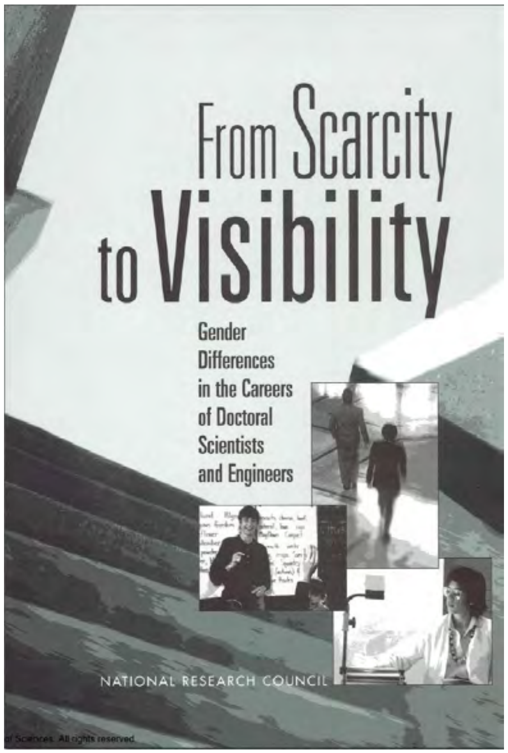
Comparing groups using predicted probabilities

J. Scott Long

Indiana University

May 9, 2006

The problem



Allison (1999): "Differences in the estimated coefficients tell us nothing about the differences in the underlying impact of x on the two groups."

Overview

1. Does the effect of x on y differ across groups?
2. Comparing β's across groups in the LRM
3. Problems with comparing β's in logit and probit models
4. Tests and CIs for comparing $\Pr(y = 1)$ across groups
5. Example of gender differences in tenure
6. Complications due to nonlinearity of models

LRM - Chow test

1. For example,

$$\text{Men:} \quad y = \alpha^m + \beta^m_{educ} educ + \beta^m_{age} age + \varepsilon$$

$$\text{Women:} \quad y = \alpha^w + \beta^w_{educ} educ + \beta^w_{age} age + \varepsilon$$

2. Do men and women have the same return for education?

$$H_0\text{: } \beta^m_{educ} = \beta^w_{educ}$$

3. We compute:

$$z = \frac{\widehat{\beta}^m_{educ} - \widehat{\beta}^w_{educ}}{\sqrt{Var\left(\widehat{\beta}^m_{educ}\right) + Var\left(\widehat{\beta}^w_{educ}\right)}}$$

Logit and probit using latent y^*

1. Chow type test is invalid for logit & probit (Allison 1999).
2. Structural model with a latent y^*:

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

3. Error is normal for probit or logistic for logit:

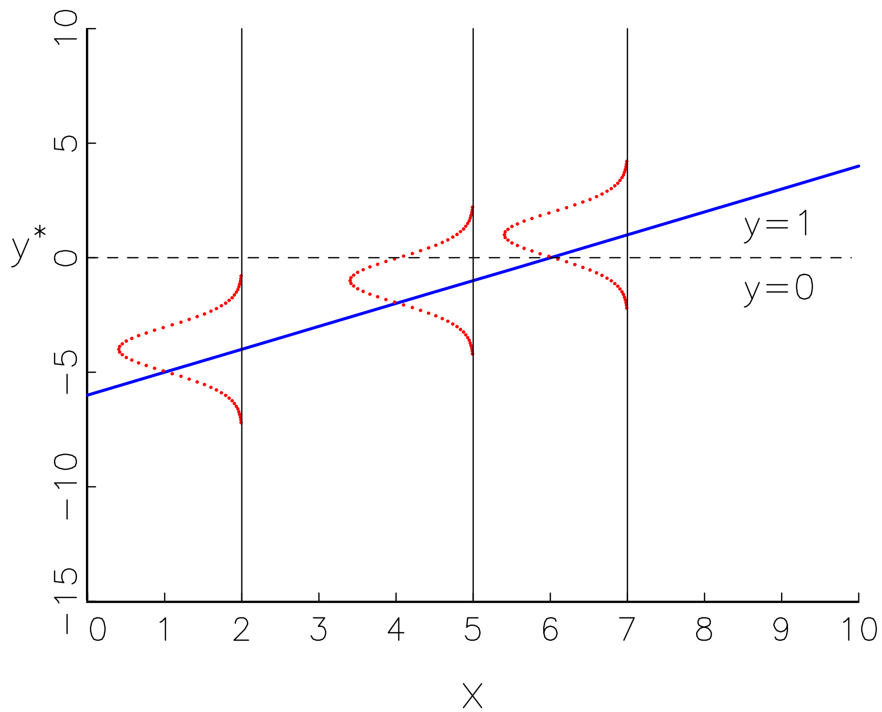
$$\varepsilon \sim f\left(0, \sigma_\varepsilon\right)$$

4. We only observed a binary y:

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

5. Graphically...

Logit and probit using latent y^*



y^* and $\Pr(y = 1)$

1. Link to observed data:

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

2. Since y^* is not observed,

β is only identified up to a scale factor.

Identification of β and σ_ε

1. Model A:

$$\begin{aligned} y_a^* &= \alpha^a + \beta_x^a x + \varepsilon_a \text{ where } \sigma_a = 1 \\ &= -6 + 1x + \varepsilon_a \end{aligned}$$

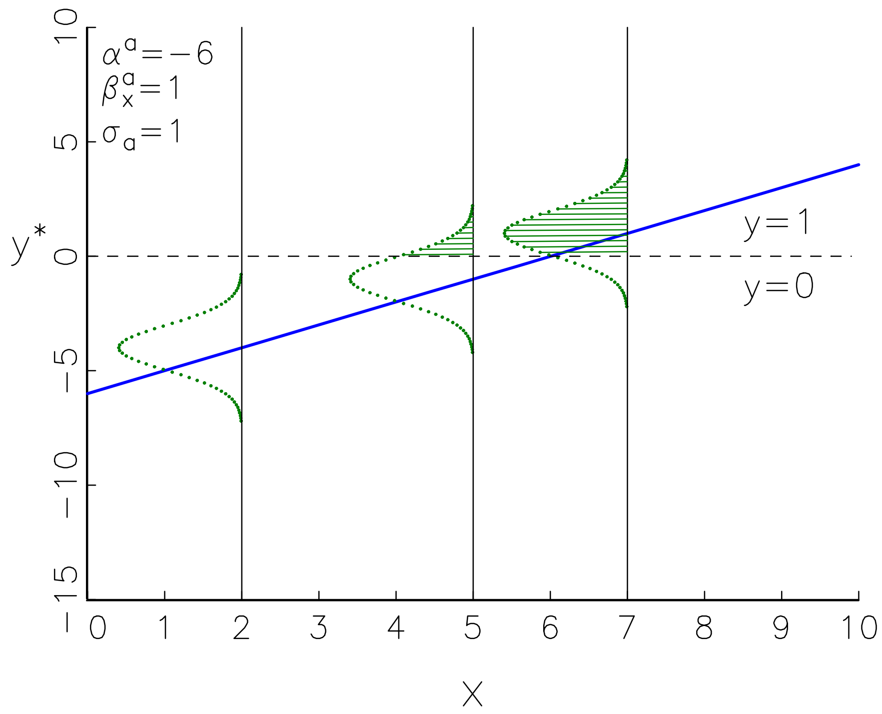
2. Model B:

$$\begin{aligned} y_b^* &= \alpha^b + \beta_x^b x + \varepsilon_b \text{ where } \sigma_b = 2 \\ &= -12 + 2x + \varepsilon_b \end{aligned}$$

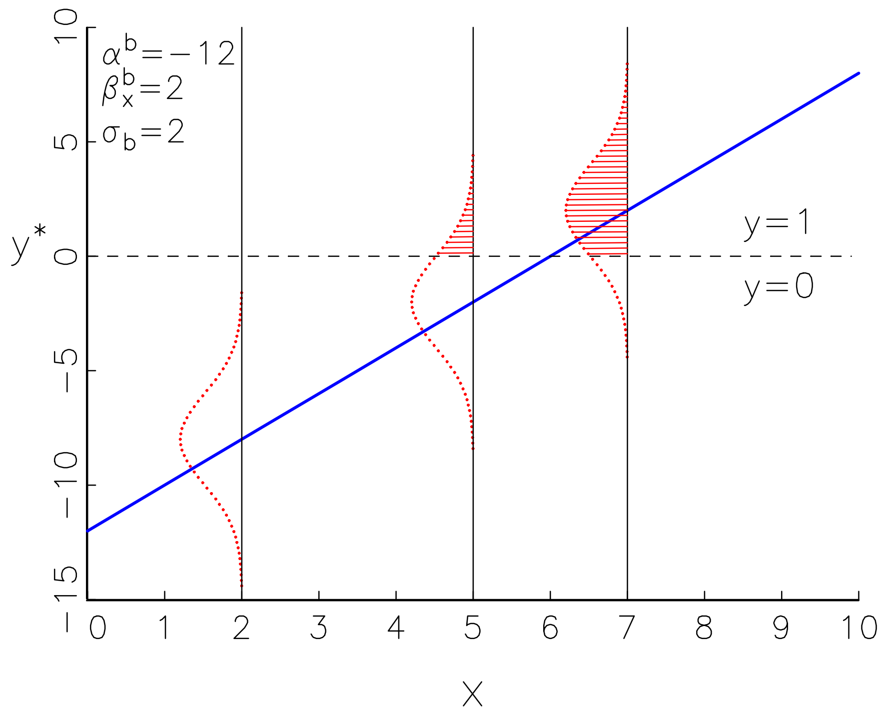
3. Where:

$$\begin{aligned} \alpha^b &= 2\alpha^a \\ \beta_x^b &= 2\beta_x^a \\ \sigma_b &= 2\sigma_a \end{aligned}$$

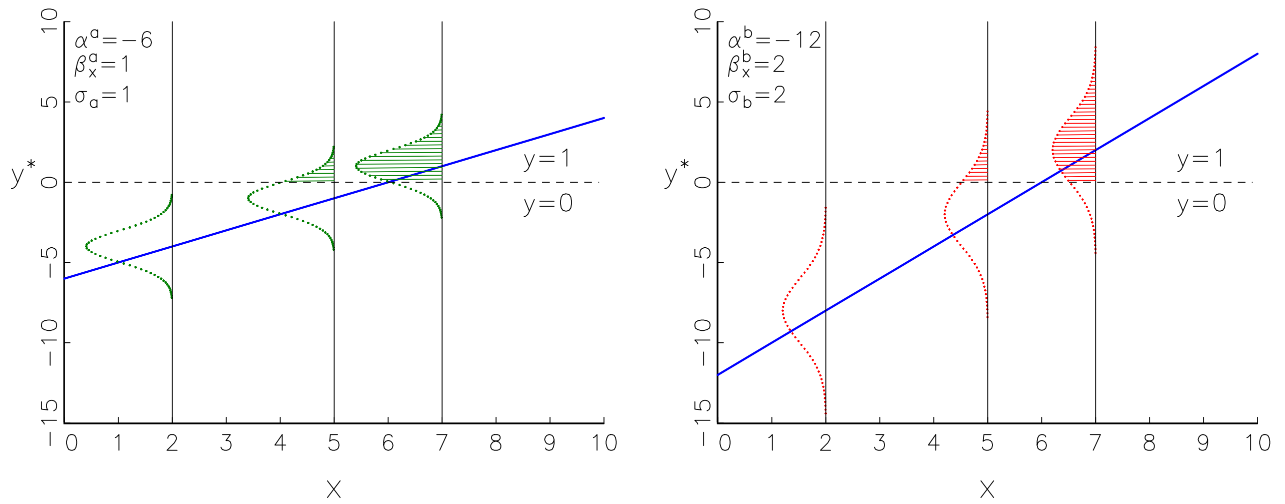
$\Pr(y = 1 \mid x)$ in model A



$\Pr(y = 1 \mid x)$ in model B



The problem and a ‘solution’



In terms of $\Pr(y = 1)$, these are empirically indistinguishable:

Case 1: A change in x of 1 when $\beta_x^a = 1$ and $\sigma_a = 1$.

Case 2: A change in x of 1 when $\beta_x^b = 2$ and $\sigma_b = 2$.

Implications for group comparisons

1. Comparing propensity for tenure for men and women:

$$\begin{aligned} \text{Men} &: \quad y^* = \alpha^m + \beta^m_{articles} articles + \varepsilon_m \\ \text{Women} &: \quad y^* = \alpha^w + \beta^w_{articles} articles + \varepsilon_w \end{aligned}$$

2. Assume the β's are equal:

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

3. Assume women have more unobserved heterogeneity:

$$\sigma_w > \sigma_m$$

4. Now we estimate the model...

Estimation

1. Probit *software* assumes that $\sigma = 1$.
2. For men, the estimated model is:

$$\begin{aligned} \frac{y^*}{\sigma_m} &= \frac{\alpha^m}{\sigma_m} + \frac{\beta^m_{articles}}{\sigma_m} articles + \frac{\varepsilon_m}{\sigma_m} \\ &= \widetilde{\alpha}^m + \widetilde{\beta}^m_{articles} articles + \widetilde{\varepsilon}_m, \ \text{sd}\left(\widetilde{\varepsilon}_m\right) = 1 \end{aligned}$$

3. For women, the estimated model is:

$$\begin{aligned} \frac{y^*}{\sigma_w} &= \frac{\alpha^w}{\sigma_w} + \frac{\beta^w_{articles}}{\sigma_w} articles + \frac{\varepsilon_w}{\sigma_w} \\ &= \widetilde{\alpha}^w + \widetilde{\beta}^w_{articles} articles + \widetilde{\varepsilon}_w, \ \text{sd}\left(\widetilde{\varepsilon}_w\right) = 1 \end{aligned}$$

Problem with Chow-type tests

1. We want to test:

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

2. But, end up testing:

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

3. And, $\widetilde{\beta}^m_{articles} = \widetilde{\beta}^w_{articles}$ does not imply $\beta^m_{articles} = \beta^w_{articles}$ unless $\sigma_m = \sigma_w$.

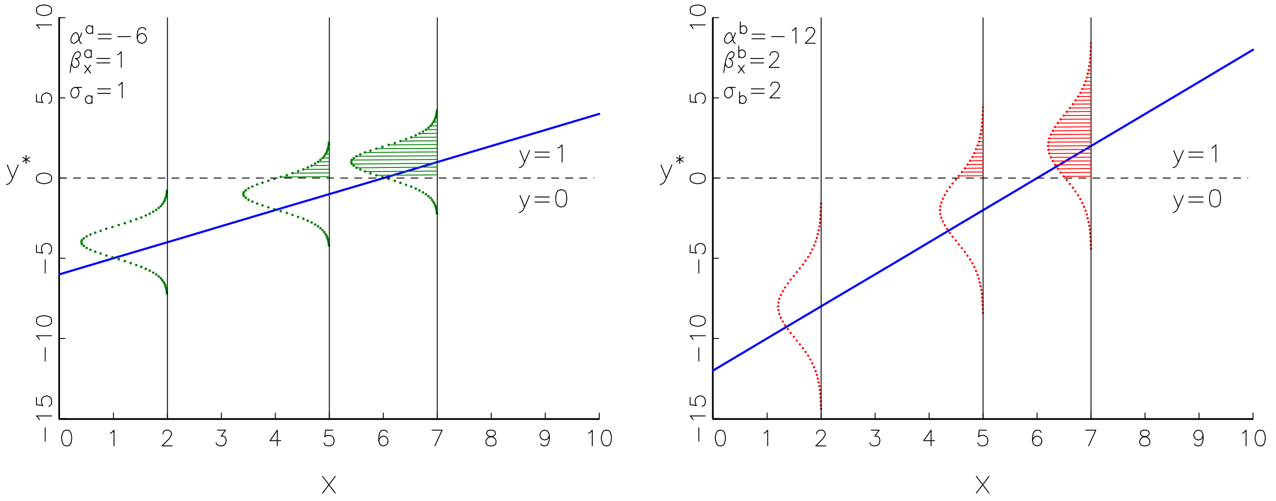
Alternatives for testing group differences

1. Allison's (1999) test of H_0: $\beta_x^m = \beta_x^w$:
   (a) Disentangles the β's and σ_ε's.
   (b) But requires that $\beta_z^m = \beta_z^w$ for some z.
2. I propose testing

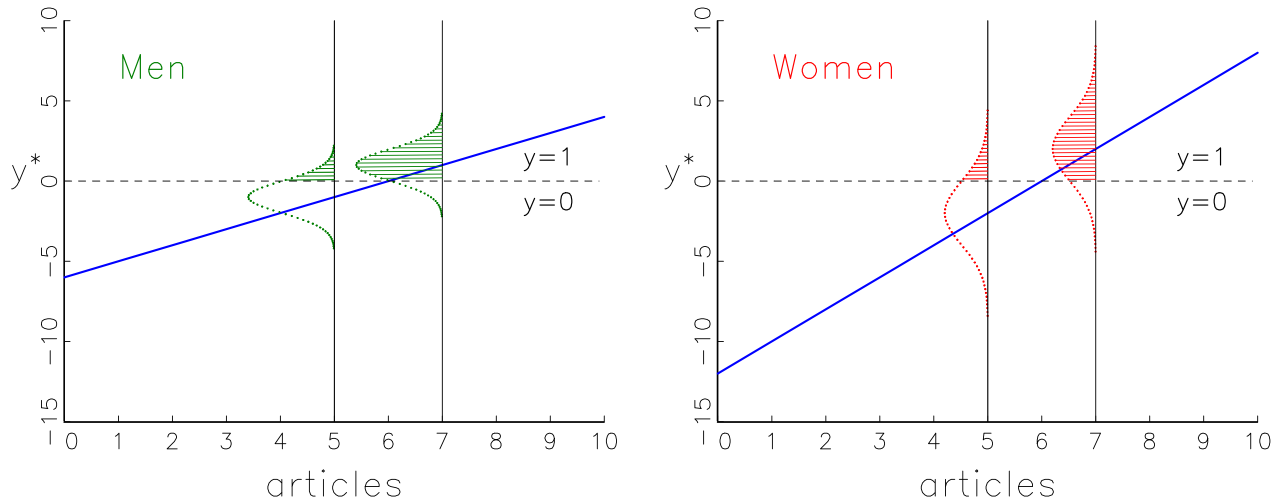
$$H_0\text{:}\;\; \Pr\left(y=1 \mid x\right)_m = \Pr\left(y=1 \mid x\right)_w$$

   since the probabilities are invariant to σ_ε.
3. Graphically...

Invariance of $\Pr(y = 1 \mid x)$



Group comparisons of Pr$(y = 1 \mid x)$



Testing $\Delta(\mathbf{x})$

1. Define:

$$\Delta(\mathbf{x}) = \Pr(y = 1 \mid \mathbf{x})_m - \Pr(y = 1 \mid \mathbf{x})_w$$

2. Then:

$$z = \frac{\widehat{\Delta}(\mathbf{x})}{\sqrt{Var\left[\widehat{\Delta}(\mathbf{x})\right]}}$$

3. Or the confidence interval:

$$\Pr(\Delta_{\mathrm{LB}} \leq \Delta(\mathbf{x}) \leq \Delta_{\mathrm{UB}}) = .95$$

4. Delta method is very fast; bootstrap is very slow; both are available with SPost's `prvalue` and `prgen`.

Comparing groups differences in $\Pr(y=1)$

1. Estimate:

$$\Pr(y=1) = F\left(\alpha^{w}w + \beta_{x}^{w}wx + \alpha^{m}m + \beta_{x}^{m}mx\right)$$

   where $w=1$ for women, else 0; $m=1$ for men, else 0; $wx = w \times x$; and $mx = m \times x$.

2. Then:

$$\begin{aligned}\Pr(y=1 \mid \mathbf{x})_{w} &= F\left(\alpha^{w} + \beta_{x}^{w}x\right)\\ \Pr(y=1 \mid \mathbf{x})_{m} &= F\left(\alpha^{m} + \beta_{x}^{m}x\right)\end{aligned}$$

3. The gender difference is:

$$\Delta(\mathbf{x}) = \Pr(y=1 \mid \mathbf{x})_{m} - \Pr(y=1 \mid \mathbf{x})_{w}$$

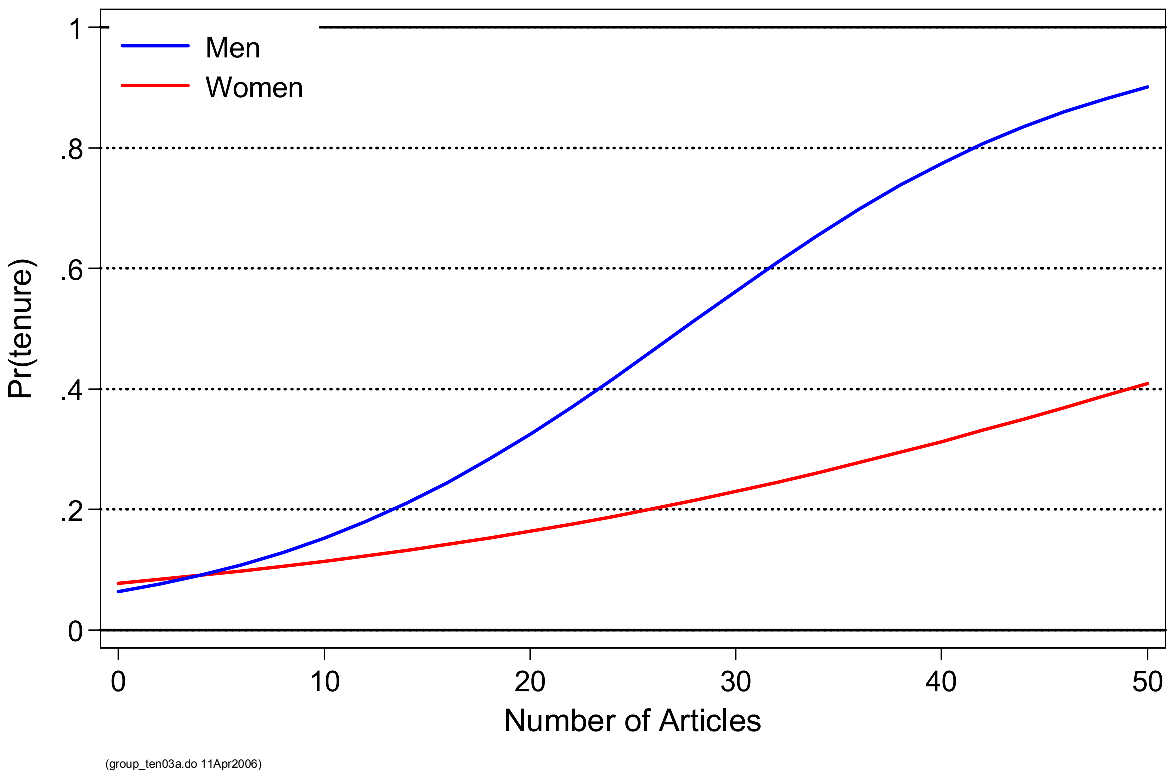
Example: gender differences in tenure

		Women		Men	
	Variable	*Mean*	*SD*	*Mean*	*SD*
Is tenured?	tenure	0.11	0.31	0.13	0.34
Year	year	3.97	2.38	3.78	2.25
Year-squared	yearsq	21.46	23.14	19.39	21.50
Bachelor's selectivity	select	5.00	1.48	4.99	1.37
Total articles	articles	7.41	7.43	6.83	5.99
Distinguished job?	distinguished	0.05	0.23	0.04	0.19
Prestige of job	prestige	2.66	0.77	2.64	0.78
Person-years		1,121		1,824	
Scientists		177		301	

M1: articles only

Variable	Women			Men		
	β	e^{β}	z	β	e^{β}	z
constant	-2.47		-18.30	-2.69		-23.00
articles	0.042	1.04	4.26	0.10	1.10	9.93
Log-lik	-375.17			-663.03		
N	1,121			1,824		

M1: Pr($tenure \mid articles$)



M1: group differences in probabilities

1. We are interested in group differences in predictions:

$$\Delta(\mathit{articles}) = \Pr(y=1 \mid \mathit{articles})_m - \Pr(y=1 \mid \mathit{articles})_w$$

2. We can test:

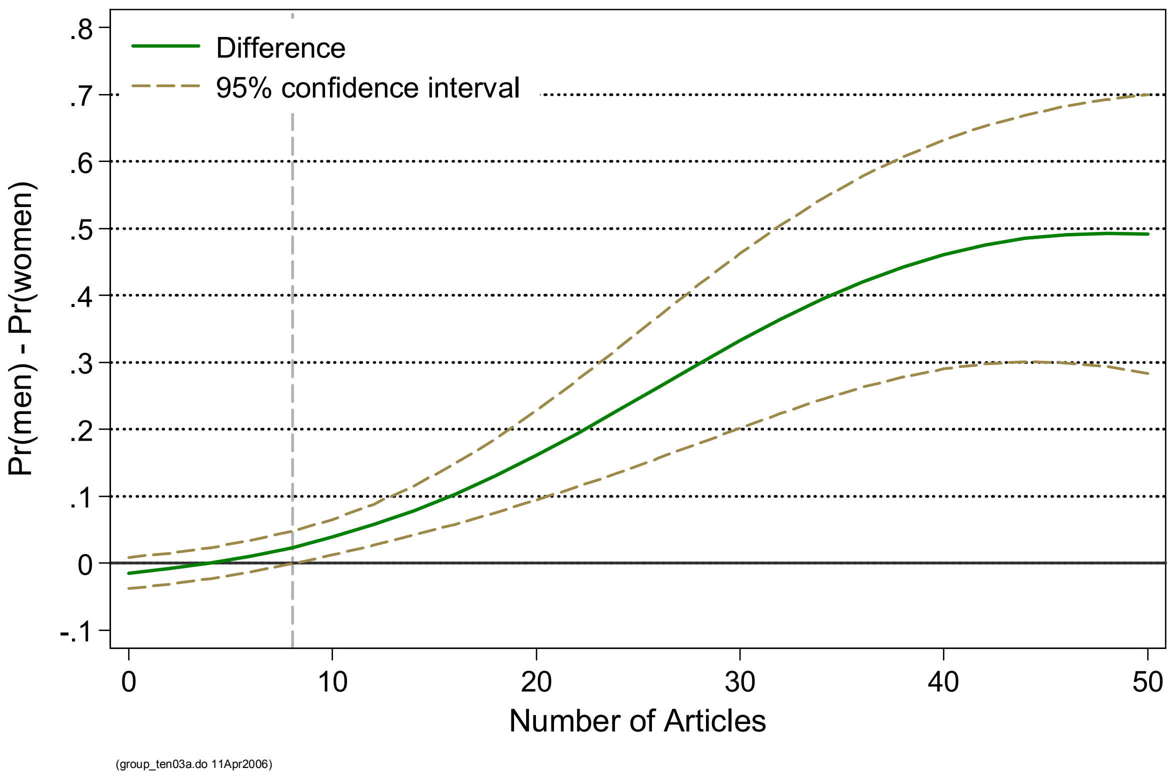
$$H_0 : \Delta(\mathit{articles}) = 0$$

3. Or:

$$\Delta(\mathit{articles})_{\text{LowerBound}}, \quad \Delta(\mathit{articles})_{\text{UpperBound}}$$

4. With one RHS variable, we can plot all comparisons.

M1: $\Delta\,(articles)$ with confidence intervals



Adding variables additional variables

1. Structural model:

$$y^* = \alpha + \beta_x x + \beta_z z + \varepsilon$$

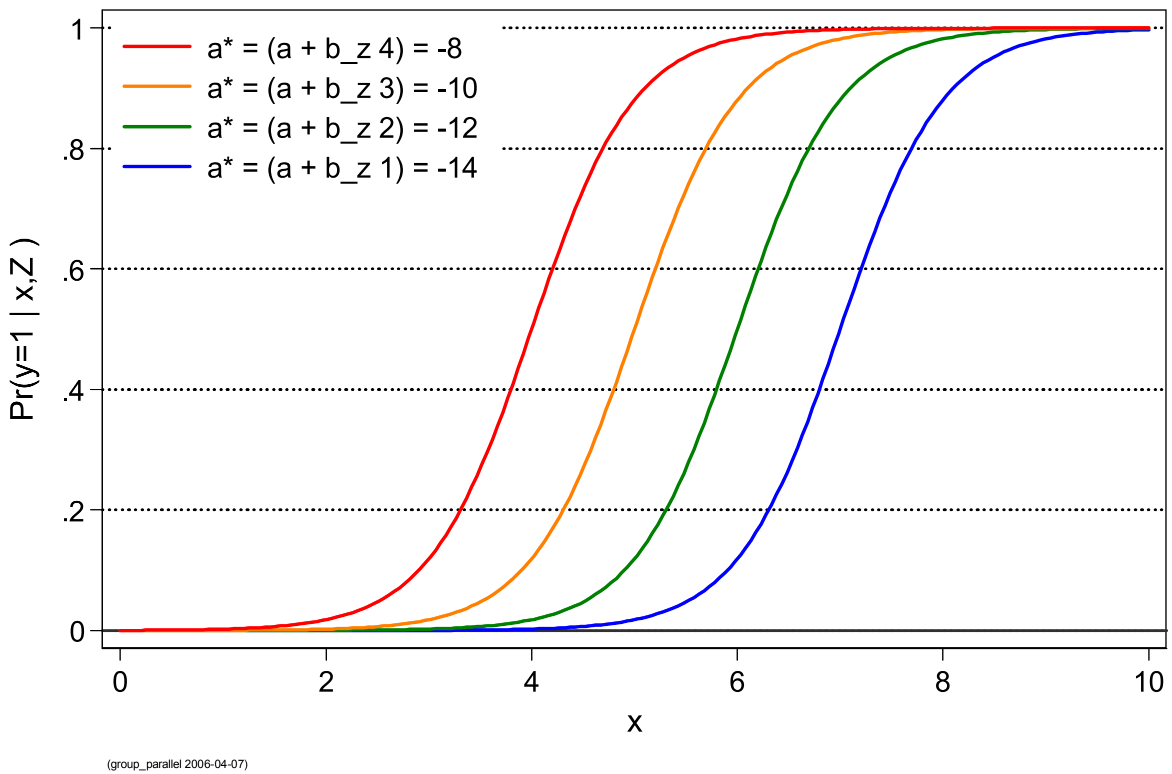
2. Setting $z = Z$ changes the intercept:

$$\begin{aligned} y^* &= \alpha + \beta_x x + \beta_z Z + \varepsilon \\ &= [\alpha + \beta_z Z] + \beta_x x + \varepsilon \\ &= \alpha^* + \beta_x x + \varepsilon \end{aligned}$$

3. Different values of z lead to different probability curves:

$$\begin{aligned} \Pr(y = 1 \mid x, z = Z) &= \Pr(\varepsilon < [\alpha^* + \beta_x x]) \\ &= F(\alpha^* + \beta_x x) \end{aligned}$$

Effect of levels of other variables



Comparing groups with added variables

1. For given $z = Z$:

$$\begin{aligned} \text{Men:} \quad & \Pr(y=1 \mid x, Z)_m = F(\alpha^{*m} + \beta_x^m x) \\ \text{Women:} \quad & \Pr(y=1 \mid x, Z)_w = F(\alpha^{*w} + \beta_x^w x) \end{aligned}$$

2. Group differences depend on z:

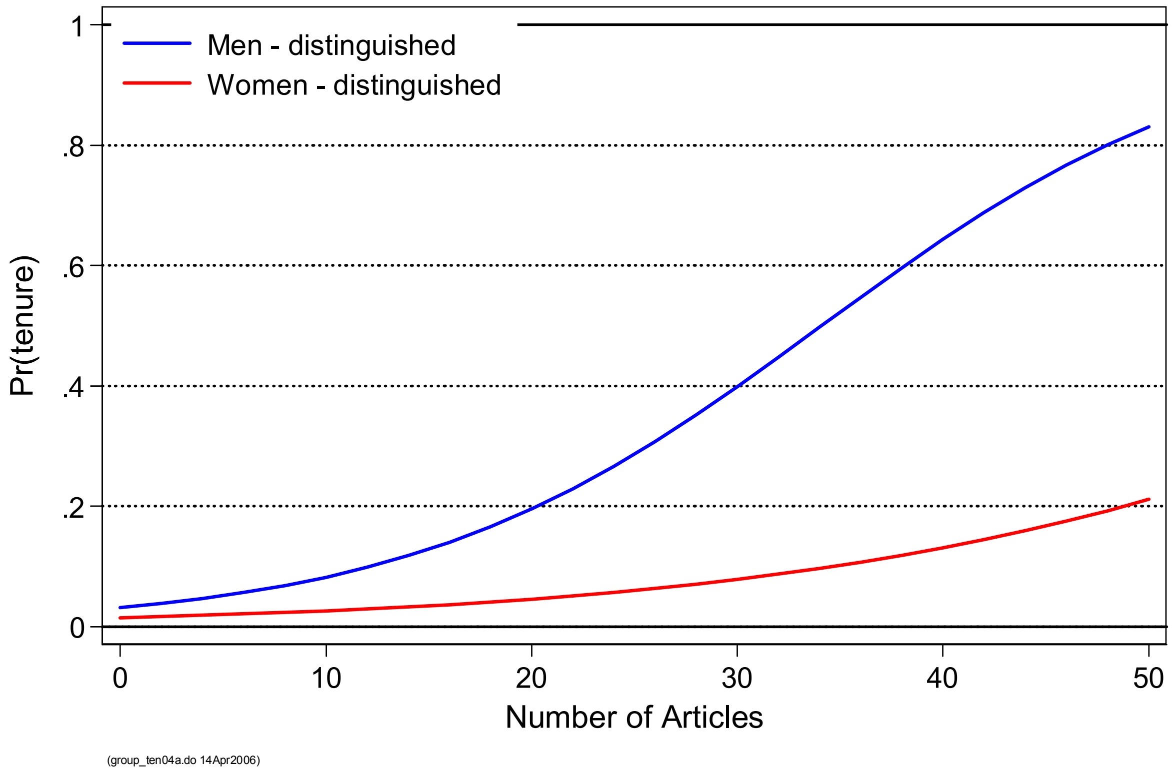
$$\Delta(x, Z) = \Pr(y=1 \mid x, Z)_m - \Pr(y=1 \mid x, Z)_w$$

3. $\Delta(x, Z)$ for a given x depends on the level of other variables.

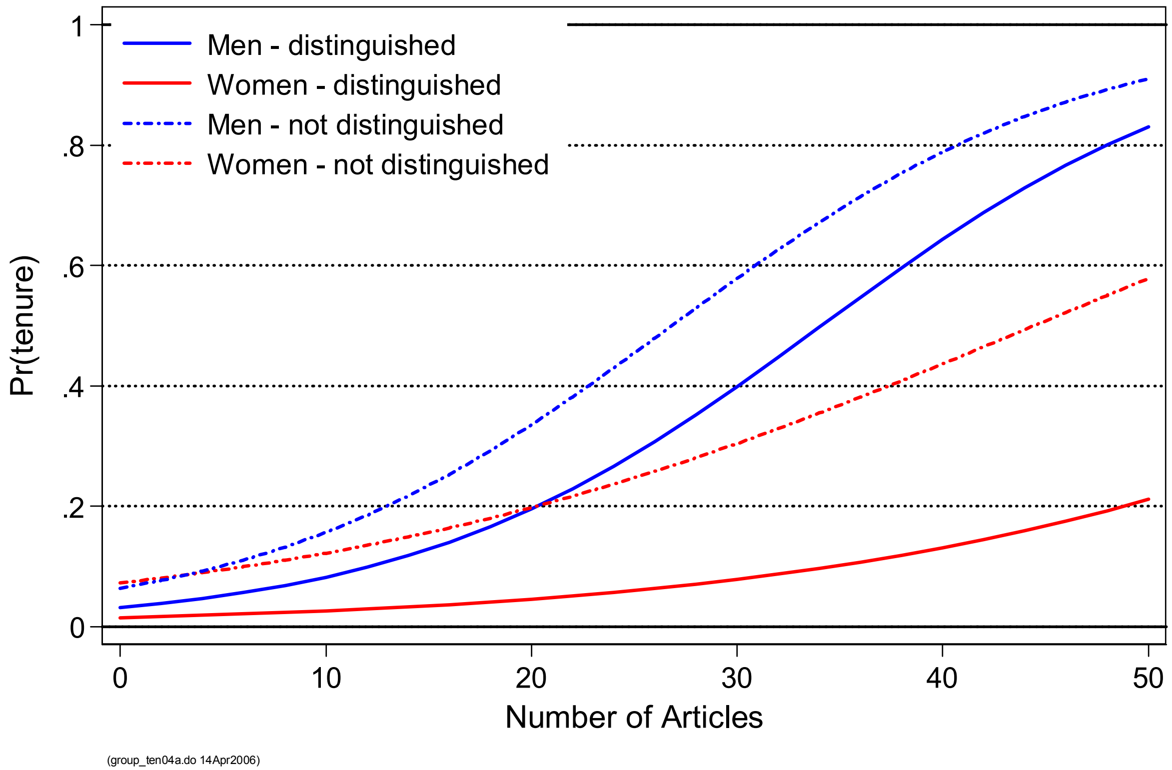
M2: adding job type

	Women			Men		
Variable	β	e^{β}	z	β	e^{β}	z
constant	-2.54		-17.90	-2.68		-22.87
articles	0.06	1.06	5.00	0.10	1.11	10.01
distinguished	-1.63	0.19	-2.43	-0.73	0.48	-1.71
Log-lik	-375.17			-663.03		
N	1,121			1,824		
Chow tests:	χ^2_{articles}=7.93, df=1, p<.01; χ^2_{dist}=1.38, df=1, p>.10.					
Allison tests:	χ^2_{articles}=15.06, df=1, p<.01 (other effects equal).					
	$\chi^2_{\text{distinguished}}$=3.54, df=1, p=.06 (other effects equal).					

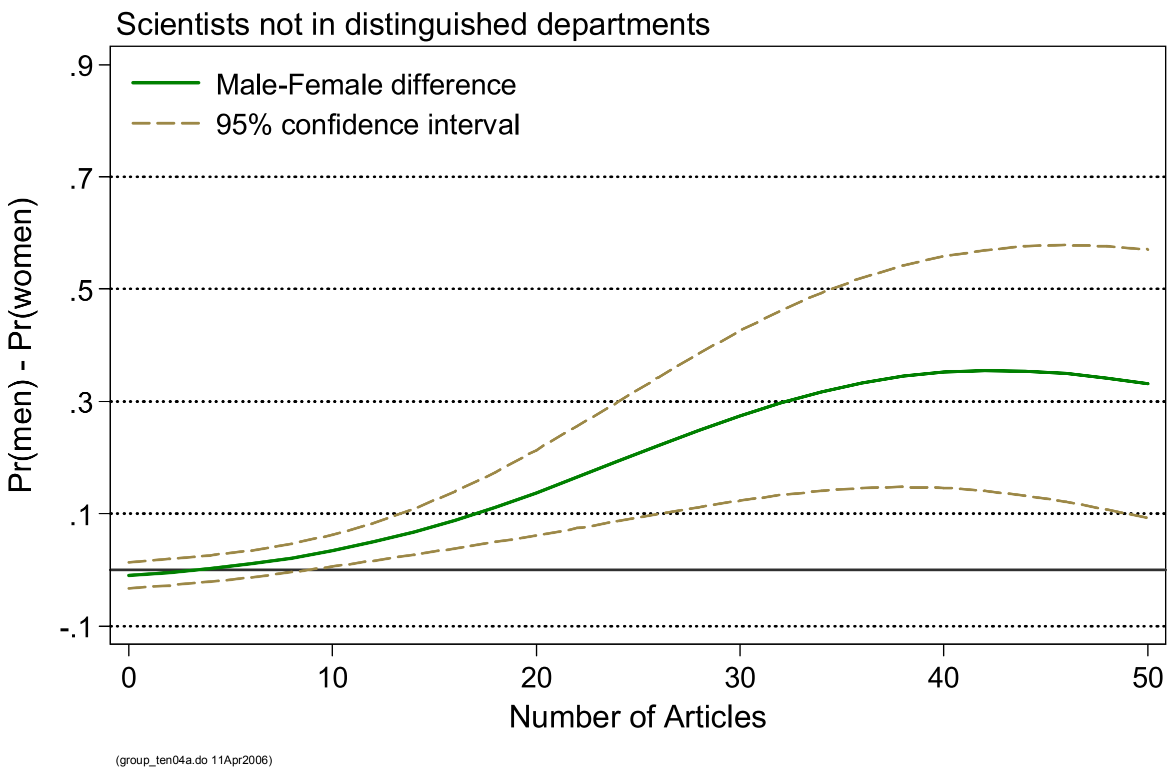
M2: Pr$(tenure \mid articles)$



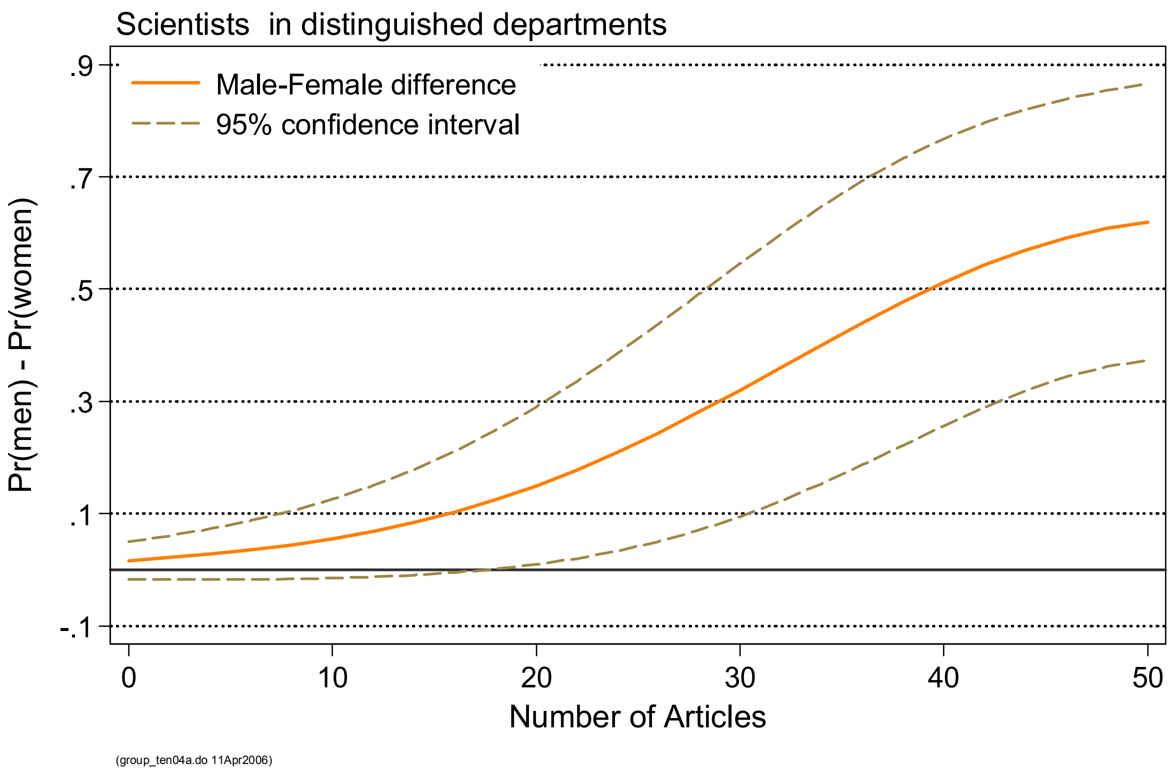
M2: Pr$(tenure \mid articles)$



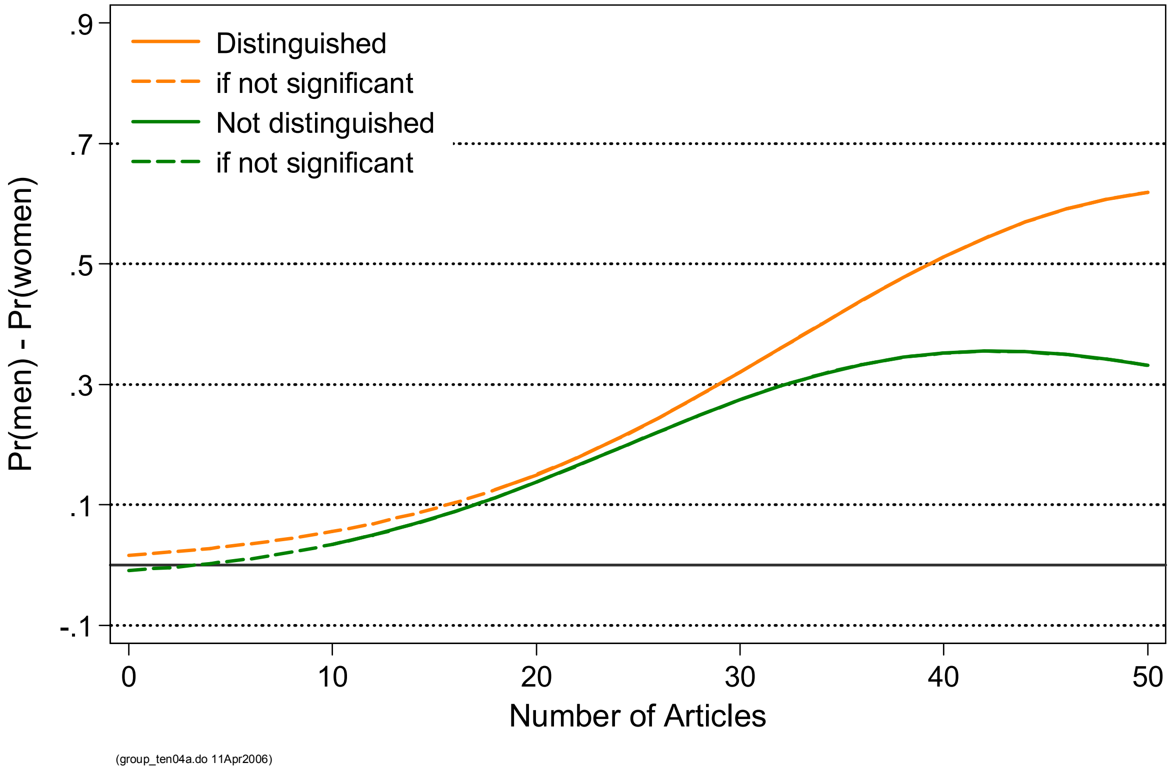
M2: $\Delta\,(articles$ **if** $not\ distinguished)$



M2: $\Delta\,(articles$ **if** $distinguished)$



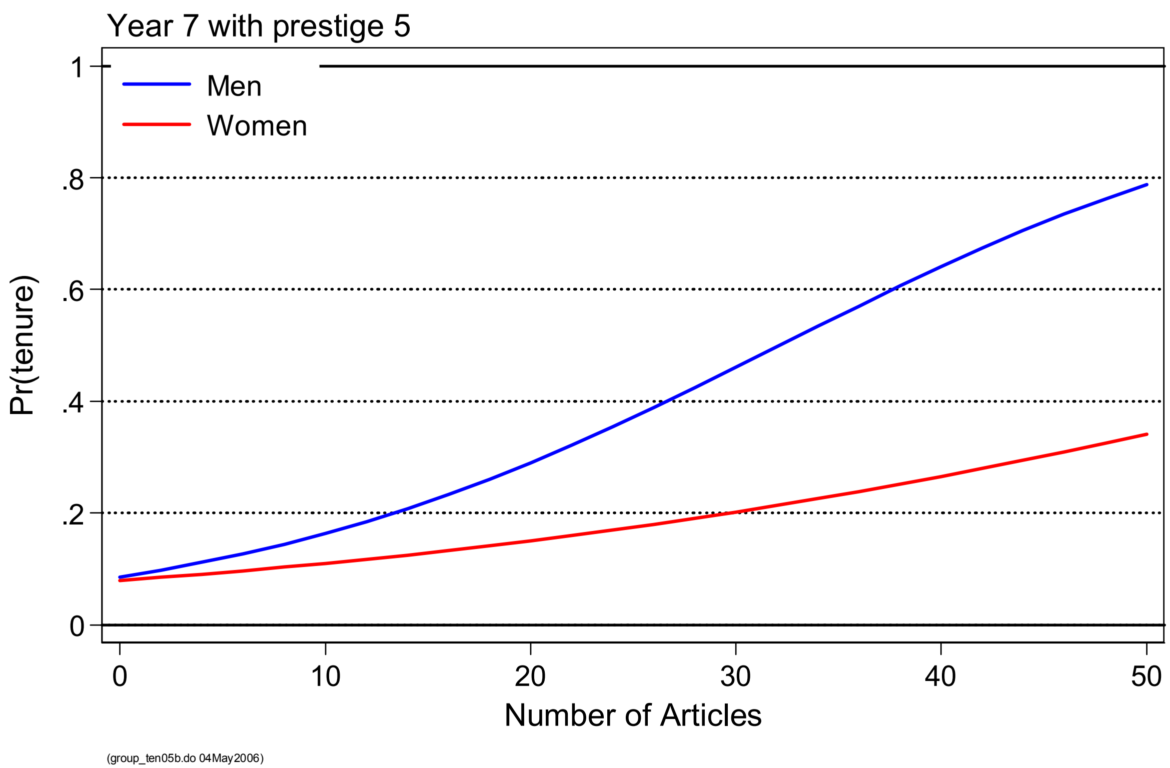
M2: $\Delta\,(articles)$ by job type



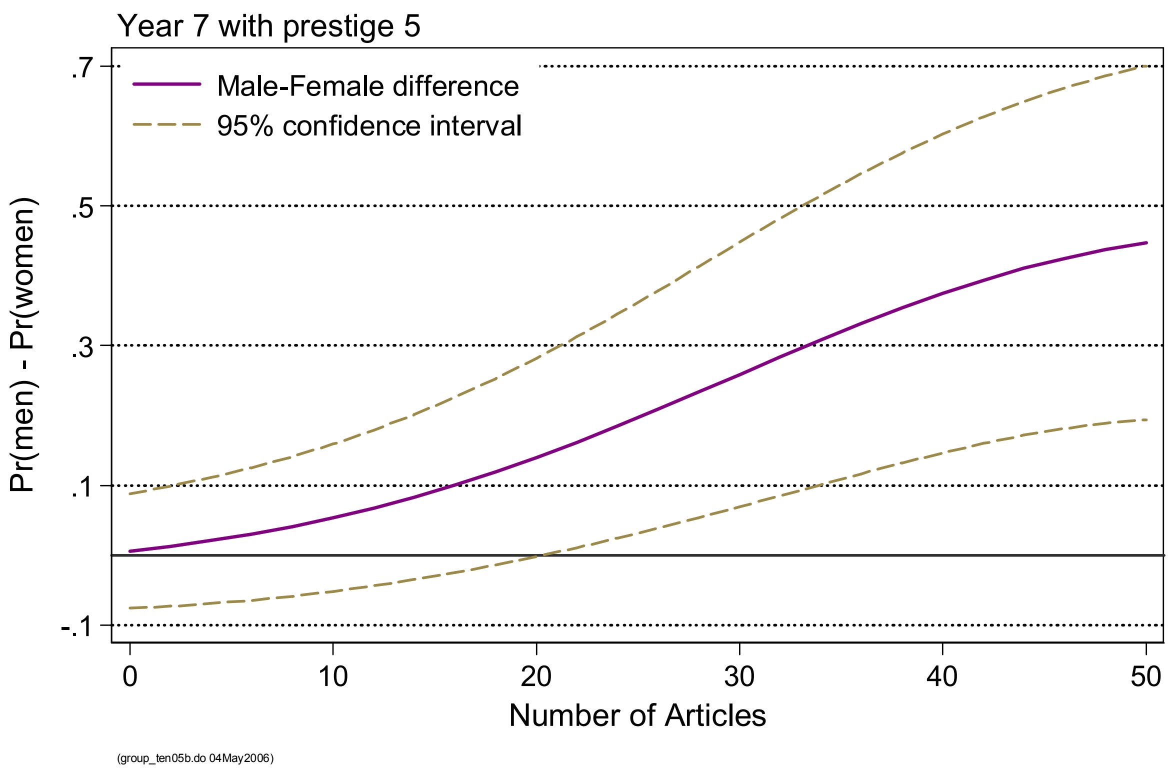
M3: full model

	Women			Men		
Variable	β	e^{β}	z	β	e^{β}	z
constant	-4.21		-6.68	-5.82		-11.55
year	0.77	—	6.12	1.07	—	9.08
yearsq	-0.04	—	-4.93	-0.07	—	-7.51
select	0.03	1.04	0.50	0.21	1.23	3.69
articles	0.04	1.04	2.98	0.07	1.08	6.84
prestige	-0.35	0.71	-2.29	-0.38	0.69	-3.64
Log-lik	-338.85			-579.23		
N	1,121			1,824		
Chow tests:	χ^2_{articles}=1.07, df=1, p=.30.					
	χ^2_{prestige}=0.03, df=1, p=.86.					
Allison tests:	χ^2_{articles}=1.73, df=1, p=.19 (other effects equal).					
	χ^2_{prestige}=1.55, df=1, p=.21 (other effects equal).					

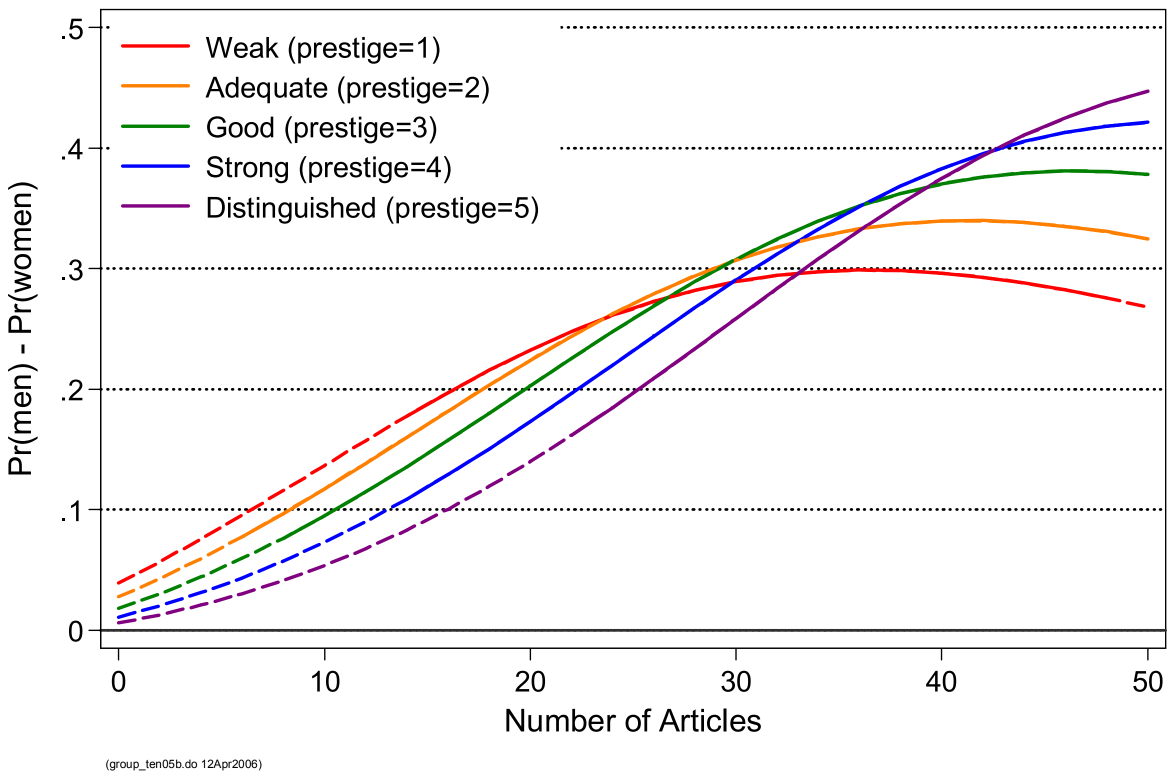
M3: Pr($tenure$ if $prestige = 5$)



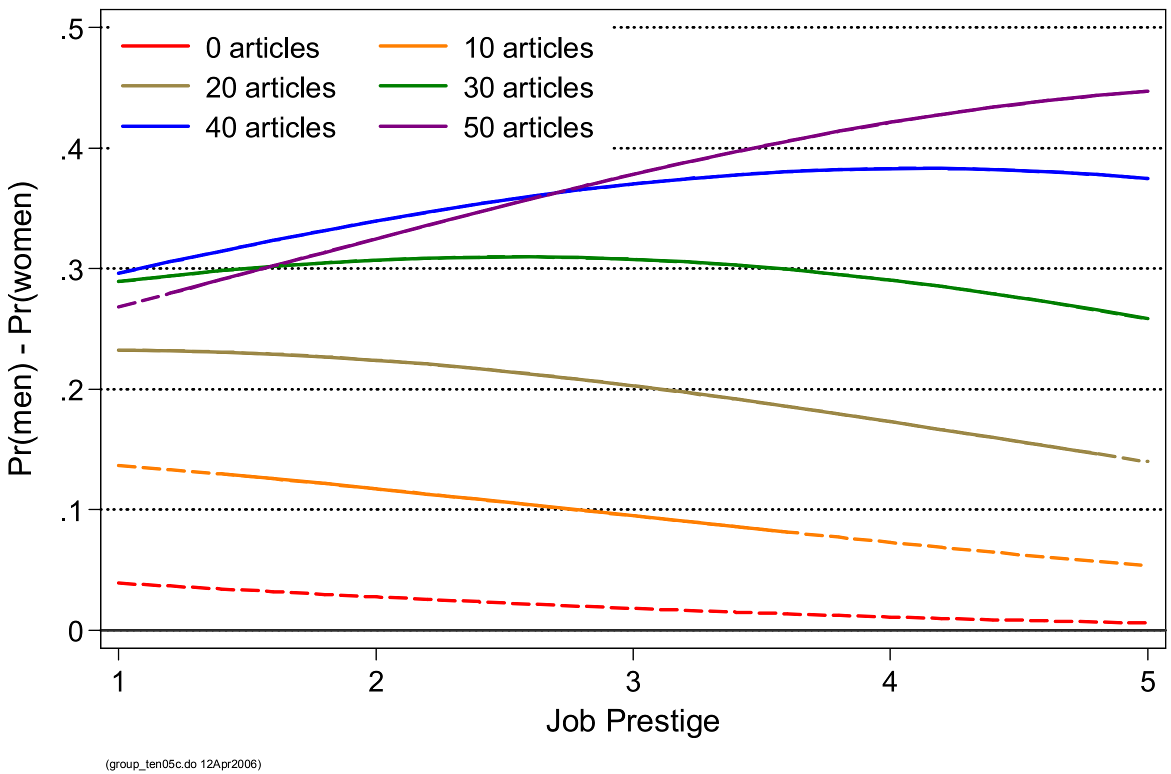
M3: $\Delta\,(articles$ **if** $prestige = 5)$



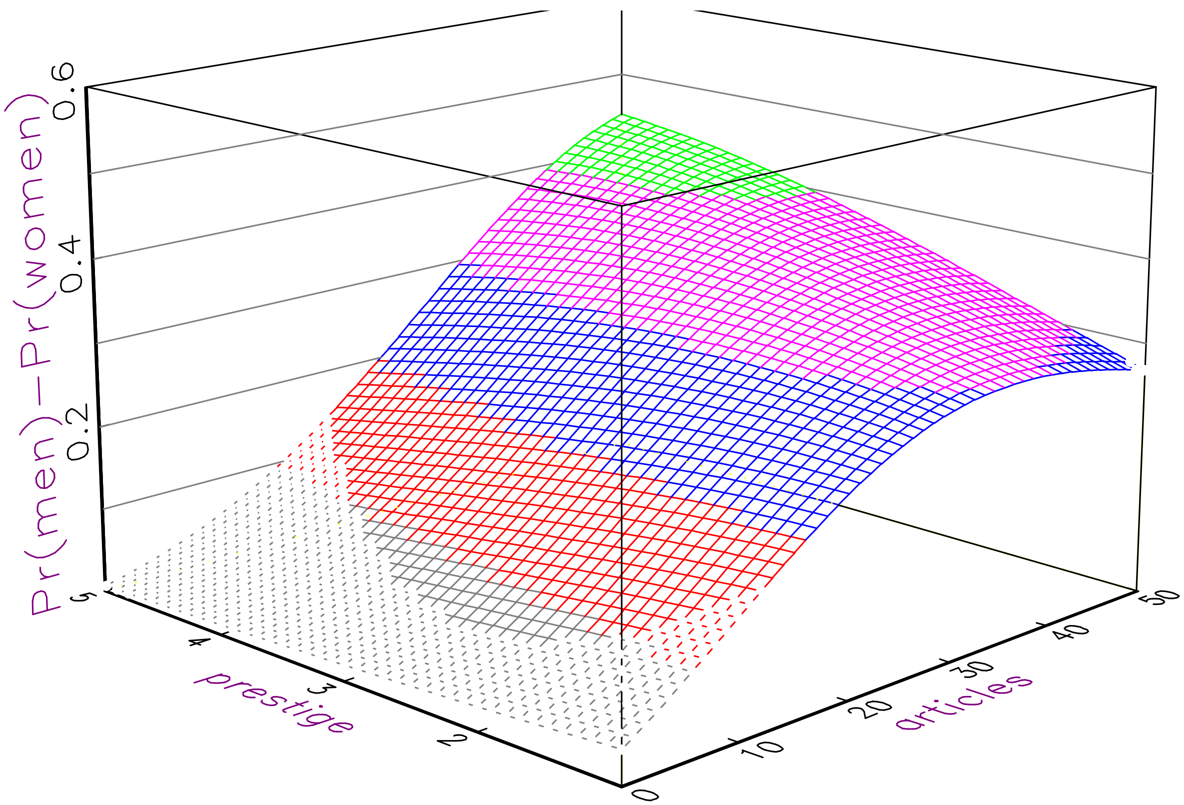
M3: $\Delta\,(articles$ by $prestige)$



M3: $\Delta\,(prestige$ by $articles)$



M3: $\Delta\,(articles,\,prestige)$



Conclusions

1. Predictions offer a general approach for comparing groups.
2. The approach cannot distinguish between different processes generating the same probabilities. For example:

$$\begin{array}{llll} \text{Case 1:} & \beta_x^m = \beta_x^w & \sigma_m \neq \sigma_w & \Rightarrow \Delta(x) = 0 \\ \text{Case 2:} & \beta_x^m \neq \beta_x^w & \sigma_m = \sigma_w & \Rightarrow \Delta(x) = 0 \end{array}$$

3. But, it provides a very detailed understanding of group differences.
4. Implications for interpreting nonlinear models.
5. In Stata, enter: `findit spost_groups` for examples.
6. www.indiana.edu/~jslsoc/research.htm

References

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. and Freese, J. 2005. Regression Models for Categorical and Limited Dependent Variables with Stata. Second Edition. College Station, TX: Stata Press.

Long, J. Scott, Paul D. Allison, and Robert McGinnis. 1993. "Rank Advancement in Academic Careers: Sex Differences and the Effects of Productivity." American Sociological Review 58:703-722.

Xu, J. and J.S. Long, 2005, Confidence intervals for predicted outcomes in regression models for categorical outcomes. The Stata Journal 5: 537-559.

Delta method for $\Delta(\mathrm{x}; \beta)$

1. Start with the predicted probability:

$$G(\boldsymbol{\beta}) = \Pr(y = 1 \mid \mathbf{x}) = F(\mathbf{x}\boldsymbol{\beta})$$

2. A Taylor series expansion of $G(\widehat{\boldsymbol{\beta}})$:

$$G(\widehat{\boldsymbol{\beta}}) \approx G(\boldsymbol{\beta}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T G'(\boldsymbol{\beta})$$

3. Where

$$\begin{aligned} G'(\boldsymbol{\beta}) &= \left[\begin{array}{cccc} \frac{\partial F(\mathbf{x}\boldsymbol{\beta})}{\partial \beta_0} & \frac{\partial F(\mathbf{x}\boldsymbol{\beta})}{\partial \beta_1} & \cdots & \frac{\partial F(\mathbf{x}\boldsymbol{\beta})}{\partial \beta_K} \end{array} \right]^T \\ &= f(\mathbf{x}\boldsymbol{\beta})\mathbf{x}^T \end{aligned}$$

4. Under standard assumptions, $G(\widehat{\boldsymbol{\beta}})$ is distributed normally around $G(\boldsymbol{\beta})$ with a variance

$$\textit{Var}\left[G(\widehat{\boldsymbol{\beta}})\right] = G'(\widehat{\boldsymbol{\beta}})^T Var(\widehat{\boldsymbol{\beta}}) G'(\widehat{\boldsymbol{\beta}})$$

5. For a difference in probability, let

$$G\left(\boldsymbol{\beta}\right) = F(\boldsymbol{\beta}|\mathbf{x}_a) - F\left(\boldsymbol{\beta}|\mathbf{x}_b\right)$$

6. Then

$$
\begin{aligned}
&Var\left[F(\widehat{\boldsymbol{\beta}}|\mathbf{x}_a) - F\left(\widehat{\boldsymbol{\beta}}|\mathbf{x}_b\right)\right] = \\
&\quad \left[\frac{\partial F(\boldsymbol{\beta}|\mathbf{x}_a)}{\partial \boldsymbol{\beta}^T} Var(\widehat{\boldsymbol{\beta}}) \frac{\partial F(\boldsymbol{\beta}|\mathbf{x}_a)}{\partial \boldsymbol{\beta}}\right] \\
&- \left[\frac{\partial F(\boldsymbol{\beta}|\mathbf{x}_a)}{\partial \boldsymbol{\beta}^T} Var(\widehat{\boldsymbol{\beta}}) \frac{\partial F(\boldsymbol{\beta}|\mathbf{x}_b)}{\partial \boldsymbol{\beta}}\right] \\
&- \left[\frac{\partial F(\boldsymbol{\beta}|\mathbf{x}_b)}{\partial \boldsymbol{\beta}^T} Var(\widehat{\boldsymbol{\beta}}) \frac{\partial F(\boldsymbol{\beta}|\mathbf{x}_a)}{\partial \boldsymbol{\beta}}\right] \\
&+ \left[\frac{\partial F(\boldsymbol{\beta}|\mathbf{x}_b)}{\partial \boldsymbol{\beta}^T} Var(\widehat{\boldsymbol{\beta}}) \frac{\partial F(\boldsymbol{\beta}|\mathbf{x}_b)}{\partial \boldsymbol{\beta}}\right]
\end{aligned}
$$

Proof of invariance of Pr$(y = 1)$ to σ_ε

1. If $N(0, \sigma = 1)$

$$\phi(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2}\right)$$

2. Then:

$$\Pr(y = 1 \mid x) = \int_{-\infty}^{\alpha+\beta x} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt \qquad (1)$$

3. If $N(0, \sigma = 2)$:

$$\phi(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{8}\right)$$

4. Then,

$$\Pr(y=1 \mid x) = \int_{-\infty}^{\sigma(\alpha+\beta x)} \frac{1}{2\sqrt{2\pi}} \exp\left(\frac{-t^2}{8}\right) dt \quad (2)$$

5. Equations 1 and 2 simply involve a linear change of variables, so the probabilities are equal.

6. If $z = g(x)$ and $x = g^{-1}(z)$, then

$$\int_B^A f(z)\, dz = \int_{g^{-1}(B)}^{g^{-1}(A)} f[g(z)] \frac{dz}{dx} dx$$

7. See Long 1997: 49-50 for an alternative proof.