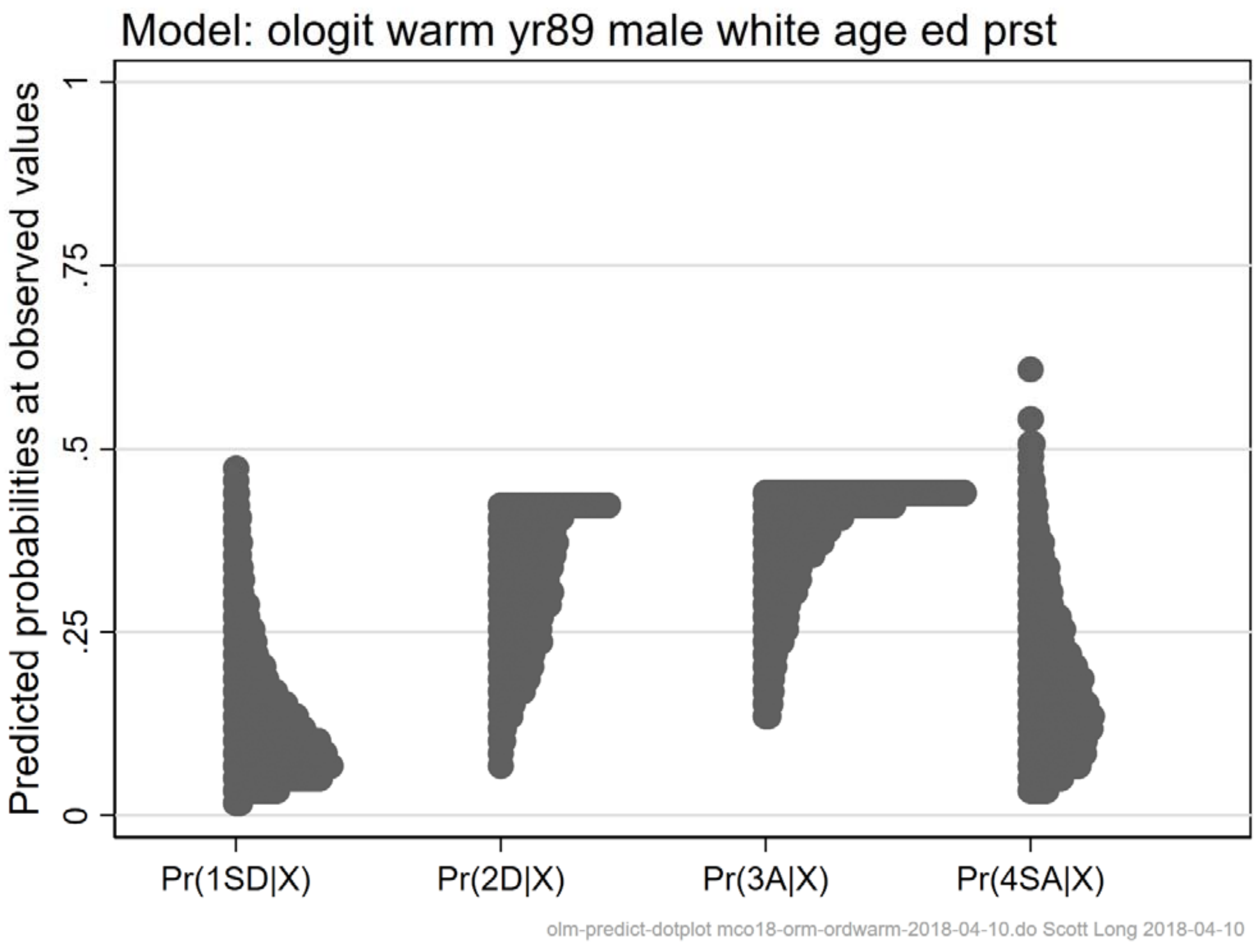


Using predictions for interpretation

1. Predictions at observed values
2. Marginal effects
3. Ideal types
4. Tables of probabilities
5. Plots of probabilities

Predictions at observed values

I start by examining the distribution of predictions.



Code for predictions at observed values

Make predictions

1. One prediction for each level of the outcome:

```
. predict OLMpr1sd OLMpr2d OLMpr3a OLMpr4sa
```

2. Add labels

```
. label var OLMpr1sd "Pr(1SD|X)"
. label var OLMpr2d  "Pr(2D|X)"
. label var OLMpr3a  "Pr(3A|X)"
. label var OLMpr4sa "Pr(4SA|X)"
```

Examine predictions

1. Summary statistics

```
. sum OLMpr1sd OLMpr2d OLMpr3a OLMpr4sa

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    OLMpr1sd |      2293    .1291898    .0827858   .0078648   .4583639
     OLMpr2d |      2293    .3152269    .0702155   .0811076   .4066651
     OLMpr3a |      2293    .3740882    .0585058   .1494467   .4274917
    OLMpr4sa |      2293    .1814951    .0976008    .018211   .5864362
```

2. Histogram

```
dotplot OLMpr1 OLMpr2 OLMpr3 OLMpr4, ylab(0(.25)1, grid gmin gmax) ///
  title(Model: ologit warm yr89 male white age ed prst, position(11)) ///
  ytitle(Predicted probabilities at observed values)
```

3. Export to PNG since EMF can be very large

```
graph export `pgm'-`graphname'.png, width(1600) replace
```

Tables of predicted probabilities: gender and age

1. When you have important categorical regressors, tables are effective.
2. Where should you hold the other variables?

   Global means are simpler but can be misleading

   Local means are harder but sometimes more realistic
3. A sensitivity analysis assesses how important this decision is.

Gender and year with global means: agree supports working women

1. Theory, past research, and the regression coefficients suggests:
   - Men are more negative than women
   - Attitudes are more positive in 1989 than 1977

A. 1977	SD	D	A	SA	
Men	0.19	0.40	0.32	0.10	
Women	0.10	0.31	0.41	0.18	
Men-Women	*0.09*	*0.09*	*-0.09*	*-0.08*	DCR(male|1977)
B. 1989	SD	D	A	SA	
Men	0.12	0.34	0.39	0.15	
Women	0.06	0.23	0.44	0.27	
Men-Women	*0.06*	*0.11*	*-0.05*	*-0.12*	DCR(male|1989)
C. 1977 to 1989	SD	D	A	SA	
Men	*-0.07*	*-0.06*	*0.07*	*0.05*	DCR(yr89|male)
Women	*-0.04*	*-0.08*	*0.03*	*0.09*	DCR(yr89|female)

Note: Other variables are held at their means.

Interpretation

TODO

XYZ

Are global means reasonable

1. The predictions are made holding all other variables at their means.
   - Predictions for men in 1989 hold variables at the same values as predictions for women in 1977
2. Is this substantively reasonable?
3. What if we make predictions using local means

   Pr(y| gender, years, Mean(x|gender, year))

Predictions with local means: Agree supports working mothers

A. 1977	SD	D	A	SA
Men	0.19	0.40	0.31	0.09
Women	0.10	0.32	0.41	0.17
Men-Women	*0.09*	*0.09*	*-0.10*	*-0.08*

B. 1989	SD	D	A	SA
Men	0.11	0.33	0.40	0.16
Women	0.06	0.23	0.44	0.27
Men-Women	*0.05*	*0.10*	*-0.04*	*-0.11*

C. 1977 to 1989	SD	D	A	SA
Men	*-0.08*	*-0.07*	*0.09*	*0.07*
Women	*-0.04*	*-0.09*	*0.03*	*0.10*

Note: Other variable are held at means for given year and gender.

Comparing predictions with global and local means

	1SD	2D	3A	4SA
1977				
Men	-0.00	-0.00	0.00	0.00
Women	-0.00	-0.01	0.00	0.01
1989				
Men	0.01	0.01	-0.01	-0.01
Women	0.00	0.00	-0.00	-0.00
1977				
Men_Women	0.00	0.01	-0.00	-0.01
1989				
Men_Women	0.01	0.01	-0.01	-0.01

1. The differences are substantively small so I am confident that conclusions are not affected by levels of the controls.
2. Which set of probabilities would you use? Why?

Code for tables of predictions

mtable with global means: Agree indicates support

1. Atspec **at(yr89=(0 1) male=(0 1))** computes predictions for all combinations of **yr89** and **male**
2. **atmeans** holds other variable at global means

```
. mtable, at(yr89=(0 1) male=(0 1)) atmeans clear

Expression: Pr(warm), predict(outcome())

          |     yr89      male      1_SD       2_D       3_A      4_SA
----------+-------------------------------------------------------------
        1 |        0         0     0.099     0.308     0.413     0.180
        2 |        0         1     0.186     0.403     0.316     0.095
        3 |        1         0     0.061     0.228     0.441     0.270
        4 |        1         1     0.119     0.339     0.390     0.151

Specified values of covariates

          |       1.
          |    white       age        ed      prst
----------+---------------------------------------
  Current |     .877      44.9      12.2      39.6
```

3. Since **i.male** is a regressor, **dydx(male)** computes DC for gender.

4. Option **atvars(1.yr89)** adds **yr89** to the table.

```
. mtable, dydx(male) at(yr89=(0 1)) atvars(1.yr89) atmeans

Expression: Marginal effect of Pr(warm), predict(outcome())

          |       1.
          |     yr89      1 SD       2 D       3 A      4 SA
----------+--------------------------------------------------
        1 |        0     0.087     0.094    -0.097    -0.085
        2 |        1     0.058     0.111    -0.050    -0.119

Specified values of covariates

          |       1.
          |     male     white       age        ed      prst
----------+--------------------------------------------------
  Current |     .465      .877      44.9      12.2      39.6
```

5. To verify my work,

```
. mchange male, at(yr89=0) atmeans brief // DCmale 1977
. mchange male, at(yr89=1) atmeans brief // DCmale 1989
```

Advanced code: Building a nicer table

```
mtable, at(yr89=0 male=1)     atmeans rowname(Men)        clear roweqnm(1977)
mtable, at(yr89=0 male=0)     atmeans rowname(Women)      below roweqnm(1977)
mtable, dydx(male) at(yr89=0) atmeans rowname(Men_Women)  below roweqnm(1977)

mtable, at(yr89=1 male=1)     atmeans rowname(Men)        below roweqnm(1989)
mtable, at(yr89=1 male=0)     atmeans rowname(Women)      below roweqnm(1989)
mtable, dydx(male) at(yr89=1) atmeans rowname(Men_Women)  below roweqnm(1989)

mtable, dydx(yr89) at(male=1) atmeans rowname(77to89)     below roweqnm(Men)
mtable, dydx(yr89) at(male=0) atmeans rowname(77to89)     below roweqnm(Women)
```

```
Expression: Marginal effect of Pr(warm), predict(outcome())

            |      1 SD        2 D        3 A       4 SA
------------+----------------------------------------
1977        |
        Men |     0.186      0.403      0.316      0.095
      Women |     0.099      0.308      0.413      0.180
  Men Women |     0.087      0.094     -0.097     -0.085
1989        |
        Men |     0.119      0.339      0.390      0.151
      Women |     0.061      0.228      0.441      0.270
  Men Women |     0.058      0.111     -0.050     -0.119
Men         |
     77to89 |    -0.067     -0.063      0.074      0.056
Women       |
     77to89 |    -0.038     -0.080      0.028      0.090

Specified values of covariates …
```

Tables using local means

1. Check means by by year and gender

```
. sort yr89 male
. by yr89 male: sum white age ed prst
-> yr89 = 1977, male = Female
    Variable |       Obs        Mean    Std. Dev.       Min        Max
       white |       718    .8760446    .3297604          0          1
         age |       718    45.19638    16.59508         19         88
          ed |       718    11.73816    2.813291          3         19
        prst |       718    37.38579    13.53379         12         78
-> yr89 = 1977, male = Male
::
-> yr89 = 1989, male = Female
::
-> yr89 = 1989, male = Male
       white |       405    .8938272    .3084397          0          1
         age |       405    43.66667    16.98412         19         89
          ed |       405    13.11358    3.368747          0         20
        prst |       405    42.16543      14.999         12         82
```

2. Combing **`if`** with **`atmeans`** makes predictions with local means

```
mtable if yr89==0 & male==1, atmeans rowname(Men)    ///
     clear roweqnm(1977)

mtable if yr89==0 & male==0, atmeans rowname(Women) ///
     below roweqnm(1977)

mtable if yr89==1 & male==1, atmeans rowname(Men)    ///
     below roweqnm(1989)

mtable if yr89==1 & male==0, atmeans rowname(Women) ///
     below roweqnm(1989)
```

Marginal effects: Discrete changes

DC at representative values (DCR and DCM)

1. Discrete change at x* is defined as:

$$\frac{\Delta \Pr\left(y=q \mid \mathbf{x}^{*}\right)}{\Delta x_k} = \Pr\left(y=q \mid \mathbf{x}^{*}, \text{End } x_k\right) - \Pr\left(y=q \mid \mathbf{x}^{*}, \text{Start } x_k\right)$$

2. Interpretation:

   When x_k changes from the *start* to *end*, the probability of *q* changes by MER, holding other variables at **x***.

3. For example,

   Being male decreases the probability of strong support for working mothers by .10, holding other variables at their means.

4. The DCR depends on
   - The regression parameters
   - The value at which x_k starts and how much it changes
   - Levels of variables that are not changing

Average DC (AME)

1. Average discrete change:

$$\text{ADC} = \frac{1}{N}\sum_{i=1}^{N}\frac{\Delta \Pr\left(y = q \mid \mathbf{x}_i\right)}{\Delta x_{ik}},$$

2. Interpretation

   On average a change of Δx_k in x_k changes the probability of *q* by ADC.

3. For example,

   On average being male decreases the probability of strongly agreeing that working mothers can be good mothers by .10.

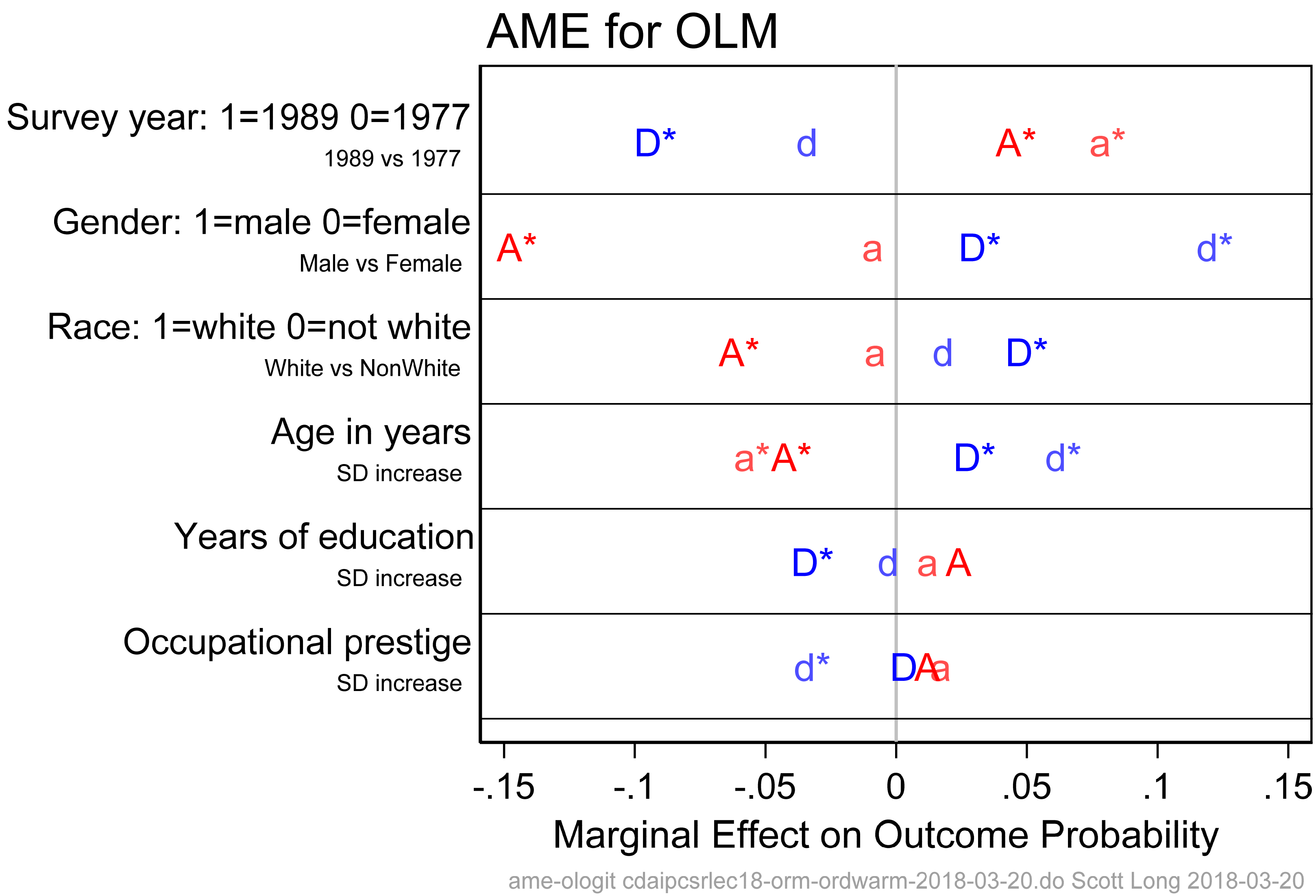
4. The ADC depends on
   - The regression parameters
   - The distribution of all regressors

ADC for all variables

Computing the effects is easy, but there is a lot of information to digest

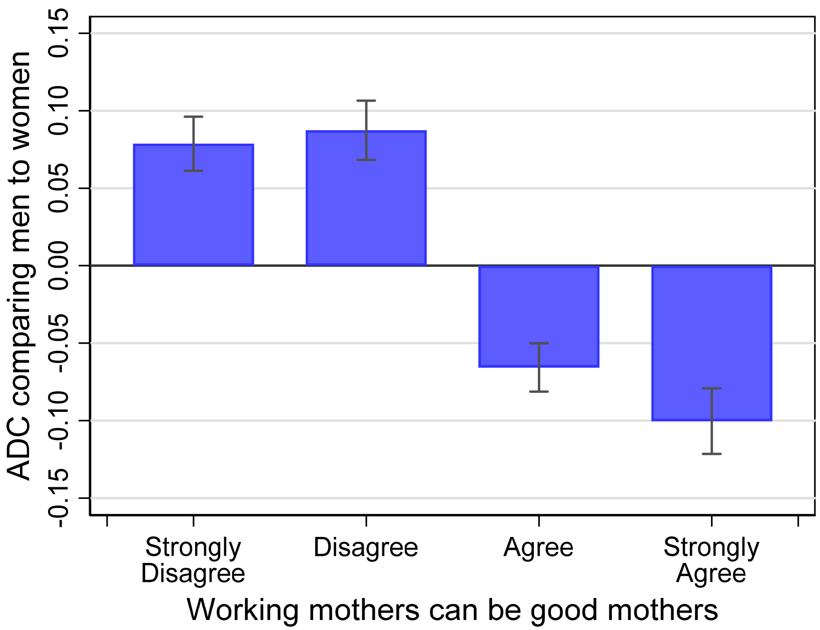
```
. mchange, amount(one sd) // see graph that follows
ologit: Changes in Pr(y) | Number of obs = 2293
Expression: Pr(warm), predict(outcome())
                   |      1 SD        2 D         3 A        4 SA
-------------------+----------------------------------------------
yr89  1989 vs 1977 |    -0.053     -0.064      0.042      0.075
           p-value |     0.000      0.000      0.000      0.000
male               |
    Male vs Female |     0.079      0.087     -0.066     -0.100
           p-value |     0.000      0.000      0.000      0.000
white              |
 White vs NonWhite |     0.038      0.048     -0.026     -0.059
           p-value |     0.000      0.001      0.000      0.002
age                |
                +1 |     0.002      0.003     -0.002     -0.003
           p-value |     0.000      0.000      0.000      0.000
               +SD |     0.043      0.038     -0.036     -0.046
           p-value |     0.000      0.000      0.000      0.000
ed                 |
                +1 |    -0.007     -0.008      0.005      0.010
           p-value |     0.000      0.000      0.000      0.000
               +SD |    -0.021     -0.026      0.015      0.031
           p-value |     0.000      0.000      0.000      0.000
::
```

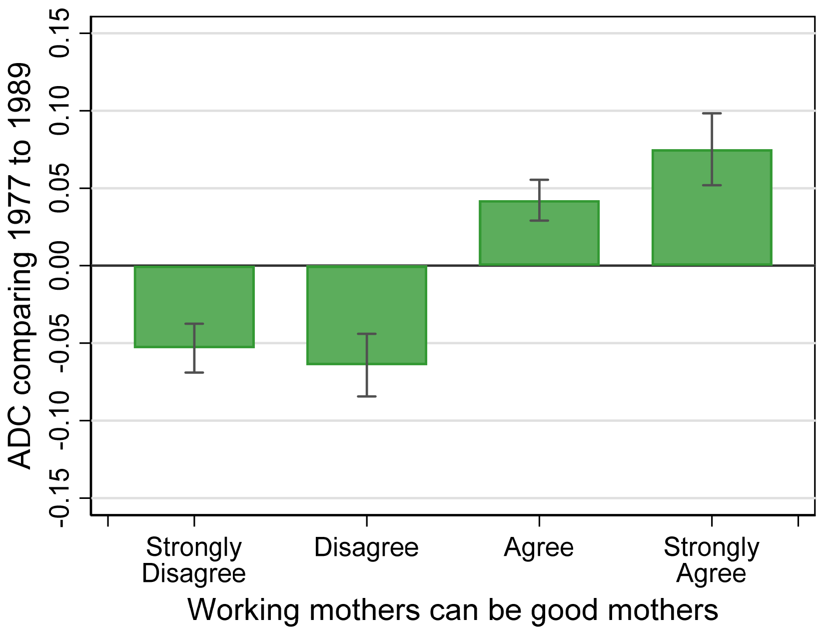
Graph to summarize effects (* indicates significant at .05 level)



Focusing on survey year and gender

1. I focus on two critical variables and use error bars to indicate precision





TODO add summary of findings

Comparing discrete changes

1. Did the effects of gender change significantly between 1977 and 1989? That is:

Ho: ADC(gender|1977,**x***) = ADC(gender|1989,**x***)

	1 SD	2 D	3 A	4 SA
ADC(male|77)	0.089	0.081	-0.083	-0.088
p-value	0.000	0.000	0.000	0.000
ADC(male|89)	0.062	0.099	-0.041	-0.119
p-value	0.000	0.000	0.000	0.000
ΔADC(male)/Δyear	-0.028	0.018	0.042	-0.032
p-value	0.000	0.000	0.000	0.000

The effects of gender changed significantly from 1977 to 1989. From 1977 to 1989 the effect of being male on strongly opposing working women as mothers got 3 points smaller, which was offset by a 2 point increase in opposing working women. Men still were less likely to agree with working women, but the effect was 4 points smaller. But, they were 3 points less to strongly endorse working women as mothers.

Computing marginal effects

Plotting DCs

Plotting mchange results with mchangeplot

```
mchange, amount(one sd)

mchangeplot, min(-.15) max(.15) gap(.05) varlabels ///
    mcol(blue blue*.7 red*.7 red) ///
    symbols(D d a A) sig(.05) ///
    title(AME for OLM, position(11)) ///
    aspect(.8) leftmargin(.5)
```

Plotting DCs with marginsplot

```
margins, dydx(male)

marginsplot, recast(bar) ciopts(color(gs5)) ///
    xlab(0.5 " "      1 `""Strongly" "Disagree""' ///
         2 "Disagree" 3 "Agree" ///
        4 `""Strongly" "Agree""' 4.5 " ") ///
    ylab(-.15(.05).15, format(%4.2f))  ///
    yline(0) plotopts(barw(0.7) color(blue*.8)) ///
    xtitle("Working mothers can be good mothers") ///
    ytitle("ADC comparing men to women") title("") ///
    scale(1.2)
```

Comparing discrete changes

Compute DC(male|year) with mtable

```
. qui mtable, dydx(male) at(yr89=0) rowname(DCmale_77) clear
. mtable, dydx(male) at(yr89=1) rowname(DCmale_89) below

             |     1 SD       2 D       3 A      4 SA
  -----------+----------------------------------------
   DCmale 77 |    0.089     0.081    -0.083    -0.088
   DCmale 89 |    0.062     0.099    -0.041    -0.119
```

Using margins since we need to post predictions

```
. margins, at(yr89=(0 1) male=(0 1)) post
::
1._predict   : Pr(warm==1), predict(pr outcome(1))
::
4._predict   : Pr(warm==4), predict(pr outcome(4))

1._at        : yr89           =          0
               male           =          0
::
4._at        : yr89           =          1
               male           =          1
```

```
             |              Delta-method
             |     Margin   Std. Err.      z    P>|z|   Annotated
-------------+------------------------------------------------------
_predict#_at |
         1 1 |   .1085645   .0077007    14.10   0.000  <= out 1: 77 women
         1 2 |   .1980111   .0116351    17.02   0.000  <= out 1: 77 men
::
         4 4 |   .1623563   .0111352    14.58   0.000  <= out 4: 89 men
-------------+------------------------------------------------------
```

lincom to estimate second difference

```
. lincom (_b[1bn._predict#1bn._at]-_b[1bn._predict#2._at]) ///
>        -(_b[1bn._predict#3._at]-_b[1bn._predict#4._at])

 ( 1)  1bn._predict#1bn._at - 1bn._predict#2._at - 1bn._predict#3._at +
       1bn._predict#4._at = 0

             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   -.027685   .0049492    -5.59   0.000    -.0373854   -.0179847
```

mlincom is easier

```
. mlincom (1-2)-(3-4)

             |  lincom   pvalue       ll       ul
-------------+------------------------------------
           1 |  -0.028    0.000   -0.037   -0.018
```

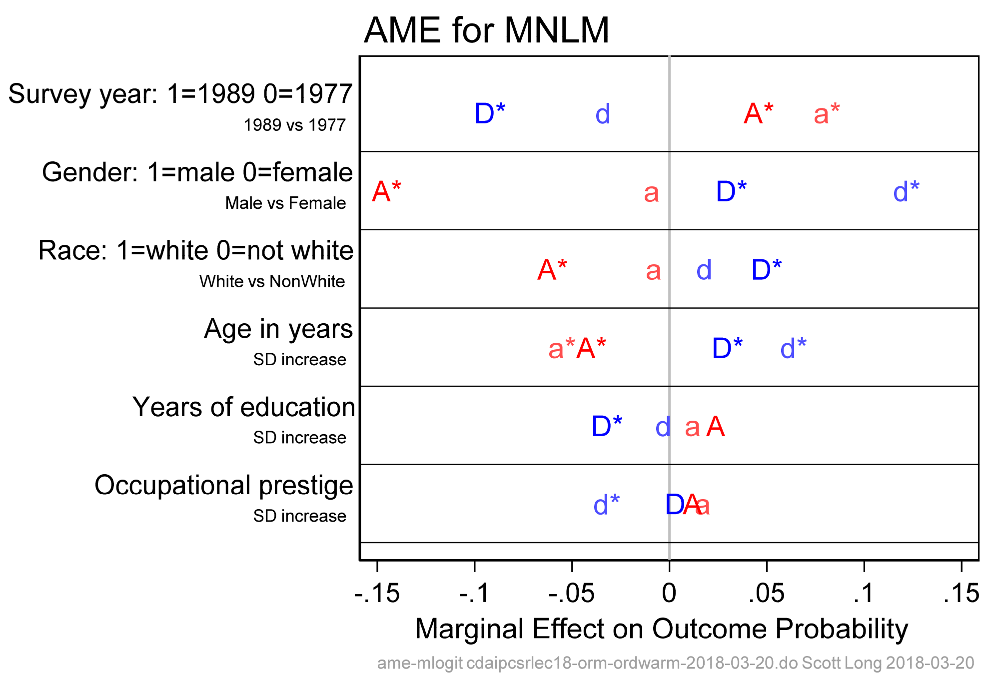
mlincom to create table of differences in effect of gender by year

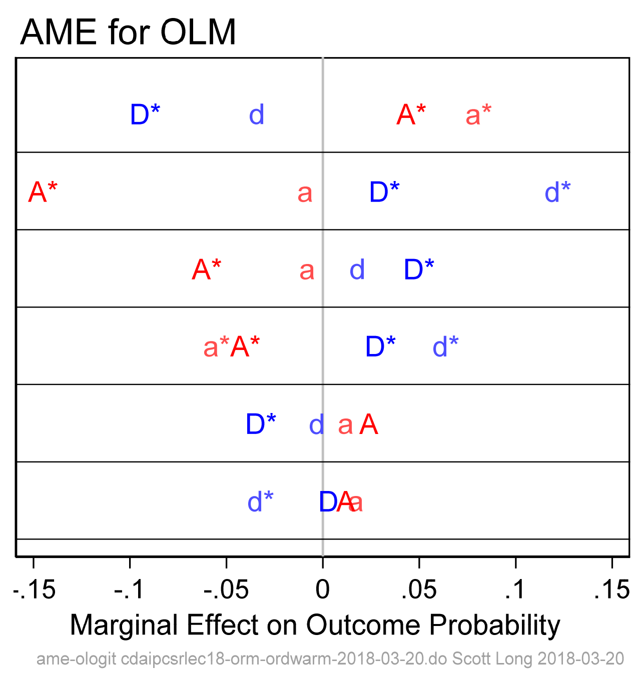
```
. qui mlincom (1-2)-(3-4),     rowname(SD) stats(est p) clear
. qui mlincom (5-6)-(7-8),     rowname(D)  stats(est p) add
. qui mlincom (9-10)-(11-12),  rowname(A)  stats(est p) add
.     mlincom (13-14)-(15-16), rowname(SA) stats(est p) add

             |   lincom    pvalue
-------------+--------------------
          SD |   -0.028     0.000
           D |    0.018     0.000
           A |    0.042     0.000
          SA |   -0.032     0.000
```

The effects of gender on all outcomes changed significantly between 1977 to 1989.

Comparing DCs for OLM and MNLM

1. Ordinal models require a specific relationship between regressors and predicted probabilities that follow from the assumption that the outcome is ordered on one dimension.
2. One way to explore if this assumption is reasonable, is to compare results to multinomial logit.
3. Later we will test differences formally





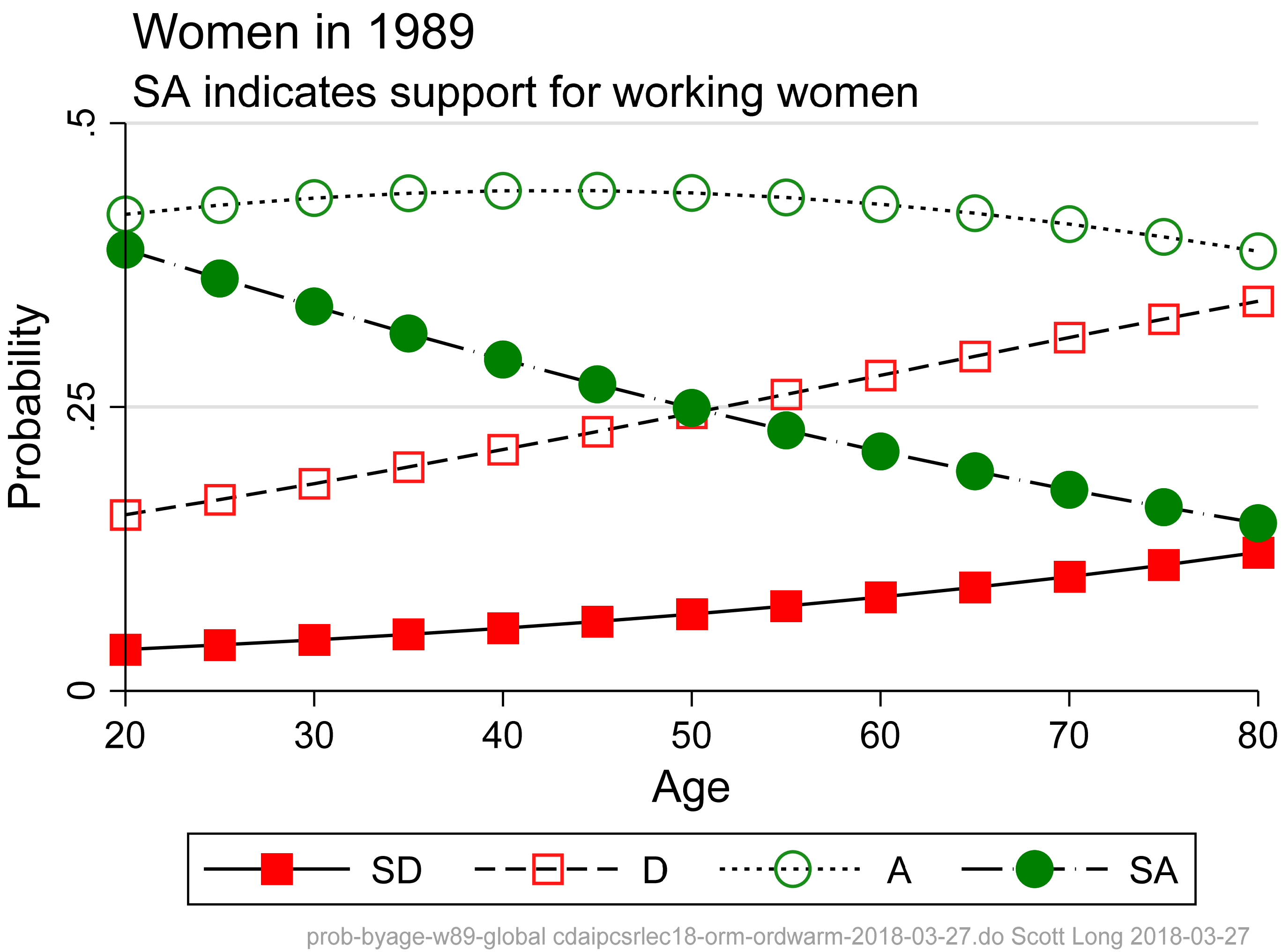
Plotting probabilities

1. Plots are more complicated than for the BRM since:
   - With <u>two</u> outcomes, you plot <u>one</u> probability
   - With <u>three</u> outcomes, you plot <u>three</u> probabilities
2. For ORM and MNLM we can plot probabilities and cumulative probabilities:

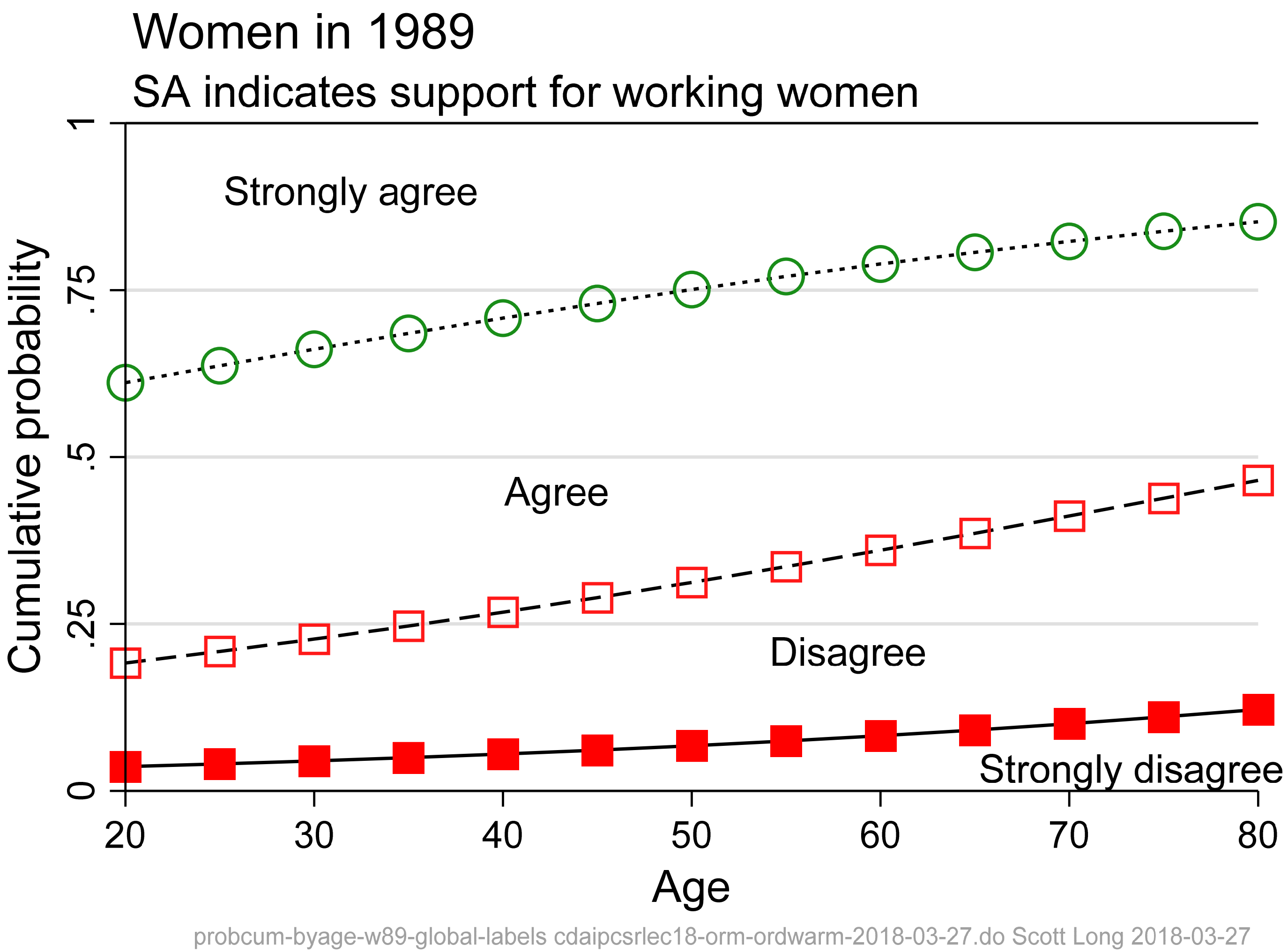
   Probability: $\Pr(y = q \mid z, \mathbf{X}^*)$ as z changes

   Cumulative probability: $\Pr(y \leq q \mid z, \mathbf{X}^*)$ as z changes
3. These are computed holding other variables constant while changing one.
   - Excepted for linked variables like x and x^2 change together

Predicted probabilities



Cumulative probabilities



Code for plots

Compute predictions with mgen

```
. mgen, at(age=(20(5)80) male=0 yr89=1) atmeans stub(W89)

Variable    Obs Unique       Mean        Min        Max  Label
----------------------------------------------------------------------------
W89pr1       13     13   .0720984   .0364676    .121933  pr(y=1_SD) from margins
W89ll1       13     13   .0586271   .0281642    .097223  95% lower limit
W89ul1       13     13   .0855696    .044771    .146643  95% upper limit
W89age       13     13         50         20         80  Age in years
W89Cpr1      13     13   .0720984   .0364676    .121933  pr(y<=1_SD)
W89pr2       13     13      .2465   .1551205   .3431755  pr(y=2_D) from margins
::
----------------------------------------------------------------------------

Specified values of covariates
                             1.
     yr89        male     white         ed        prst
------------------------------------------------------
        1           0  .8765809   12.21805    39.58526
label var W89pr1 "SD"
label var W89pr2 "D"
label var W89pr3 "A"
label var W89pr4 "SA"
```

Globals with characteristics for symbols

```
global warmsym "mcol(red red*.9 green*.9 green) "
global warmsym "$warmsym msym(s sh Oh O) msiz(3.5 3.5 3 3)" // symbol details
```

Plot predicted probabilities

```
graph twoway connected W89pr1 W89pr2 W89pr3 W89pr4 W89age, $warmsym ///
    title(Women in 1989, pos(11))  ///
    subtitle("SA indicates support for working women", pos(11)) ///
    xtitle("Age") xlab(20(10)80) ///
    ytitle("Probability") ylab(0(.25).50, grid gmin gmax) ///
    plotregion(margin(zero) lcol(white)) legend(rows(1)) scale(*1.1)
```

Plot cumulative probabilities with interior text labels

```
graph twoway connected W89Cpr1 W89Cpr2 W89Cpr3 W89age, $warmsym ///
    title(Women in 1989, pos(11)) $warmsym ///
    subtitle("SA indicates support for working women", pos(11)) ///
    xtitle("Age") xlab(20(10)80) yline(1, lcol(black) lwid(*.7)) ///
    ytitle("Cumulative probability") ylab(0(.25)1, grid gmin gmax) ///
    text(.035 65 "Strongly disagree", place(e)) ///
    text(.21 54 "Disagree", place(e)) ///
    text(.45 40 "Agree", place(e)) ///
    text(.90 25 "Strongly agree", place(e)) legend(off) ///
    plotregion(margin(zero) lcol(white)) scale(*1.1)
```

The parallel regression assumption

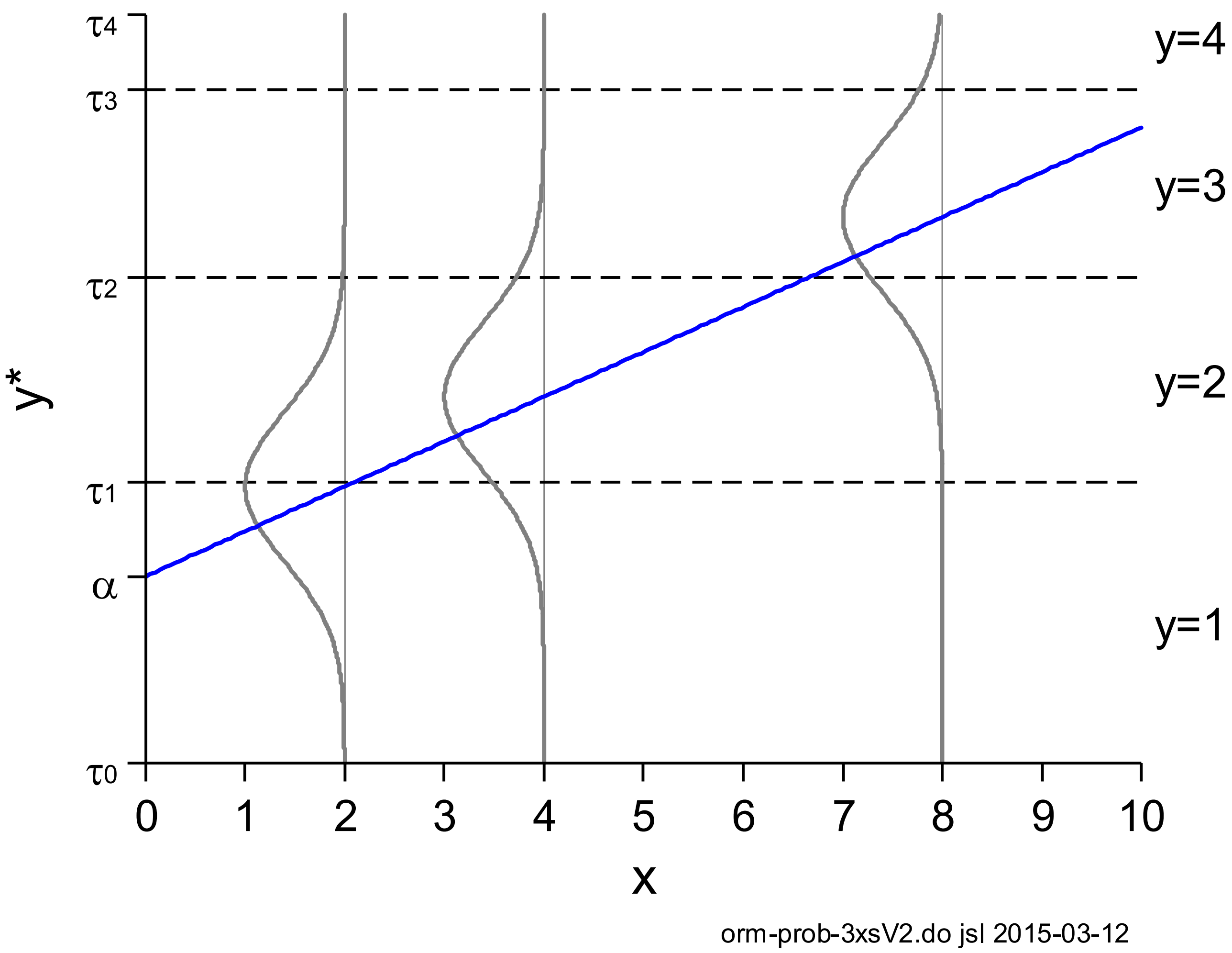
ORM can be thought of as constrained binary regressions

1. The outcome y has J categories.
2. J-1 binary variables are created:
   - $y_j = 1$ if $y \leq j$, else 0.
3. J-1 binary logits or probits are run on y_1, y_2, …
4. Slopes are constrained to be equal in all regressions.
5. The intercepts are allowed to differ.

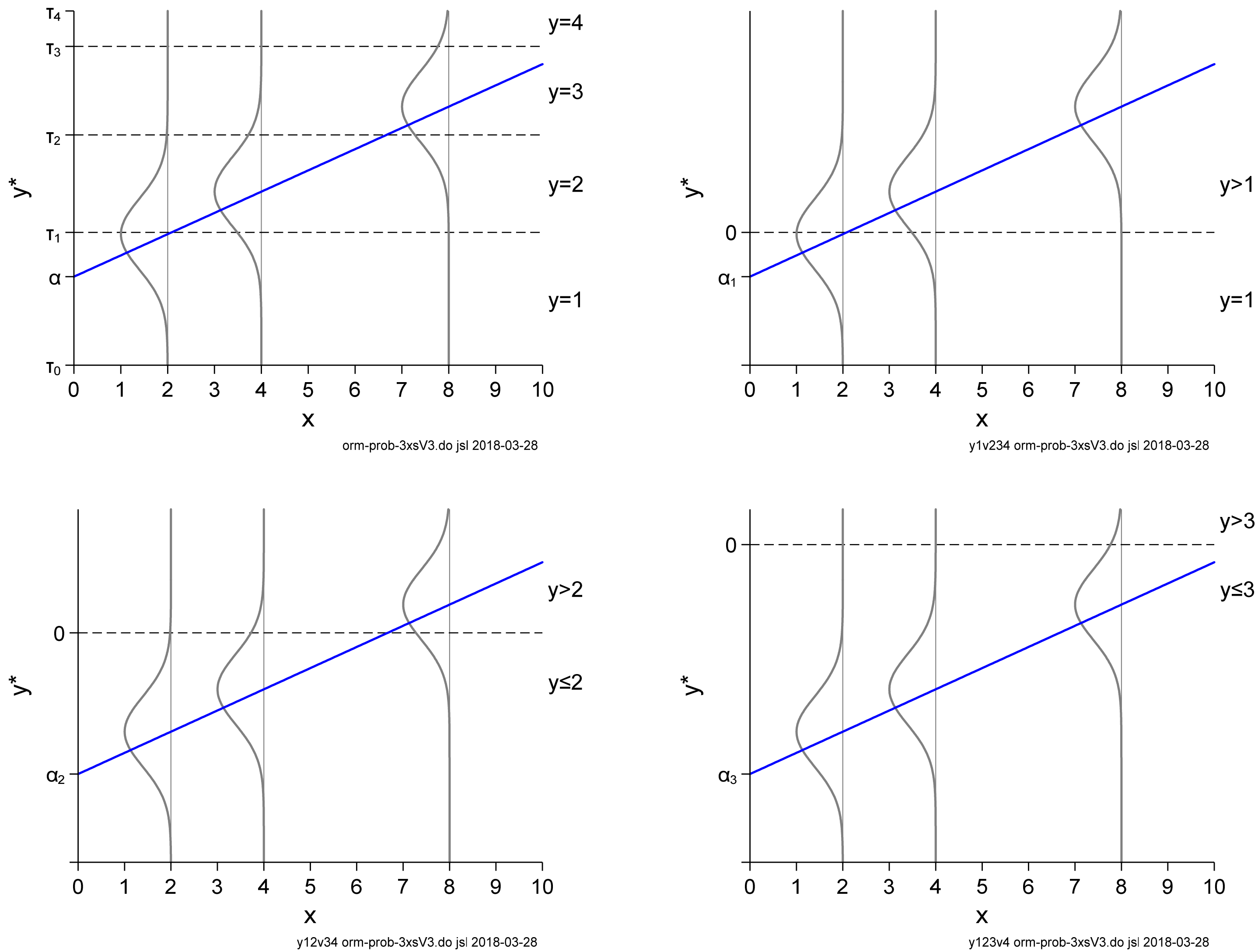
Proportional odds assumption

This property is called the proportional odds assumption since the ORs are the same in each BRM

The y* structural model for the ORM

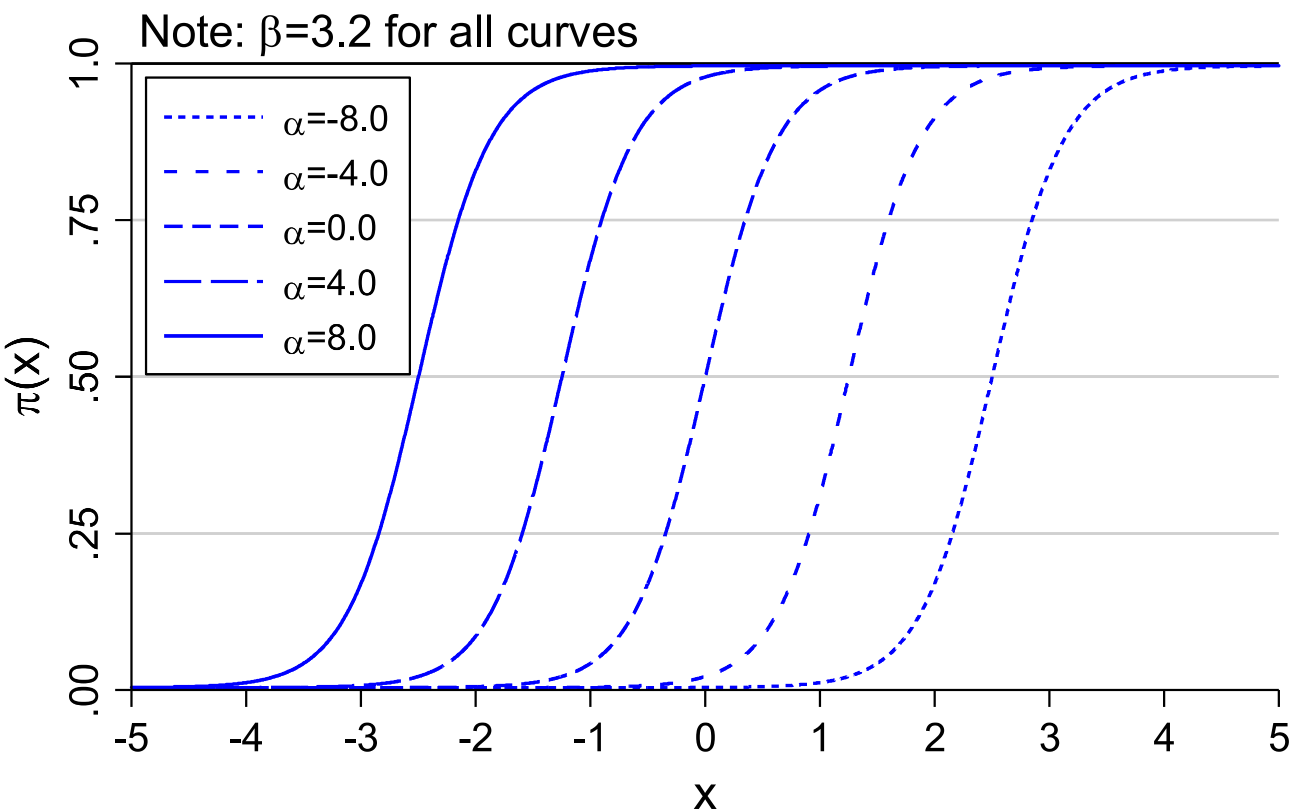


Different dichotomizations change the intercept, but not the slope



Parallel probability curves

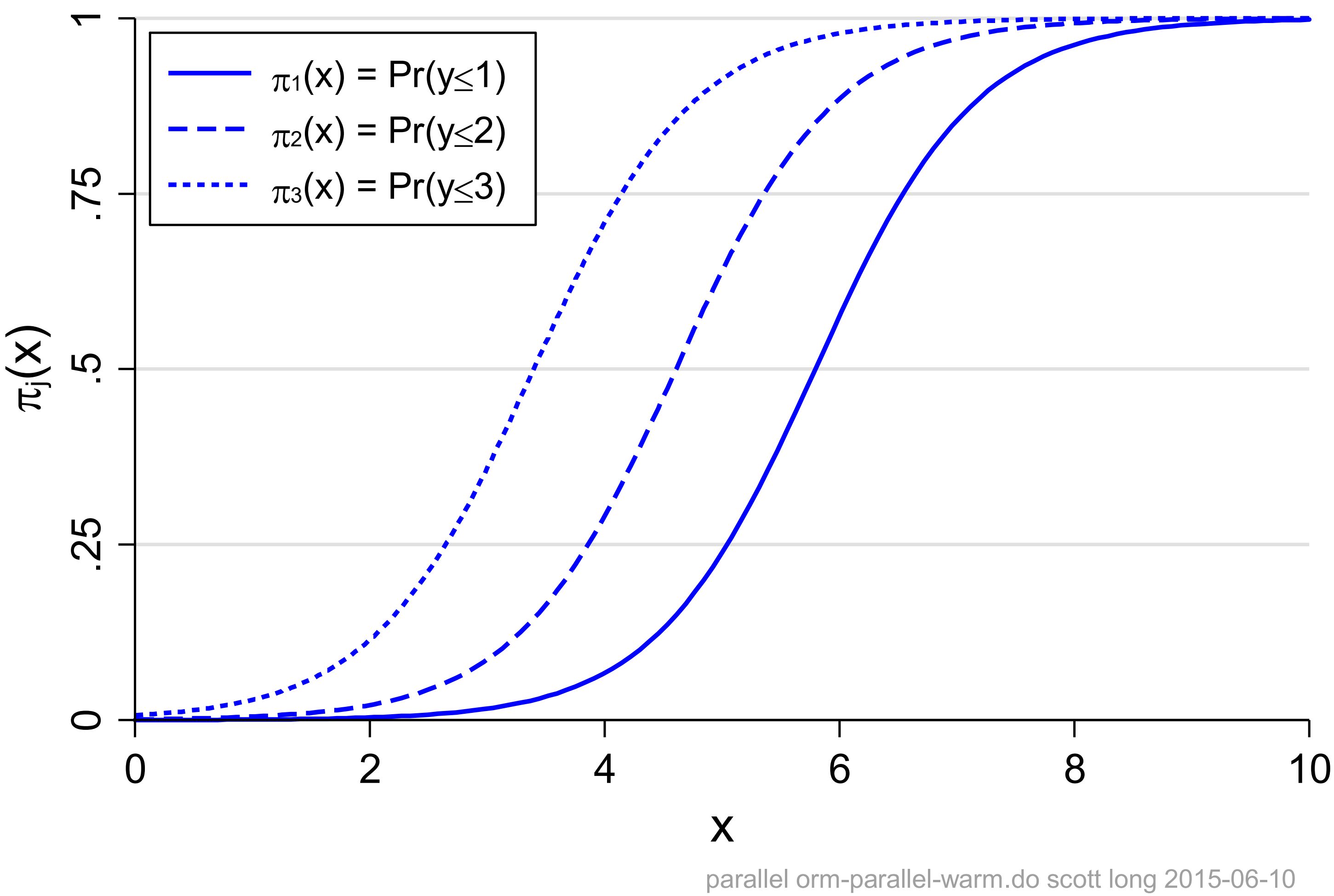
1. For the BRM we learned that if the intercept changes, the curve shifts
   - o That is, the curves are parallel



2. The ORM implies a set of parallel curves

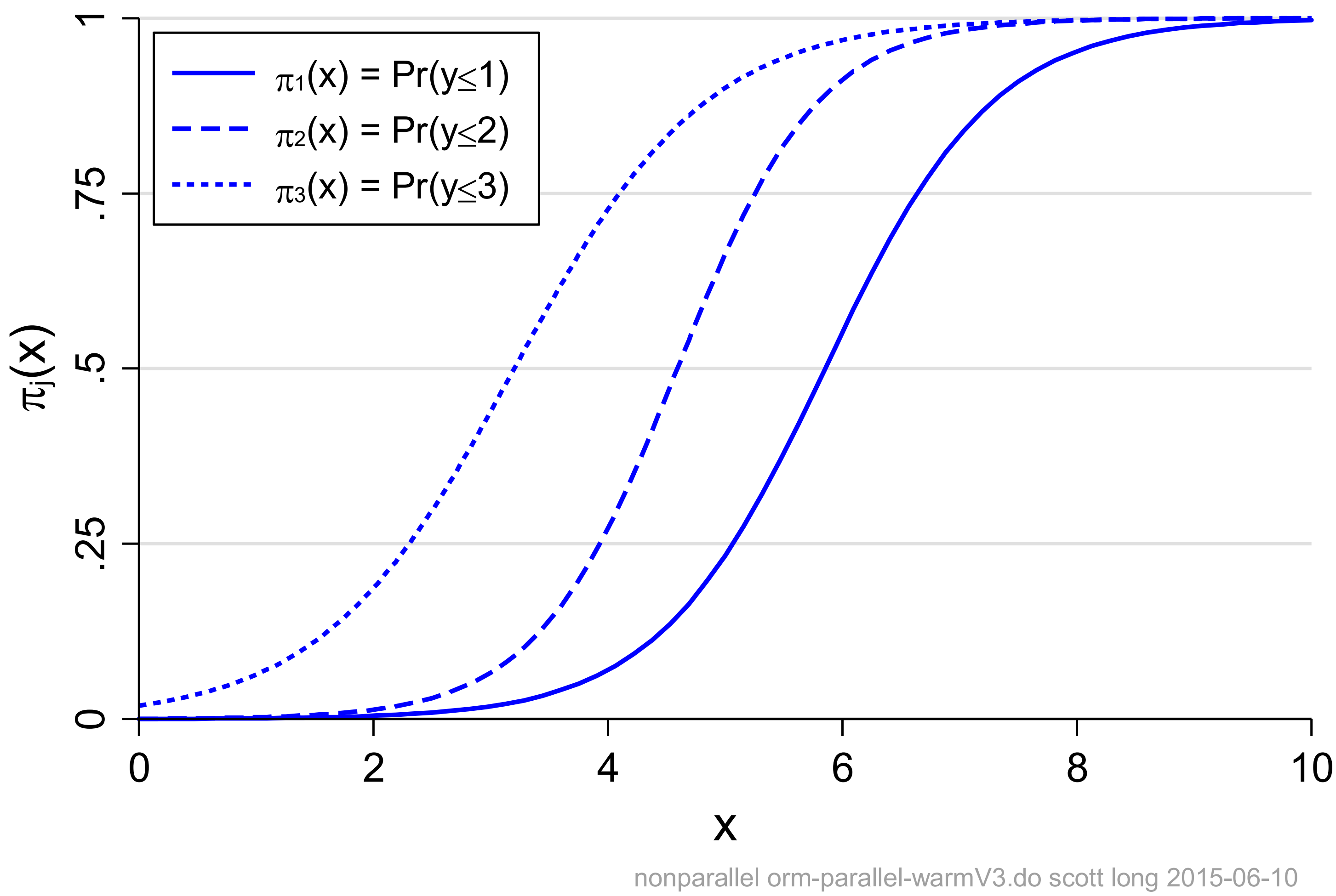
Parallel regressions: ORM implies BRM on y≤j with equal slopes

$\Pr(y{\color{red}j}\,|\,x) = \Pr(y \le {\color{red}j}\,|\,x) = F(\alpha_{\color{red}j} + \gamma_1 x_1 + \gamma_2 x_2)$



Without parallel regressions: BRM on y≤j without constraints

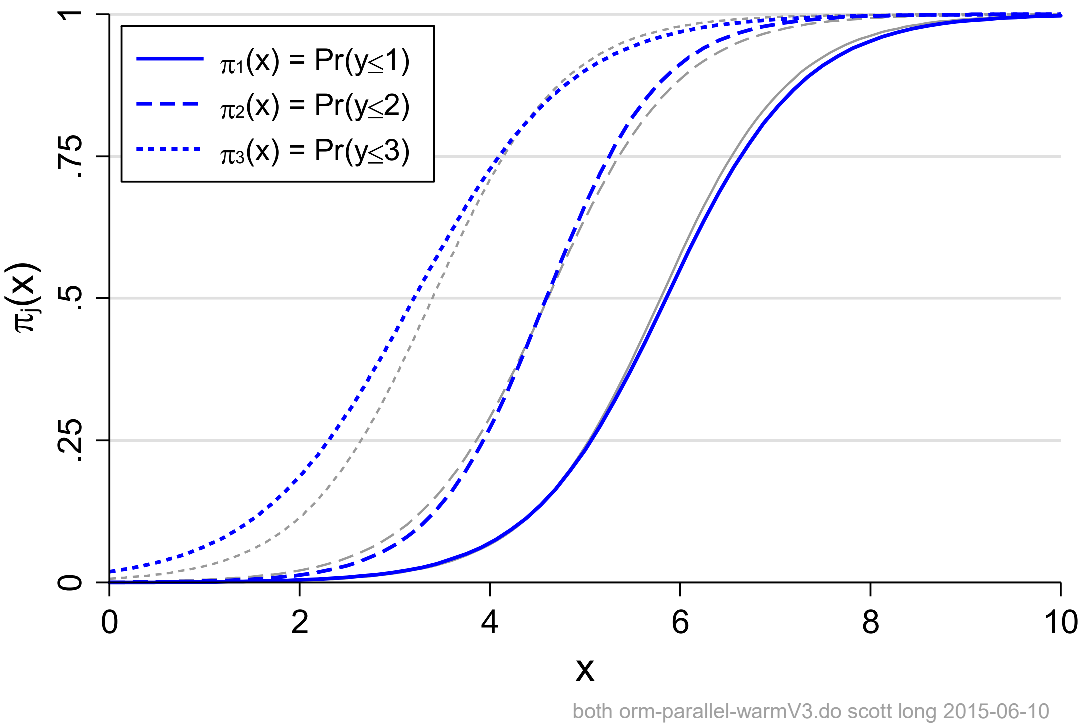
$\Pr(y\,j\,|\,x) = \Pr(y \leq j\,|\,x) = F(\alpha_j + \gamma_{j1}x_1 + \gamma_{j1}x_2)$



Informal assessment of parallel regressions

1. Informally compare slopes from J-1 binary regressions to those from ORM.

Variable	lt2	lt3	lt4	ologit
Odds ratios				
yr89	0.381***	0.568***	0.727**	1.689***
male	1.357*	1.995***	2.956***	0.480***
white	1.738*	1.369*	1.481*	0.676***
age	1.017***	1.026***	1.019***	0.979***
ed	0.901***	0.949**	0.944*	1.069***
prst	1.001	0.991*	0.994	1.006



Formal tests of parallel regressions

1. The Brant test formally compares the coefficients from BLM and ORM
2. LR tests can compare **`gologit2`** estimates with **`ologit`** estimates
3. Tests often reject the parallel regression assumption
   - Rejection <u>might</u> indicate a violation of the parallel regression assumption
   - Rejection could be caused by other violations of the model assumptions
   - The tests might be sensitive to sample size, small categories, or other factors. That is, they might not have good properties
4. I rely more on compare the results from an ordinal model to a model that is not ordinal, such as MNLM or GOLM.

Brant test of parallel regressions

```
Brant Test of Parallel Regression Assumption

    Variable |       chi2    p>chi2    df
-------------+--------------------------
         All |      49.18     0.000    12
-------------+--------------------------
        yr89 |      13.01     0.001     2  <= these are key variables
        male |      22.24     0.000     2  <= so I need to check them
       white |       1.27     0.531     2
         age |       7.38     0.025     2
          ed |       4.31     0.116     2
        prst |       4.33     0.115     2
----------------------------------------

A significant test statistic provides evidence that the parallel
regression assumption has been violated.
```

LR test of parallel regressions

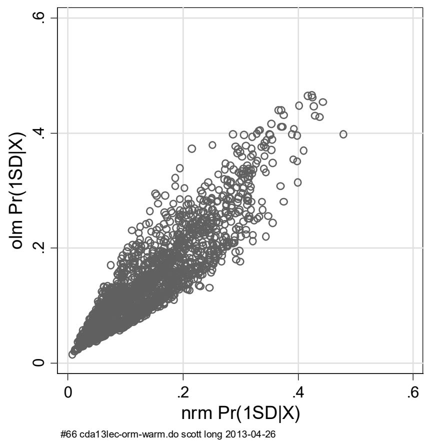
```
. gologit2 warm yr89 male white age ed prst, nolog
. est store gologit2
. lrtest ologit gologit2, force // force since not estimated with same command

Likelihood-ratio test                                 LR chi2(12) =     49.20
(Assumption: ologit nested in gologit2)               Prob > chi2 =    0.0000
```

Comparing ologit and mlogit predictions

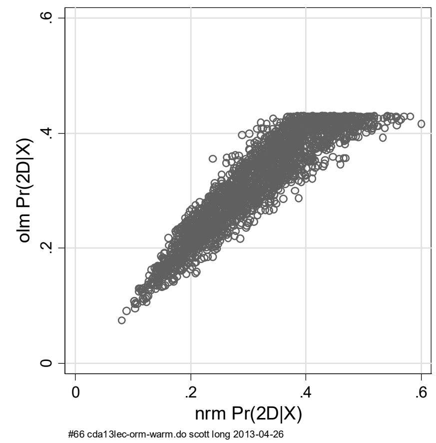
1. Before accepting or rejecting the OLM based on a test of parallel regressions, examine whether predictions from the OLM differ from those in a model that <u>does not impose ordinality</u>.
2. I typically compare OLM to MNLM. You could also compare to the generalized ordered logit model.
3. In general, the results are similar.
   - Whether differences are substantively important depend on your research goals and require more careful evaluation.

Comparing in sample predictions

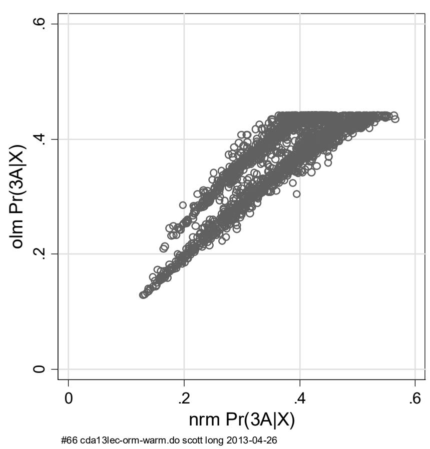
Correlation: 0.90



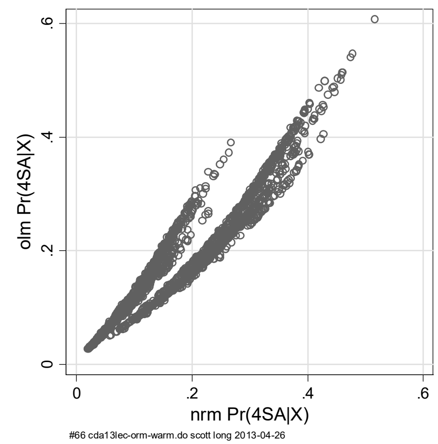
Correlation: 0.92



Correlation: 0.86



Correlation: 0.95



Code for examining parallel regressions assumption

Comparing coefficients from binary logits

```
ologit warm i.yr89 i.male i.white age ed prst
gen warmlt2 = warm<=1 & !missing(warm)
logit warmlt2 i.yr89 i.male i.white age ed prst
gen warmlt3 = warm<=2 & !missing(warm)
logit warmlt3 i.yr89 i.male i.white age ed prst
gen warmlt4 = warm<=3 & !missing(warm)
logit warmlt4 i.yr89 i.male i.white age ed prst
```

Brant test of parallel regressions

```
brant, detail
```

LR test of parallel regressions

```
ologit warm yr89 male white age ed prst
est store ologitmodel
gologit2 warm yr89 male white age ed prst, nolog
est store gologit2model
* force since compare models from different estimation commands
lrtest ologit gologit2, force
```

Comparing in sample predictions

```
ologit warm i.yr89 i.male i.white age ed prst
predict OLMpr1 OLMpr2 OLMpr3 OLMpr4
mlogit warm i.yr89 i.male i.white age ed prst
predict NRMpr1 NRMpr2 NRMpr3 NRMpr4
pwcorr OLMpr1 NRMpr1
pwcorr OLMpr2 NRMpr2
pwcorr OLMpr3 NRMpr3
pwcorr OLMpr4 NRMpr4
```

Plot predictions

```
scatter OLMpr1 NRMpr1
::
```

Compare other predictions

1. Compare plots from **mgen**
2. Compare DCs from **mchange**

Modeling political party

This example from the American National Election Study taught me to never assume an outcome is ordinal.

Party ID	Freq.	Percent	Cum.
StrDem	266	19.25	19.25
Dem	427	30.90	50.14
Indep	151	10.93	61.07
Rep	369	26.70	87.77
StrRep	169	12.23	100.00
Total	1,382	100.00	

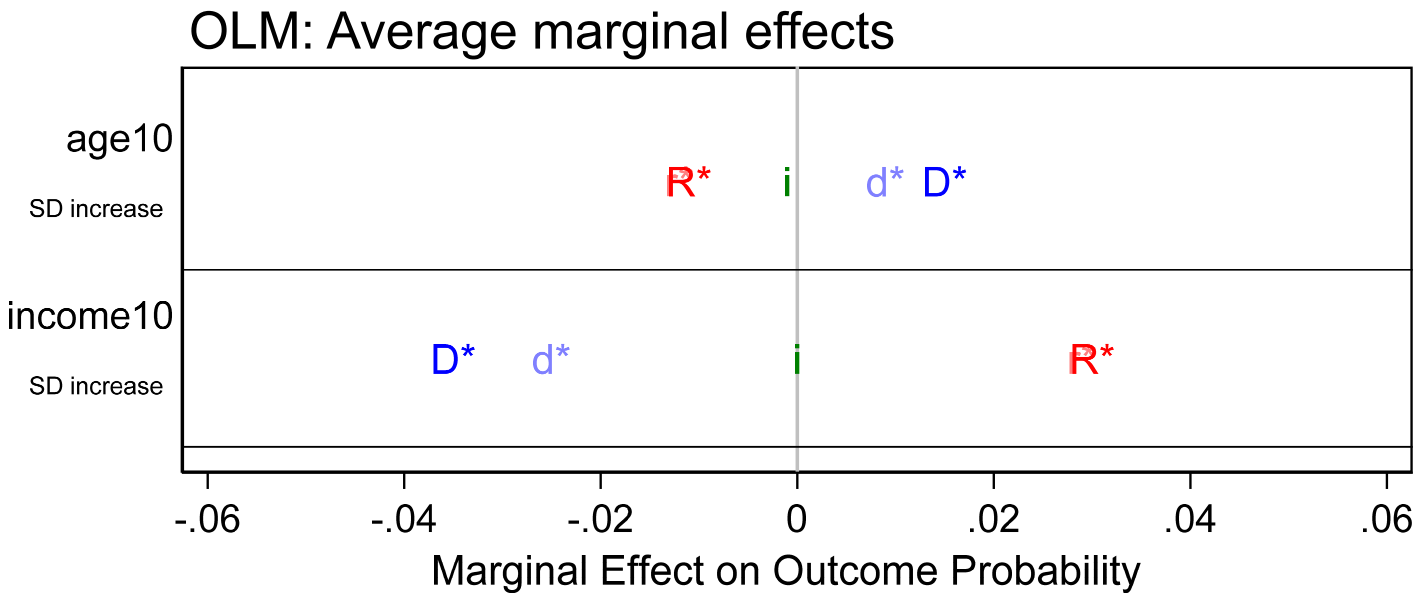
Variable	Obs	Mean	Std. Dev.	Min	Max
party	1382	2.817656	1.342787	1	5
age	1382	45.94645	16.78311	18	91
income	1382	37.45767	27.78148	1.5	131.25
black	1382	.1374819	.34448	0	1
female	1382	.4934877	.5001386	0	1
educ					
hs only	1382	.5803184	.4936854	0	1
college	1382	.2590449	.4382689	0	1

OLM

```
. ologit party age10 income10 i.black i.female i.educ

Ordered logistic regression                     Number of obs     =      1,382
                                                LR chi2(6)        =     212.68
                                                Prob > chi2       =     0.0000
Log likelihood = -2010.1976                     Pseudo R2         =     0.0502

------------------------------------------------------------------------------
       party |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       age10 |  -.0635915   .0312192    -2.04   0.042      -.12478   -.002403
    income10 |   .0961097   .0200567     4.79   0.000     .0567993     .13542
::
```

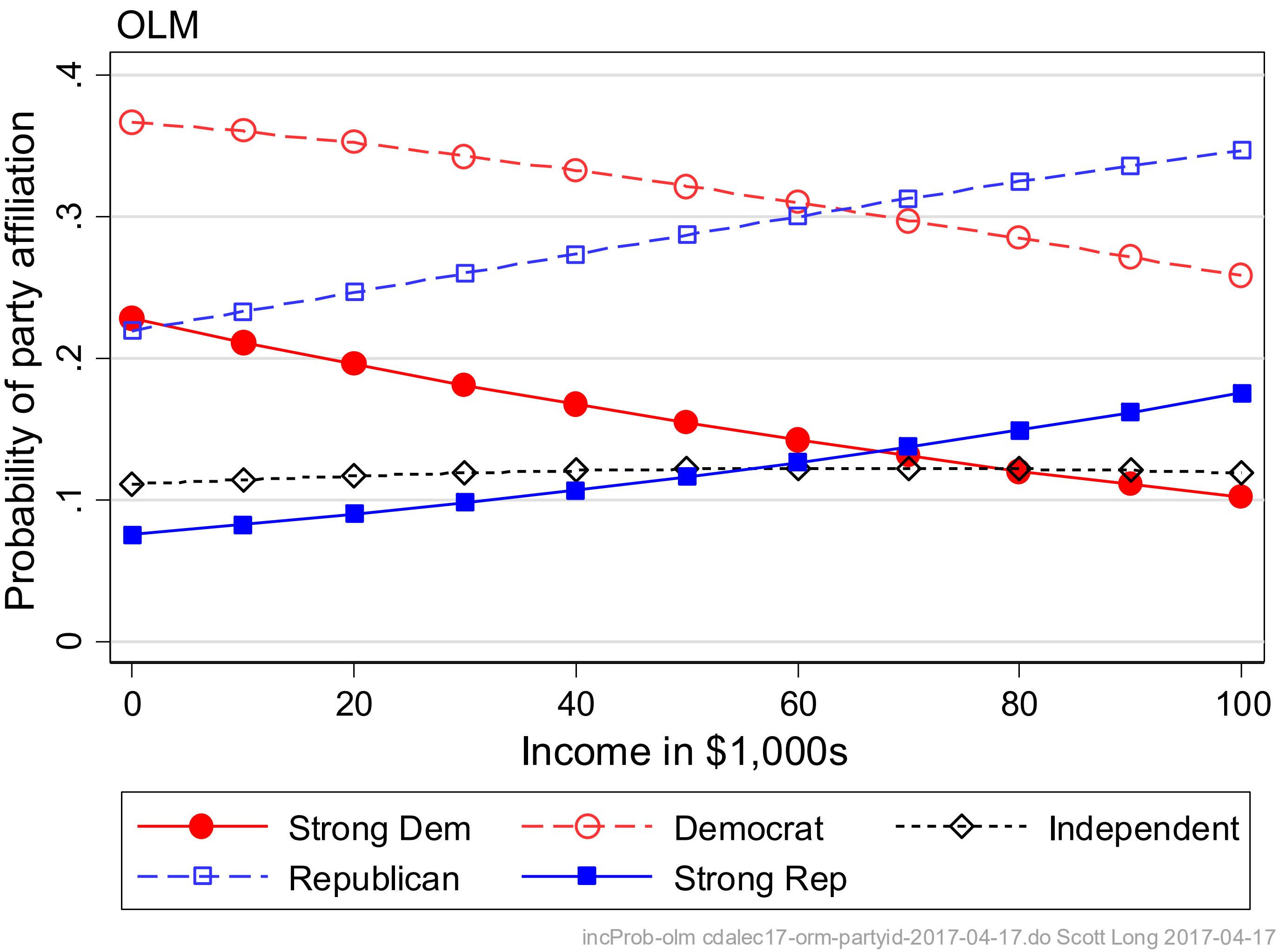
1. Substantive research would be concerned with the effects of all regressors.
2. For methodological reasons, I will focus on age and income
   - **`age10`** is years in ten-year units white is signifcant at the .04 level.
   - **`income10`** is income in $10,000 units which is significant at the .001 levels.
3. Consider the ADC(age) and ADC(income)

OLM: Average discrete change

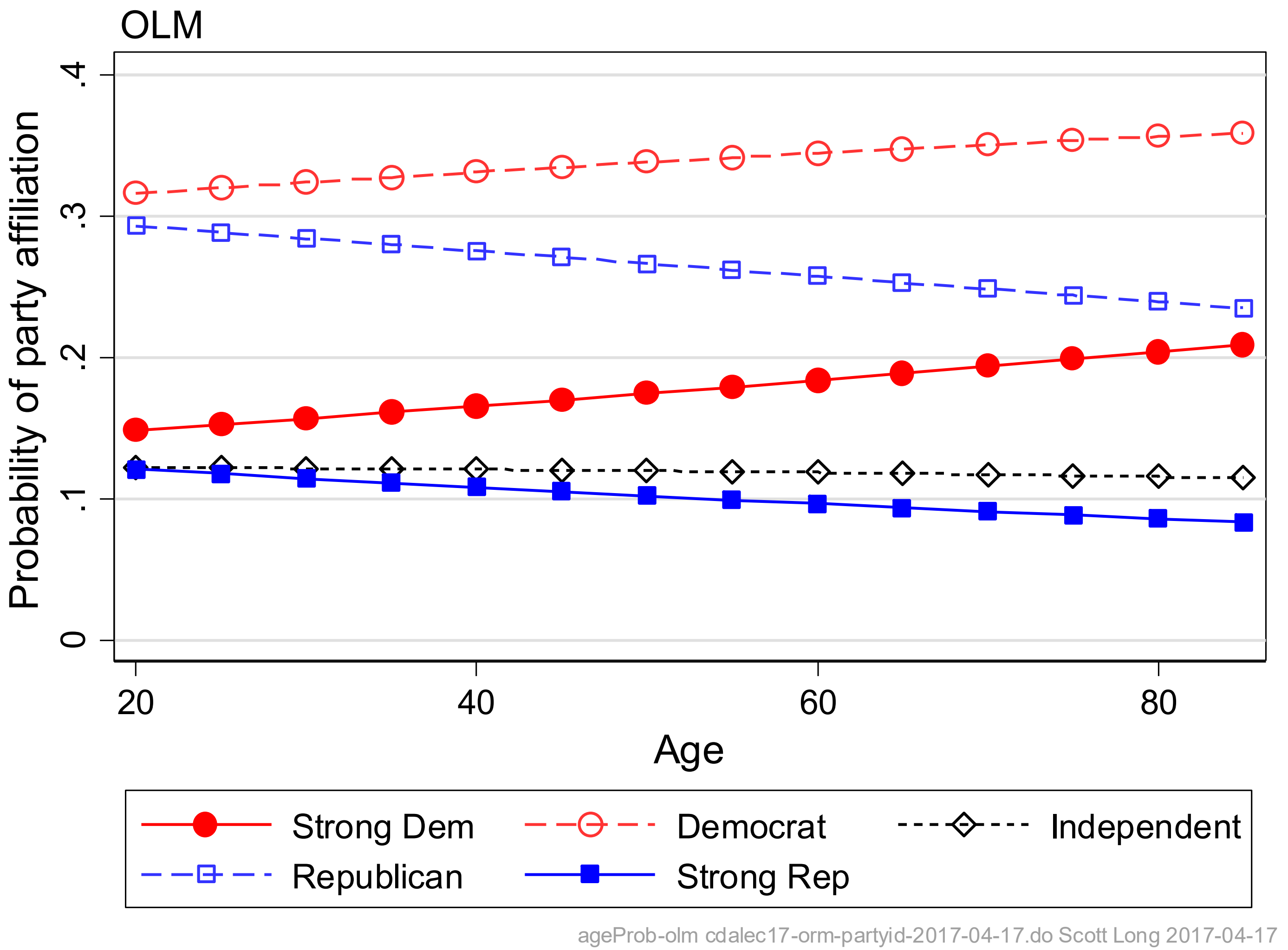


1. Income significantly decreases the probabilities of both Democratic affiliations and significantly increases the probabilities of both Republican affiliations.
2. The effects of age are smaller, with age significantly increases the probabilities of both Democratic affiliations and significantly decreases the probabilities of both Republican affiliations.
3. Graphs of predicted probabilities show that since the relationships are nearly linear, the ADCs are good summaries of the effects.

OLM: Predicted probabilities by income



OLM: Predicted probabilities by age



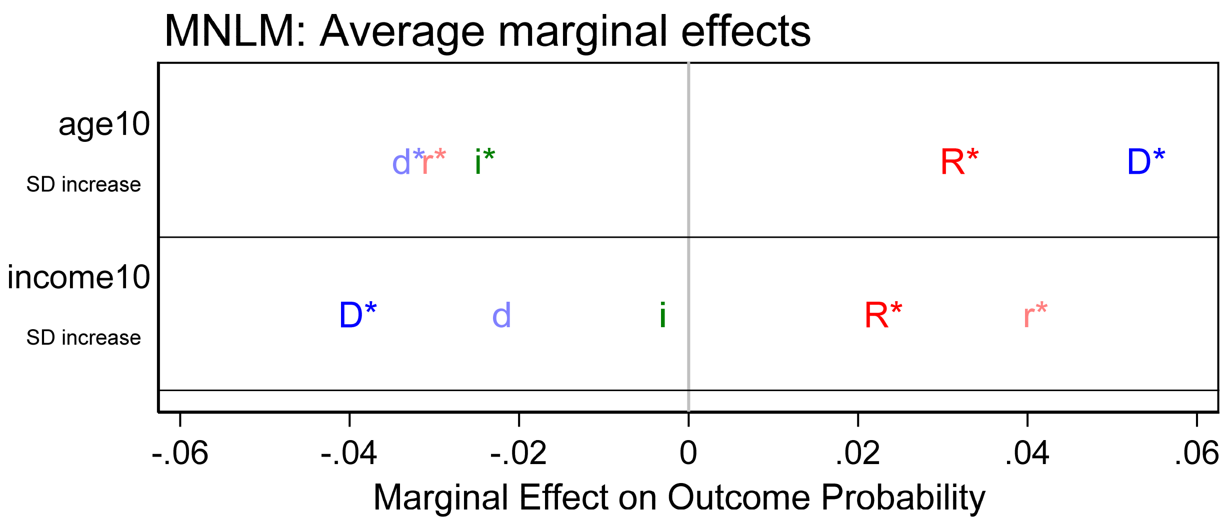
MNLM model

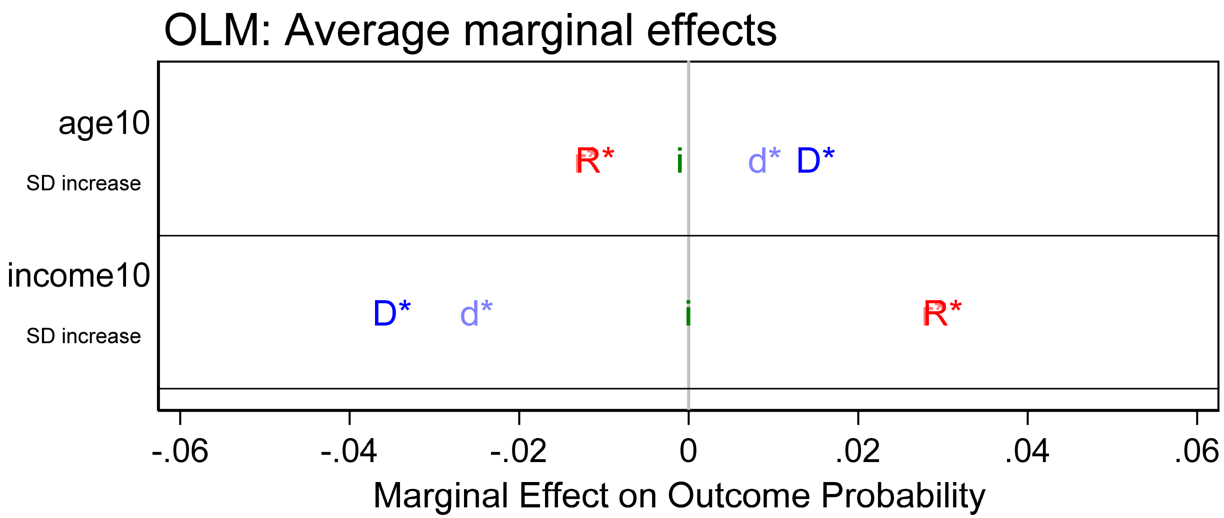
1. We fit this model in the lecture on MNLM.
2. Both age and income are were significant at the .001 levels:

```
Wald tests for independent variables (N=1382)

  Ho: All coefficients associated with given variable(s) are 0

                 |      chi2     df    P>chi2
-----------------+-------------------------
           age10 |    43.815      4     0.000       p=.042 for OLM
        income10 |    22.985      4     0.000       p=.000 for OLM
```

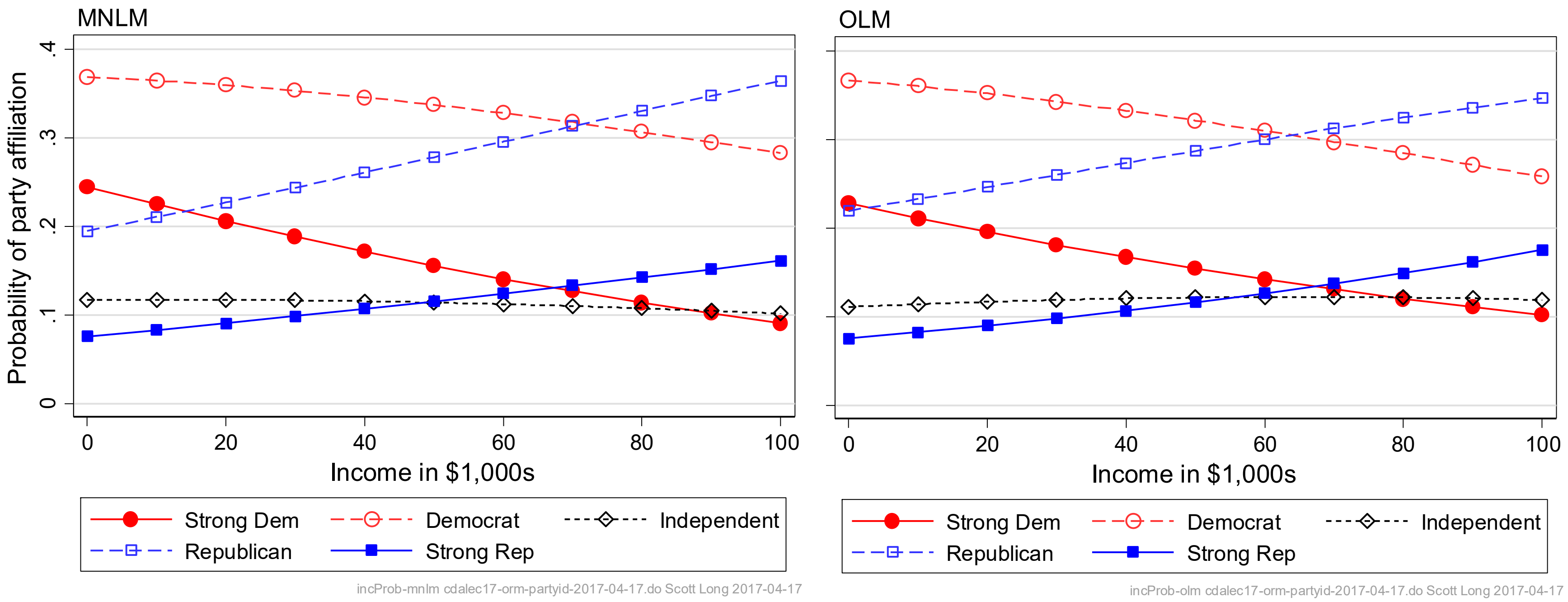
3. Let’s compare the ADCs.

MNLM vs OLM: ADC of age and income are very different



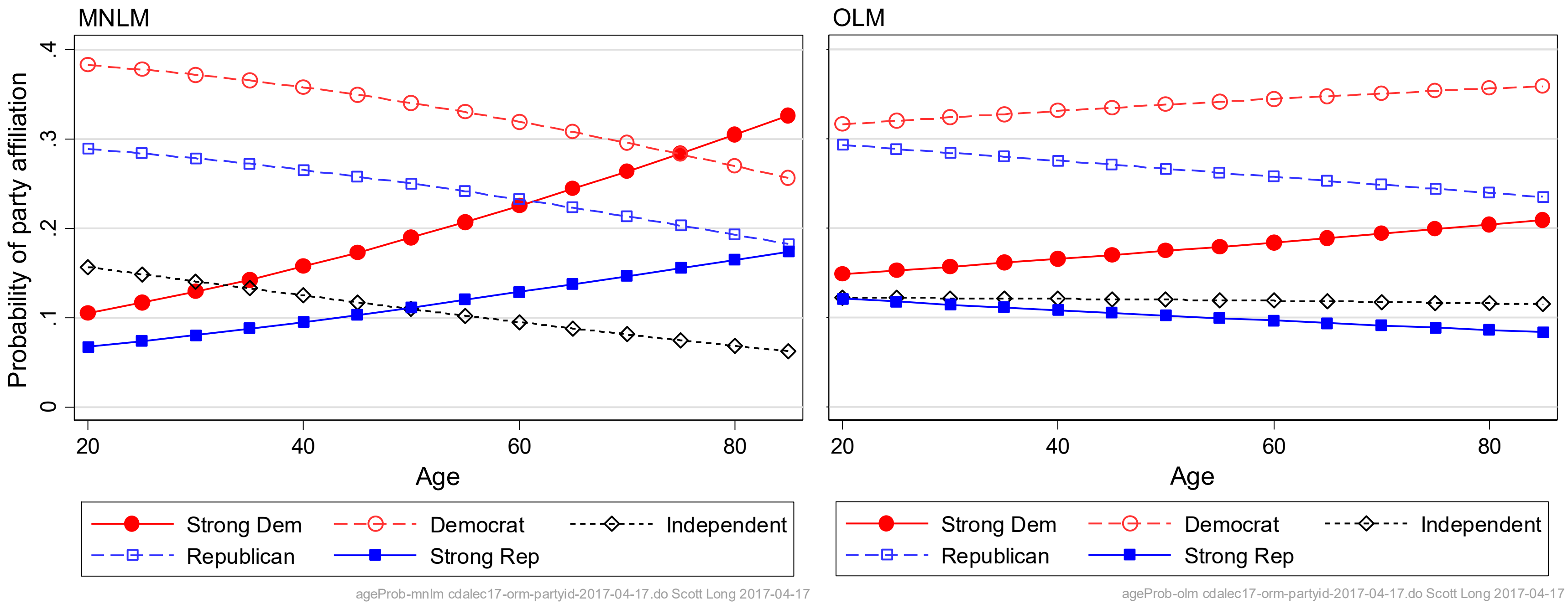


OLM & MNLM: Plots for income



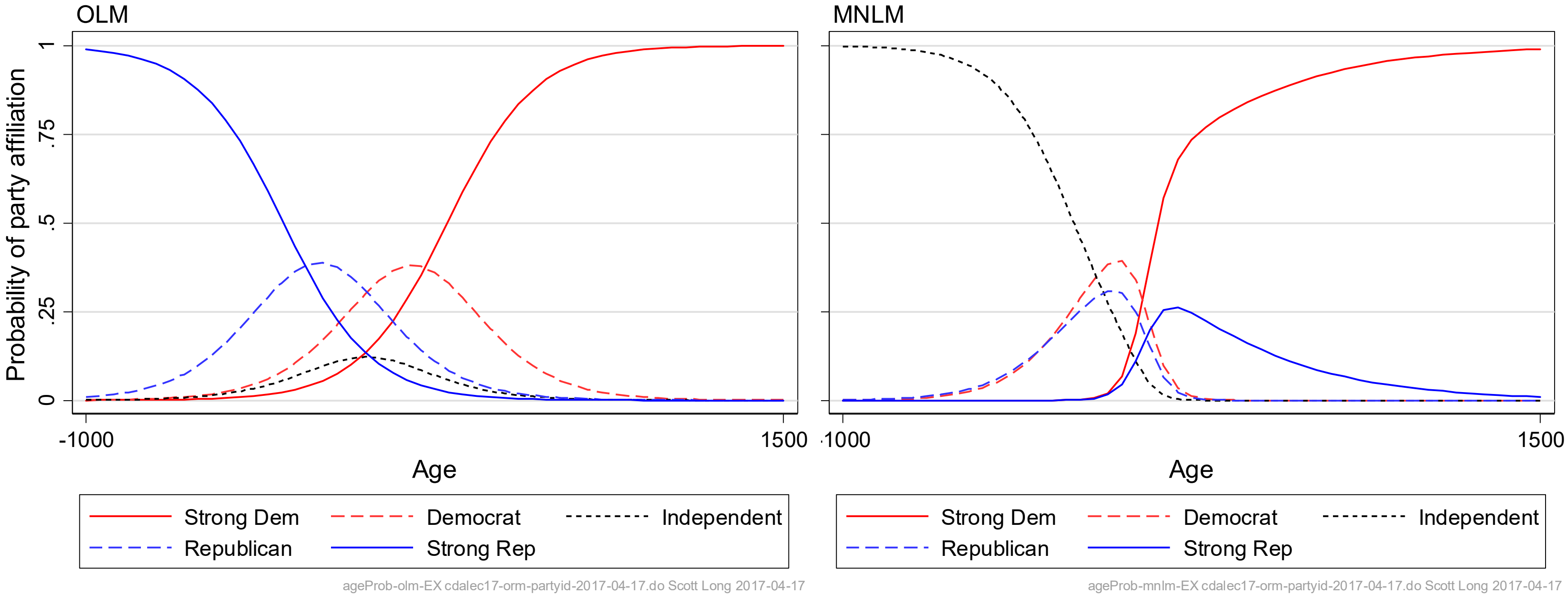
Unless you need very precise predictions, the results are eqivalent.

OLM & MNLM: Plots for age



1. The conclusions are very different.
2. What causes the difference?

Extending the range of predictions



1. Ordinal models must have the pattern described by Anderson
   - This is shown on the left
   - This <u>forces</u> the probabilities of end categories to either always get larger or always get smaller
2. Nominal models do not have this constraint on predictions

* Comparing MEs from ologit and mlogit

1. To test if results from OLM & MNLM differ, we estimate them simultaneously
2. Create two copies of the outcome variable to trick **gsem**

```
clonevar partyOLM = party
clonevar partyMNL = party
```

3. Estimate both models simultaneously:

```
gsem (partyOLM <- c.age10 c.income10 i.black i.female i.educ, ologit) ///
     (partyMNL <- c.age10 c.income10 i.black i.female i.educ, mlogit), ///
     vce(robust)
est store gsemmodel
```

Comparing DCM(age+SD)

	lincom	se	pvalue
StrDem			
OLM	0.015	0.008	0.064
MNL	0.054	0.010	0.000
Difference	-0.039***	0.007	0.000
Dem			
OLM	0.012	0.006	0.057
MNL	-0.030	0.014	0.028
Difference	0.042***	0.013	0.001
Indep			
OLM	-0.002	0.001	0.099
MNL	-0.026	0.010	0.010
Difference	0.024**	0.010	0.015
Repub			
OLM	-0.015	0.008	0.063
MNL	-0.026	0.013	0.050
Difference	0.011	0.012	0.348
StrRep			
OLM	-0.010	0.005	0.055
MNL	0.027	0.009	0.002
Difference	-0.037***	0.007	0.000

Comparing DCM(inc+sd)

	lincom	se	pvalue
StrDem			
OLM	-0.038	0.008	0.000
MNL	-0.046	0.015	0.002
Difference	0.009	0.013	0.500
Dem			
OLM	-0.029	0.006	0.000
MNL	-0.021	0.016	0.190
Difference	-0.008	0.015	0.619
Indep			
OLM	0.004	0.001	0.001
MNL	-0.003	0.012	0.799
Difference	0.007	0.012	0.554
Repub			
OLM	0.037	0.008	0.000
MNL	0.048	0.013	0.000
Difference	-0.010	0.011	0.363
StrRep			
OLM	0.025	0.005	0.000
MNL	0.023	0.008	0.004
Difference	0.002	0.005	0.727

Code

Code to compare DCM's for age

Set up margins command

```
sum age10
local mnage = r(mean)
local sdage  = r(sd)
local start = `mnage'-(`sdage'/2)
local end   = `mnage'+(`sdage'/2)
```

Compute predictions

```
. margins, at(age=(`start' `end')) atmeans post
::
1._predict   : Predicted mean (1.partyOLM), predict(pr outcome(partyOLM 1))
::
10._predict  : Predicted mean (5.partyMNL), predict(pr outcome(partyMNL 5))

1._at        : age10            =     3.75549
               income10         =    3.745767 (mean)
               0.black          =    .8625181 (mean)
               1.black          =    .1374819 (mean)
               0.female         =    .5065123 (mean)
               1.female         =    .4934877 (mean)
               1.educ           =    .1606368 (mean)
               2.educ           =    .5803184 (mean)
               3.educ           =    .2590449 (mean)
```

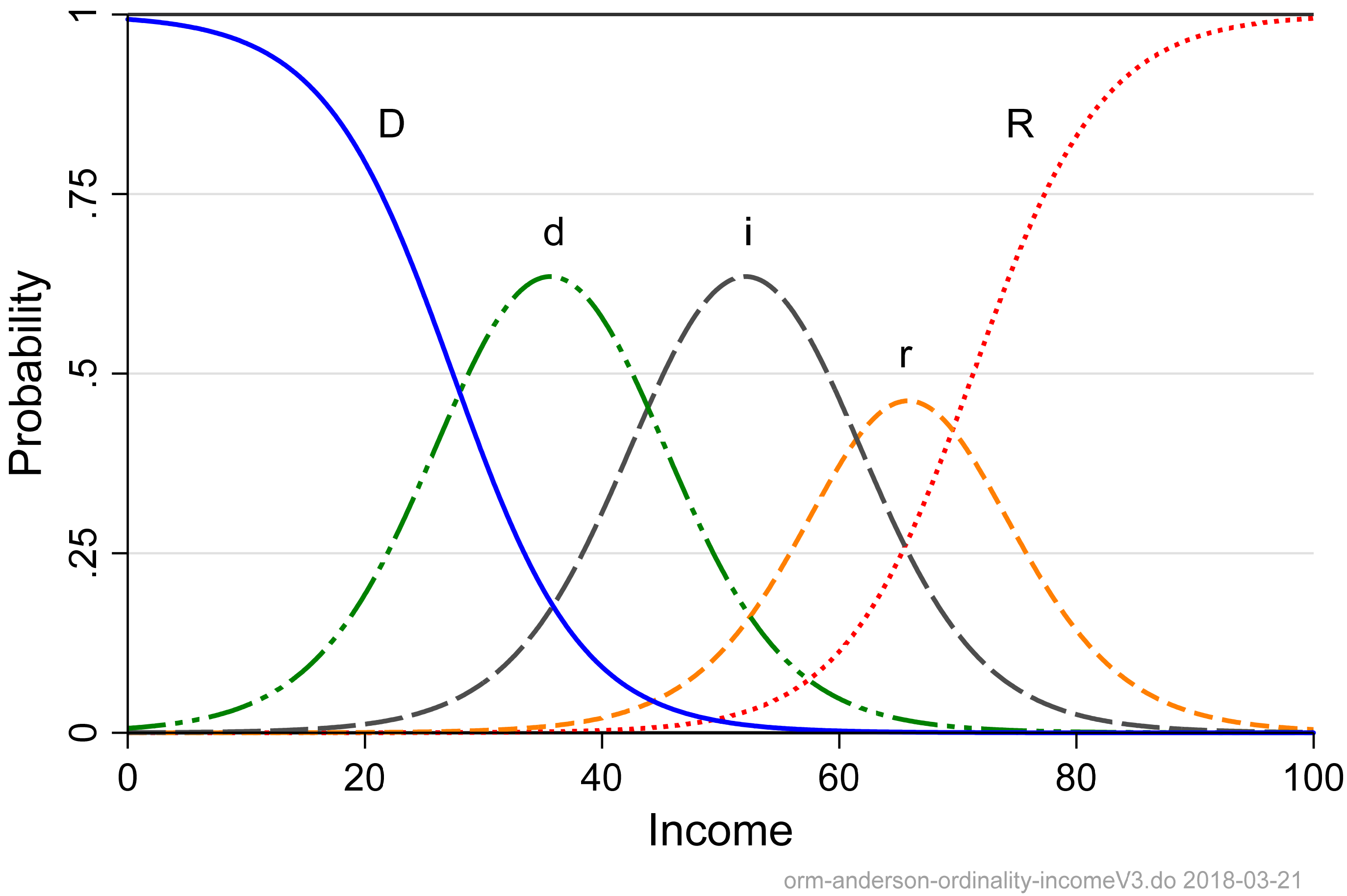
```
::
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_predict#_at |
        1 1  |   .1631553   .0100932    16.16   0.000     .1433729    .1829377
        1 2  |   .1782558   .0118232    15.08   0.000     .1550827    .2014288
        2 1  |   .3292235   .0137389    23.96   0.000     .3022958    .3561512
        2 2  |   .3407948   .0143541    23.74   0.000     .3126612    .3689283
::
        9 1  |   .2686426   .0163059    16.48   0.000     .2366836    .3006017
        9 2  |   .2425236   .0156883    15.46   0.000     .2117751     .273272
       10 1  |   .0913318   .0111995     8.15   0.000     .0693812    .1132825
       10 2  |   .1187061   .0126255     9.40   0.000     .0939606    .1434516
------------------------------------------------------------------------------
```

Computing DC's for each model and then second differences

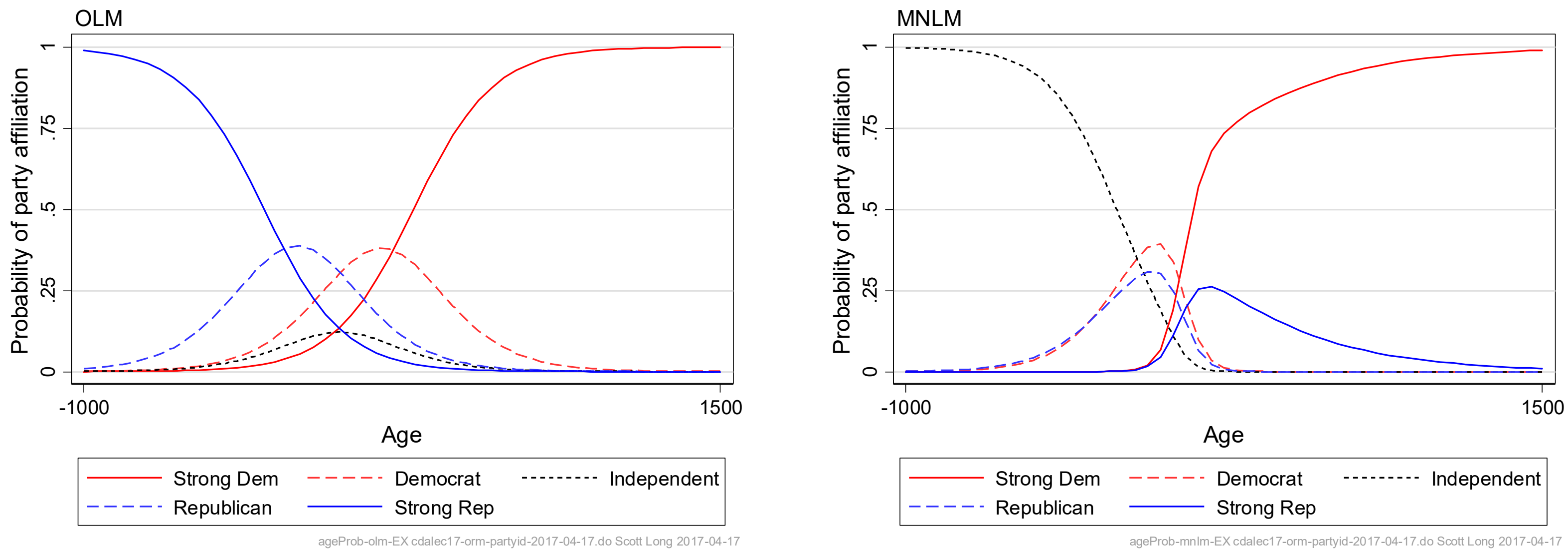
```
qui {
    mlincom 2-1,             rowname("StrDem:OLM") stat(est se p) clear
    mlincom 12-11,           rowname("StrDem:MNL") stat(est se p) add
    mlincom (2-1)-(12-11),   rowname("StrDem:Difference") stat(est se p) add
::
    mlincom 10-9,            rowname("StrRep:OLM") stat(est se p) add
    mlincom 20-19,           rowname("StrRep:MNL") stat(est se p) add
    mlincom (10-9)-(20-19),  rowname("StrRep:Difference") stat(est se p) add
}
mlincom, stat(est se p) twidth(15) dec(3) ///
    title("DC age for SD Change at Mean Across OLM and MNL")
```

Ordinal or nominal?

1. I find ordinal regression models are often overly restrictive even if the outcome <u>seems</u> to be ordinal.



2. Ordinality is relaxed in the MNLM and the generalized ordered logit model
3. I prefer MNLM because I find the results more intuitive
4. Returning to our example of political party, compare what happens in the OLM and MNLM if we extend the range of age:



5. Any time you are use an ordinal regression model, compare the result to those from a model that is not ordinal.

Overview of ordinal LHS

1. If you are using LRM for ordinal outcomes, consider y* standardized coefficients from the ORM (below).
   - If you *must* use the LRM, at least verify that the conclusions are consistent with those from the ORM.
2. Before using ordinal models, consider whether your variable is ordinal as Stevens defined it.
   - Categories are ranked on a <u>single dimension</u>
3. <u>Always do a sensitivity analysis</u> before accepting the results of ORM.
   - Compare results to those from a nominal model (MNLM or GOLM).
4. Even if you don't find the ORM useful for your work, this model is the foundation for the IRT and Rasch models for ordinal indicators.

Other methods of interpretation

1. I prefer methods based on predictions and marginal effects
2. Two other approaches can be used
   - Regression coefficients for y*
   - Odds ratios for the odds of larger values of the response

Interpretation with marginal change in y*

1. If you are plan to use LRM with ordinal outcomes, this is a better solution.
   - Sometimes it gives you similar results, but sometimes they differ.
2. In the ORM,

$$y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \varepsilon$$

3. The scale of y* is not identified, so use fully standardized ory*-standardized coefficients.
4. To estimate $Var(\hat{y}^*)$

$$Var\left(\hat{y}^*\right) = Var\left(\hat{\beta}x + e\right) = Var\left(\hat{\beta}x\right) + Var\left(e\right) + 2Cov\left(\hat{\beta}x, e\right)$$

$$= \hat{\beta}^2 Var\left(x\right) + Var\left(e\right) + 0 \quad \text{where } Var\left(e\right) \text{ is assumed.}$$

5. Generalizing:

$$\hat{\sigma}^2_{y^*} = \hat{\boldsymbol{\beta}}' \mathrm{Var}\left(\mathbf{x}\right) \hat{\boldsymbol{\beta}} + \mathrm{Var}\left(\varepsilon\right)$$

6. The y* standardized coefficient

$$\beta_k^{Sy^*} = \frac{\beta_k}{\sigma_{y^*}}$$

For a unit increase in x_k, y^* *is expected to increase by* $\beta_k^{Sy^*}$ *standard deviations holding other variables constant.*

7. The fully standardized coefficient is:

$$\beta_k^S = \frac{\sigma_k \beta_k}{\sigma_{y^*}} = \sigma_k \beta_k^{Sy^*}$$

For a standard deviation increase in x_k, y^* *is expected to increase by* β_k^S *standard deviations, holding other variables constant.*

y* standardized coefficients

```
. ologit warm i.yr89 i.male i.white age ed prst
. listcoef, help std

ologit (N=2293): Unstandardized and standardized estimates

  Observed SD:   0.9282
    Latent SD:   1.9411

          |        b         z     P>|z|    bStdX    bStdY   bStdXY     SDofX
----------+----------------------------------------------------------------
    yr89 |   0.5239    6.557    0.000    0.257    0.270    0.132    0.490
    male |  -0.7333   -9.343    0.000   -0.366   -0.378   -0.188    0.499
   white |  -0.3912   -3.304    0.001   -0.129   -0.202   -0.066    0.329
     age |  -0.0217   -8.778    0.000   -0.364   -0.011   -0.187   16.779
      ed |   0.0672    4.205    0.000    0.212    0.035    0.109    3.161
    prst |   0.0061    1.844    0.065    0.088    0.003    0.045   14.492
---------------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
   bStdX = x-standardized coefficient
   bStdY = y-standardized coefficient
  bStdXY = fully standardized coefficient
   SDofX = standard deviation of X. listcoef, help std
```

Interpretations follow…

Examples

	b	z	P>|z|	bStdX	bStdY	bStdXY	SDofX
1.yr89	0.52390	6.557	0.000	0.2566	0.2699	0.1322	0.4897

In 1989 support was .27 standard deviations higher than in 1977, holding other variables constant.

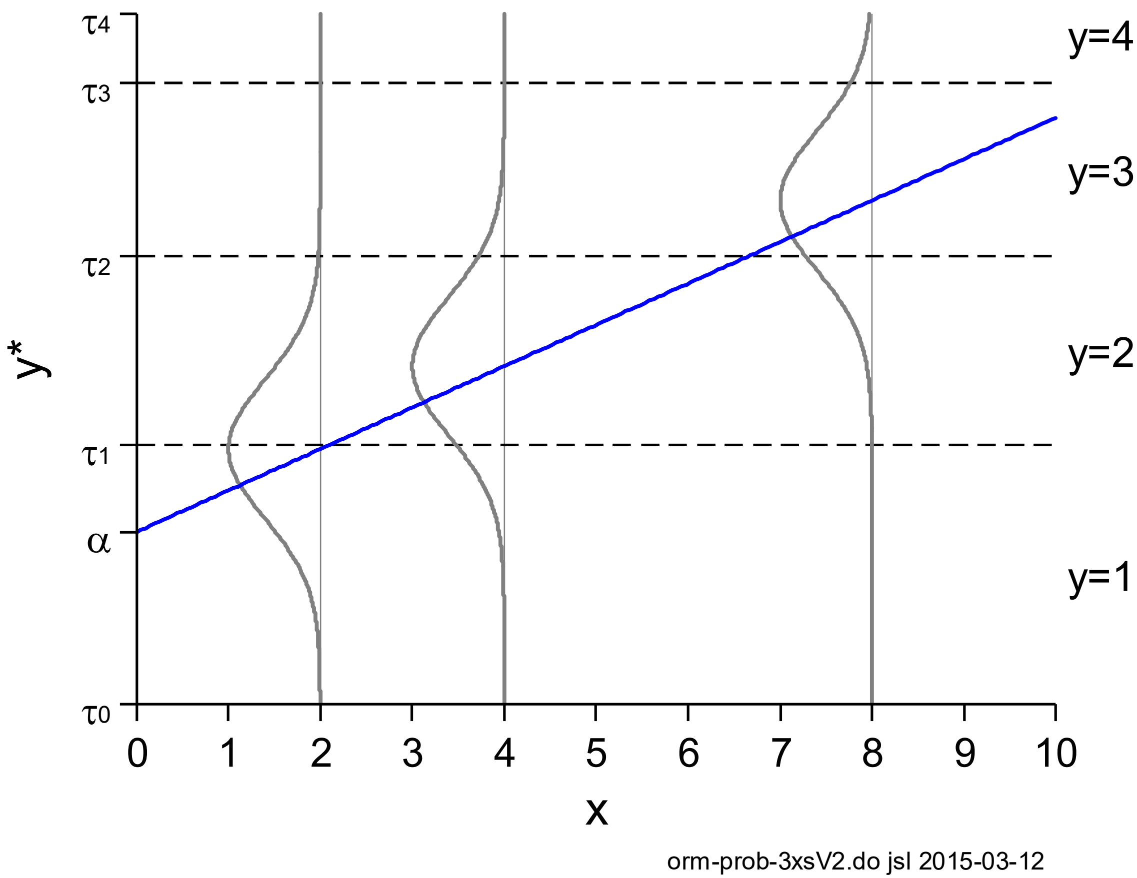
	b	z	P>|z|	bStdX	bStdY	bStdXY	SDofX
age	-0.02167	-8.778	0.000	-0.3635	-0.0112	-0.1873	16.7790

Each additional year of age decreases support by .01 standard deviations, holding other variables constant. Alternatively, each additional ten years of age decreases support by .11 standard deviations (=10x-.011), holding other variables constant.

	b	z	P>|z|	bStdX	bStdY	bStdXY	SDofX
ed	0.06717	4.205	0.000	0.2123	0.0346	0.1094	3.1608

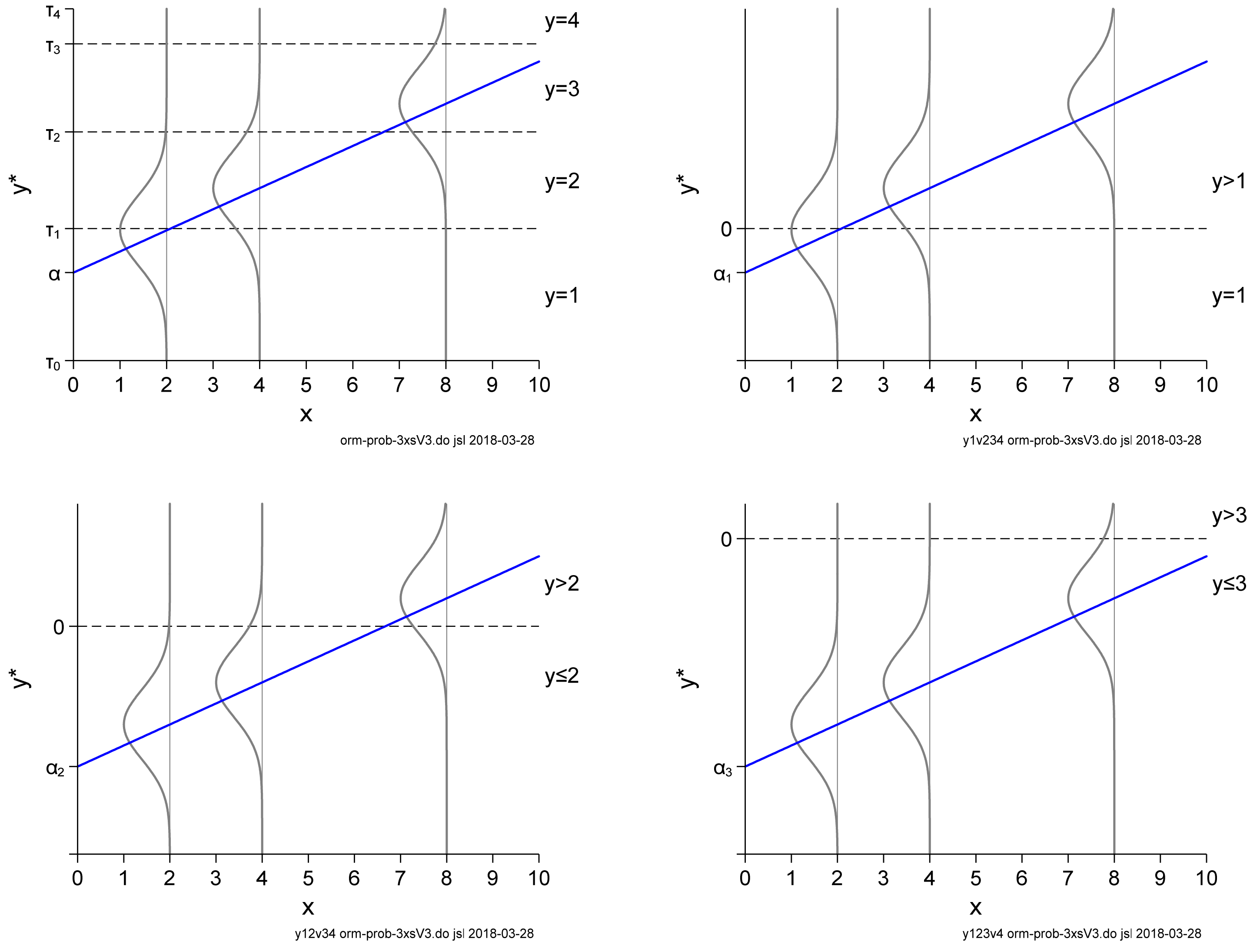
Each standard deviation increase in education increases support by .11 standard deviations, holding other variables constant.

Odds ratios for the OLM



1. $\exp(\beta_x)$ is the same for any dichotomization of y.

2. Dichotomizing any place and the slope is unchanged



3. The cumulative probability equals:

$$\Pr(y \le q \mid \mathbf{x}) = \sum_{j=1}^{q} \Pr(y = j \mid \mathbf{x}) \quad \text{for } q = 1, J-1$$

4. For S+D+A,

$$\begin{aligned}
\Pr(\text{SD+D+A} \mid \mathbf{x}) &= \Pr(y \le 3) \\
&= \Pr(y = 1 \mid \mathbf{x}) + \Pr(y = 2 \mid \mathbf{x}) + \Pr(y = 3 \mid \mathbf{x}) \\
&= \left[\Lambda(\tau_1 - \mathbf{x}_i\boldsymbol{\beta}) - \Lambda(\tau_0 - \mathbf{x}_i\boldsymbol{\beta})\right] + \left[\Lambda(\tau_2 - \mathbf{x}_i\boldsymbol{\beta}) - \Lambda(\tau_1 - \mathbf{x}_i\boldsymbol{\beta})\right] \\
&\quad + \left[\Lambda(\tau_3 - \mathbf{x}_i\boldsymbol{\beta}) - \Lambda(\tau_2 - \mathbf{x}_i\boldsymbol{\beta})\right] \\
&= \left[\Lambda(\tau_3 - \mathbf{x}_i\boldsymbol{\beta}) - \Lambda(\tau_0 - \mathbf{x}_i\boldsymbol{\beta})\right] = \Lambda(\tau_3 - \mathbf{x}_i\boldsymbol{\beta})
\end{aligned}$$

5. In general,

$$\Pr(y \le q \mid \mathbf{x}) = \Lambda(\tau_q - \mathbf{x}\boldsymbol{\beta})$$

6. This is a BLM with intercept $\alpha_q = (\tau_q - \beta_0)$ and slopes $\boldsymbol{\beta}^* = -\boldsymbol{\beta}$:

$$\Pr(y \le q \mid \mathbf{x}) = \Lambda(\tau_q - \mathbf{x}\boldsymbol{\beta}) = \Lambda(\alpha_q + \mathbf{x}\boldsymbol{\beta}^*)$$

7. The odds of $y \le q$ versus $y > q$ given **x** is

$$\Omega_q(\mathbf{x}) = \frac{\Pr(y \le q \mid \mathbf{x})}{\Pr(y > q \mid \mathbf{x})} = \exp(\tau_q - \mathbf{x}\boldsymbol{\beta}) = \exp(\alpha_q + \mathbf{x}\boldsymbol{\beta}^*)$$

8. The odds ratio for a change in x_k:

$$\frac{\Omega_q(\mathbf{x}, x_k + 1)}{\Omega_q(\mathbf{x}, x_k)} = \exp(\beta_k^*)$$

9. Interpretation

For a unit increase in x_k the odds of lower outcomes compared to higher outcomes change by the factor $\exp(\beta_k^*)$, holding other variables constant.

Odds ratios for supporting working mothers

NOTE: Odds of higher compared to lower outcome

$$\Omega\left(\frac{P(SD)}{P(D,A,SA)}\right)=\Omega\left(\frac{P(SD,D)}{P(A,SA)}\right)=\Omega\left(\frac{P(SD,D,A)}{P(SA)}\right).$$

```
. listcoef, help

ologit (N=2293): Factor change in odds

  Odds of: >m vs <=m

             |        b         z     P>|z|       e^b   e^bStdX     SDofX
-------------+-------------------------------------------------------------
  yr89: 1989 |   0.5239     6.557     0.000     1.689     1.292     0.490
  male: Male |  -0.7333    -9.343     0.000     0.480     0.694     0.499
white: White |  -0.3912    -3.304     0.001     0.676     0.879     0.329
         age |  -0.0217    -8.778     0.000     0.979     0.695    16.779
          ed |   0.0672     4.205     0.000     1.069     1.237     3.161
        prst |   0.0061     1.844     0.065     1.006     1.092    14.492
---------------------------------------------------------------------------
       b = raw coefficient
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
   SDofX = standard deviation of X
```

Interpretations on the next page...

<u>Interpretation using odds ratios</u>

	b	z	P>|z|	e^b	e^bStdX	SDofX
yr89						
1989	0.5239	6.557	0.000	1.689	1.292	0.490

From 1977 to 1989, the odds of being more positive toward working women increased by a factor of 1.7, holding other variables constant.

	b	z	P>|z|	e^b	e^bStdX	SDofX
male						
Male	-0.7333	-9.343	0.000	0.480	0.694	0.499

Being male decreases the odds of having more favorable attitudes toward working women by a factor of .48, holding other variables constant.

warm	b	z	P>|z|	e^b	e^bStdX	SDofX
age	-0.02167	-8.778	0.000	0.9786	0.6952	16.7790

As standard deviation increase in age, 17 years, decreases the odds of supporting working mothers by a factor of .70, holding other variables constant.

Other ordinal models

The continuation ratio model (**`ocratio`**)

1. Outcomes are a progression of events or stages in a process
2. If $\Pr\left(y=m\mid\mathbf{x}\right)$ is the probability of being in stage m given **x**,
3. The probability of being in stage m or later is

$$\Pr\left(y\geq m\mid\mathbf{x}\right)=\sum_{j=m}^{J}\Pr\left(y=j\mid\mathbf{x}\right)$$

4. The odds,

$$\frac{\Pr\left(y=m\mid\mathbf{x}\right)}{\Pr\left(y>m\mid\mathbf{x}\right)}=\exp\left(\tau_m-\mathbf{x}\boldsymbol{\beta}\right)$$

5. Predicted probabilities

$$\Pr\left(y=m\mid\mathbf{x}\right)=\frac{\exp\left(\tau_m-\mathbf{x}\boldsymbol{\beta}\right)}{\prod_{j=1}^{m}\left[1+\exp\left(\tau_j-\mathbf{x}\boldsymbol{\beta}\right)\right]}\text{ for } m=1\text{ to } J-1$$

The generalized ordered regression model (`glogit2`)

1. GOLM relaxes the assumption of equal β's over outcome categories and the model is not ordinal!
2. Define

$$\Omega_{y\le q}(\mathbf{x}) = \frac{\Pr(y \le q \mid \mathbf{x})}{\Pr(y > q \mid \mathbf{x})}$$

3. The GOLM removes the restriction of equal βs from the OLM:

$$\ln \Omega_{y\le q}(\mathbf{x}) = \tau_q - \mathbf{x}\boldsymbol{\beta}_q \quad \text{for } q = 1, J-1$$

4. In terms of odds

$$\Omega_{y\le q}(\mathbf{x}) = \exp(\tau_q - \mathbf{x}\boldsymbol{\beta}_q) \quad \text{for } q = 1, J-1 .$$

5. Predicted probabilities

$$\Pr(y = j \mid \mathbf{x}) = \frac{\exp(\tau_j - \mathbf{x}\boldsymbol{\beta}_j)}{1 + \exp(\tau_j - \mathbf{x}\boldsymbol{\beta}_j)} - \frac{\exp(\tau_{j-1} - \mathbf{x}\boldsymbol{\beta}_{j-1})}{1 + \exp(\tau_{j-1} - \mathbf{x}\boldsymbol{\beta}_{j-1})}$$

6. Software does not enforce that predictions are between 0 and 1.

7. Consider ORs from ologit and gologit2 for political affiliation

```
Odds of:               OLM        GOLM      | Odds of:   MNLM
---------------------------------------------------------------
SD vs D+I+R+SD         0.89**     0.81***   | SD vs D    1.25***
SD+D vs I+R+SD         0.89**     0.94*     | D vs I     1.07
SD+D+I vs R+SD         0.89**     0.99      | I vs R     0.94
SD+D+I+R vs +SD        0.89**     1.14**    | R vs SR    0.82***
```

How do you interpret the OR's for the GOLM? What is your conclusion?

β1a Generalized Marginal Effects

TODO: add more interpretations

Overview

1. The standard definition of a discrete changes is:

$$\frac{\Delta\pi(\mathbf{x})}{\Delta x_k(\text{start} \rightarrow \text{end})} = (x_k = \text{end}, \mathbf{x} = \mathbf{x}^*) - \pi(x_k = \text{start}, \mathbf{x} = \mathbf{x}^*)$$

For a change in x_k from start *to* end*, the probability changes by $\Delta\pi/\Delta x_k$, holding other variables at the specified values.*

2. Sometimes it makes sense to:
   - Change more than one variable at a time
   - Make different size changes for different variables
   - Change a component of an interation variable

Review of DC

DCM: change at the center of the data

$$\text{DCM}(x_k) = \frac{\Delta\pi(\mathbf{x} = \bar{\mathbf{x}})}{\Delta x_k(\text{start} \to \text{end})} = \pi(x_k = \text{end}, \mathbf{x} = \bar{\mathbf{x}}) - \pi(x_k = \text{start}, \mathbf{x} = \bar{\mathbf{x}})$$

For someone who is average on all variables, increasing x_k *from start to end changes the probability by* DCM(x_k)

ADC: average change

$$\text{ADC}(x_k) = \frac{1}{N}\sum_{i=1}^{N} \frac{\Delta\pi(\mathbf{x} = \mathbf{x}_i)}{\Delta x_{ik}(\text{start}_i \to \text{end}_i)}$$

On average, increasing x_k *from start to end changes the probability by* ADC(x_k).

Generalized effects

Here are some extensions to the standard definition of a DC

1. Average changes in a subsample
2. Proportional change in x_k
3. Change in x_k and x_j based on their observed relationship
   - Change height and what you predict the change in weight to be
4. Change of a component in a multiplicative measure
   - Change weight only as a component of BMI
5. Two or more mathematically linked variables
   - Change age and age-squared

Example: Health and Retirement Survey

```
. use hrs-gme-analysis2, clear
(hrs-gme-analysis2.dta | Health & Retirement Study GME sample | 2016-04-08)

. codebook diabetes female hsdegree age white bmi weight height, compact

Variable      Obs Unique      Mean       Min       Max  Label
-------------------------------------------------------------------------------
diabetes    16071      2  .2047166         0         1  Respondent has diabetes?
female      16071      2  .5684774         0         1  Is female?
hsdegree    16071      2  .7624914         0         1  Has high school degree?
age         16071     49  69.29276        53       101  Age
white       16071      2  .7715761         0         1  Is white respondent?
bmi         16071   2469  27.89787  10.57755   82.6728  Body mass index
weight      16071    258  174.9041        73       400  Weight in pounds
height      16071     37  66.28847        48        89  Height in inches
-------------------------------------------------------------------------------
```

Body mass index: $BMI = \frac{weight_{kg}}{height_m^2} = \frac{703 \times weight_{lb}}{height_{in}^2}$

Diabetes

Given the diseases burden of diabetes, small changes in Pr(diabetes) are important.

Models

1. Two models that vary in how body mass is included
2. **Mbmi**: uses the BMI index

```
logit diabetes c.bmi i.white c.age##c.age i.female i.hsdegree
estimates store Mbmi
```

3. **Mwt**: uses height and weight

```
logit diabetes c.weight c.height ///
     i.white c.age##c.age i.female i.hsdegree
estimates store Mwt
```

Estimates

```
    Variable |    Mbmi            Mwt
-------------+--------------------------
         bmi |    1.1046*
             |     0.000
      weight |                 1.0165*
      height |                 0.9299*
             |
       white |
      White  |    0.5412*      0.5313*

         age |    1.3091*      1.3093*
 c.age#c.age |    0.9983*      0.9983*
::
::
-------------+--------------------------
         aic |  14937.47     14920.55
         bic |  14991.26     14982.03
----------------------------------------
Note: # significant at .05 level; * at the .001 level.
```

ADC(bmi+sd) and ADC(white)

```
. estimates restore Mbmi

. mchange bmi white

logit: Changes in Pr(y) | Number of obs = 16071

Expression: Pr(diabetes), predict(pr)

                    |     Change    p-value
--------------------+----------------------
bmi                 |
                +SD |      0.097      0.000
white               |
 White vs Non-white |     -0.099      0.000
```

Interpretation

Increasing BMI by one standard deviation on average increases the probability of diabetes .097 ($p < .001$).

On average, the probability of diabetes is .099 less for white respondents than non-white respondents ($p < .001$).

Computing DC with =gen()

margins, at(varnm=gen(exp))

1. To compute predictions at the observed values:

```
margins, at( bmi = gen(bmi) )
```

2. Predictions at 1 more than the observed value of **bmi**

```
margins, at( bmi = gen(bmi + 1) )
```

3. Computing predictions at observed bmi + standard deviation

```
quietly sum bmi // summary statics
local sd = r(sd) // retrieve standard deviation
margins, at( bmi = gen(bmi + `sd') )
```

4. Computing both average predictions

```
margins, at( bmi = gen(bmi) ) at( bmi = gen(bmi + 1) )
```

ADC(bmi+sd)

1. Compute at observed bmi and observed + sd

```
Expression   : Pr(diabetes), predict()

1._at        : bmi              = bmi
2._at        : bmi              = bmi+5.770835041238605

             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .2047166   .0030338    67.48   0.000     .1987704    .2106627
          2  |   .3017056    .005199    58.03   0.000     .2915159    .3118954
------------------------------------------------------------------------------
```

2. Compute ADC(bmi+sd)

```
. mlincom 2-1, stats(all)

             |   lincom        se    zvalue    pvalue        ll        ul
-------------+------------------------------------------------------------
           1 |    0.097     0.004    27.208     0.000     0.090     0.104
```

On average, increasing BMI by one standard deviation, about 6 points, increases the probability of diabetes by .097 (p < .001).

Proportional change in x_k

1. Body mass involves a comparison of height and weight
2. Here we include them as separate regressors rather than as BMI

```
logit diabetes c.weight c.height ///
    i.white c.age##c.age i.female i.hsdegree
estimates store Mwt
```

3. ADC(weight+25) increases weight by 25 pounds

   : 25 pounds is a 25% increase if you weigh 100 pounds

   : 25 pounds is an 8% increase if you weigh 300 pounds

4. Does increasing weight <u>proportionally</u> make more substantive sense?
5. We compute
   - ADC(weight+25) increasing weight by 25 pounts
   - ADC(weight*1.14) increasing weight by 14 percent

Fixed change in x_k : ADC(weight+25)

1. Compute ADC(weight+25)

```
. qui estimates restore Mwt

. mtable, gen(PRadd) ///
>     at(weight=gen(weight)) at(weight=gen(weight+25)) post

Expression: Pr(diabetes), predict()

           |     Pr(y)
 ----------+---------
         1 |     0.205
         2 |     0.271

. mlincom 2-1

             |    lincom     pvalue         ll         ul
-------------+----------------------------------------
           1 |     0.067      0.000      0.062      0.071
```

Proportional change in x_k : ADC(weight*1.14)

1. A simple change computes ADC(weight * 1.14)

```
. mtable, gen(PRpct) ///
>     at(weight=gen(weight)) at(weight=gen(weight*1.14)) post

Expression: Pr(diabetes), predict()

          |    Pr(y)
----------+---------
        1 |    0.205
        2 |    0.273

. mlincom 2-1

             |    lincom     pvalue         ll         ul
-------------+-----------------------------------------
           1 |     0.068      0.000      0.063      0.073
```

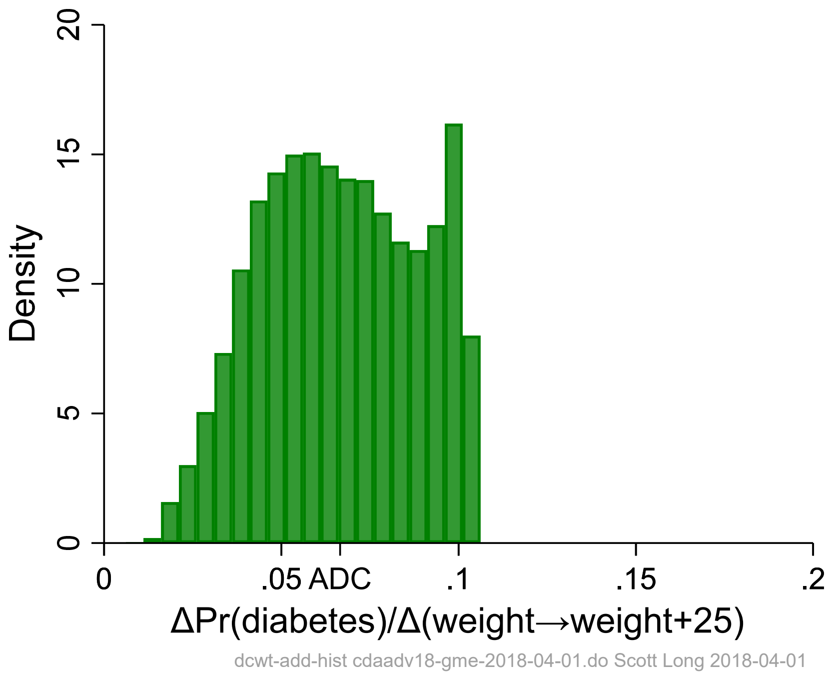
Does ADC(weight*1.14) = ADC(weight+25)?

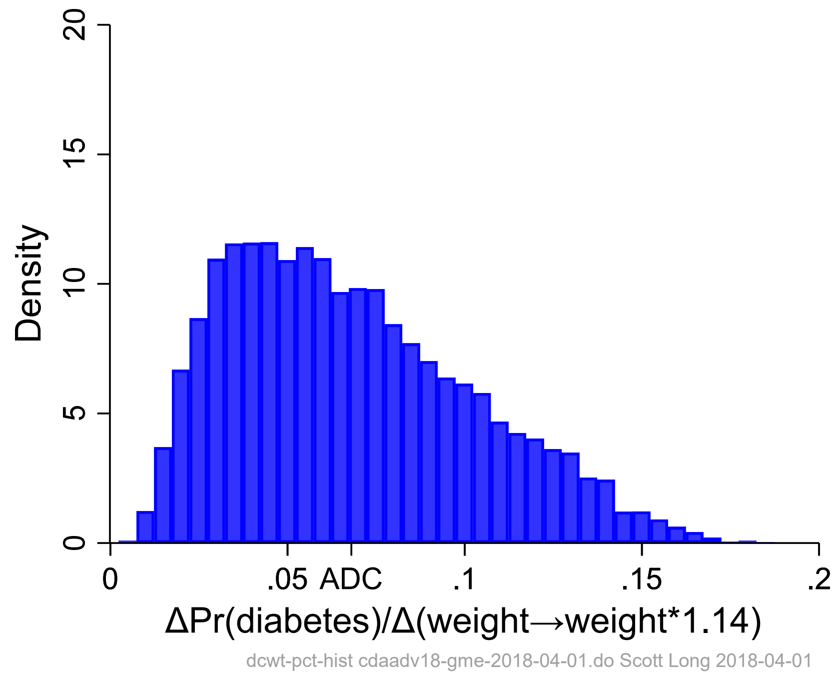
```
. qui estimates restore Mwt
. mtable, at(weight=gen(weight)) /// observed
>         at(weight=gen(weight+25)) /// + 15
>         at(weight=gen(weight*1.14)) post // + 14%

Expression: Pr(diabetes), predict()

           |    Pr(y)
 ----------+---------
         1 |    0.205
         2 |    0.271
         3 |    0.273

. qui mlincom (2-1),         rowname(ADC add) clear
. qui mlincom (3-1),         rowname(ADC pct) add
.     mlincom (2-1) - (3-1), rowname(Difference) add dec(5)

             |   lincom    pvalue        ll        ul
-------------+---------------------------------------
     ADC add |  0.06659   0.00000   0.06197   0.07120
     ADC pct |  0.06801   0.00000   0.06332   0.07271
  Difference | -0.00142   0.00000  -0.00163  -0.00122
```

A 14 percent increases in weight has a small but significant larger average effect on diabetes than increasing weight by 25 pounds (p<.001).

Are they really that similar?

1. The ADCs are similar but the distributions are quite different





2. Does this have implications for health policy?

Discrete change with linked variables

Mathematically linked variables

1. With polynomials multiple variables must change together

$$\frac{\Delta\pi(\mathbf{x})}{\Delta \text{age}(50 \rightarrow 60)} = \pi(\text{age=}60, \text{agesq=}60^2) - \pi(\text{age=}50, \text{agesq=}50^2)$$

2. Using factor syntax, **`margins`** handles this automatically

Substantively linked variables

1. Sometimes it makes sense to change multiple variables that are not mathematically linked
2. If two people have the same body mass, is the larger person more likely to have diabetes (i.e., taller and proportionally heavier)?
3. I compute an effect where height and weight change proportionally
   - Use height to predict weight
   - Use at(...=gen()) to change height and weight together

Linked variables: ADC(height, weight)

1. Regress weight on height and height-squared (to get a higher R^2)

```
. regress weight c.height##c.height, noci
::
           weight |      Coef.   Std. Err.      t    P>|t|
------------------+----------------------------------------
           height |  -6.338708    1.61073    -3.94   0.000
                  |
c.height#c.height |   .0855799   .0120867     7.08   0.000
                  |
            _cons |   217.5991    53.5548     4.06   0.000
-----------------------------------------------------------
```

2. Save the estimates

```
. scalar b0 = _b[_cons]
. scalar b1 = _b[height]
. scalar b2 = _b[c.height#c.height]
```

3. Predict weight from height

```
weighthat = b0 + b1*height + b2*height#height
```

4. Make predictions:

```
. qui estimates restore Mwt
. mtable, post commands ///
>     at(height=gen(height) weight=gen(weight))              /// observed
>     at(height=gen(height+6)                                 /// height + 6
>        weight=gen(b0 + b1*(height+6) + b2*((height+6)^2))) //  => weight

Expression: Pr(diabetes), predict()

           |    Pr(y)
 ----------+---------
         1 |    0.205
         2 |    0.208

. mlincom 2-1

             |   lincom    pvalue        ll         ul
-------------+----------------------------------------
           1 |    0.004     0.601    -0.010      0.017
```

There is no evidence that being physically larger without greater body mass contributes to the incidence of diabetes.

Decomposing the effect of BMI

1. The BMI index measures relative weight

$$\text{BMI} = 703 \times \frac{weight_{lb}}{height_{in}^2} = 703 \times weight \times height^{-1} \times height^{-1}$$

2. With BMI in the model, can we compute the effect of weight change?
   - Why do this? DC(weight) is clearer to patients than DC(bmi)

3. Create components of BMI

```
generate heightinv = 1/height
label var heightinv "1/height"
generate factor = 703
label var factor "scale factor to convert from metric"
```

4. These models are identical

```
logit diabetes c.factor#c.weight#c.heightinv#c.heightinv ///
    i.white c.age##c.age i.female i.hsdegree
estimates store MbmiFV
logit diabetes c.bmi i.white c.age##c.age i.female i.hsdegree
estimates store Mbmi
```

5. The estimates are identical (so are the standard errors)

```
    Variable |    Mbmi          Mbmifv
-------------+--------------------------
    c.factor#|
    c.weight#|
c.height_inv#|
c.height_inv |              1.1045533
             |
         bmi | 1.1045533
             |
       white |
      White  |  .5411742     .5411742
::
```

6. Factor syntax makes it possible to change **weight** as a component

```
. qui estimates restore Mbmifv

. mchange weight, amount(sd) delta(25)

logit: Changes in Pr(y) | Number of obs = 16071

Expression: Pr(diabetes), predict(pr)

             |    Change    p-value
-------------+----------------------
weight       |
      +delta |     0.065      0.000

Average predictions

             |        No   Diabetes
-------------+----------------------
  Pr(y|base) |     0.795      0.205

1: Delta equals 25.
```

Conclusions on GME

- Other examples
- More interpretation

TODO Conclusions

1. LRM, BRM, ORM, MNLM, PRM, NBRM, and ZIP/ZINB are building blocks for models in many areas
2. Extensions add panels, hierarchies, clustering, survey sampling, and more
3. The basic structure of the models stays the same
4. Since the models are nonlinear, the challenge is to determine what is substantively important and to find the best way to summarize the results
5. Alternative strategies need to be tried to find the most convincing approach
6. Remember Neal Henry’s sage advice:

    Don't let the numbers get in the way of the data.