Statistical Modeling: Applications

Intermediate Social Statistics

Week 7 & 8 (1 & 8 March 2016)

Andy Eggers

Ordinal probit application: Hainmueller and Hiscox 2010

Two economic explanations for (variation in) anti-immigrant sentiment:

- **Labor market competition** → natives should oppose immigrants with skill levels similar to their own
- **Fiscal burden** → rich natives should be more opposed to low-skilled immigrants than poor natives (especially where immigrants use a lot of public services)

(Random whether respondent gets A or B)

Hainmueller and Hiscox ask a sample of US respondents either

A. Do you agree or disagree that the US should allow more **highly skilled immigrants** from other countries to come and live here?

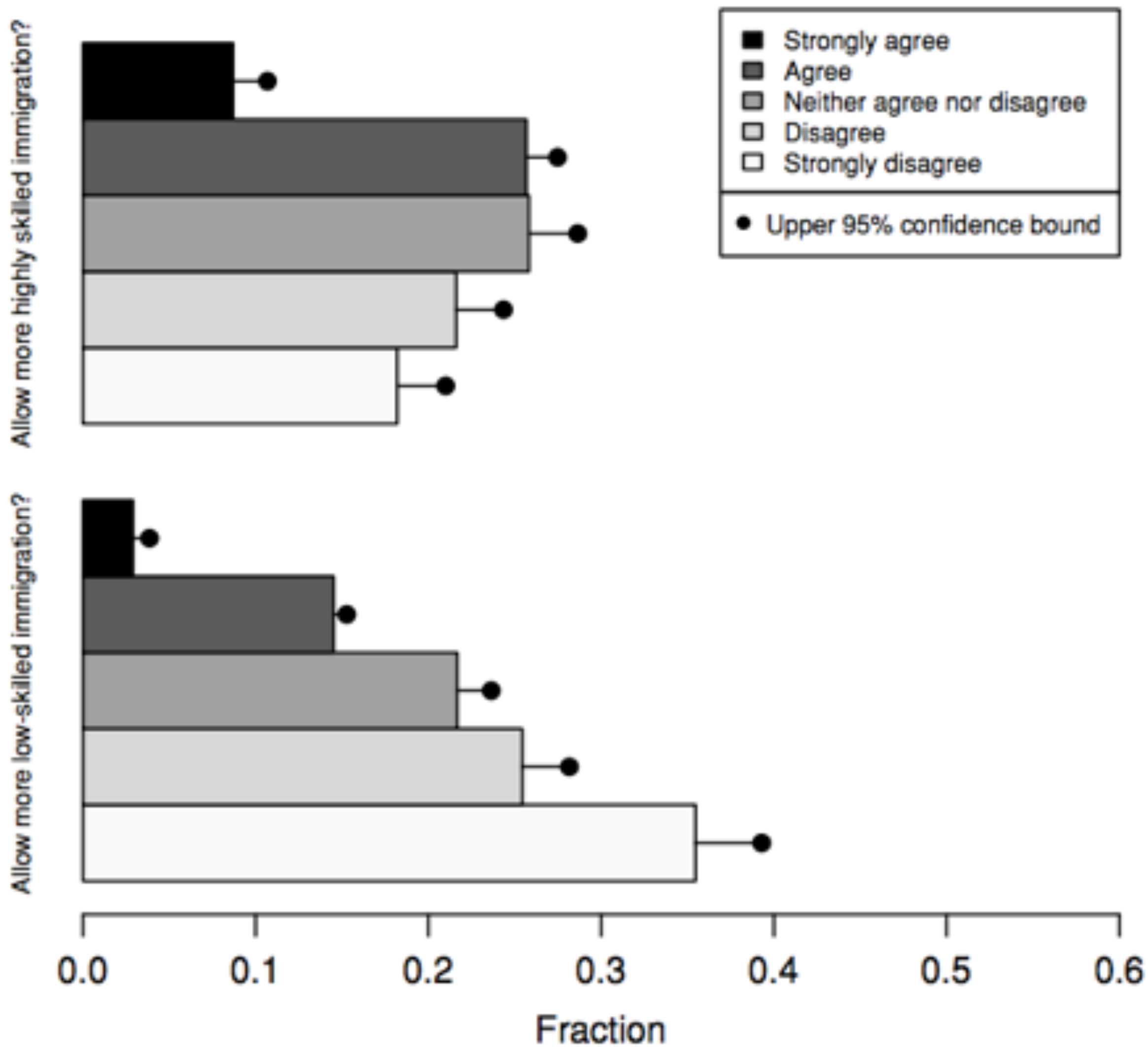
B. Do you agree or disagree that the US should allow more **low-skilled immigrants** from other countries to come and live here?

Hainmueller and Hiscox (2010)

FIGURE 2. Support for Highly Skilled and Low-skilled Immigration

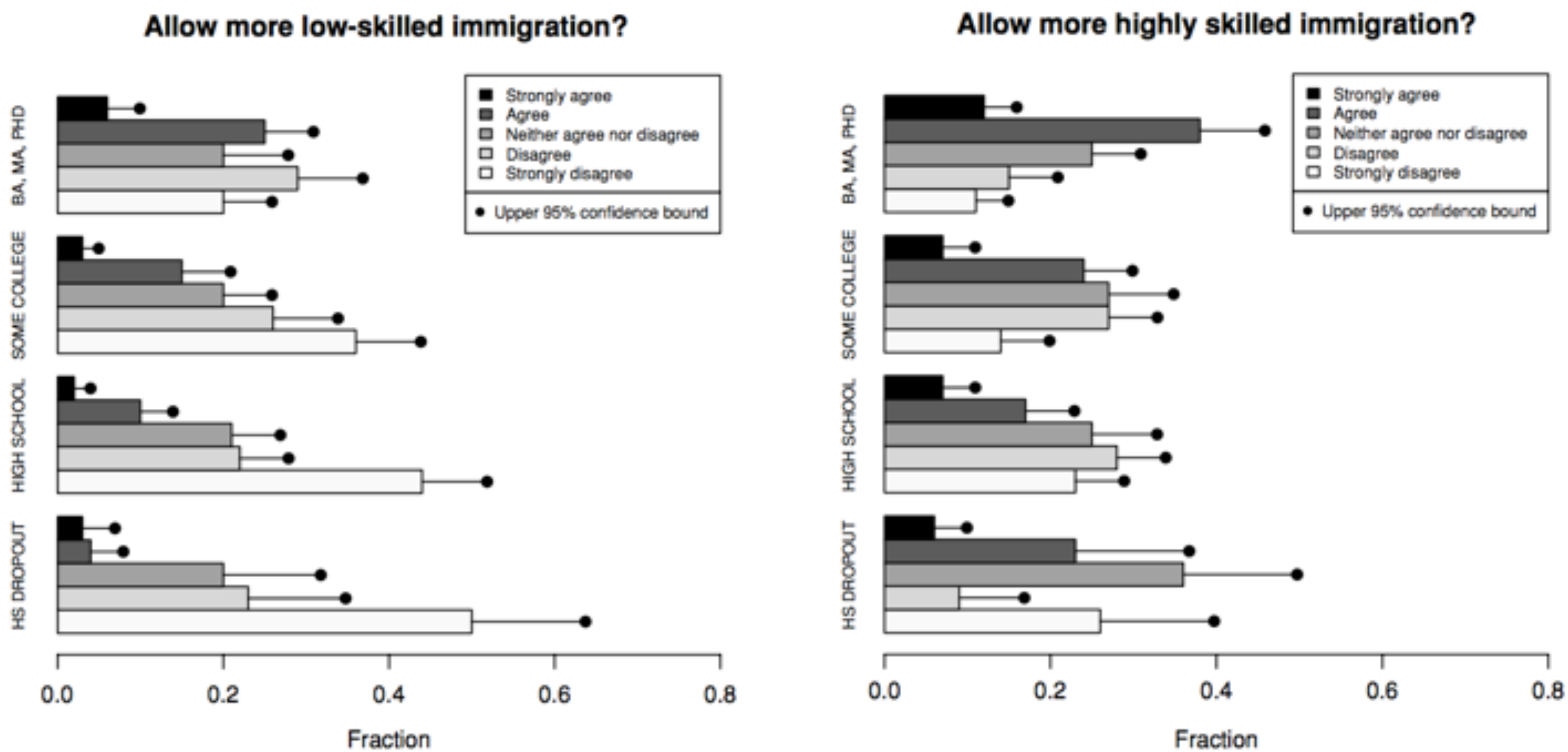
Question A (allow more high-skilled immigration?)

Question B (allow more low-skilled immigration?)



Hainmueller and Hiscox (2010)

FIGURE 3. Support for Highly Skilled and Low-skilled Immigration by Respondents' Skill Level



Ordered probit

Motivations:

- **Predict** ordered outcome Y
- Characterize the determinants of a **latent variable** Y* (e.g. support for immigration) underlying ordered outcome Y

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

Ordered probit

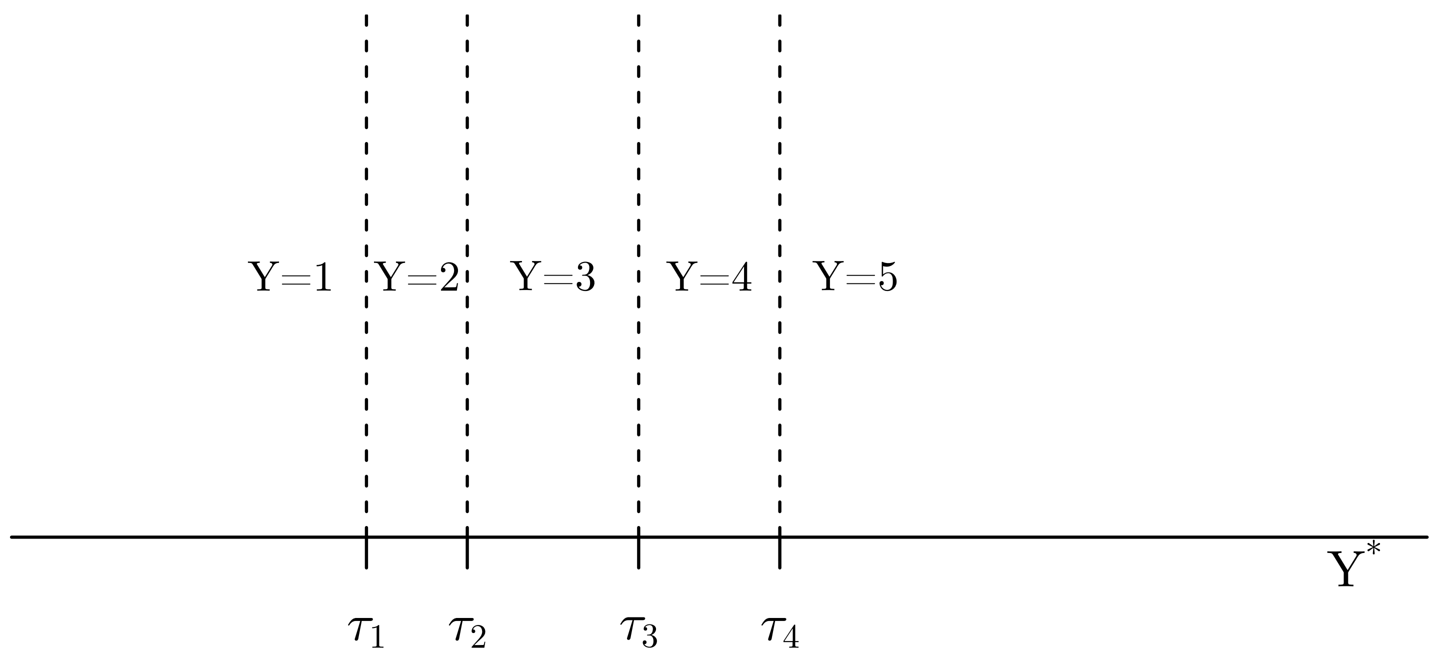
Suppose we observed Y* (support for immigration), which perfectly predicts the response given:

$$Y = \begin{cases} 1, & \text{if } Y^* \leq \tau_1 \\ 2, & \text{if } Y^* \in (\tau_1, \tau_2] \\ 3, & \text{if } Y^* \in (\tau_2, \tau_3] \\ 4, & \text{if } Y^* \in (\tau_3, \tau_4] \\ 5, & \text{if } Y^* > \tau_4 \end{cases}$$

Ordered probit

Suppose we observed Y* (support for immigration), which perfectly predicts the response given:

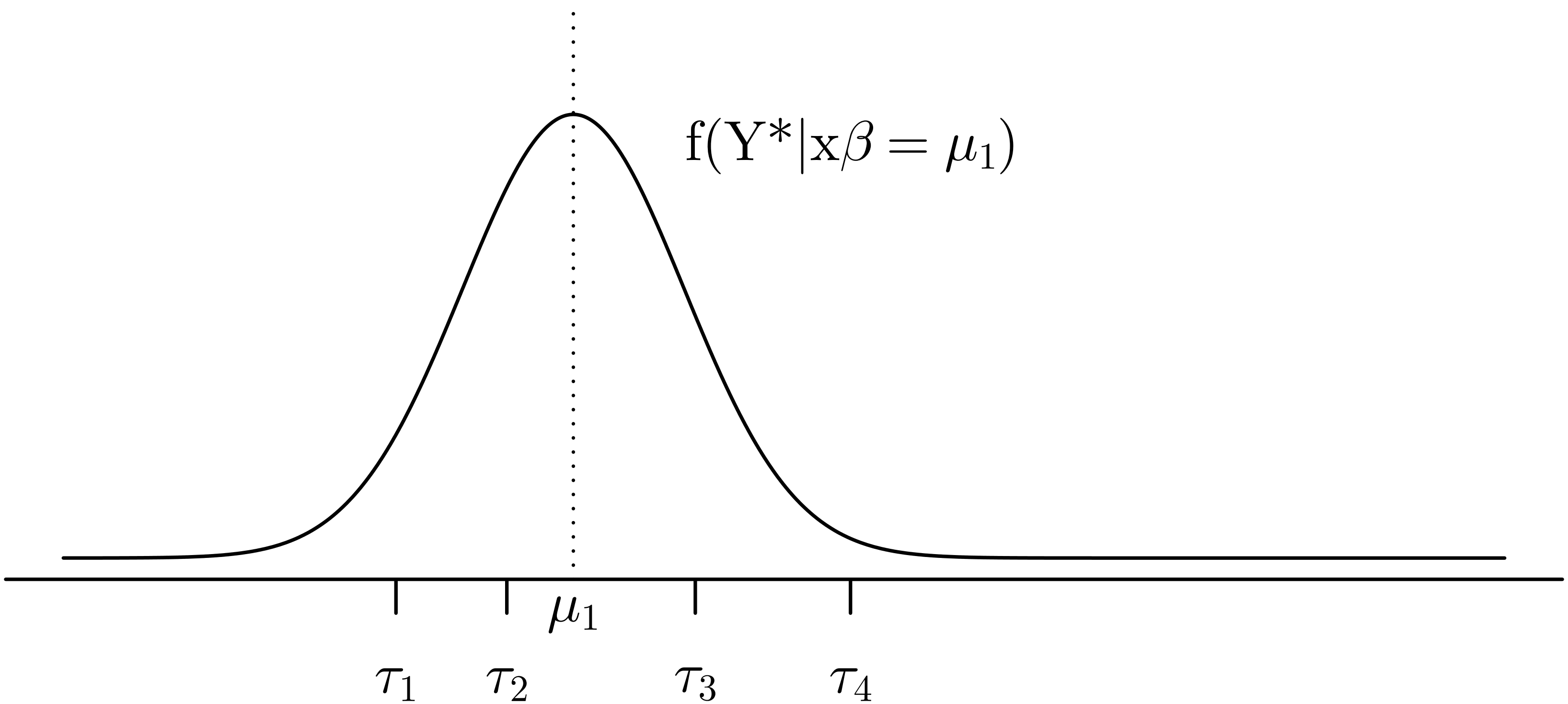
$$Y = \begin{cases} 1, & \text{if } Y^* \leq \tau_1 \\ 2, & \text{if } Y^* \in (\tau_1, \tau_2] \\ 3, & \text{if } Y^* \in (\tau_2, \tau_3] \\ 4, & \text{if } Y^* \in (\tau_3, \tau_4] \\ 5, & \text{if } Y^* > \tau_4 \end{cases}$$



Ordered probit

We don't observe Y*, but we postulate that it is a linear function of covariates, plus error:

$$Y^* = x\beta + \epsilon$$
$$\epsilon \sim N(0,1)$$

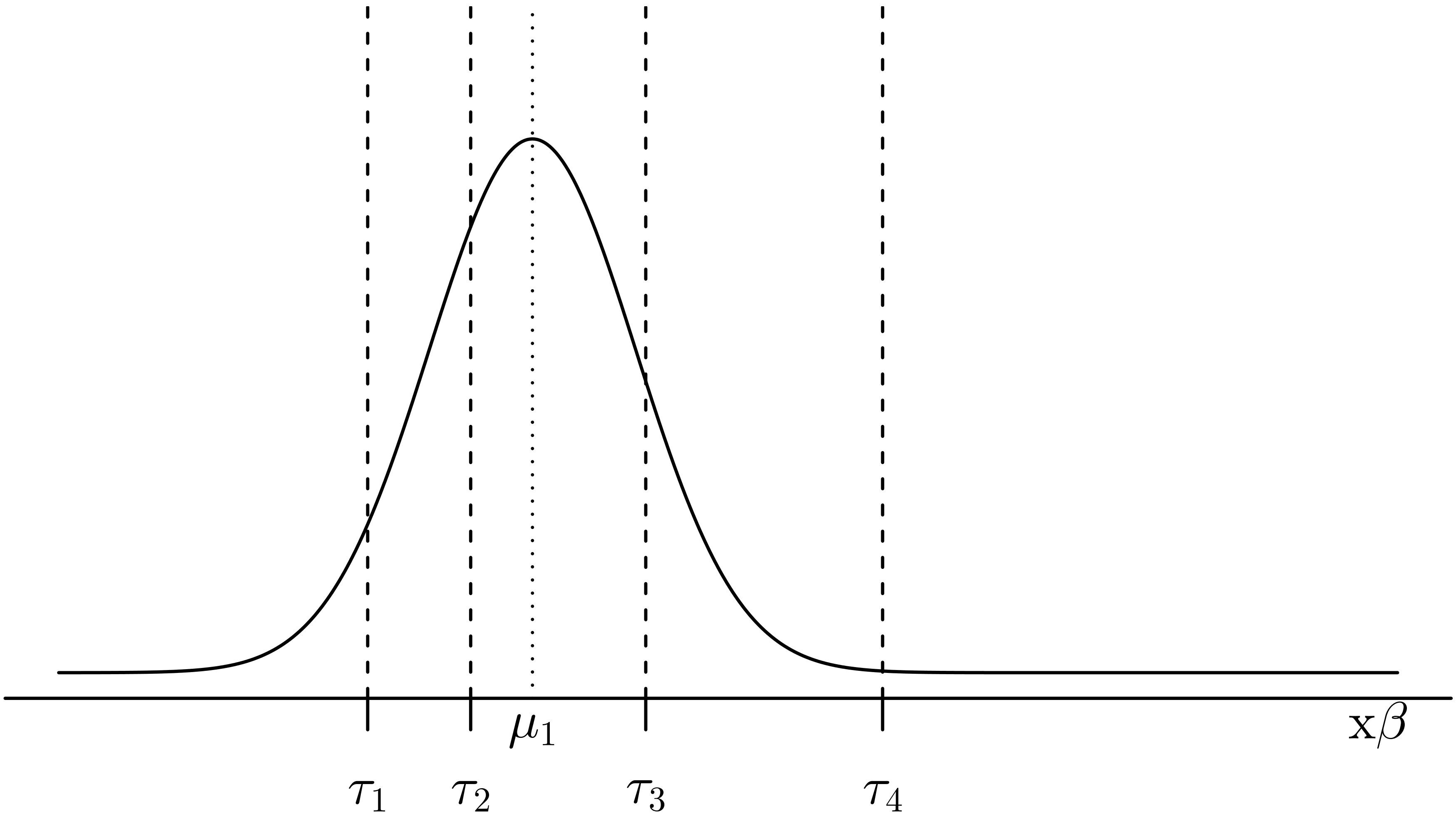


Ordered probit: visualization

That implies that given $\tau_1, \tau_2, \tau_3, \tau_4$ and $\mu_i = x_i\beta$ we know the probability of each outcome:

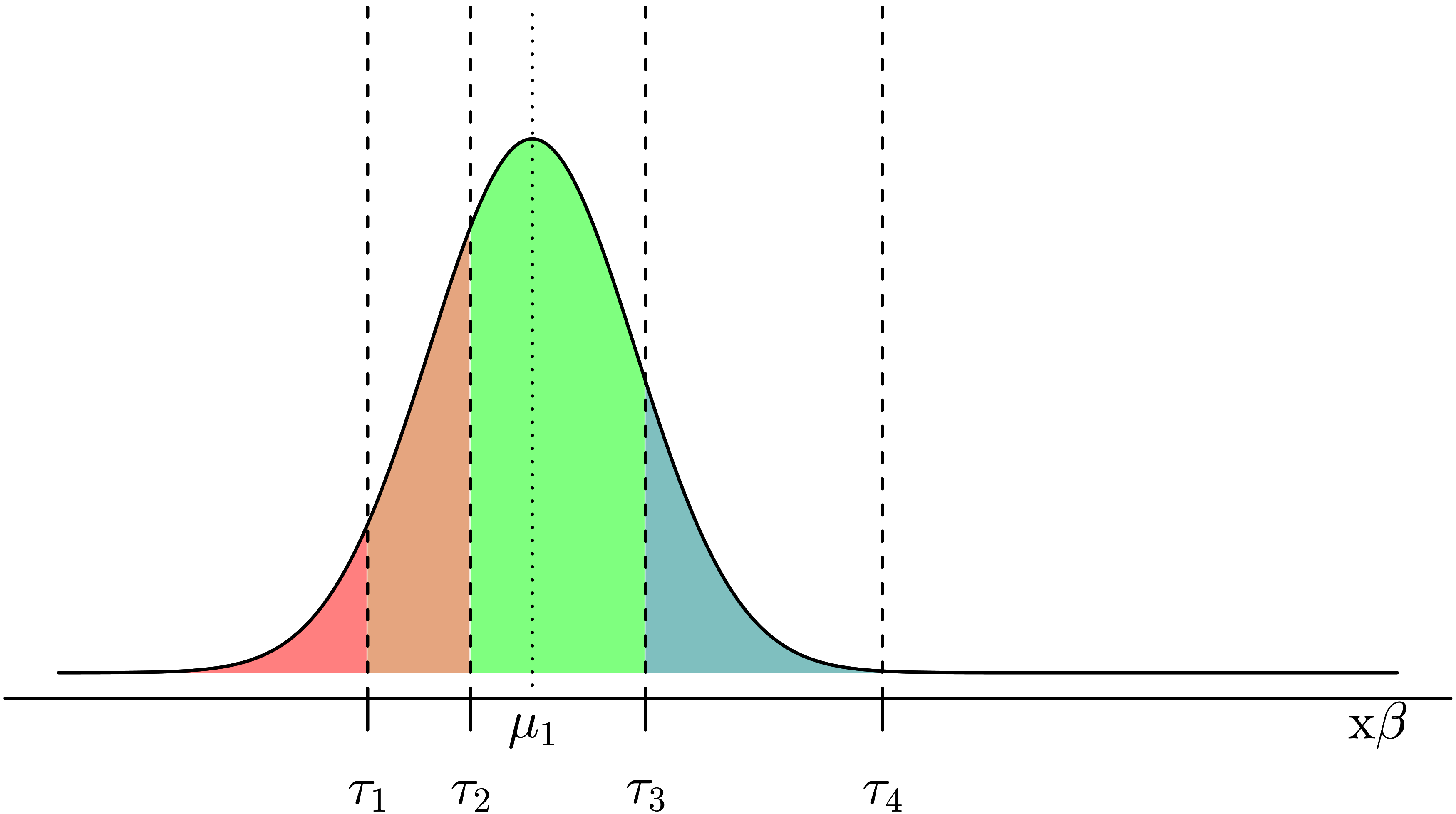
Ordered probit: visualization

That implies that given $\tau_1, \tau_2, \tau_3, \tau_4$ and $\mu_i = x_i\beta$ we know the probability of each outcome:

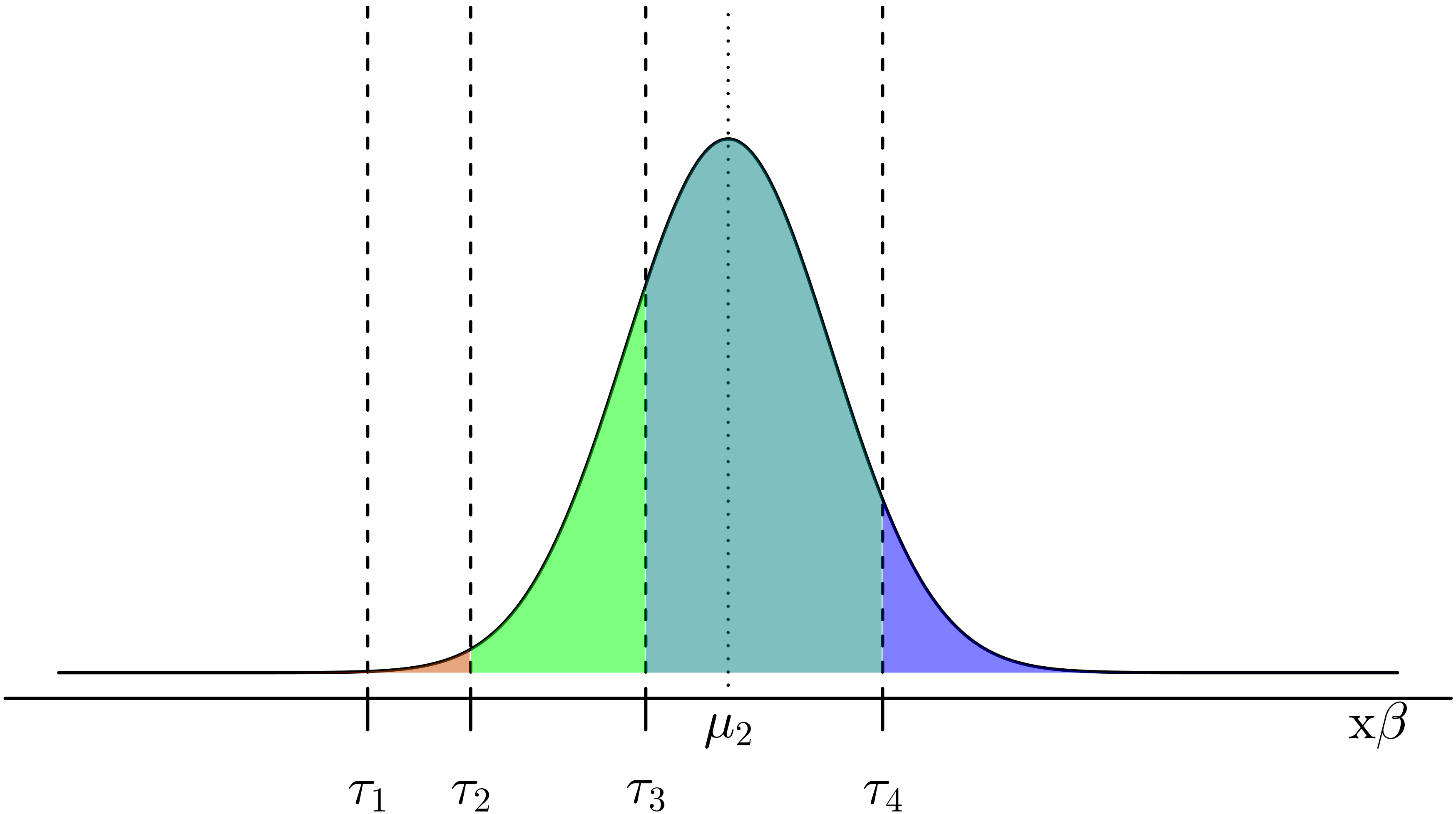


Ordered probit: visualization

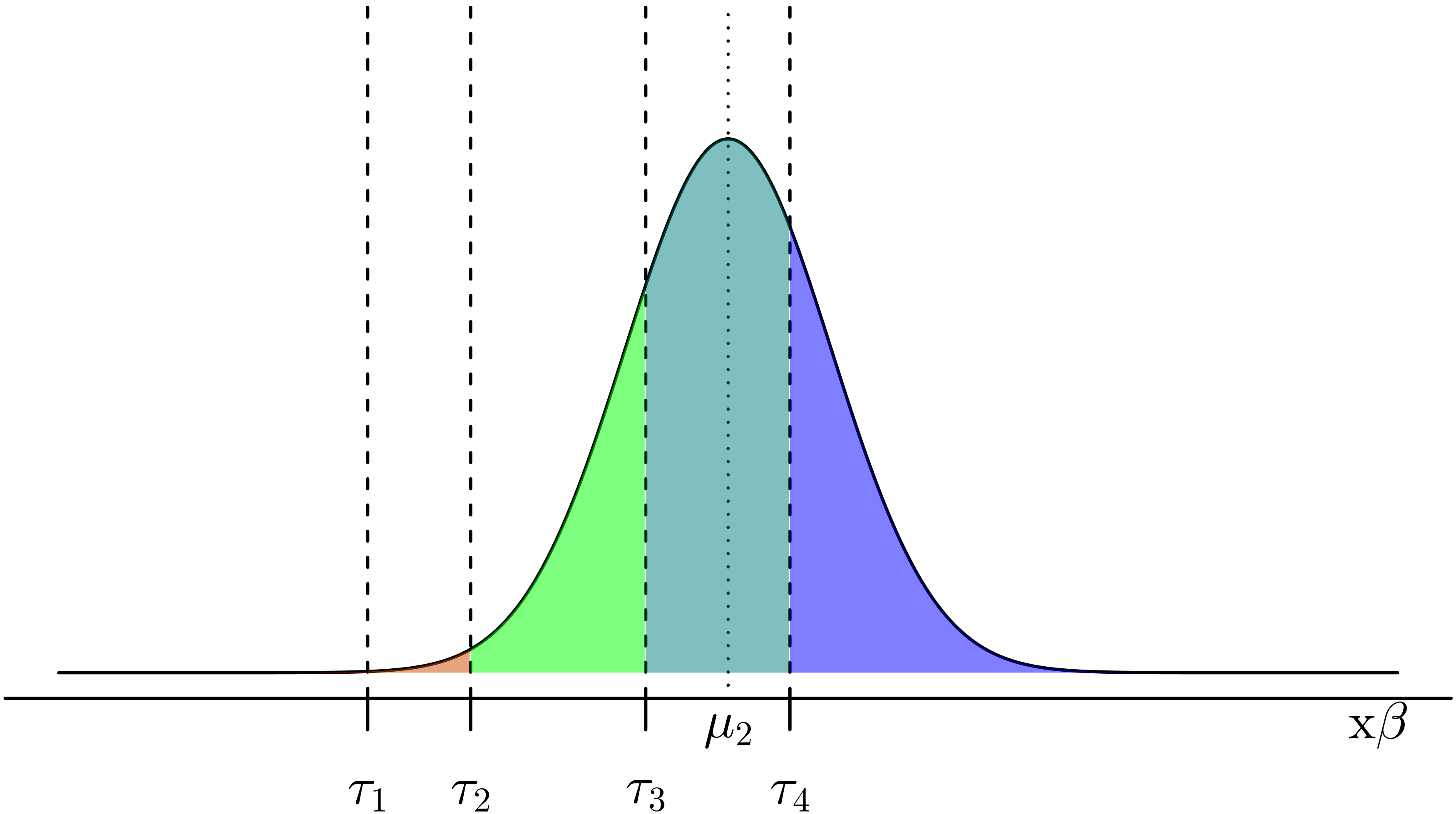
That implies that given $\tau_1, \tau_2, \tau_3, \tau_4$ and $\mu_i = x_i\beta$ we know the probability of each outcome:



Ordered probit: visualization (2)

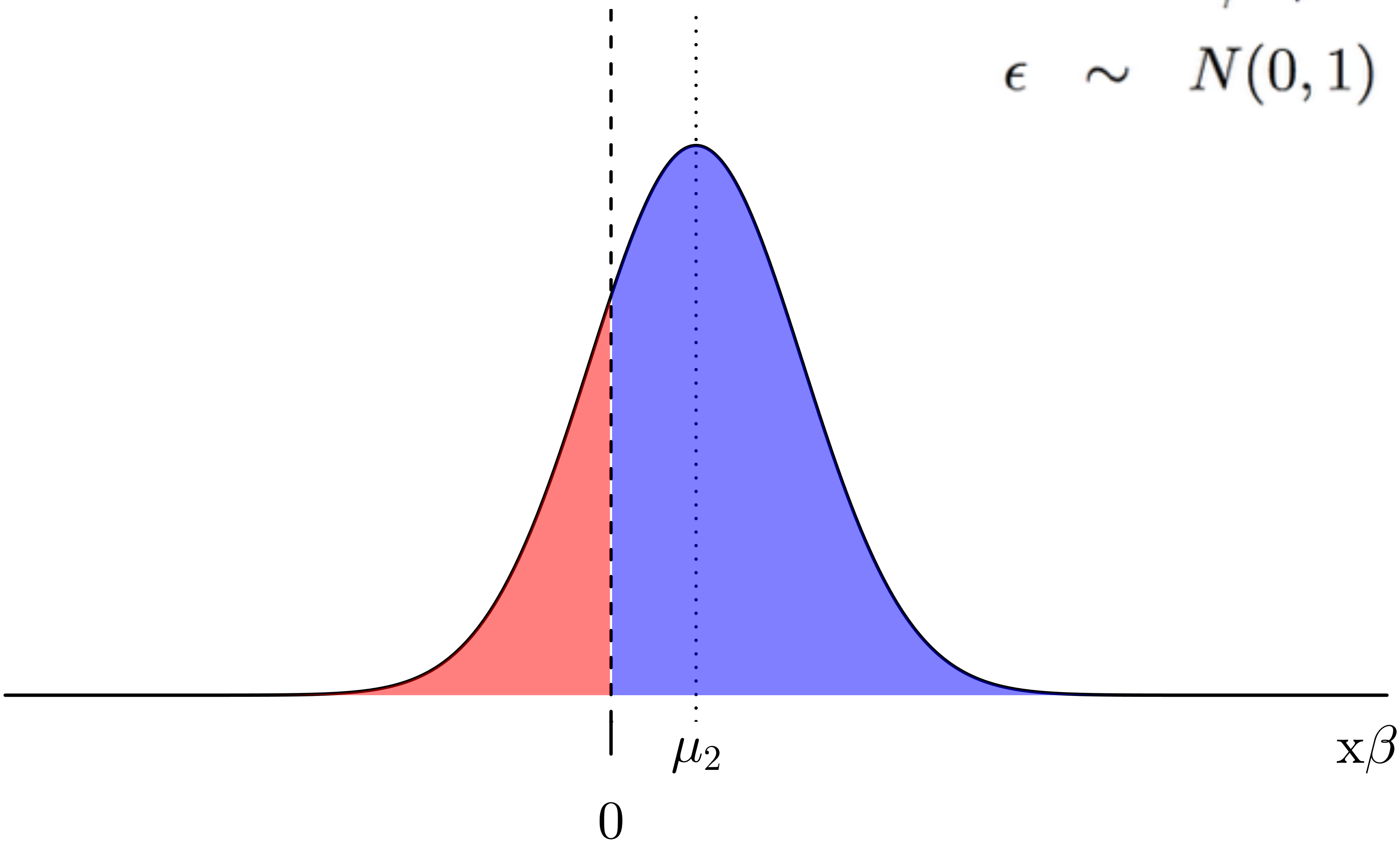


Ordered probit: visualization (3)



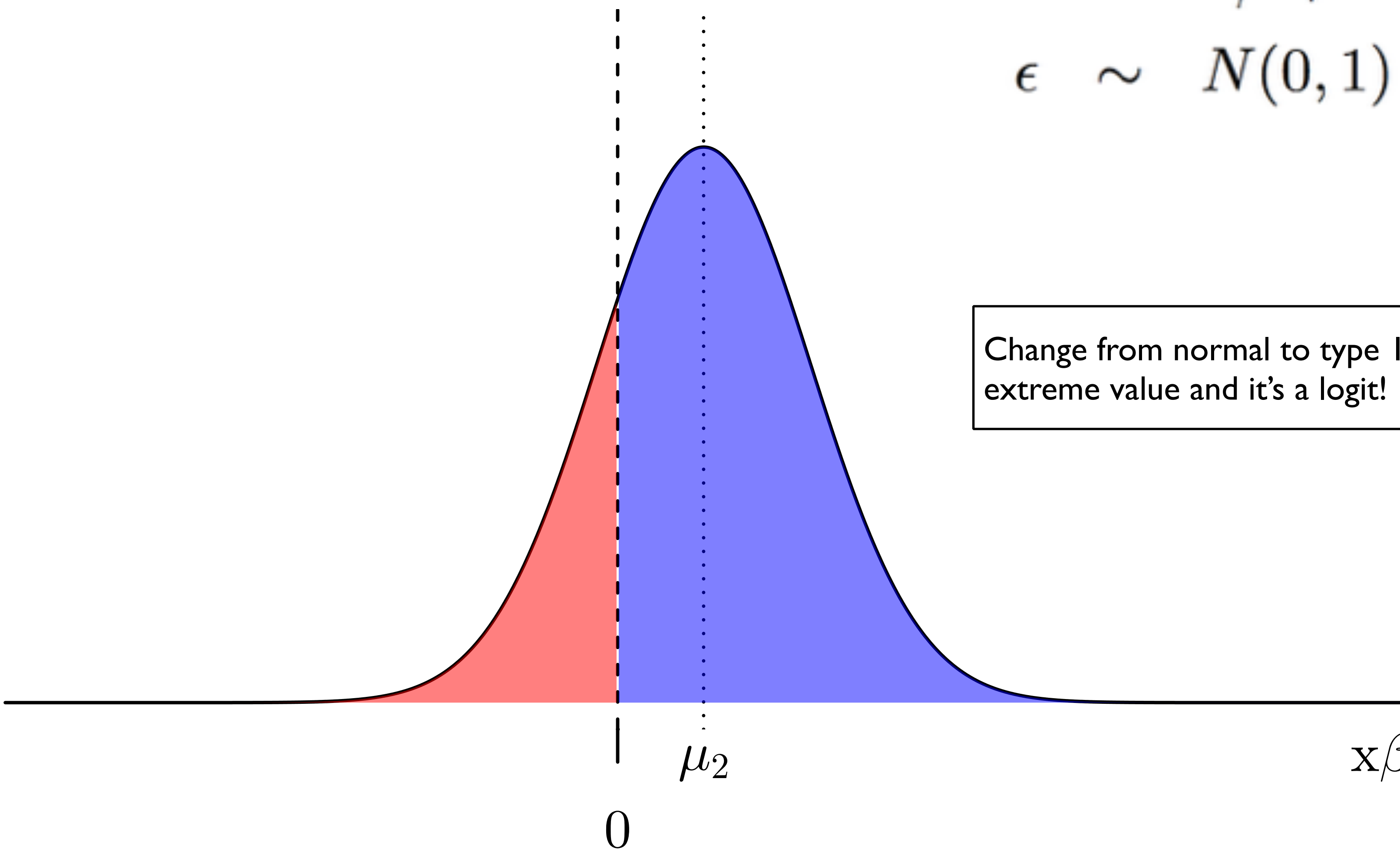
Binary probit: a special case with single threshold at 0

$$Y^* = x\beta + \epsilon$$
$$\epsilon \sim N(0,1)$$



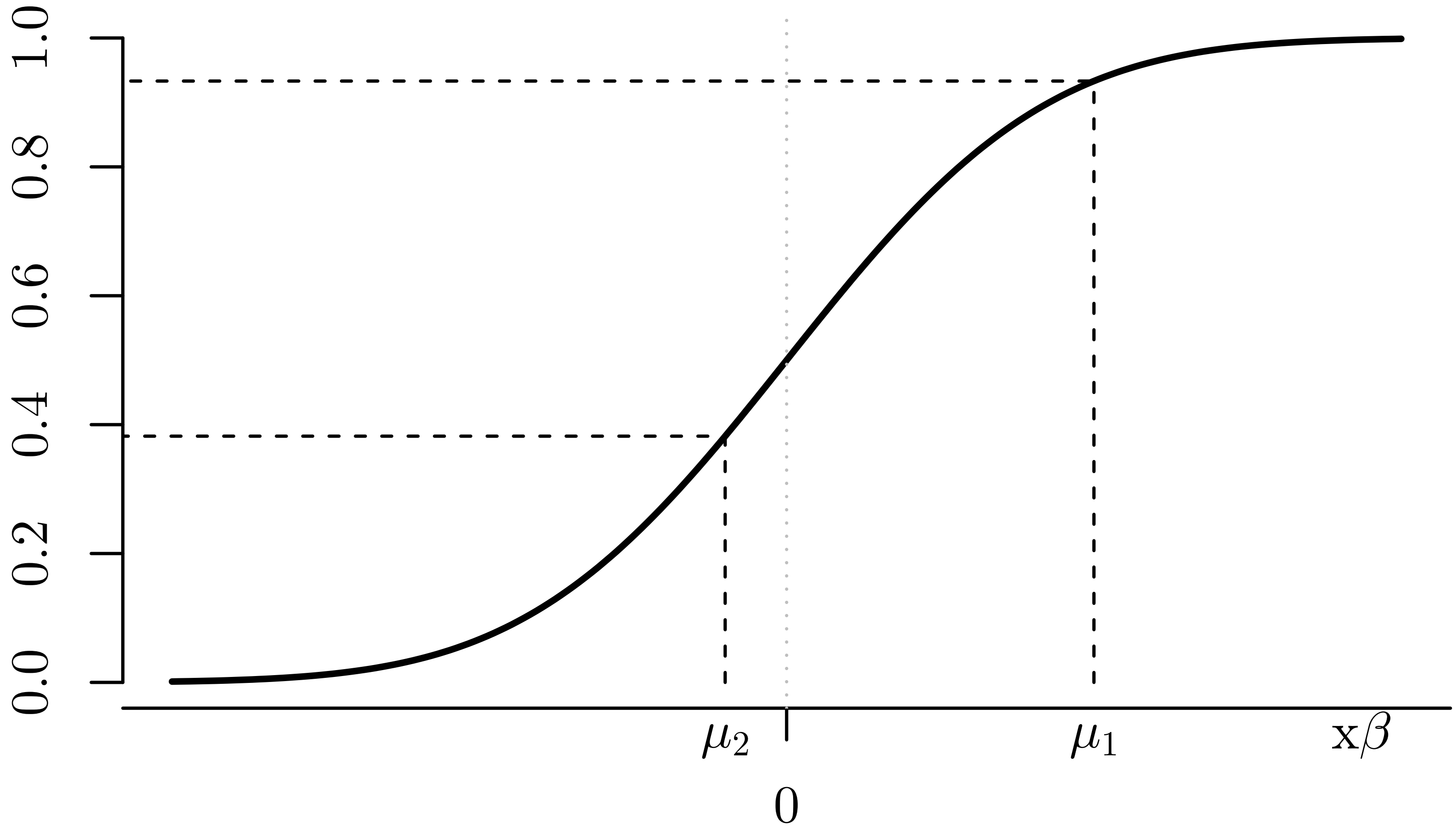
Binary probit: a special case with single threshold at 0

$$Y^* = x\beta + \epsilon$$
$$\epsilon \sim N(0,1)$$



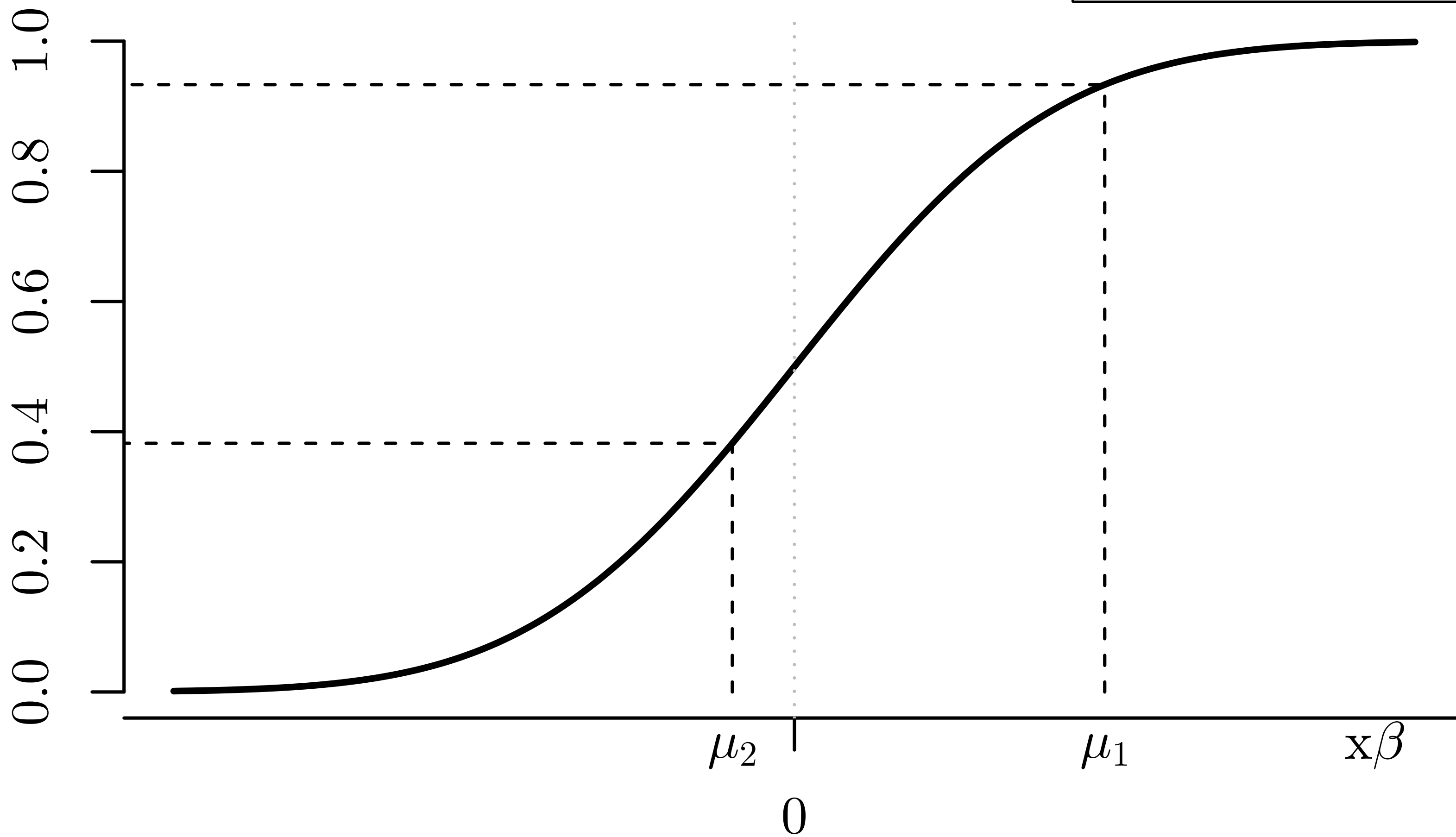
Change from normal to type 1 extreme value and it's a logit!

Binary probit: a special case with single threshold at 0



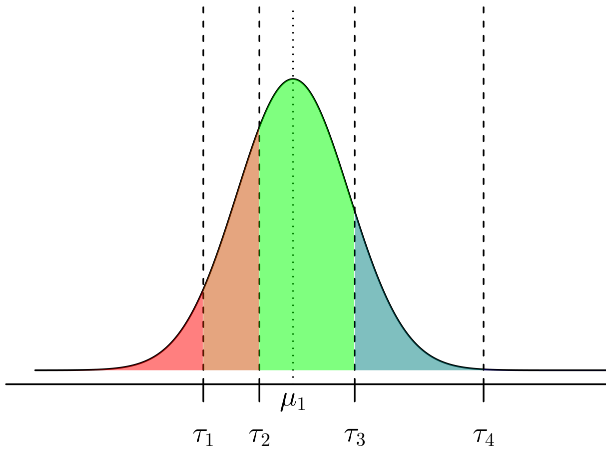
Binary probit: a special case with single threshold at 0

Change from normal CDF to type I extreme value and it's a logit!



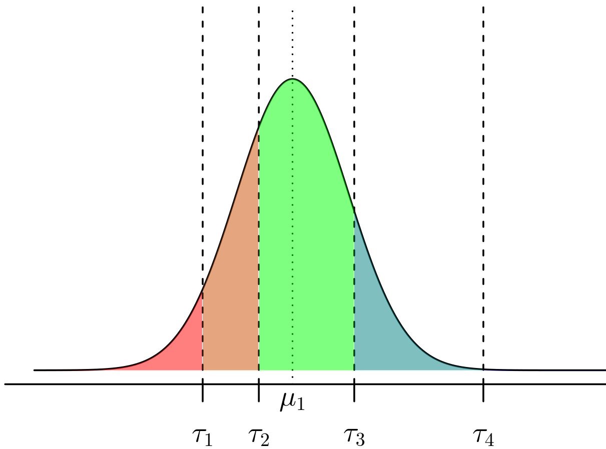
Ordered probit: assumptions

What are the key assumptions of the standard ordered probit model? In what circumstances would these assumptions not hold? What might we miss?



Ordered probit: assumptions

What are the key assumptions of the standard ordered probit model? In what circumstances would these assumptions not hold? What might we miss?



Some key points:

- model does not permit "polarization" of responses due to given X
  - if Xβ implies outcome j, then increasing Xβ makes outcomes below j less likely and outcomes above j more likely
  - (no different from OLS, other GLMs in that respect)
- standard model does not permit given X affecting probability of outcome 1 vs outcome 2 without affecting outcome 3, etc (but could imagine making cutoffs a function of covariates?)

Ordered probit: estimation

Ordered probit: estimation

How do we estimate β and $\tau_1, \tau_2, \tau_3, \tau_4$?

Ordered probit: estimation

How do we estimate β and $\tau_1, \tau_2, \tau_3, \tau_4$?

Stata: `oprobit depvar [indepvars] [weight] [, options]`

Ordered probit: estimation

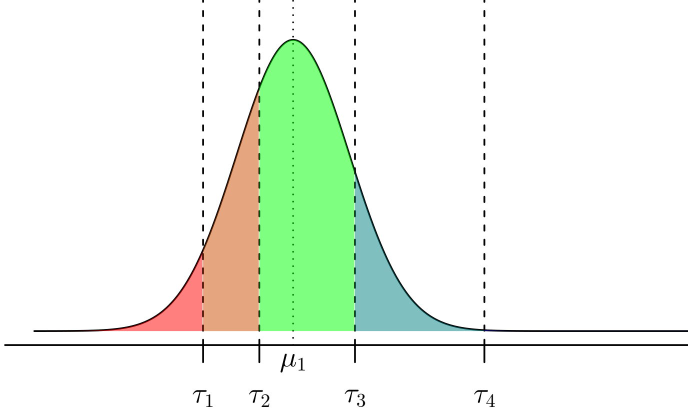
How do we estimate β and $\tau_1, \tau_2, \tau_3, \tau_4$?

Stata: `oprobit depvar [indepvars] [weight] [, options]`

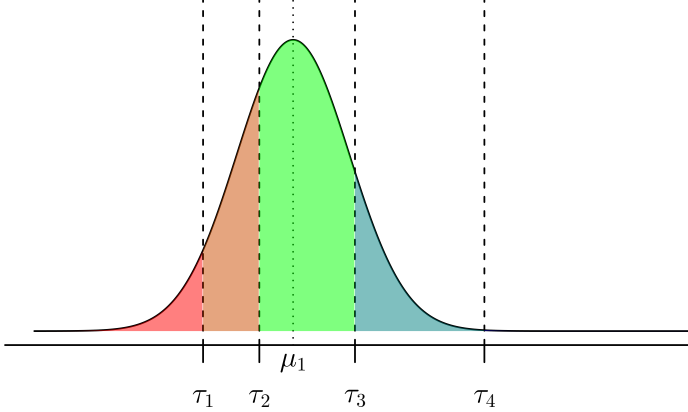
```
. oprobit sh_both hskframe ppeducat hskeduc xx* [pweight=weight1]

Iteration 0:   log pseudolikelihood = -2418.2933
Iteration 1:   log pseudolikelihood = -2306.2688
Iteration 2:   log pseudolikelihood = -2306.1887
Iteration 3:   log pseudolikelihood = -2306.1887

Ordered probit regression                       Number of obs     =      1,589
                                                Wald chi2(8)      =     158.52
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -2306.1887               Pseudo R2         =     0.0464
```

sh_both	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
hskframe	.7261249	.2025688	3.58	0.000	.3290974	1.123152
ppeducat	.2683796	.0484328	5.54	0.000	.1734531	.3633061
hskeduc	-.0653202	.0667142	-0.98	0.328	-.1960777	.0654373
xxfemale	-.1771998	.0644352	-2.75	0.006	-.3034904	-.0509092
xxppagecat	-.0110243	.0196088	-0.56	0.574	-.0494569	.0274083
xxWhite	-.374742	.0990717	-3.78	0.000	-.5689189	-.1805651
xxBlack	-.4720909	.1352577	-3.49	0.000	-.7371911	-.2069907
xxHispanic	.0627729	.2058409	0.30	0.760	-.3406679	.4662136
/cut1	-.114744	.1910944			-.4892822	.2597941
/cut2	.5613041	.1905945			.1877457	.9348625
/cut3	1.254911	.1907666			.8810152	1.628807
/cut4	2.258038	.2003352			1.865388	2.650688

Ordered probit: estimation

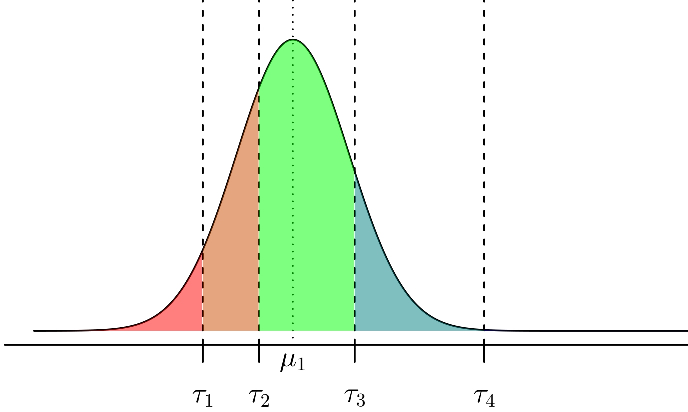


Ordered probit: estimation



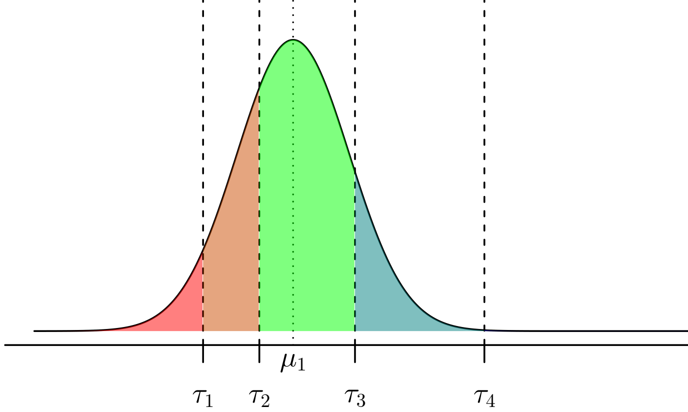
Think about what Stata is doing. Can you relate it to last week's Poisson activity? Notice any potential problems?

Ordered probit: estimation



Think about what Stata is doing. Can you relate it to last week's Poisson activity? Notice any potential problems?

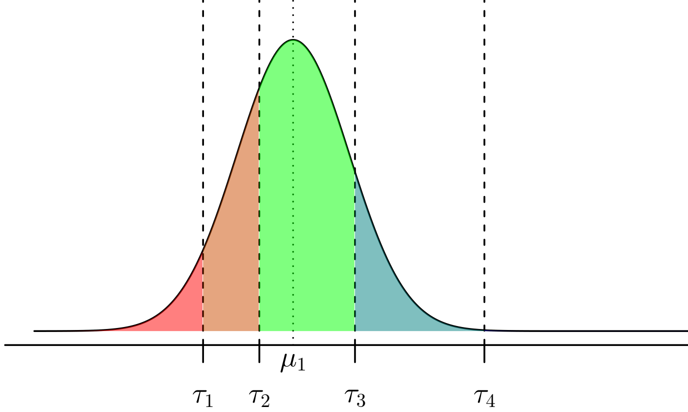
Ordered probit: estimation



Think about what Stata is doing. Can you relate it to last week's Poisson activity? Notice any potential problems?

Parameters are **unidentified** (no unique solution) unless we assume $\sigma^2 = 1$ and either

Ordered probit: estimation

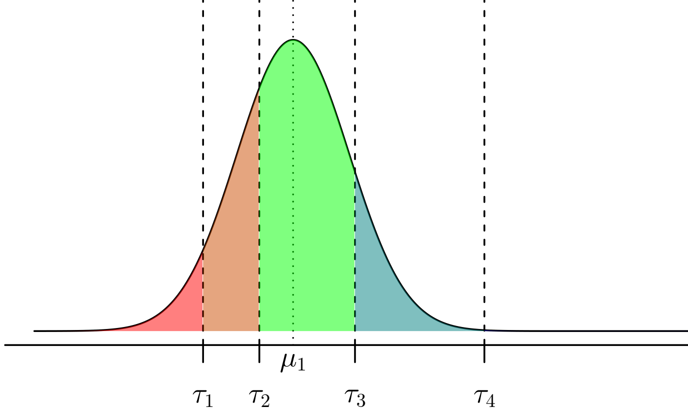


Think about what Stata is doing. Can you relate it to last week's Poisson activity? Notice any potential problems?

Parameters are **unidentified** (no unique solution) unless we assume $\sigma^2 = 1$ and either

- constrain cutoffs, e.g. $\tau_1 = 0$, or

Ordered probit: estimation



Think about what Stata is doing. Can you relate it to last week's Poisson activity? Notice any potential problems?

Parameters are **unidentified** (no unique solution) unless we assume $\sigma^2 = 1$ and either

- constrain cutoffs, e.g. $\tau_1 = 0$, or
- drop intercept (that is what Stata does automatically)

Back to Hainmueller and Hiscox

To explicitly test the labor market competition argument, we estimate the systematic component of the ordered probit model with the specification.

$$\mu_i = \alpha + \gamma \text{HSKFRAME}_i + \delta(\text{HSKFRAME}_i \cdot \text{EDUCATION}_i) + \theta \text{EDUCATION}_i + Z_i \psi,$$

where the parameter γ is the lower-order term on the treatment indicator that identifies the premium that natives attach to highly skilled immigrants relative to low-skilled immigrants. The parameter δ captures how the premium for highly skilled immigration varies conditional on the skill level of the respondent.

Z_i contains controls: 7 age bracket dummies, gender dummy, 4 race dummies

"Notice that because the randomization orthogonalized HSKFRAME with respect to Z, the exact covariate choice does not affect the results of the main coefficients of interest." p.70

Hainmueller and Hiscox: ordered probit results

TABLE 1. Individual Support for Highly Skilled and Low-skilled Immigration—Test of the Labor Market Competition Model

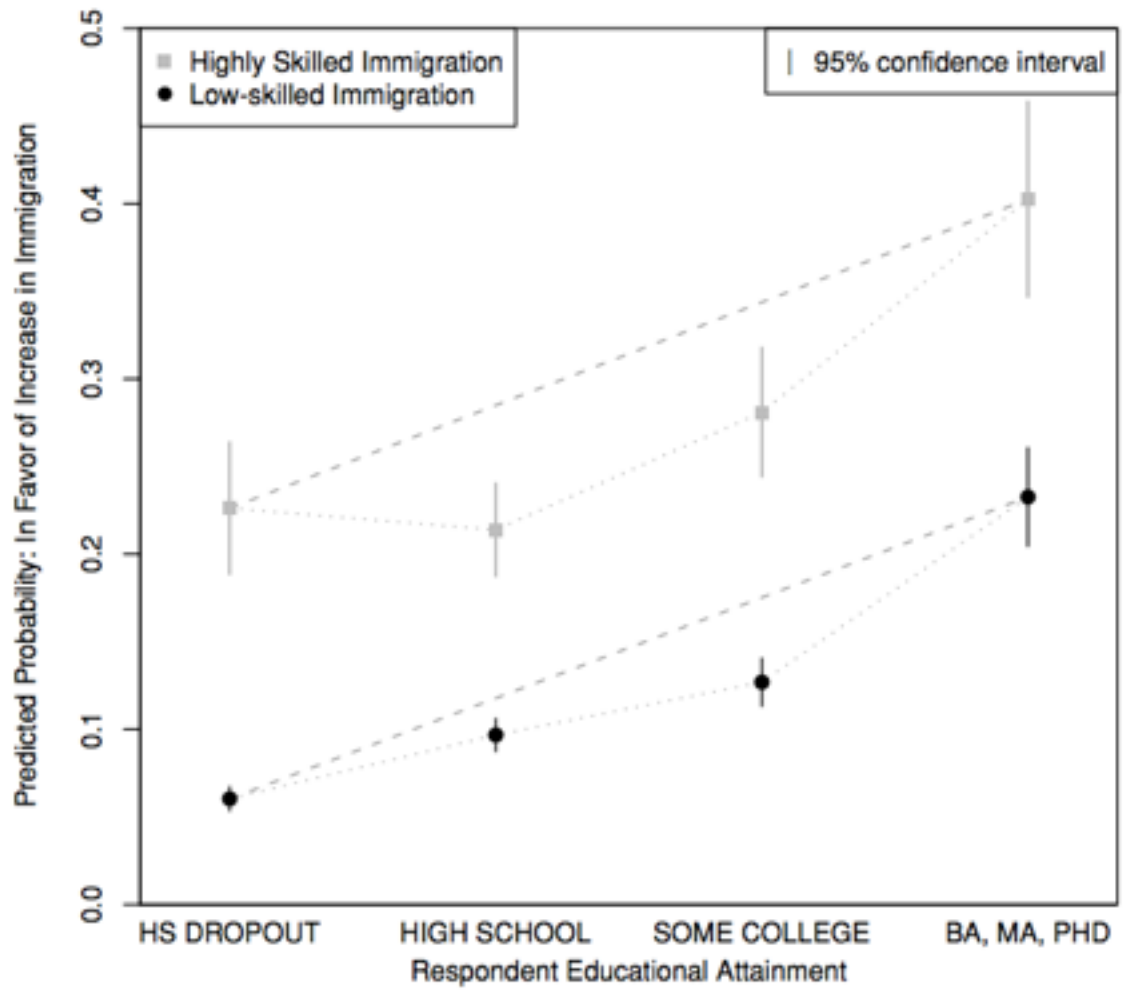
	In Favor of: High Skilled Immigration	In Favor of: Low-skilled Immigration	In Favor of: Immigration				
	(1)	(2)	(3)	(4)	(5)	(6) labor force in	(7) labor force out
Dependent Variable							
EDUCATION	0.21	0.27		0.27		0.33	0.19
	(0.05)	(0.05)		(0.05)		(0.06)	(0.07)
HSKFRAME			0.54	0.73	0.56	0.73	0.64
			(0.07)	(0.20)	(0.12)	(0.28)	(0.29)
HSKFRAME·EDUCATION				−0.07		−0.08	0.00
				(0.07)		(0.09)	(0.11)
HS DROPOUT					−0.41		
					(0.18)		
HSKFRAME·HS DROPOUT					0.24		
					(0.25)		
HIGH SCHOOL					−0.16		
					(0.12)		
HSKFRAME·HIGH SCHOOL					−0.05		
					(0.17)		
BA DEGREE					0.41		
					(0.12)		
HSKFRAME·BA DEGREE					−0.08		
					(0.16)		
(*N*)	798	791	1589	1589	1589	946	643
Covariates	x	x	x	x	x	x	x

Order Probit Coefficients shown with standard errors in parentheses. All models include a set of the covariates age, gender, and race (coefficients not shown here). The reference category for the set of education dummies is SOME COLLEGE (respondents with some college education).

Hainmueller and Hiscox: logit results

To give some sense of the substantive magnitudes involved, we simulate the predicted probability of supporting an increase in immigration (answers "somewhat agree" and "strongly agree" that the U.S. should allow more immigration) for the median respondent (a white woman aged 45) for all four skill levels and both immigration types based on the least restrictive model (model five in Table 1).

FIGURE 4. Support for Highly Skilled and Low-skilled Immigration by Respondents' Skill Level



Hainmueller and Hiscox: presentation

ACTIVITY!!!

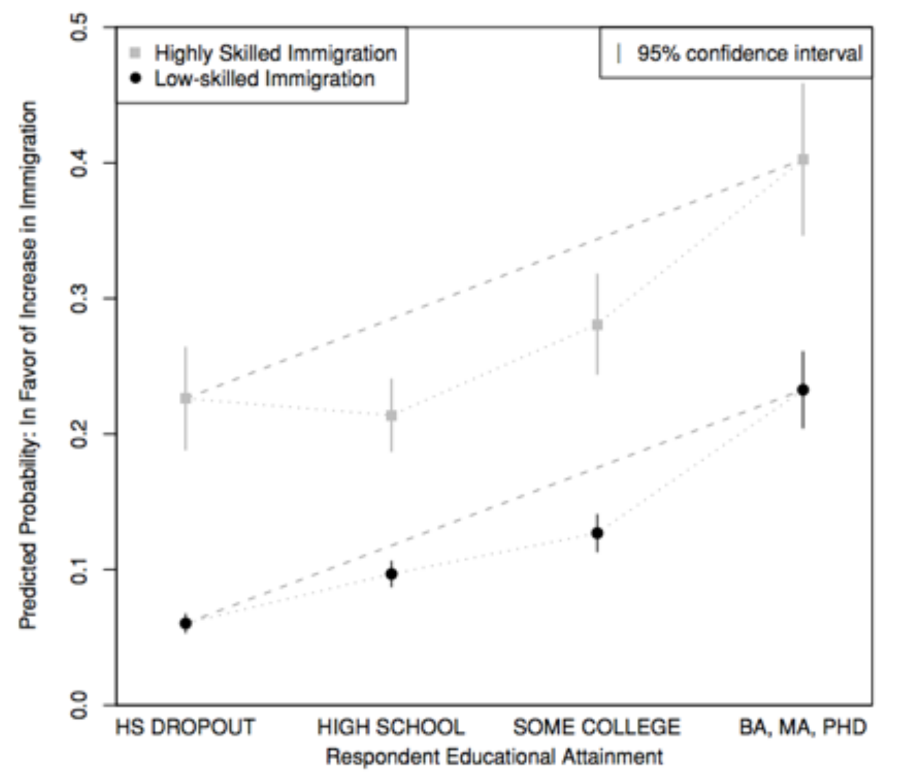
How could Hainmueller and Hiscox have graphically summarized the findings of their ordered probit regression (rather than switching to a binary outcome)?

TABLE 1. Individual SupLow-skilled Immigration—Test Labor Market Competition

			In Favor of: Immigration
	(3)	(4)	(5)
Dependent Variable			
EDUCATION		0.27 (0.05)	
HSKFRAME	0.54 (0.07)	0.73 (0.20)	0.56 (0.12)
HSKFRAME·EDUCATION		−0.07 (0.07)	
HS DROPOUT			−0.41 (0.18)
HSKFRAME·HS DROPOUT			0.24 (0.25)
HIGH SCHOOL			−0.16 (0.12)
HSKFRAME·HIGH SCHOOL			−0.05 (0.17)
BA DEGREE			0.41 (0.12)
HSKFRAME·BA DEGREE			−0.08 (0.16)
(*N*)	1589	1589	1589
Covariates	x	x	x

Order Probit Coefficients shown wAll models include a set of the covariate race (coefficients not shown here). ¹ education dummies is SOME COLLEGI some college education).

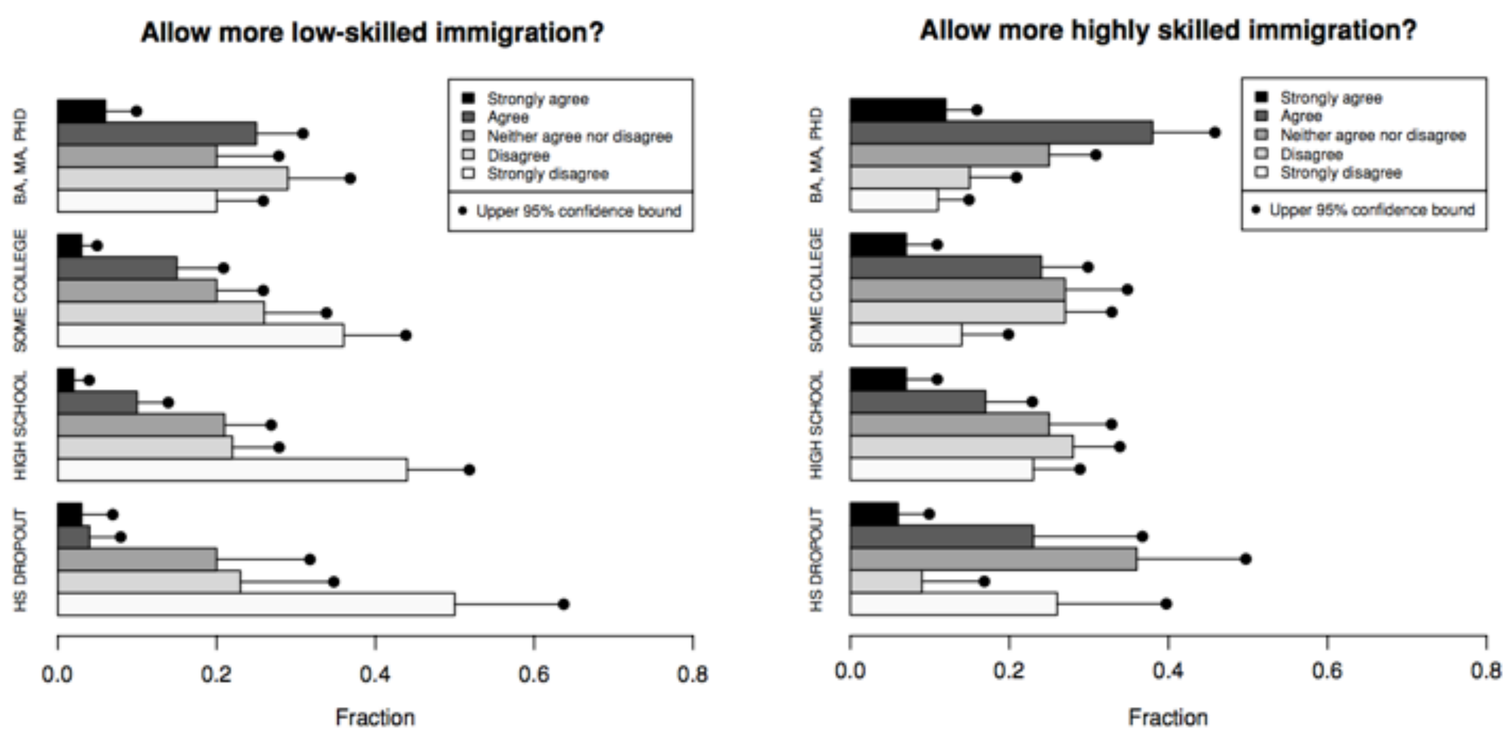
ıpport for Highly Skilled and Low-skilled Immigration by Respondents' Sl



Hainmueller and Hiscox: presentation SOLUTION?

One option: like Figure 3 but with predicted probabilities from the model.

FIGURE 3. Support for Highly Skilled and Low-skilled Immigration by Respondents' Skill Level



Hainmueller and Hiscox: presentation

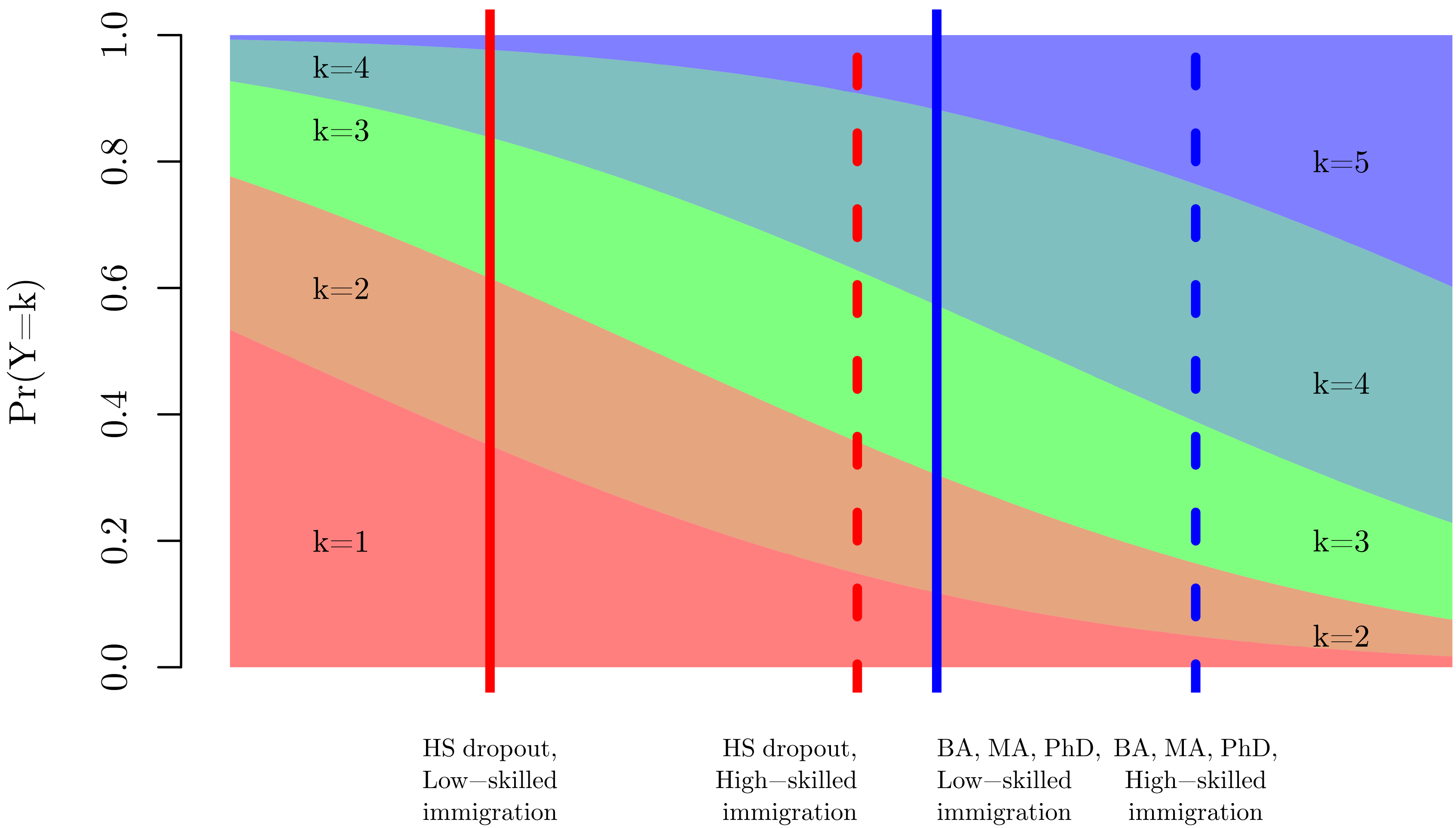
SOLUTION?

Another option: predicted probabilities at various values of Xβ, with some predicted values of Xβ shown

Hainmueller and Hiscox: presentation

SOLUTION?

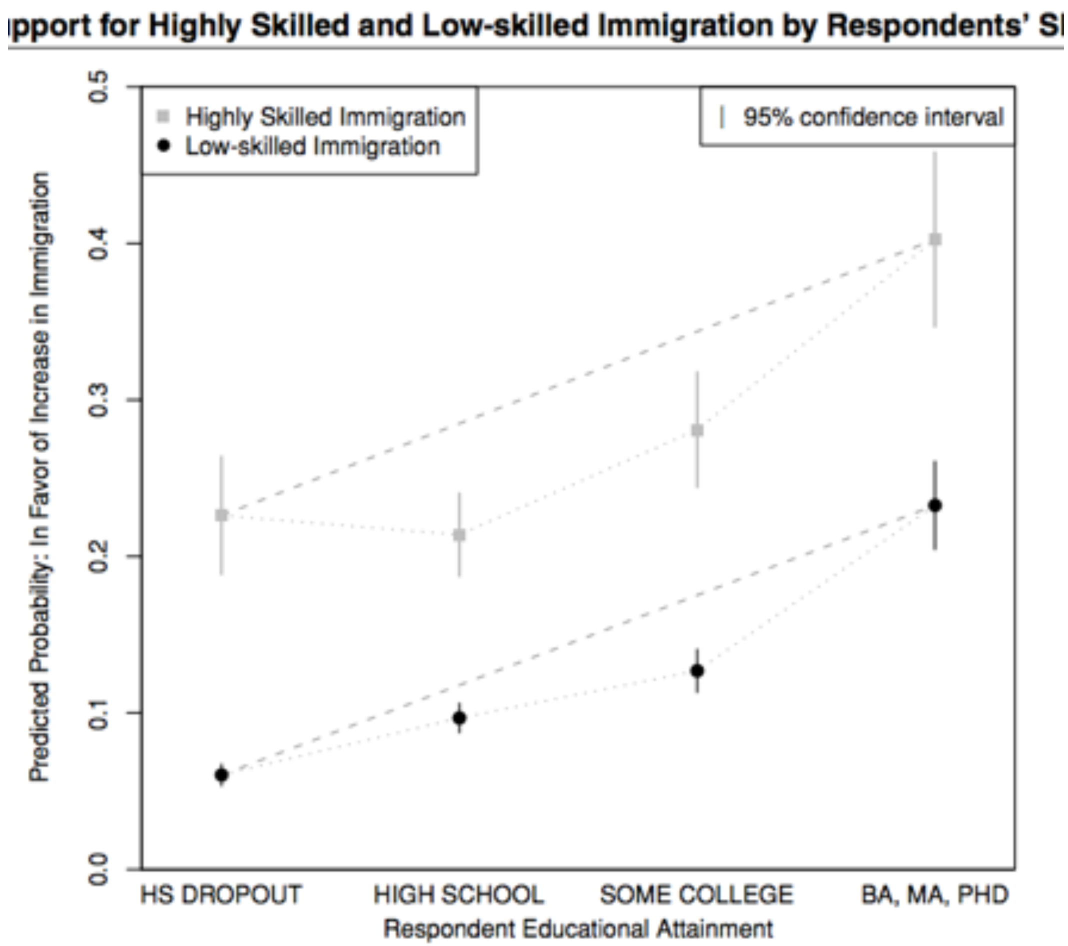
Another option: predicted probabilities at various values of Xβ, with some predicted values of Xβ shown



Why do we need logit?

ACTIVITY!!!

Consider H&H's logit analysis: support for more immigration **(binary)** as function of education, type of immigration.

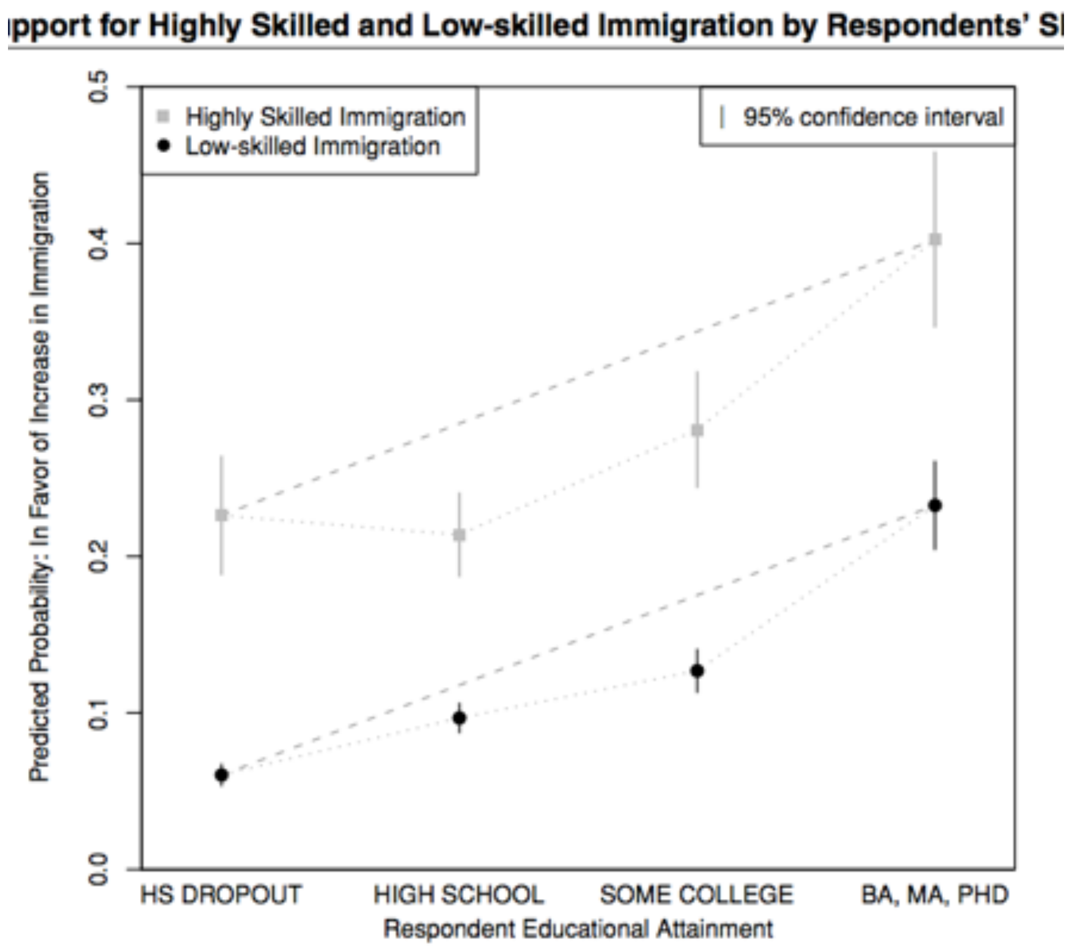


Why do we need logit?

ACTIVITY!!!

Consider H&H's logit analysis: support for more immigration **(binary)** as function of education, type of immigration.

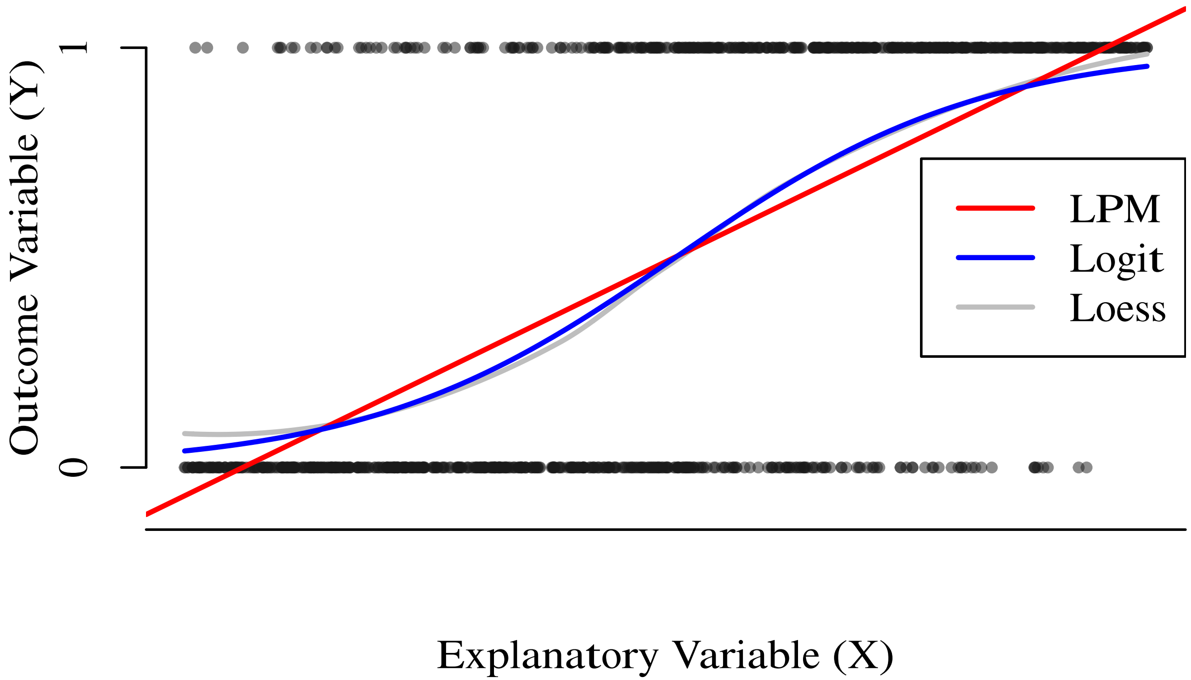
Why not estimate a **linear probability model** (LPM)?



$$\text{SUPPORT}_i = \alpha + \gamma \text{HSKFRAME}_i + \delta \text{HSKFRAME}_i \times \text{EDUCATION}_i + \theta \text{EDUCATION}_i + Z_i \psi$$

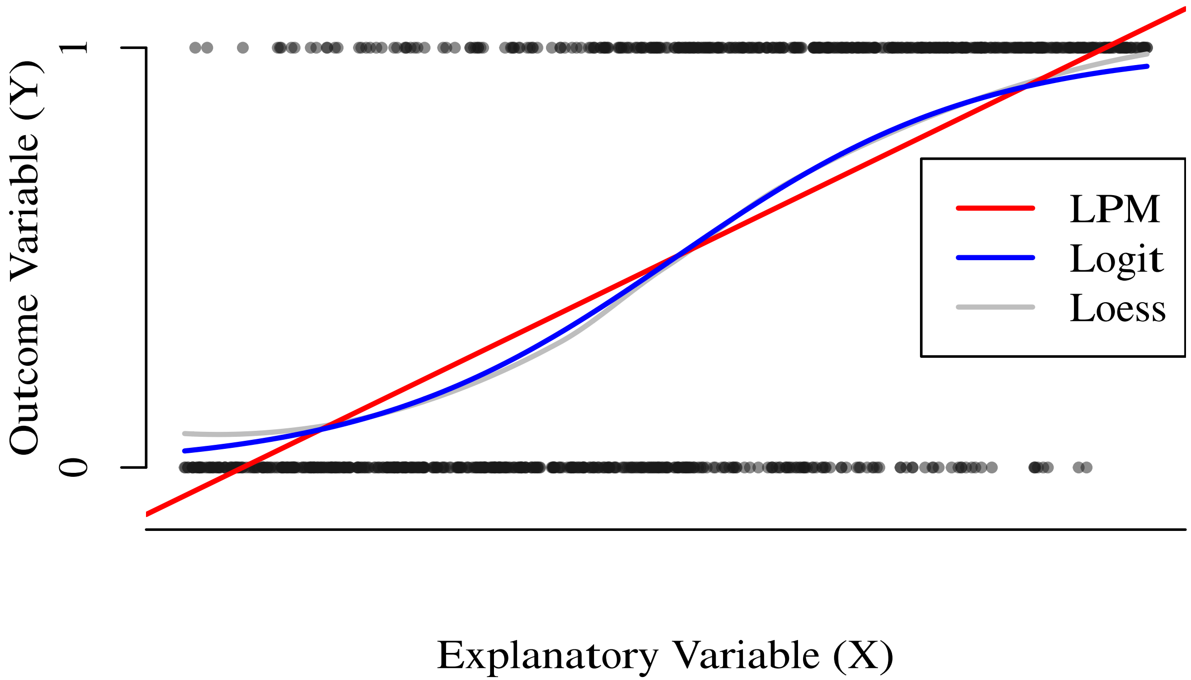
The usual case against the linear probability model (LPM)

SOLUTION?



The usual case against the linear probability model (LPM)

SOLUTION?



- *Predictions outside the range of dependent variable*
- *Heteroskedasticity (violates OLS assumption)*
- *Non-normal errors (violates OLS assumption)*
- *Unrealistic for probability to be linear in X*

The defense of the LPM: responses to critiques SOLUTION?

The defense of the LPM: responses to critiques SOLUTION?

- *Predictions outside the range of dependent variable*

The defense of the LPM: responses to critiques SOLUTION?

- *Predictions outside the range of dependent variable*
  - Is prediction (for outliers) the goal?

The defense of the LPM: responses to critiques SOLUTION?

- *Predictions outside the range of dependent variable*
  - Is prediction (for outliers) the goal?
- *Heteroskedasticity (violates OLS assumption)*

The defense of the LPM: responses to critiques SOLUTION?

- *Predictions outside the range of dependent variable*
  - Is prediction (for outliers) the goal?
- *Heteroskedasticity (violates OLS assumption)*
  - See Huber-White standard errors, other corrections for heteroskedasticity (`robust` option in Stata)

The defense of the LPM: responses to critiques SOLUTION?

- *Predictions outside the range of dependent variable*
  - Is prediction (for outliers) the goal?
- *Heteroskedasticity (violates OLS assumption)*
  - See Huber-White standard errors, other corrections for heteroskedasticity (`robust` option in Stata)
- *Non-normal errors (violates OLS assumption)*

The defense of the LPM: responses to critiques SOLUTION?

- *Predictions outside the range of dependent variable*
  - Is prediction (for outliers) the goal?
- *Heteroskedasticity (violates OLS assumption)*
  - See Huber-White standard errors, other corrections for heteroskedasticity (`robust` option in Stata)
- *Non-normal errors (violates OLS assumption)*
  - That assumption is necessary for inference (i.e. valid standard errors) in small samples, but not asymptotically (see MHE section 3.1), and not for approximating the CEF

The defense of the LPM: responses to critiques SOLUTION?

- *Predictions outside the range of dependent variable*
  - Is prediction (for outliers) the goal?
- *Heteroskedasticity (violates OLS assumption)*
  - See Huber-White standard errors, other corrections for heteroskedasticity (`robust` option in Stata)
- *Non-normal errors (violates OLS assumption)*
  - That assumption is necessary for inference (i.e. valid standard errors) in small samples, but not asymptotically (see MHE section 3.1), and not for approximating the CEF
- *Unrealistic for probability to be linear in X*

The defense of the LPM: responses to critiques SOLUTION?

- *Predictions outside the range of dependent variable*
  - Is prediction (for outliers) the goal?
- *Heteroskedasticity (violates OLS assumption)*
  - See Huber-White standard errors, other corrections for heteroskedasticity (`robust` option in Stata)
- *Non-normal errors (violates OLS assumption)*
  - That assumption is necessary for inference (i.e. valid standard errors) in small samples, but not asymptotically (see MHE section 3.1), and not for approximating the CEF
- *Unrealistic for probability to be linear in X*
  - Yes, especially when probabilities are near 1 or 0 **(ceiling and floor effects)**; but is probit the right form?

The defense of the LPM: continued

SOLUTION?

The defense of the LPM: continued

SOLUTION?

- **Advantage of LPM**: ease of interpretation

The defense of the LPM: continued

SOLUTION?

- **Advantage of LPM**: ease of interpretation
  - *Is that just because you don't understand log odds?*

The defense of the LPM: continued

SOLUTION?

- **Advantage of LPM**: ease of interpretation
  - *Is that just because you don't understand log odds?*
- **Disadvantage of logit/probit**:

The defense of the LPM: continued

SOLUTION?

- **Advantage of LPM**: ease of interpretation
  - *Is that just because you don't understand log odds?*
- **Disadvantage of logit/probit**:
  - Doesn't directly give the Average Treatment Effect

The defense of the LPM: continued

SOLUTION?

- **Advantage of LPM**: ease of interpretation
  - *Is that just because you don't understand log odds?*
- **Disadvantage of logit/probit**:
  - Doesn't directly give the Average Treatment Effect
  - Can convert logit/probit estimates to something equivalent, and in simulations that is the same as the LPM estimate

The defense of the LPM: continued

SOLUTION?

- **Advantage of LPM**: ease of interpretation
  - *Is that just because you don't understand log odds?*
- **Disadvantage of logit/probit**:
  - Doesn't directly give the Average Treatment Effect
  - Can convert logit/probit estimates to something equivalent, and in simulations that is the same as the LPM estimate
  - Other estimates are sensitive omitted variables — even *those uncorrelated with treatment* (Carina Mood, Eur. Soc. Rev. 2010)

The defense of the LPM: continued

SOLUTION?

- **Advantage of LPM**: ease of interpretation
  - *Is that just because you don't understand log odds?*
- **Disadvantage of logit/probit**:
  - Doesn't directly give the Average Treatment Effect
  - Can convert logit/probit estimates to something equivalent, and in simulations that is the same as the LPM estimate
  - Other estimates are sensitive omitted variables — even *those uncorrelated with treatment* (Carina Mood, Eur. Soc. Rev. 2010)
- When interest is in coefficient on binary variable (e.g. treatment),

The defense of the LPM: continued

SOLUTION?

- **Advantage of LPM**: ease of interpretation
  - *Is that just because you don't understand log odds?*
- **Disadvantage of logit/probit**:
  - Doesn't directly give the Average Treatment Effect
  - Can convert logit/probit estimates to something equivalent, and in simulations that is the same as the LPM estimate
  - Other estimates are sensitive omitted variables — even *those uncorrelated with treatment* (Carina Mood, Eur. Soc. Rev. 2010)
- When interest is in coefficient on binary variable (e.g. treatment),
  - CEF is linear with respect to variable of interest

The defense of the LPM: continued

SOLUTION?

- **Advantage of LPM**: ease of interpretation
  - *Is that just because you don't understand log odds?*
- **Disadvantage of logit/probit**:
  - Doesn't directly give the Average Treatment Effect
  - Can convert logit/probit estimates to something equivalent, and in simulations that is the same as the LPM estimate
  - Other estimates are sensitive omitted variables — even *those uncorrelated with treatment* (Carina Mood, Eur. Soc. Rev. 2010)
- When interest is in coefficient on binary variable (e.g. treatment),
  - CEF is linear with respect to variable of interest
  - Logit vs LPM matters only if particular kind of covariate imbalance

The defense of the LPM: continued

SOLUTION?

Gailmard pp 171-2

"If the CEF is linear, as it is for a saturated model, [OLS] gives the CEF.… If the CEF is non-linear, [OLS] approximates the CEF. Usually it does it pretty well. Obviously, the LPM won't give the true marginal effects from the right nonlinear model. But then, the same is true for the 'wrong' nonlinear model! The fact that we have a probit, a logit, and the LPM [shows] that we don't know what the 'right' model is. Hence, there is a lot to be said for sticking to a linear regression function as compared to a fairly arbitrary choice of a non-linear one! Nonlinearity per se is a red herring."

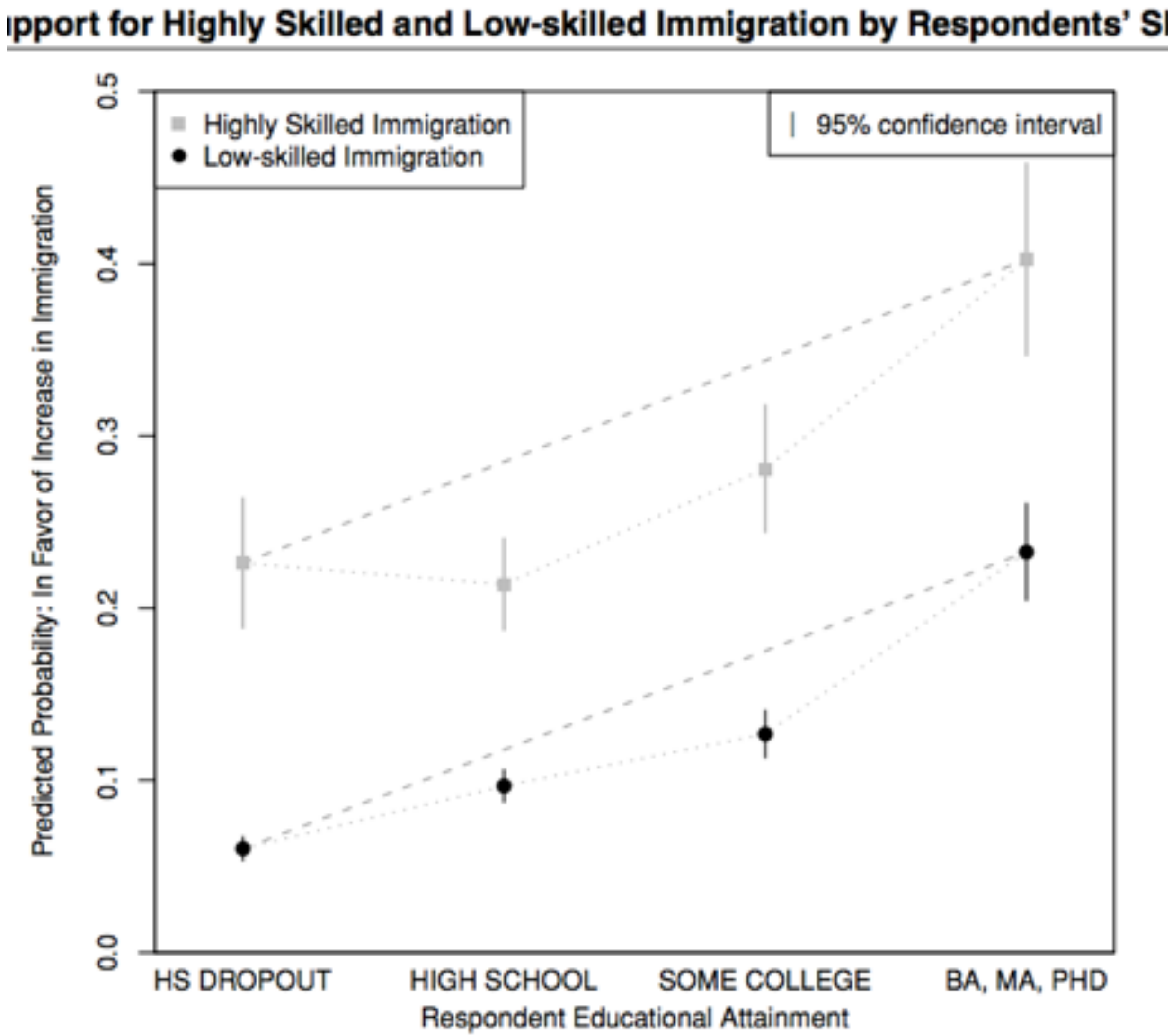
Steve Pischke

from MHE blog http://www.mostlyharmlesseconometrics.com/2012/07/probit-better-than-lpm/

The defense of the LPM: continued

SOLUTION?

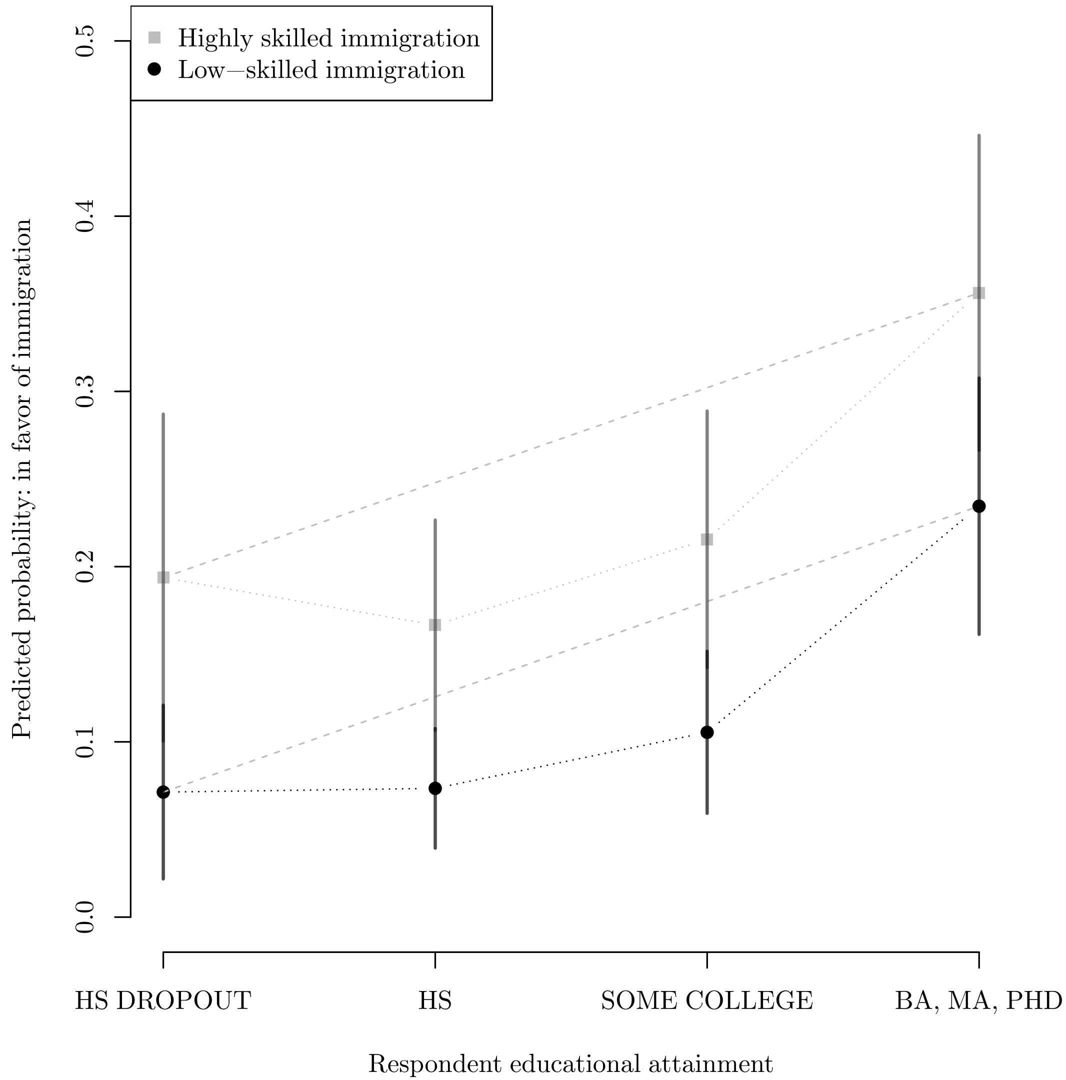
Original Figure 4
(based on logit)



The defense of the LPM: continued

SOLUTION?

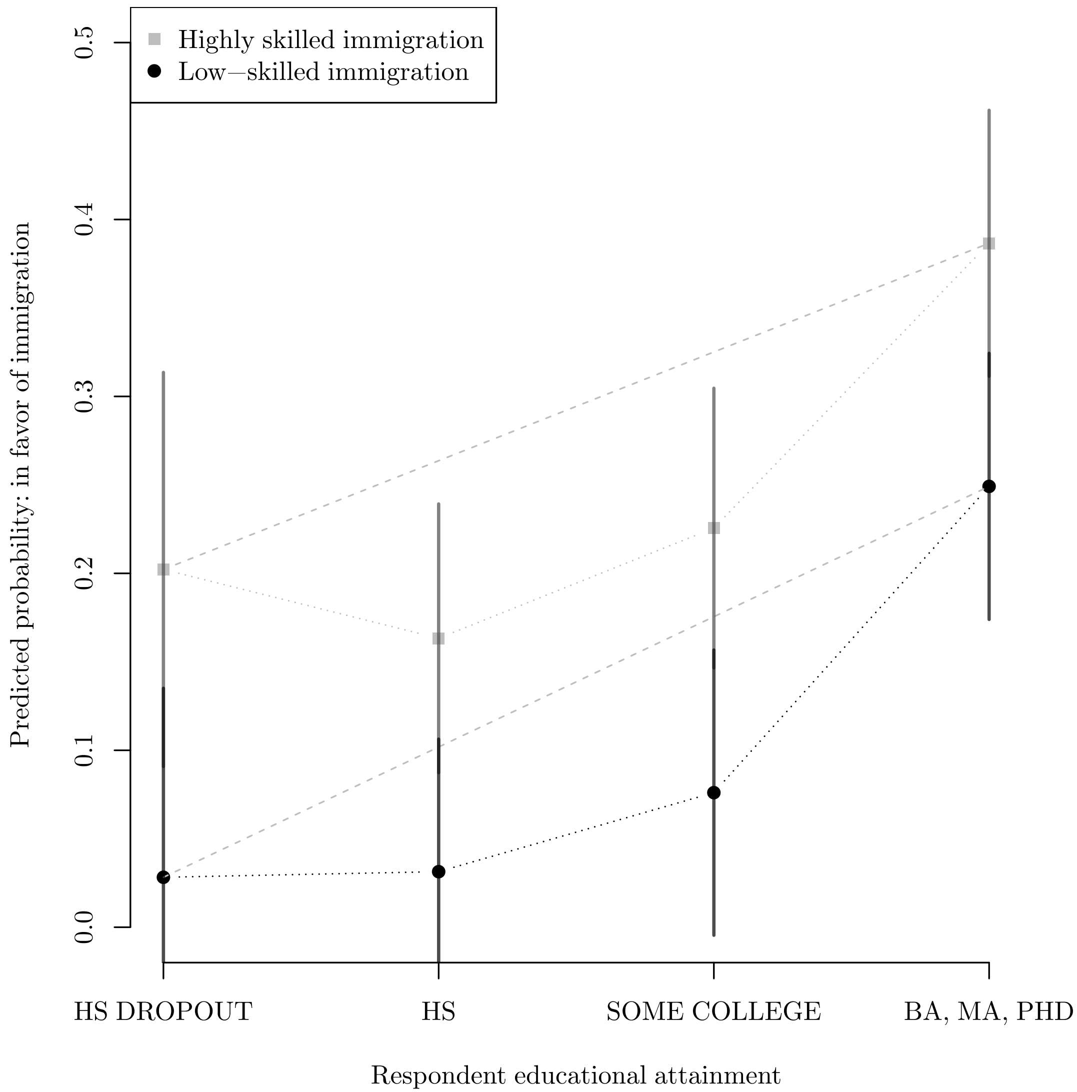
My Figure 4 (based on logit)



The defense of the LPM: continued

SOLUTION?

My Figure 4 (based on LPM)



Why do we need ordered probit?

ACTIVITY!!!!

Consider H&H's ordered probit analysis: support for more immigration **(five categories)** as function of education, type of immigration.

Why do we need ordered probit?

ACTIVITY!!!

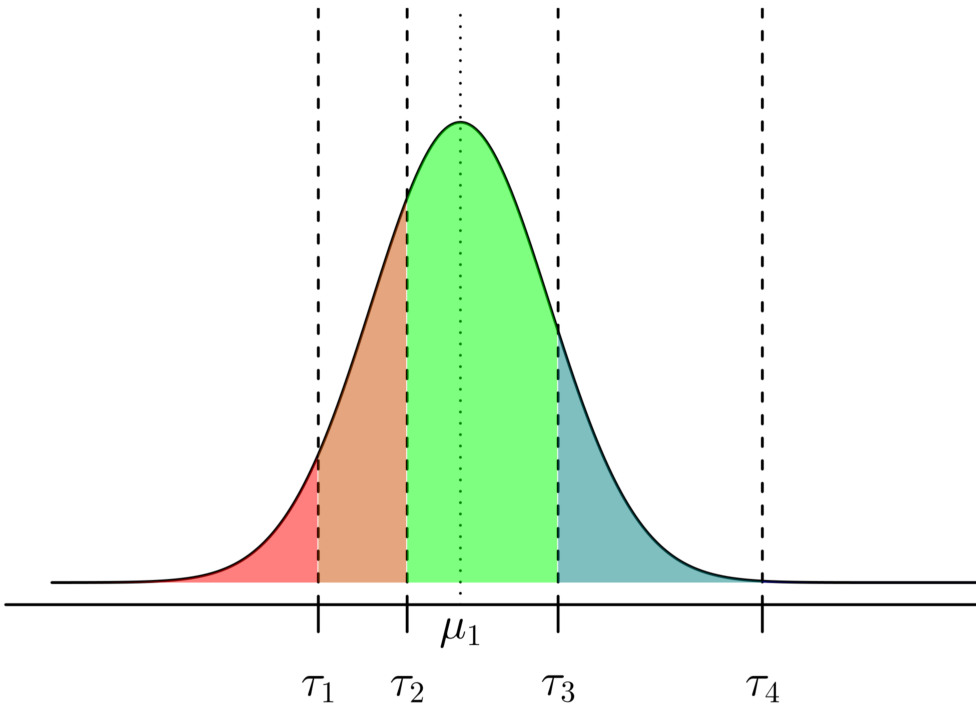
Consider H&H's ordered probit analysis: support for more immigration **(five categories)** as function of education, type of immigration.

Why not just estimate a **linear regression** where the DV is 1-5 scores?

Why do we need ordered probit?

ACTIVITY!!!

Consider H&H's ordered probit analysis: support for more immigration **(five categories)** as function of education, type of immigration.



Why not just estimate a **linear regression** where the DV is 1-5 scores?

Why do we need ordered probit?

SOLUTION?

- Some reviewers (still) ask for it
- Could produce predicted probabilities separately for each category
- Ceiling and floor effects: if nonlinearity is a problem in LPM, it could be here too
- More generally: outcome scores may not be linear in covariates

Introduction to measurement/scaling models

	Bill 1	Bill 2	Bill 3	…
Legislator 1	Y	Y		…
Legislator 2	Y	N	N	…
Legislator 3		N	N	…
Legislator 4	Y	Y	Y	…
…	…	…	…	…

Introduction to measurement/scaling models

Suppose we had voting data like that. What could you do with it?

	Bill 1	Bill 2	Bill 3	…
Legislator 1	Y	Y		…
Legislator 2	Y	N	N	…
Legislator 3		N	N	…
Legislator 4	Y	Y	Y	…
…	…	…	…	…

	Word 1	Word 2	Word 3	…
Article 1	0	14	2	…
Article 2	1	8	0	…
Article 3	0	7	1	…
Article 4	2	3	0	…
…	…	…	…	…

Or text data like that. What could you do with it?

	Word 1	Word 2	Word 3	…
Article 1	0	14	2	…
Article 2	1	8	0	…
Article 3	0	7	1	…
Article 4	2	3	0	…
…	…	…	…	…

	Candidate1	Candidate2	Candidate3	…
Interest group 1	0	$5,000	0	…
Interest group 2	$1,000	$1,000	0	…
Interest group 3	0	0	$10,000	…
Interest group 4	$500	0	0	…
…	…	…	…	…

Or political contribution data like that. What could you do with it?

	Candidate1	Candidate2	Candidate3	…
Interest group 1	0	$5,000	0	…
Interest group 2	$1,000	$1,000	0	…
Interest group 3	0	0	$10,000	…
Interest group 4	$500	0	0	…
…	…	…	…	…

Common structure

Common structure

	these	that	those
Article 1	0	14	2
Article 2	1	8	0
Article 3	0	7	1
Article 4	2	3	0

Data is **grouped**:

- many legislators, many bills
- many speakers, many words.

Though it probably didn't come in that format originally!

Common structure

Data is **grouped**:

- many legislators, many bills
- many speakers, many words.

Though it probably didn't come in that format originally!

	these	that	those
Article 1	0	14	2
Article 2	1	8	0
Article 3	0	7	1
Article 4	2	3	0

Article 1: That that "that" that. That that those; that that that that. That that that that those.

Article 2: That that/these that that! That that that.

Article 3: That that those — that that that that that.

Article 4: These that! These that that.

Common structure

Data is **grouped**:

- many legislators, many bills
- many speakers, many words.

Though it probably didn't come in that format originally!

Article 1: That that "that" that. That that those; that that that that. That that that that those.

Article 2: That that/these that that! That that that.

Article 3: That that those — that that that that that.

Article 4: These that! These that that.

	these	that	those
Article 1	0	14	2
Article 2	1	8	0
Article 3	0	7	1
Article 4	2	3	0

Article	Word	Count
1	these	0
2	these	1
3	these	0
4	these	2
1	that	14
2	that	8
3	that	7
4	that	3
1	those	2
2	those	0
3	those	1
4	those	0

A simpler example

ACTIVITY!!!

	Vote on Bill 2	x (Ideology score)
Legislator 1	1	34
Legislator 2	0	67
Legislator 3	0	49
Legislator 4	1	12
...	...	...

A simpler example

ACTIVITY!!!

Suppose we had voting data on just one bill, and maybe a covariate.

	Vote on Bill 2	X (Ideology score)
Legislator 1	1	34
Legislator 2	0	67
Legislator 3	0	49
Legislator 4	1	12
...	...	...

A simpler example

ACTIVITY!!!

Suppose we had voting data on just one bill, and maybe a covariate.

	Vote on Bill 2	x (Ideology score)
Legislator 1	1	34
Legislator 2	0	67
Legislator 3	0	49
Legislator 4	1	12
…	…	…

Q: How could you relate x to vote in a simple way via LPM, probit, or logit?

What would this tell you?

A simpler example

ACTIVITY!!!!

Suppose we had voting data on just one bill, and maybe a covariate.

	Vote on Bill 2	x (Ideology score)
Legislator 1	1	34
Legislator 2	0	67
Legislator 3	0	49
Legislator 4	1	12
…	…	…

Q: How could you relate x to vote in a simple way via LPM, probit, or logit?

What would this tell you?

A: Regress vote on x.

- LPM: $\alpha + \beta x_i$ is the predicted probability conditional on x_i
- Probit: $\Phi(\alpha + \beta x_i)$ (Normal CDF) is the predicted probability conditional on x_i
- Logit: $\alpha + \beta x_i$ is the log odds conditional on x_i

A slightly less simple example

ACTIVITY!!!

	Bill	Vote	X (Ideology score)
Legislator 1	2	1	34
Legislator 2	2	0	67
Legislator 3	2	0	49
Legislator 4	2	1	12
…	…	…	…
Legislator 1	1	0	34
Legislator 2	1	1	67
Legislator 3	1	0	49
Legislator 4	1	0	12
…	…	…	…

A slightly less simple example

ACTIVITY!!!

How could we extend this to more than one bill?

	Bill	Vote	X (Ideology score)
Legislator 1	2	1	34
Legislator 2	2	0	67
Legislator 3	2	0	49
Legislator 4	2	1	12
...	...	...	...
Legislator 1	1	0	34
Legislator 2	1	1	67
Legislator 3	1	0	49
Legislator 4	1	0	12
...	...	...	...

A slightly less simple example

ACTIVITY!!!

How could we extend this to more than one bill?

	Bill	Vote	x (Ideology score)
Legislator 1	2	1	34
Legislator 2	2	0	67
Legislator 3	2	0	49
Legislator 4	2	1	12
…	…	…	…
Legislator 1	1	0	34
Legislator 2	1	1	67
Legislator 3	1	0	49
Legislator 4	1	0	12
…	…	…	…

Regress vote on

- x (ideology score)
- a dummy (indicator variable) for each bill, and
- the interactions between x and the bill dummies.

Result is a intercept α_j and slope β_j for each bill.

- LPM: $\alpha_j + \beta_j x_i$ is the predicted probability legislator i would vote for bill j
- Probit and Logit: same pattern as previous (simple) example

What does β_j tell you?

Doing the seemingly impossible

ACTIVITY!!!

	Bill	Vote	x (Ideology score)
Legislator 1	2	1	?
Legislator 2	2	0	?
Legislator 3	2	0	?
Legislator 4	2	1	?
…	…	…	…
Legislator 1	1	1	?
Legislator 2	1	1	?
Legislator 3	1		?
Legislator 4	1	1	?
…	…	…	…

Doing the seemingly impossible

ACTIVITY!!!

Now suppose the ideology score was missing. What now?

	Bill	Vote	x (Ideology score)
Legislator 1	2	1	?
Legislator 2	2	0	?
Legislator 3	2	0	?
Legislator 4	2	1	?
…	…	…	…
Legislator 1	1	1	?
Legislator 2	1	1	?
Legislator 3	1		?
Legislator 4	1	1	?
…	…	…	…

Doing the seemingly impossible

ACTIVITY!!!

Now suppose the ideology score was missing. What now?

	Bill	Vote	x (Ideology score)
Legislator 1	2	1	?
Legislator 2	2	0	?
Legislator 3	2	0	?
Legislator 4	2	1	?
…	…	…	…
Legislator 1	1	1	?
Legislator 2	1	1	?
Legislator 3	1		?
Legislator 4	1	1	?
…	…	…	…

Statistical model is the same as if x was observed, but x becomes an additional parameter to estimate.

This works because the same legislator votes on many bills; x is estimated based on recurring patterns of voting behavior.

The (generative) statistical model: same as probit

The (generative) statistical model: same as probit

Let $y_{ij} \in \{0, 1\}$ indicate i's vote on bill j.
Let y^*_{ij} indicate i's (unobserved) propensity to vote for bill j: i votes for j if $y^*_{ij} > 0$.

The (generative) statistical model: same as probit

Let $y_{ij} \in \{0, 1\}$ indicate i's vote on bill j.
Let y^*_{ij} indicate i's (unobserved) propensity to vote for bill j: i votes for j if $y^*_{ij} > 0$.

Model for y^*_{ij}:

$$y^*_{ij} = \alpha_j + \beta_j x_i + \epsilon_{ij}$$

The (generative) statistical model: same as probit

Let $y_{ij} \in \{0, 1\}$ indicate i's vote on bill j.
Let y^*_{ij} indicate i's (unobserved) propensity to vote for bill j: i votes for j if $y^*_{ij} > 0$.

Model for y^*_{ij}:

$$y^*_{ij} = \alpha_j + \beta_j x_i + \epsilon_{ij}$$

where

- x_i is i's ideology
- β_j relates ideology to voting behavior: are voters with higher x_i more or less likely to vote for the bill?
- α_j indicates the general attractiveness of voting for the bill, conditional on ideology
- $\epsilon_{ij} \sim \mathrm{N}(0, 1)$: other factors affect voting

The (generative) statistical model: same as probit

Let $y_{ij} \in \{0, 1\}$ indicate i's vote on bill j.
Let y^*_{ij} indicate i's (unobserved) propensity to vote for bill j: i votes for j if $y^*_{ij} > 0$.

Model for y^*_{ij}:

$$y^*_{ij} = \alpha_j + \beta_j x_i + \epsilon_{ij}$$

where

- x_i is i's ideology
- β_j relates ideology to voting behavior: are voters with higher x_i more or less likely to vote for the bill?
- α_j indicates the general attractiveness of voting for the bill, conditional on ideology
- $\epsilon_{ij} \sim \mathrm{N}(0, 1)$: other factors affect voting

Then

$$\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$$

where $\Phi(\cdot)$ is the standard normal CDF.

Estimation (1)

Estimation (1)

$$\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$$

where $\Phi(\cdot)$ is the standard normal CDF.

Estimation (1)

$$\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$$

where $\Phi(\cdot)$ is the standard normal CDF.

Recall that x_i is unobserved: so **this is a probit regression with no covariates**. Seems impossible!

Estimation (1)

$$\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$$

where $\Phi(\cdot)$ is the standard normal CDF.

Recall that x_i is unobserved: so **this is a probit regression with no covariates**. Seems impossible!

But imagine making guesses for α_j, β_j, and x_i. Because each x_i appears many times in the likelihood (i.e. the same legislator votes on many bills), some guesses would be better than others (i.e. would yield a higher value for the likelihood). Maximize the likelihood!

Estimation (1)

$$\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$$

where $\Phi(\cdot)$ is the standard normal CDF.

Recall that x_i is unobserved: so **this is a probit regression with no covariates**. Seems impossible!

But imagine making guesses for α_j, β_j, and x_i. Because each x_i appears many times in the likelihood (i.e. the same legislator votes on many bills), some guesses would be better than others (i.e. would yield a higher value for the likelihood). Maximize the likelihood!

Q: How many parameters are you estimating, given n legislators and k bills?

Estimation (1)

$$\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$$

where $\Phi(\cdot)$ is the standard normal CDF.

Recall that x_i is unobserved: so **this is a probit regression with no covariates**. Seems impossible!

But imagine making guesses for α_j, β_j, and x_i. Because each x_i appears many times in the likelihood (i.e. the same legislator votes on many bills), some guesses would be better than others (i.e. would yield a higher value for the likelihood). Maximize the likelihood!

Q: How many parameters are you estimating, given n legislators and k bills?

A: Each of k bills has an α_j and a β_j; each of n legislators has an x_i → 2k + n.

Estimation (1)

$$\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$$

where $\Phi(\cdot)$ is the standard normal CDF.

Recall that x_i is unobserved: so **this is a probit regression with no covariates**. Seems impossible!

But imagine making guesses for α_j, β_j, and x_i. Because each x_i appears many times in the likelihood (i.e. the same legislator votes on many bills), some guesses would be better than others (i.e. would yield a higher value for the likelihood). Maximize the likelihood!

Q: How many parameters are you estimating, given n legislators and k bills?

A: Each of k bills has an α_j and a β_j; each of n legislators has an x_i → 2k + n.

Q: How many data points do you have, given n legislators and k bills?

Estimation (1)

$$\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$$

where $\Phi(\cdot)$ is the standard normal CDF.

Recall that x_i is unobserved: so **this is a probit regression with no covariates**. Seems impossible!

But imagine making guesses for α_j, β_j, and x_i. Because each x_i appears many times in the likelihood (i.e. the same legislator votes on many bills), some guesses would be better than others (i.e. would yield a higher value for the likelihood). Maximize the likelihood!

Q: How many parameters are you estimating, given n legislators and k bills?

A: Each of k bills has an α_j and a β_j; each of n legislators has an x_i → 2k + n.

Q: How many data points do you have, given n legislators and k bills?

A: n × k

Estimation (2)

Estimation (2)

As with ordinal probit, "identification" is an issue: different combinations of parameters would yield exactly the same likelihood.

Recall likelihood is based on: $\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$

Estimation (2)

As with ordinal probit, "identification" is an issue: different combinations of parameters would yield exactly the same likelihood.

Recall likelihood is based on: $\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$

See any issues?

Estimation (2)

As with ordinal probit, "identification" is an issue: different combinations of parameters would yield exactly the same likelihood.

Recall likelihood is based on: $\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$

See any issues?

- if you double all the x_i values and halve all the β_j values you get the same likelihood.

Estimation (2)

As with ordinal probit, "identification" is an issue: different combinations of parameters would yield exactly the same likelihood.

Recall likelihood is based on: $\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$

See any issues?

- if you double all the x_i values and halve all the β_j values you get the same likelihood.
- if you multiply all the x_i values and all the β_j values by -1 you get the same likelihood.

Estimation (2)

As with ordinal probit, "identification" is an issue: different combinations of parameters would yield exactly the same likelihood.

Recall likelihood is based on: $\Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_j x_i)$

See any issues?

- if you double all the x_i values and halve all the β_j values you get the same likelihood.
- if you multiply all the x_i values and all the β_j values by -1 you get the same likelihood.

Solution: various constraints (e.g. "Corbyn's x_i must be negative, and the standard deviation of the x_i values must be 1.")

Estimation (3)

Estimation (3)

If you want to investigate models like this:

Estimation (3)

If you want to investigate models like this:

- Stata: irt (item response theory)

Estimation (3)

If you want to investigate models like this:

- Stata: irt (item response theory)
- R: wnominate() in wnominate package; ideal() in pscl package (Bayesian); see also fastideal (Imai et al)

Estimation (3)

If you want to investigate models like this:

- Stata: irt (item response theory)
- R: wnominate() in wnominate package; ideal() in pscl package (Bayesian); see also fastideal (Imai et al)

CRAN Task View: Psychometric Models and Methods

Maintainer: Patrick Mair
Contact: mair at fas.harvard.edu
Version: 2016-01-31

Psychometrics is concerned with theory and techniques of psychological measurement. Psychometricians have also worked collaboratively with those in the field of statistics and quantitative methods to develop improved ways to organize, analyze, and scale corresponding data. Since much functionality is already contained in base R and there is considerable overlap between tools for psychometry and tools described in other views, particularly in SocialSciences, we only give a brief overview of packages that are closely related to psychometric methodology.

Please let me know if I have omitted something of importance, or if a new package or function should be mentioned here.

Item Response Theory (IRT):

- The eRm package fits extended Rasch models, i.e. the ordinary Rasch model for dichotomous data (RM), the linear logistic test model (LLTM), the rating scale model (RSM) and its linear extension (LRSM), the partial credit model (PCM) and its linear extension (LPCM) using conditional ML estimation. Missing values are allowed.
- The package ltm also fits the simple RM. Additionally, functions for estimating Birnbaum's 2- and 3-parameter models based on a marginal ML approach are implemented as well as the graded response model for polytomous data, and the linear multidimensional logistic model.
- TAM fits unidimensional and multidimensional item response models and also includes multifaceted models, latent regression models and options for drawing plausible values.
- The mirt allows for the analysis of dichotomous and polytomous response data using unidimensional and multidimensional latent trait models under the IRT paradigm. Exploratory and confirmatory models can be estimated with quadrature (EM) or stochastic (MHRM) methods. Confirmatory bi-factor and two-tier analyses are available for modeling item testlets. Multiple group analysis and mixed effects designs also are available for detecting differential item functioning and modelling item and person covariates.
- IRTShiny provides an interactive shiny application for IRT analysis.
- The mcIRT package provides functions to estimate the Nominal Response Model and the Nested Logit Model. Both are models to examine multiple-choice items and other polytomous response formats. Some additional uni- and multidimensional item response models (especially for locally dependent item responses) and some exploratory methods (DETECT, LSDM, model-based reliability) are included in sirt.
- The pcIRT estimates the multidimensional polytomous Rasch model and the Mueller's continuous rating scale model.
- Thurstonian IRT models can be fitted with the kcirt package.
- MultiLCIRT estimates IRT models under (1) multidimensionality assumption, (2) discreteness of latent traits, (3) binary and ordinal polytomous items.
- Conditional maximum likelihood estimation via the EM algorithm and information-criterion-based model selection in

Use of scaling models beyond legislative voting

Use of scaling models beyond legislative voting

- Measuring student ability and question difficulty in educational testing (origin of item response theory)

Use of scaling models beyond legislative voting

- Measuring student ability and question difficulty in educational testing (origin of item response theory)
- Measuring ideology of contributors and ideological appeal of candidates using campaign contribution data (Bonica)

Use of scaling models beyond legislative voting

- Measuring student ability and question difficulty in educational testing (origin of item response theory)
- Measuring ideology of contributors and ideological appeal of candidates using campaign contribution data (Bonica)
- Measuring ideology of parties and ideological use of words using text of party manifestos (Slapin & Proksch, wordfish)

Use of scaling models beyond legislative voting

- Measuring student ability and question difficulty in educational testing (origin of item response theory)
- Measuring ideology of contributors and ideological appeal of candidates using campaign contribution data (Bonica)
- Measuring ideology of parties and ideological use of words using text of party manifestos (Slapin & Proksch, wordfish)
- Measuring ideology of groups of citizens (e.g. French women) using responses to survey questions (Caughey & Warshaw, group IRT)

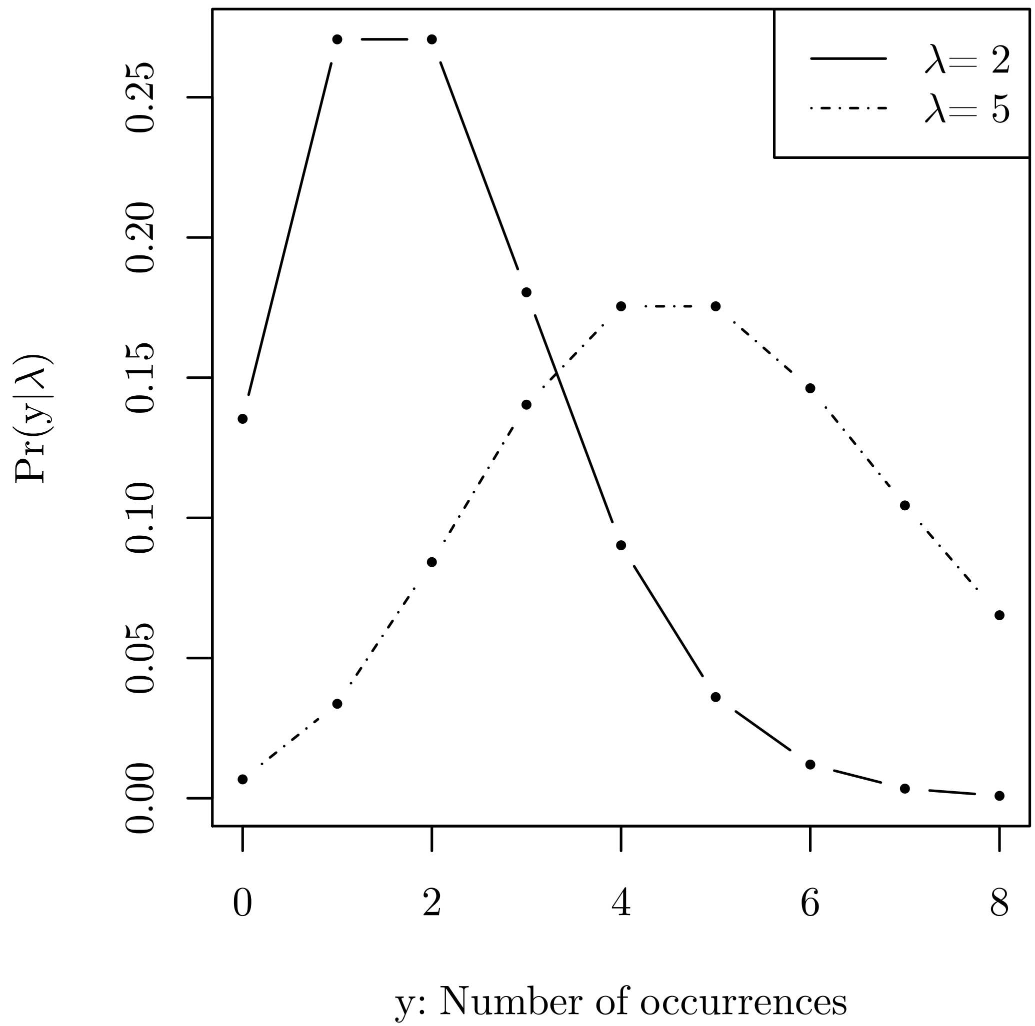
Use of scaling models beyond legislative voting

- Measuring student ability and question difficulty in educational testing (origin of item response theory)
- Measuring ideology of contributors and ideological appeal of candidates using campaign contribution data (Bonica)
- Measuring ideology of parties and ideological use of words using text of party manifestos (Slapin & Proksch, wordfish)
- Measuring ideology of groups of citizens (e.g. French women) using responses to survey questions (Caughey & Warshaw, group IRT)
- Measuring judges' ideology and how it changes over time (Martin & Quinn)

Scaling text: wordfish

Scaling text: wordfish

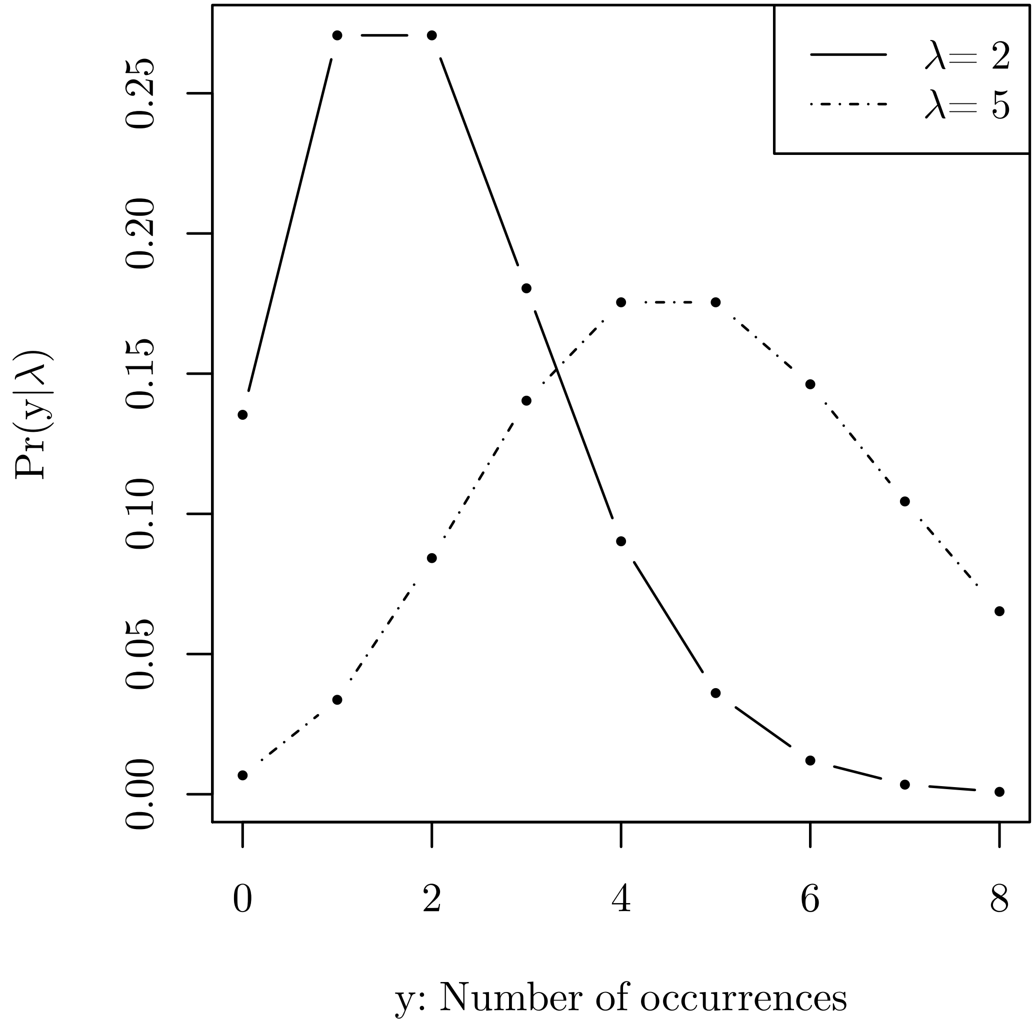
Recall Poisson distribution for counts.



Scaling text: wordfish

Recall Poisson distribution for counts.

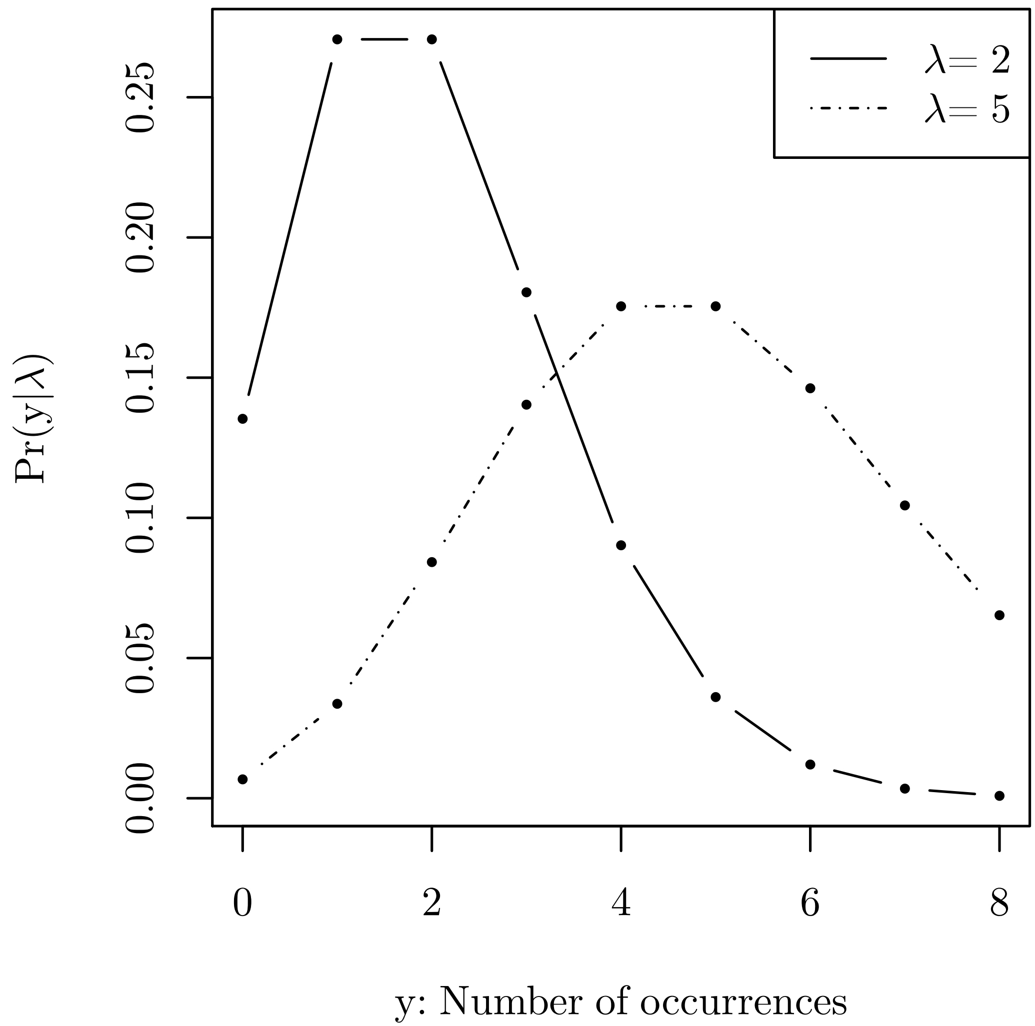
Think about rates at which parties use different words in party manifestos.



Scaling text: wordfish

Recall Poisson distribution for counts.

Think about rates at which parties use different words in party manifestos.



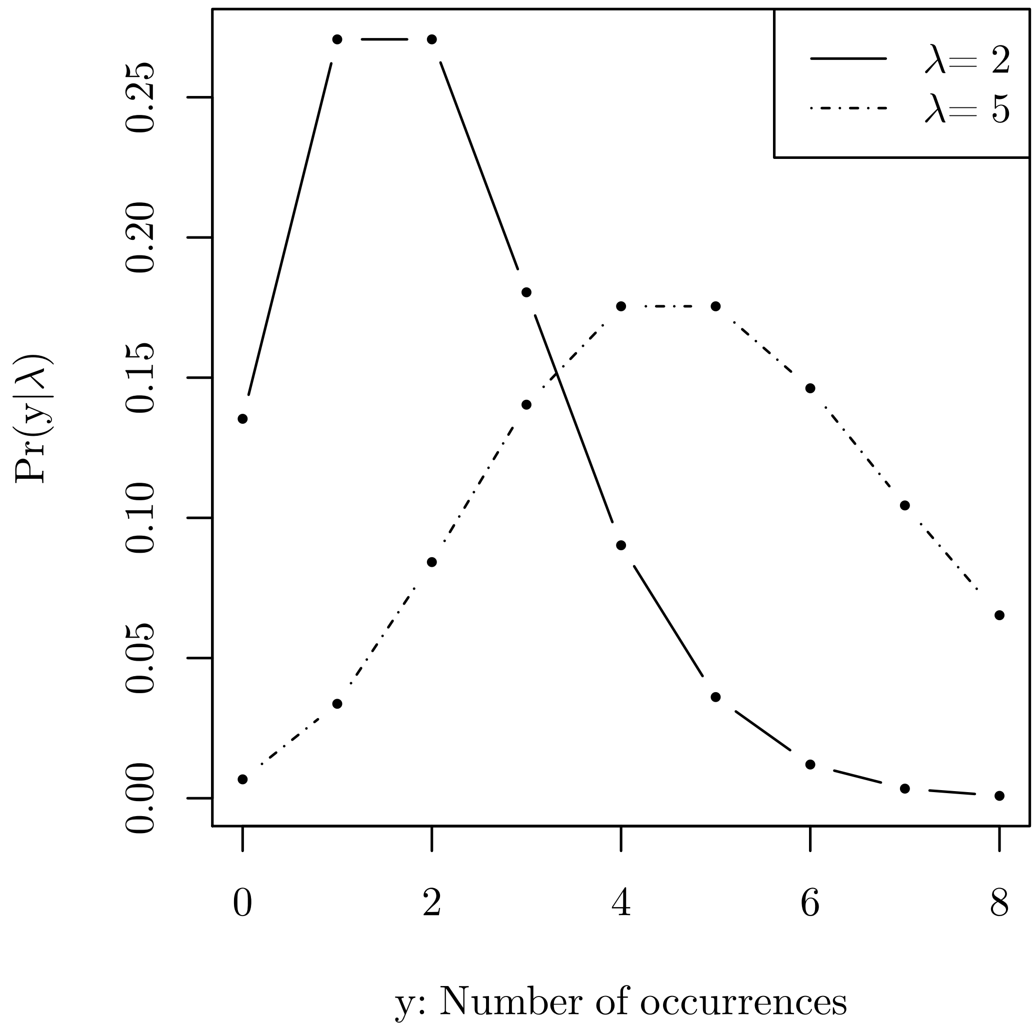
Scaling text: wordfish

Recall Poisson distribution for counts.

Think about rates at which parties use different words in party manifestos.

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it} + \psi_j + \beta_j \omega_{it}}$$



Scaling text: wordfish

Recall Poisson distribution for counts.

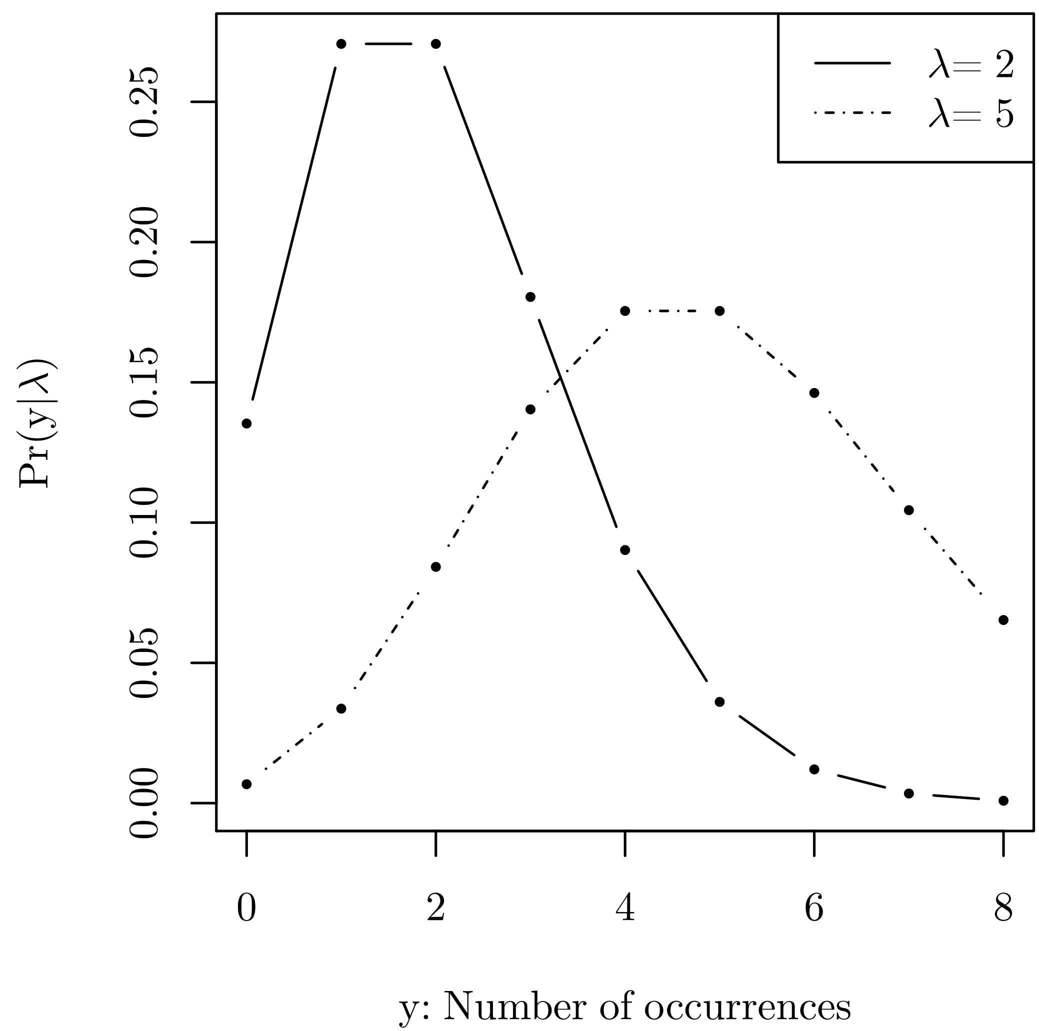
Think about rates at which parties use different words in party manifestos.

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it} + \psi_j + \beta_j \omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t



Scaling text: wordfish

ACTIVITY!!!

Scaling text: wordfish

ACTIVITY!!!

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it} + \psi_j + \beta_j \omega_{it}}$$

Scaling text: wordfish

ACTIVITY!!!

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it} + \psi_j + \beta_j \omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t

Scaling text: wordfish

ACTIVITY!!!

Consider the words "and" and "deficit".

Q: What values of ψ_j and β_j would you expect for these words?

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it}+\psi_j+\beta_j\omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t

Scaling text: wordfish

ACTIVITY!!!

Consider the words "and" and "deficit".

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it}+\psi_j+\beta_j\omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t

Q: What values of ψ_j and β_j would you expect for these words?

A: For the word "and":

Scaling text: wordfish

ACTIVITY!!!

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it} + \psi_j + \beta_j \omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t

Consider the words "and" and "deficit".

Q: What values of ψ_j and β_j would you expect for these words?

A: For the word "and":

- high ψ_j, because it is a common word

Scaling text: wordfish

ACTIVITY!!!

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it} + \psi_j + \beta_j \omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t

Consider the words "and" and "deficit".

Q: What values of ψ_j and β_j would you expect for these words?

A: For the word "and":

- high ψ_j, because it is a common word
- small (in magnitude) β_j because its frequency is not likely to differ between parties

Scaling text: wordfish

ACTIVITY!!!

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it}+\psi_j+\beta_j\omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t

Consider the words "and" and "deficit".

Q: What values of ψ_j and β_j would you expect for these words?

A: For the word "and":

- high ψ_j, because it is a common word
- small (in magnitude) β_j because its frequency is not likely to differ between parties

For the word "deficit":

Scaling text: wordfish

ACTIVITY!!!

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it} + \psi_j + \beta_j \omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t

Consider the words "and" and "deficit".

Q: What values of ψ_j and β_j would you expect for these words?

A: For the word "and":

- high ψ_j, because it is a common word
- small (in magnitude) β_j because its frequency is not likely to differ between parties

For the word "deficit":

- lower ψ_j

Scaling text: wordfish

ACTIVITY!!!

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it}+\psi_j+\beta_j\omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t

Consider the words "and" and "deficit".

Q: What values of ψ_j and β_j would you expect for these words?

A: For the word "and":

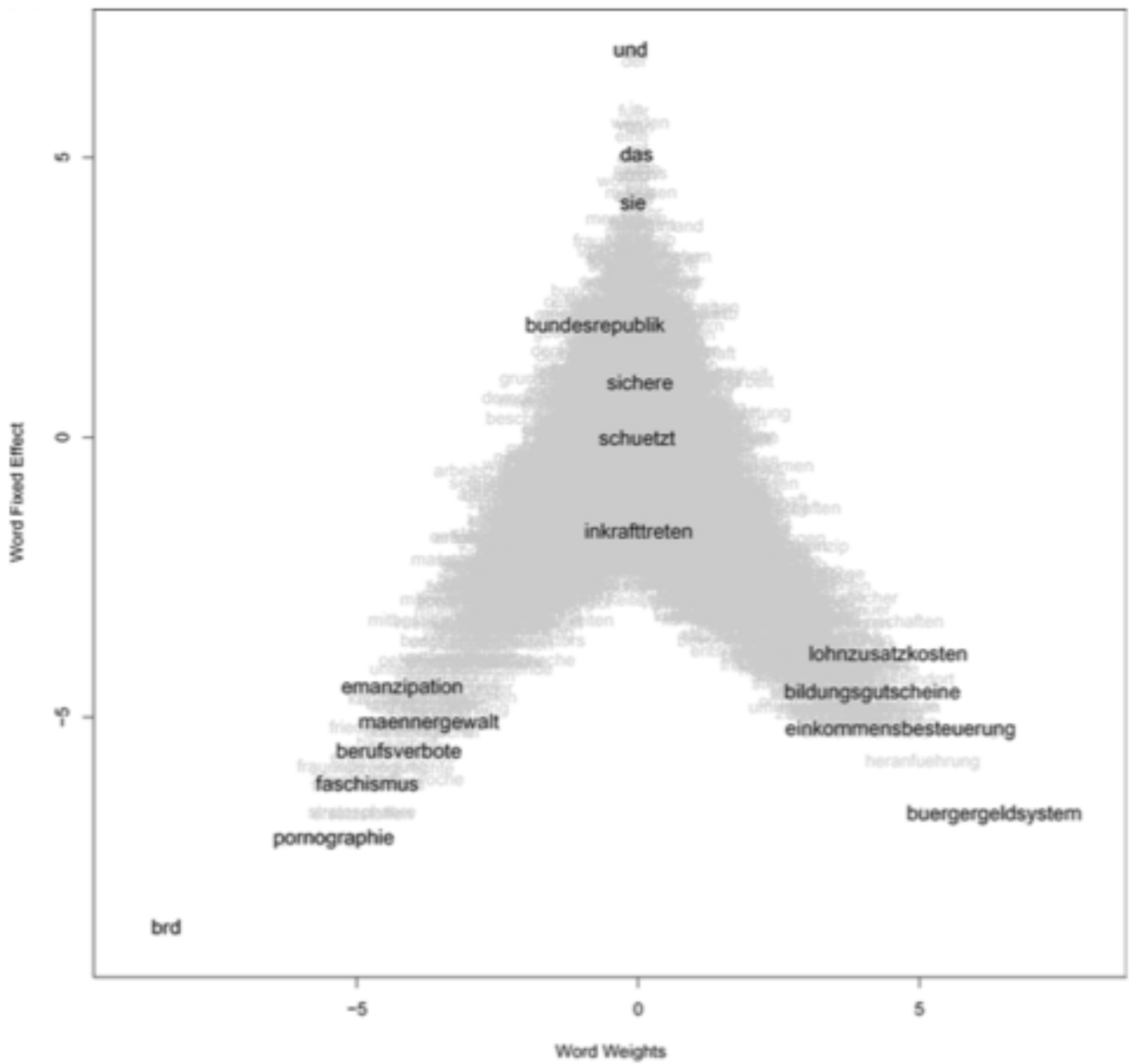
- high ψ_j, because it is a common word
- small (in magnitude) β_j because its frequency is not likely to differ between parties

For the word "deficit":

- lower ψ_j
- larger (in magnitude) β_j; for example, if the right talks about "deficits" more frequently and party positions are oriented so that right is positive, β_j should be large and positive.

Eiffel Tower of words

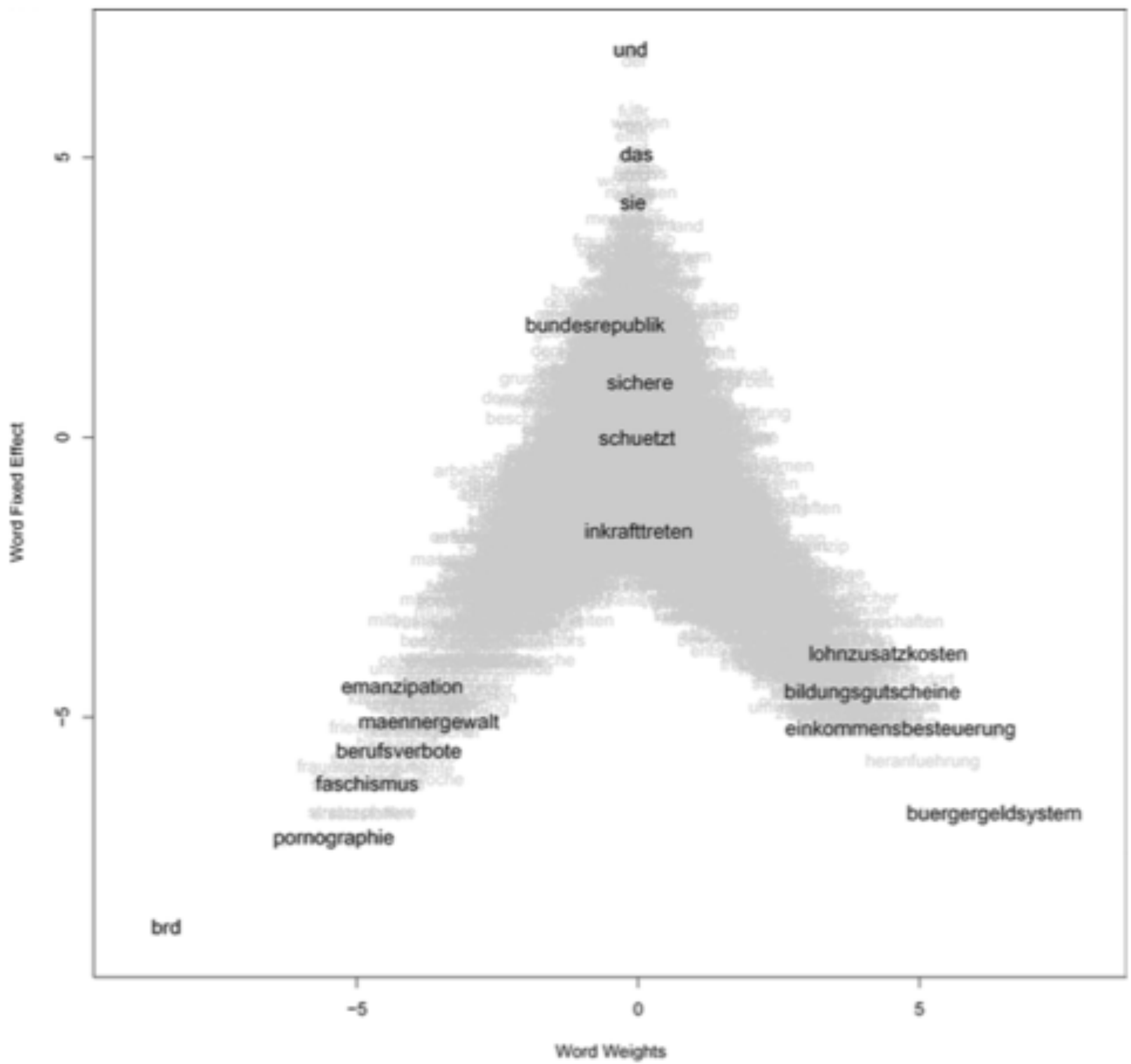
FIGURE 2 **Word Weights vs. Word Fixed Effects. Left-Right Dimension, Germany 1990–2005 (Translations given in text)**



Eiffel Tower of words

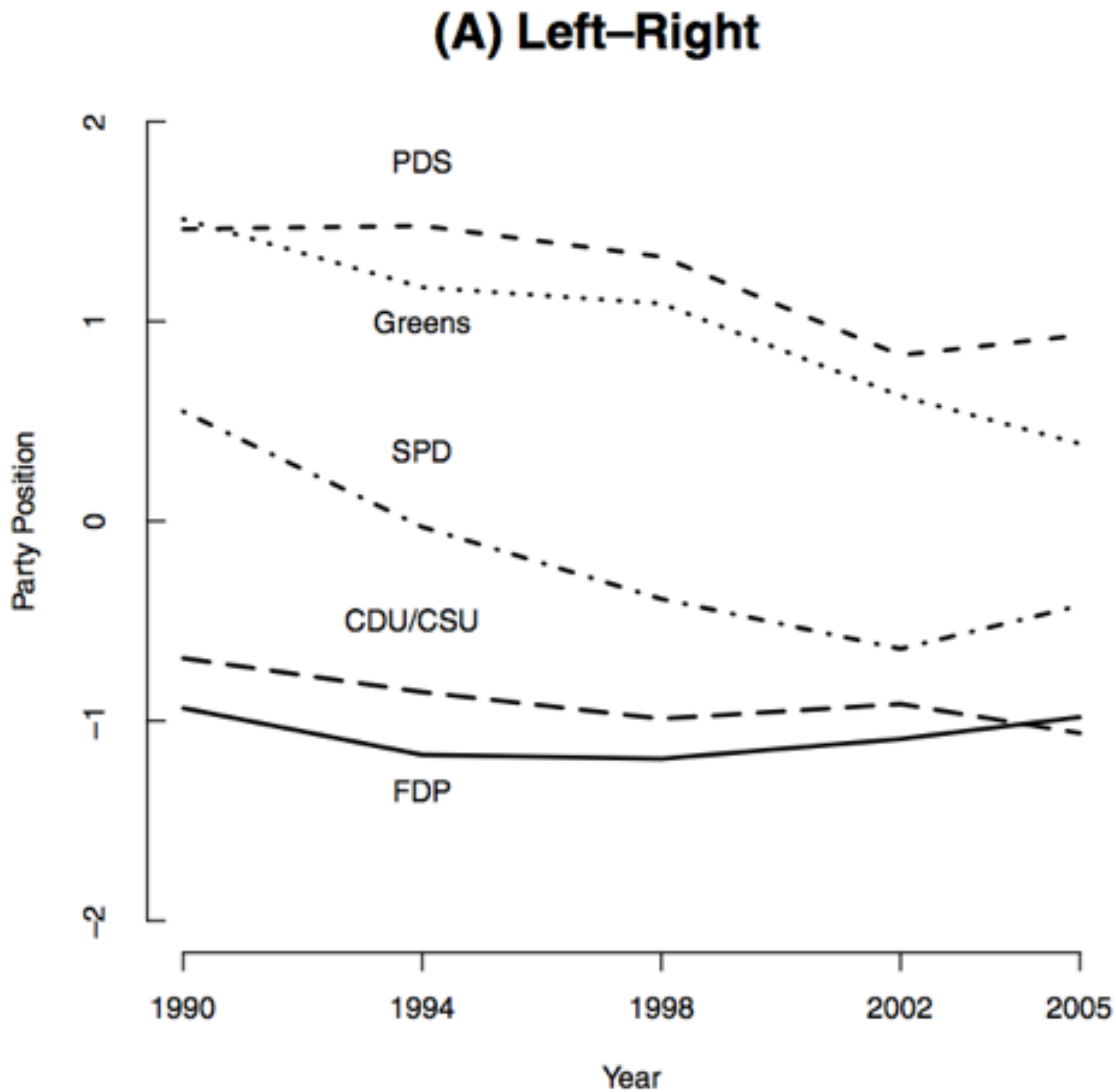
Slapin and Proksch, 2008

FIGURE 2 **Word Weights vs. Word Fixed Effects. Left-Right Dimension, Germany 1990–2005 (Translations given in text)**



Estimated party positions in Germany

Slapin and Proksch, 2008



Variations to be aware of

The **underlying model**:

- In IRT approaches, behavior is monotonic in x_i: the further right you are, the more likely you are to vote for a conservative measure
- In other approaches (e.g. Bonica 2013 on PAC contributions; Solomon and Messing 2015 on Facebook likes), behavior depends on proximity: the closer I am to the candidate the more likely I am to contribute/like

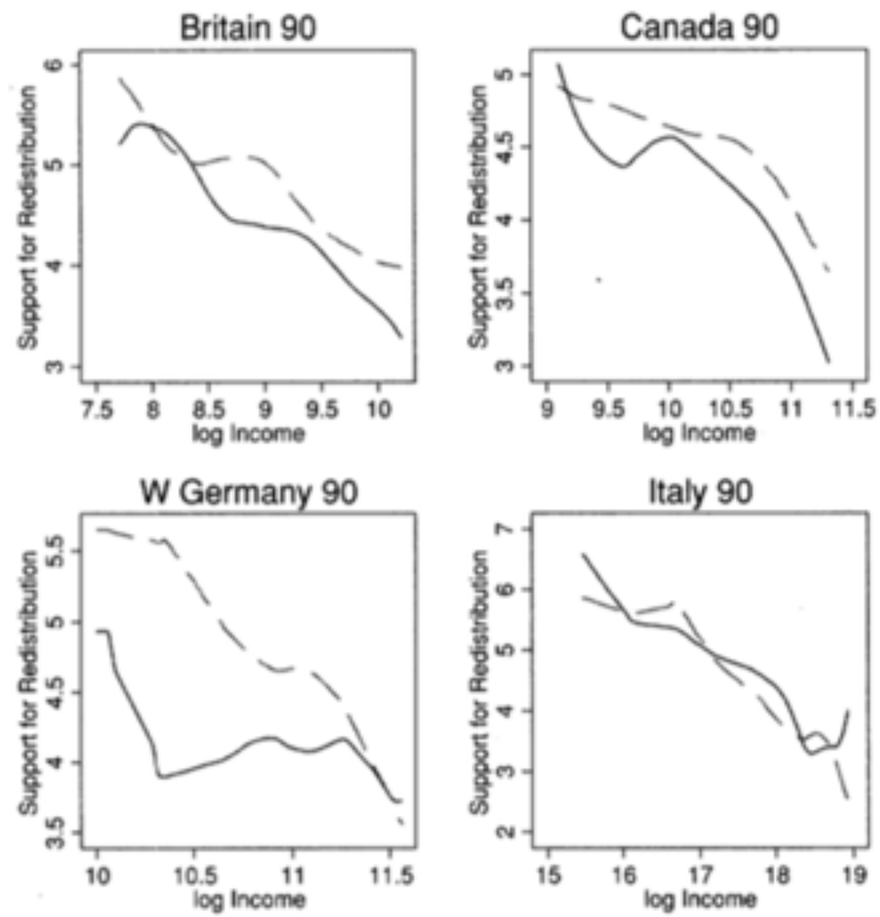
Level of aggregation:

- Classic uses are about estimating x for each individual: student ability, legislator ideology, etc
- Caughey and Warshaw 2015 estimate a group-level x based on sparse survey data

Other interesting uses of statistical modeling

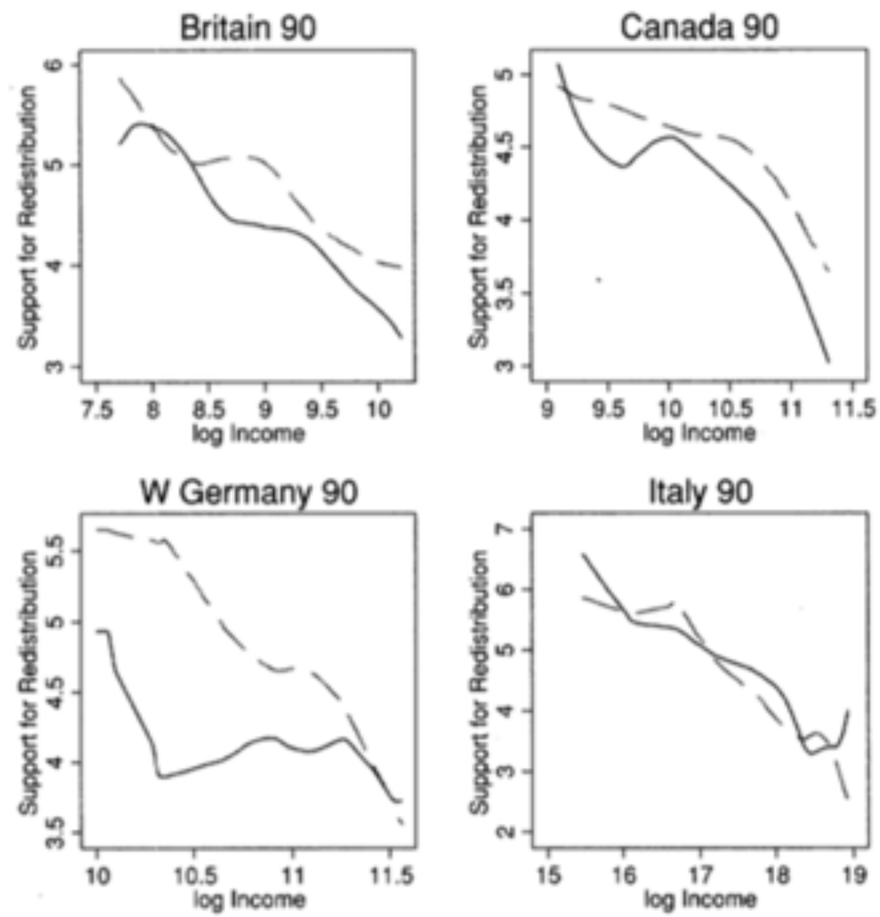
- "Small-area estimation": How can we estimate the average preference of each legislative district (e.g. on same-sex marriage) with a survey that only has 5-10 respondents per district? (**MRP**: Multilevel regression and post-stratification)
- Topic modeling in text: what "topics" are being discussed in a corpus? How much does each document participate in each topic?

You can say a lot without a statistical model!



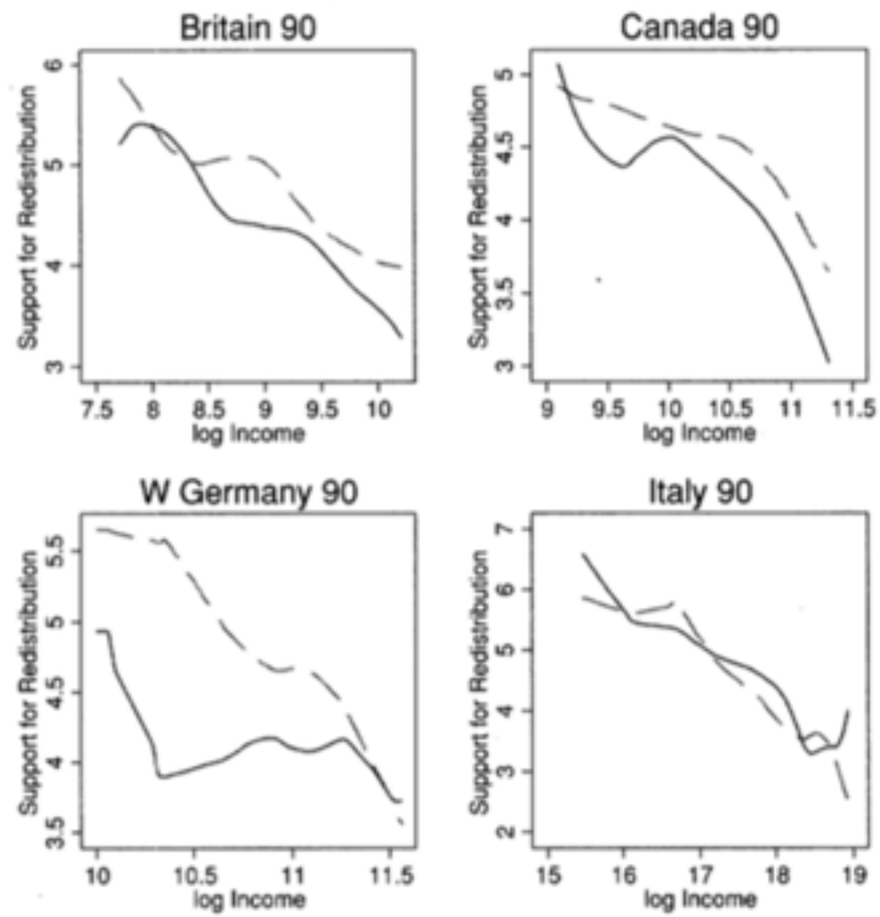
You can say a lot without a statistical model!

At right: Kernel regressions of support for redistribution as function of income for WVS respondents who were "Very Proud" and "Less Proud" of their country



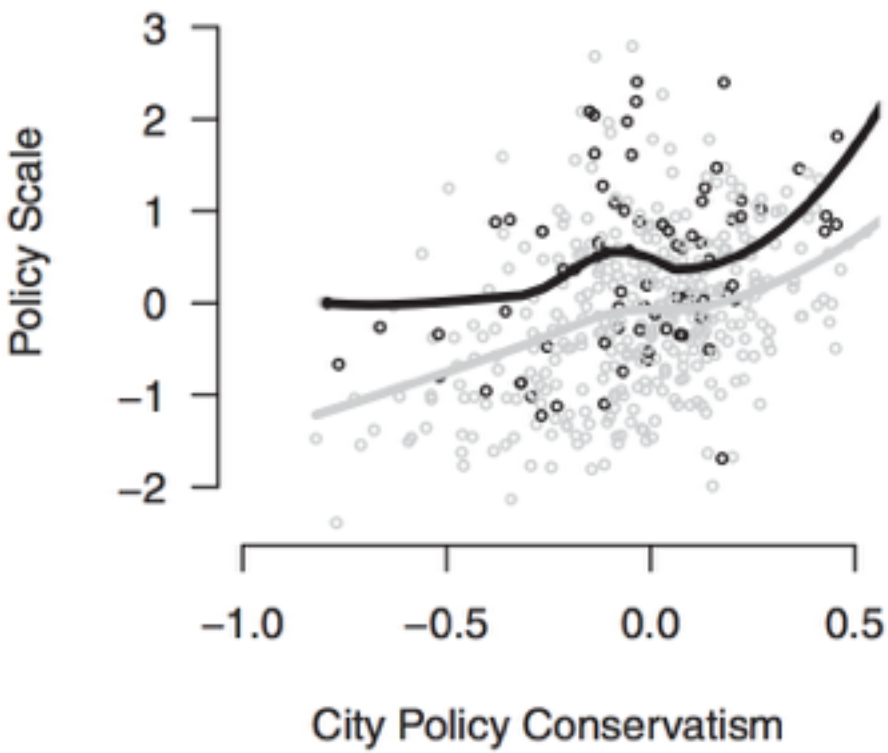
You can say a lot without a statistical model!

At right: Kernel regressions of support for redistribution as function of income for WVS respondents who were “Very Proud” and “Less Proud” of their country

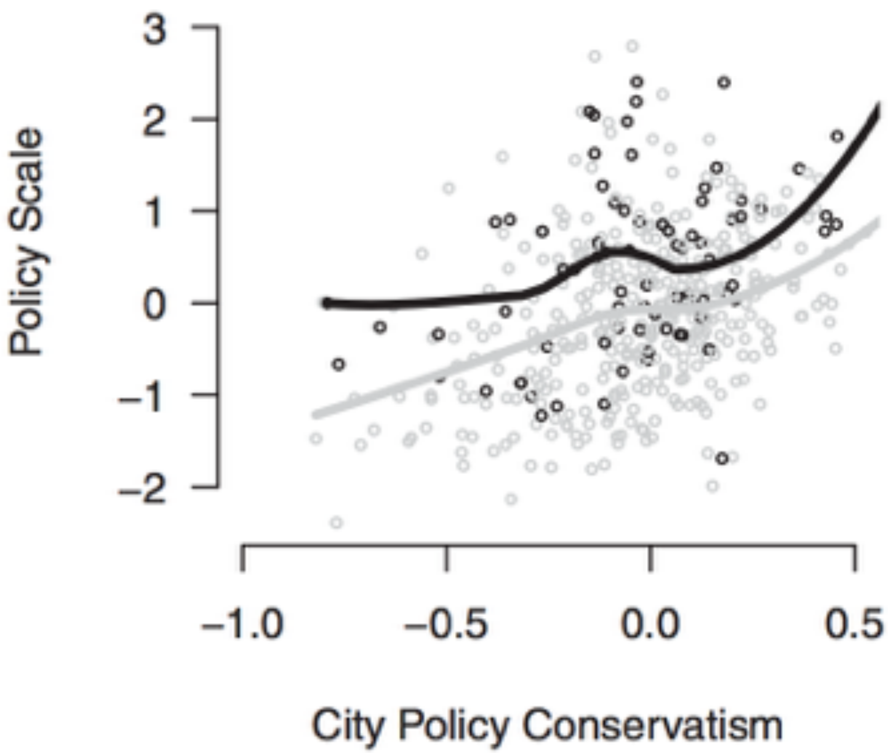


Shayo (2009) “A model of social identity” APSR.

And the best statistics are easy to miss

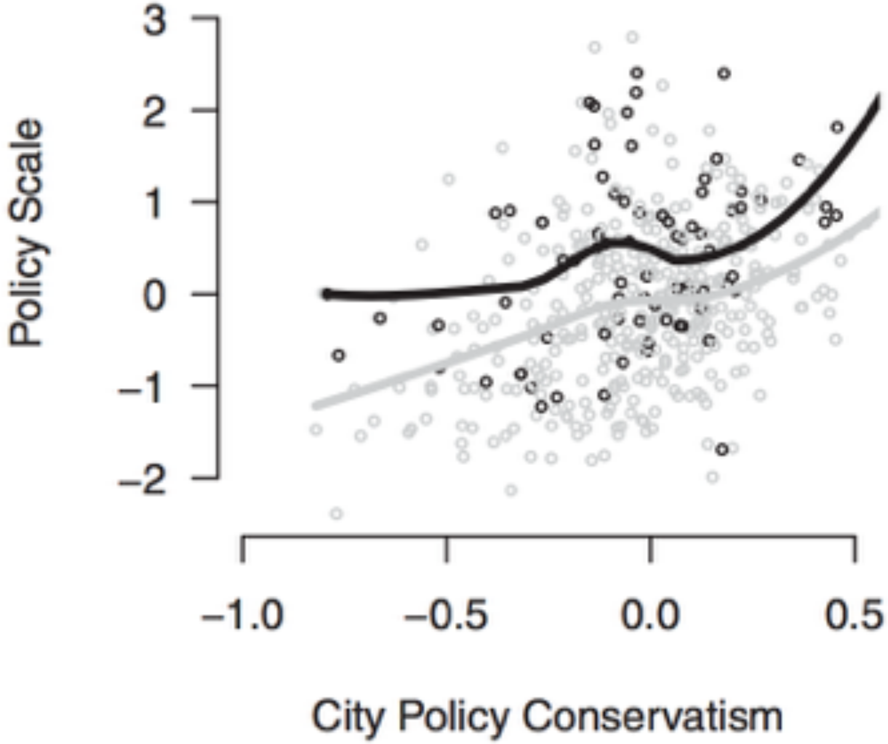


And the best statistics are easy to miss



No explicit model in comparing responsiveness, but note:

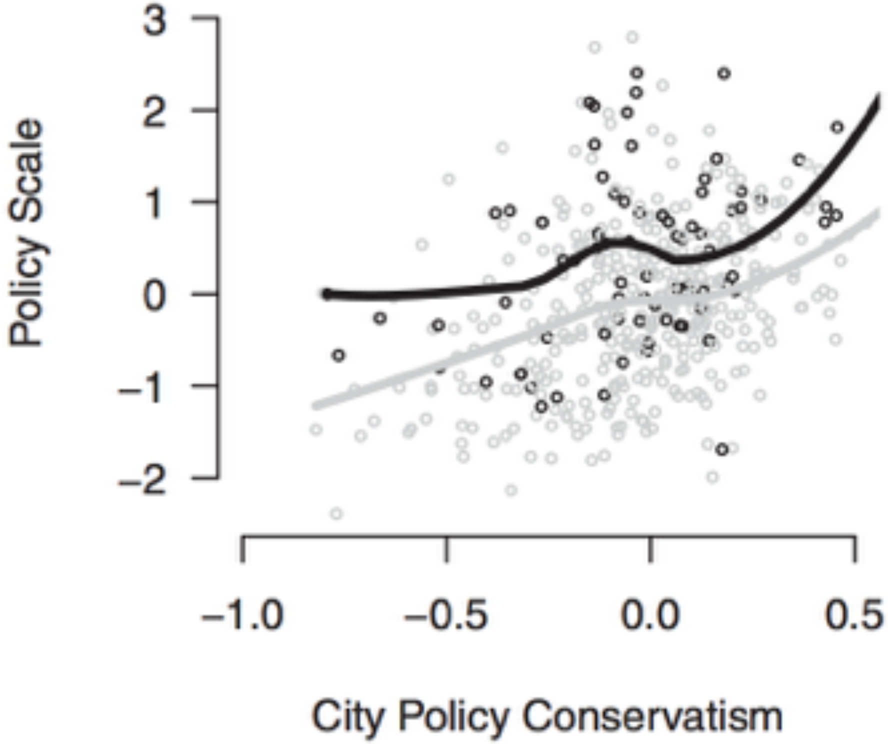
And the best statistics are easy to miss



No explicit model in comparing responsiveness, but note:

- policy conservatism (public opinion) based on

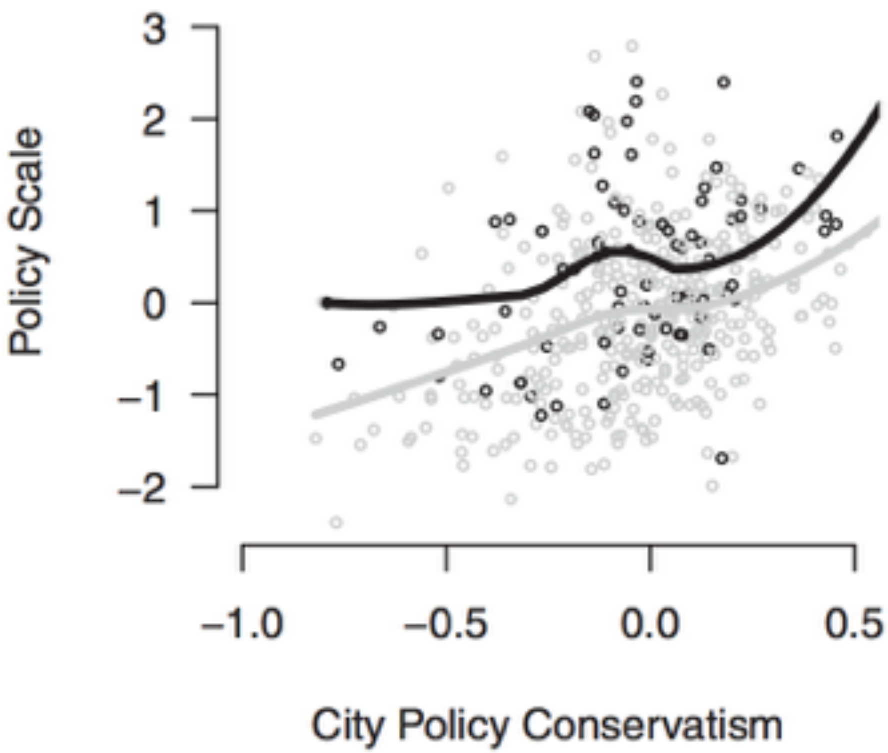
And the best statistics are easy to miss



No explicit model in comparing responsiveness, but note:

- policy conservatism (public opinion) based on
    - IRT estimates for survey respondents

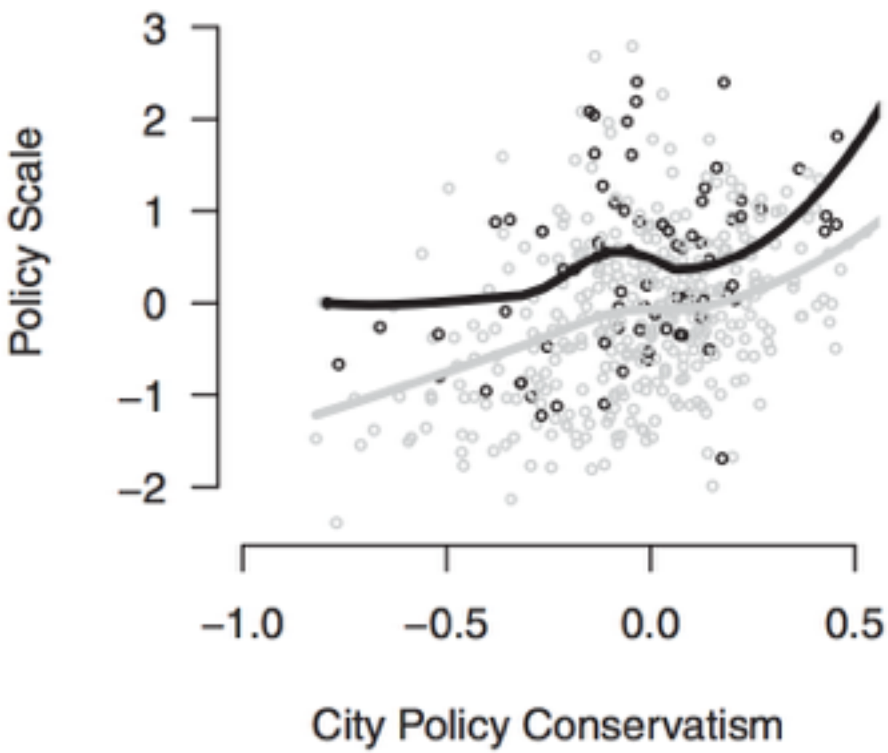
And the best statistics are easy to miss



No explicit model in comparing responsiveness, but note:

- policy conservatism (public opinion) based on
  - IRT estimates for survey respondents
  - with averages estimated for each city by MRP

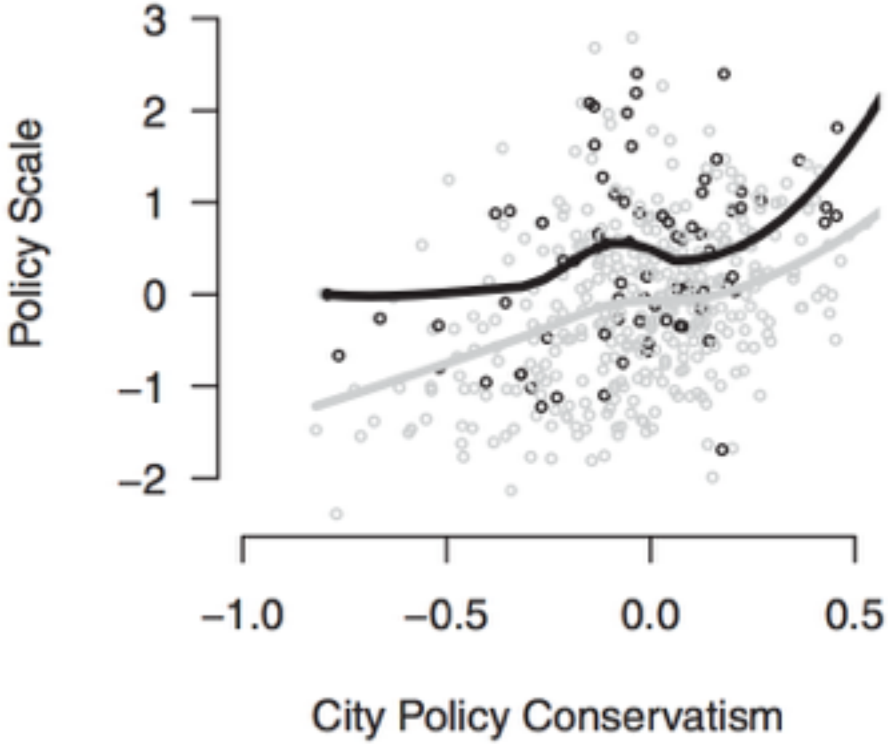
And the best statistics are easy to miss



No explicit model in comparing responsiveness, but note:

- policy conservatism (public opinion) based on
  - IRT estimates for survey respondents
  - with averages estimated for each city by MRP
- municipal policy estimated via IRT from list of policy questions

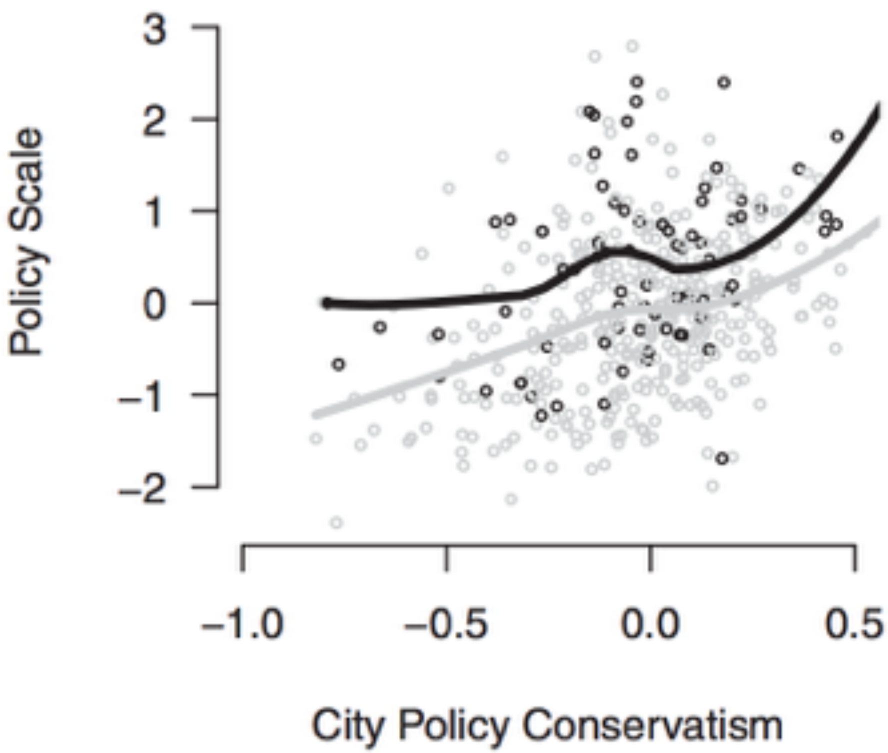
And the best statistics are easy to miss



No explicit model in comparing responsiveness, but note:

- policy conservatism (public opinion) based on
  - IRT estimates for survey respondents
  - with averages estimated for each city by MRP
- municipal policy estimated via IRT from list of policy questions
- lowess lines for elected mayor vs manager here based on entropy balancing (generalization of matching due to Hainmueller)

And the best statistics are easy to miss

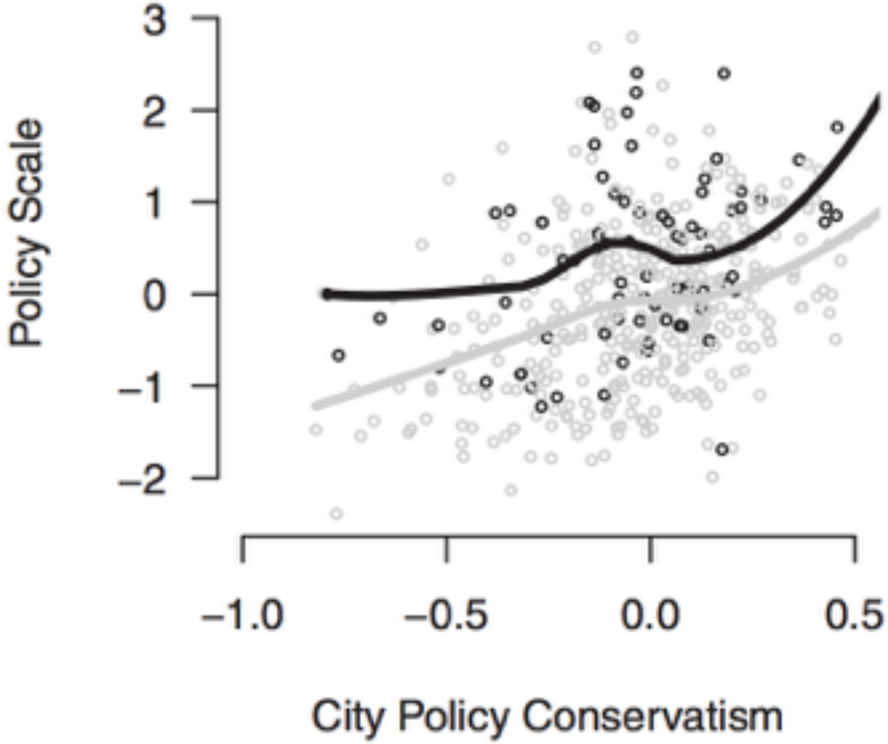


Tausanovitch and Warshaw, "Representation in Municipal Government", APSR 2014:

No explicit model in comparing responsiveness, but note:

- policy conservatism (public opinion) based on
  - IRT estimates for survey respondents
  - with averages estimated for each city by MRP
- municipal policy estimated via IRT from list of policy questions
- lowess lines for elected mayor vs manager here based on entropy balancing (generalization of matching due to Hainmueller)

And the best statistics are easy to miss



Tausanovitch and Warshaw, "Representation in Municipal Government", APSR 2014:
Showing how municipal policy (y) varies with municipal public opinion (x) in cities with elected mayor (black) vs. manager (gray)

No explicit model in comparing responsiveness, but note:

- policy conservatism (public opinion) based on
  - IRT estimates for survey respondents
  - with averages estimated for each city by MRP
- municipal policy estimated via IRT from list of policy questions
- lowess lines for elected mayor vs manager here based on entropy balancing (generalization of matching due to Hainmueller)

General advice

General advice

- Keep it simple

General advice

- Keep it simple
- Keep it linked to an interesting research question

General advice

- Keep it simple
- Keep it linked to an interesting research question
- Keep it visual: before (and after) running a model, look at the data!

General advice

- Keep it simple
- Keep it linked to an interesting research question
- Keep it visual: before (and after) running a model, look at the data!
- Don't be restricted by your inabilities and **especially** your ignorance: you need to recognize when statistics could help

General advice

- Keep it simple
- Keep it linked to an interesting research question
- Keep it visual: before (and after) running a model, look at the data!
- Don't be restricted by your inabilities and **especially** your ignorance: you need to recognize when statistics could help
- Learn to program: at least one of Stata, R, python, ruby

General advice

- Keep it simple
- Keep it linked to an interesting research question
- Keep it visual: before (and after) running a model, look at the data!
- Don't be restricted by your inabilities and **especially** your ignorance: you need to recognize when statistics could help
- Learn to program: at least one of Stata, R, python, ruby
- There are many ways to contribute. Choose some combination of:
    - better data
    - better design (e.g. causal inference)
    - better measurement
    - better theory

  Often one of these makes possible another.