Lecture 9: Logit/Probit

Prof. Sharyn O'Halloran

Sustainable Development U9611

Econometrics II

Review of Linear Estimation

- So far, we know how to handle linear estimation models of the type:

$$Y = \beta_0 + \beta_1 {*} X_1 + \beta_2 {*} X_2 + \ldots + \varepsilon \equiv \mathbf{X}\beta + \varepsilon$$

- Sometimes we had to transform or add variables to get the equation to be linear:
  - Taking logs of Y and/or the X's
  - Adding squared terms
  - Adding interactions
- Then we can run our estimation, do model checking, visualize results, etc.

Nonlinear Estimation

- In all these models Y, the dependent variable, was continuous.
  - Independent variables could be dichotomous (dummy variables), but not the dependent var.
- This week we'll start our exploration of non-linear estimation with dichotomous Y vars.
- These arise in many social science problems
  - Legislator Votes: Aye/Nay
  - Regime Type: Autocratic/Democratic
  - Involved in an Armed Conflict: Yes/No

Link Functions

- Before plunging in, let's introduce the concept of a <u>link function</u>
  - This is a function linking the actual Y to the estimated Y in an econometric model
- We have one example of this already: logs
  - Start with $Y = \mathbf{X}\beta + \varepsilon$
  - Then change to $\log(Y) \equiv Y' = \mathbf{X}\beta + \varepsilon$
  - Run this like a regular OLS equation
  - Then you have to "back out" the results

Link Functions

- Before plunging in, let's introduce the concept of a <u>link function</u>
  - This is a function linking the actual Y to the estimated Y in an econometric model
- We have one example of this already: logs
  - Start with $Y = \mathbf{X}\beta + \varepsilon$
  - Then change to $\log(Y) \equiv Y' = \mathbf{X}\beta + \varepsilon$
  - Run this like a regular OLS equation
  - Then you have to "back out" the results

Different β's here

Link Functions

- If the coefficient on some particular X is β, then a 1 unit $\Delta X \rightarrow \beta \cdot \Delta(Y') = \beta \cdot \Delta[\log(Y))]$

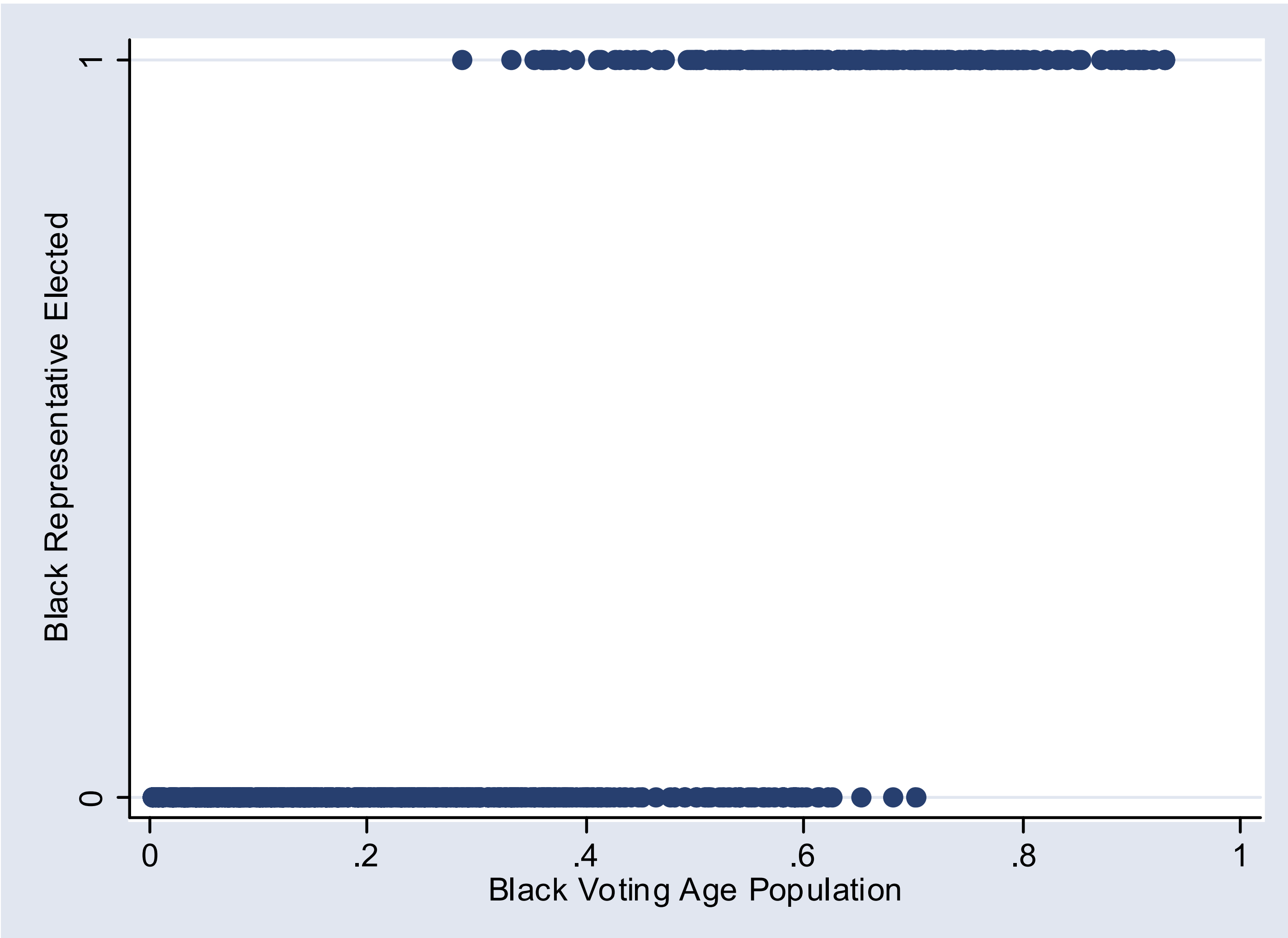
  $$= e^{\beta} \cdot \Delta(Y)$$

  - Since for small values of β, $e^{\beta} \approx 1+\beta$, this is almost the same as saying a β% increase in Y
  - (This is why you should use natural log transformations rather than base-10 logs)
- In general, a link function is some $F(\cdot)$ s.t.
  - $F(Y) = \mathbf{X}\beta + \varepsilon$
- In our example, $F(Y) = \log(Y)$

Dichotomous Independent Vars.

- How does this apply to situations with dichotomous dependent variables?
  - I.e., assume that $Y_i \in \{0,1\}$
- First, let's look at what would happen if we tried to run this as a linear regression
- As a specific example, take the election of minorities to the Georgia state legislature
  - Y = 0: Non-minority elected
  - Y = 1: Minority elected

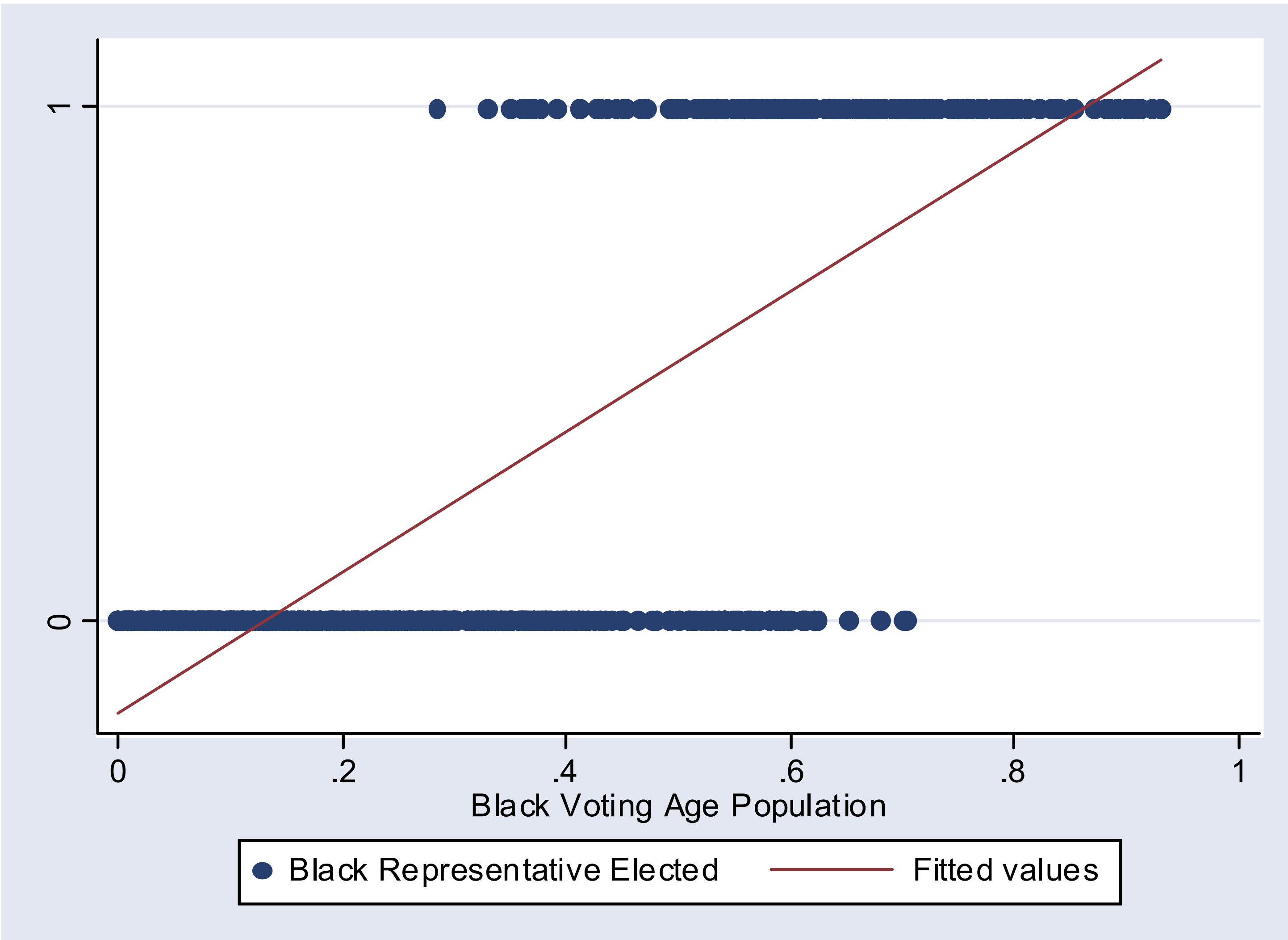
Dichotomous Independent Vars.



The data look like this.

The only values Y can have are 0 and 1

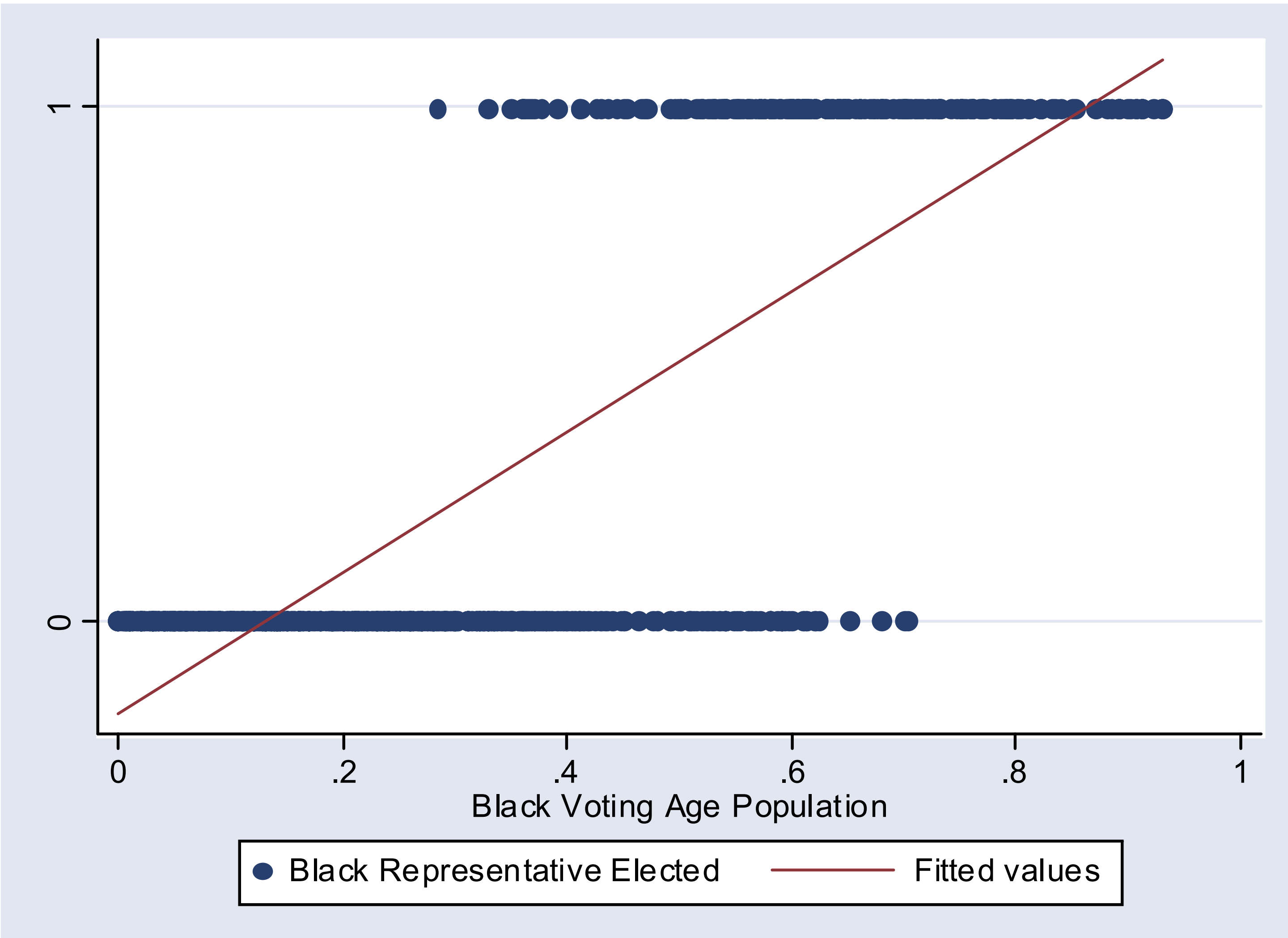
Dichotomous Independent Vars.



And here's a linear fit of the data

Note that the line goes below 0 and above 1

Dichotomous Independent Vars.



The line doesn't fit the data very well.

And if we take values of Y between 0 and 1 to be probabilities, this doesn't make sense

Redefining the Dependent Var.

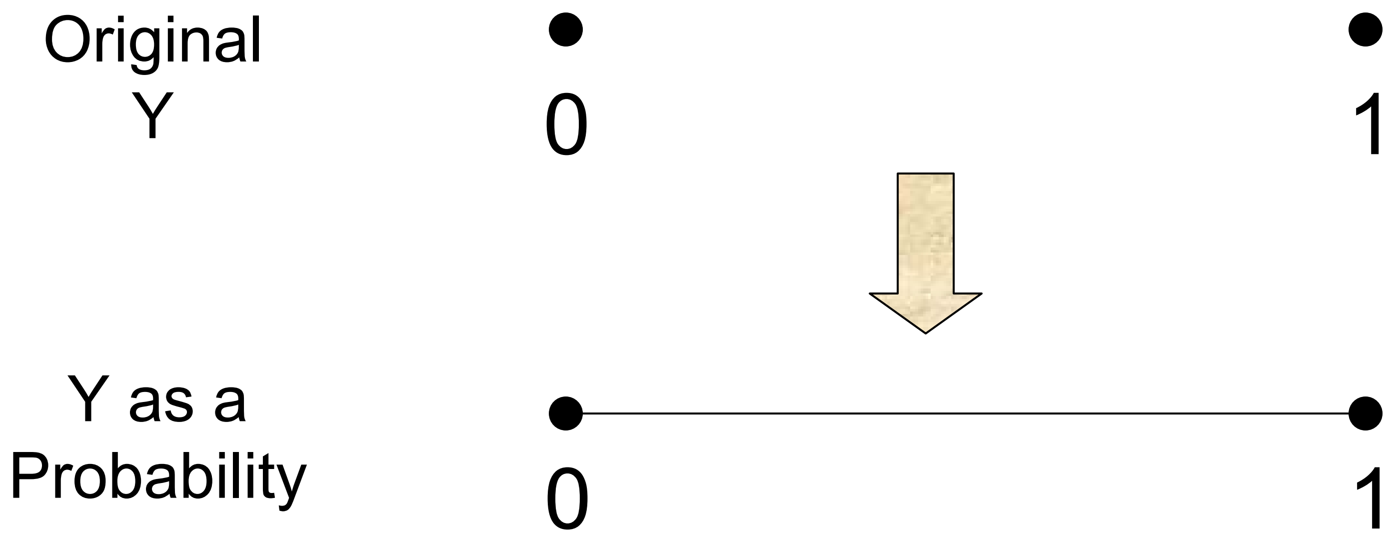
- How to solve this problem?
- We need to transform the dichotomous Y into a continuous variable $Y' \in (-\infty,\infty)$
- So we need a link function F(Y) that takes a dichotomous Y and gives us a continuous, real-valued Y′
- Then we can run

$$F(Y) = Y' = \mathbf{X}\beta + \varepsilon$$

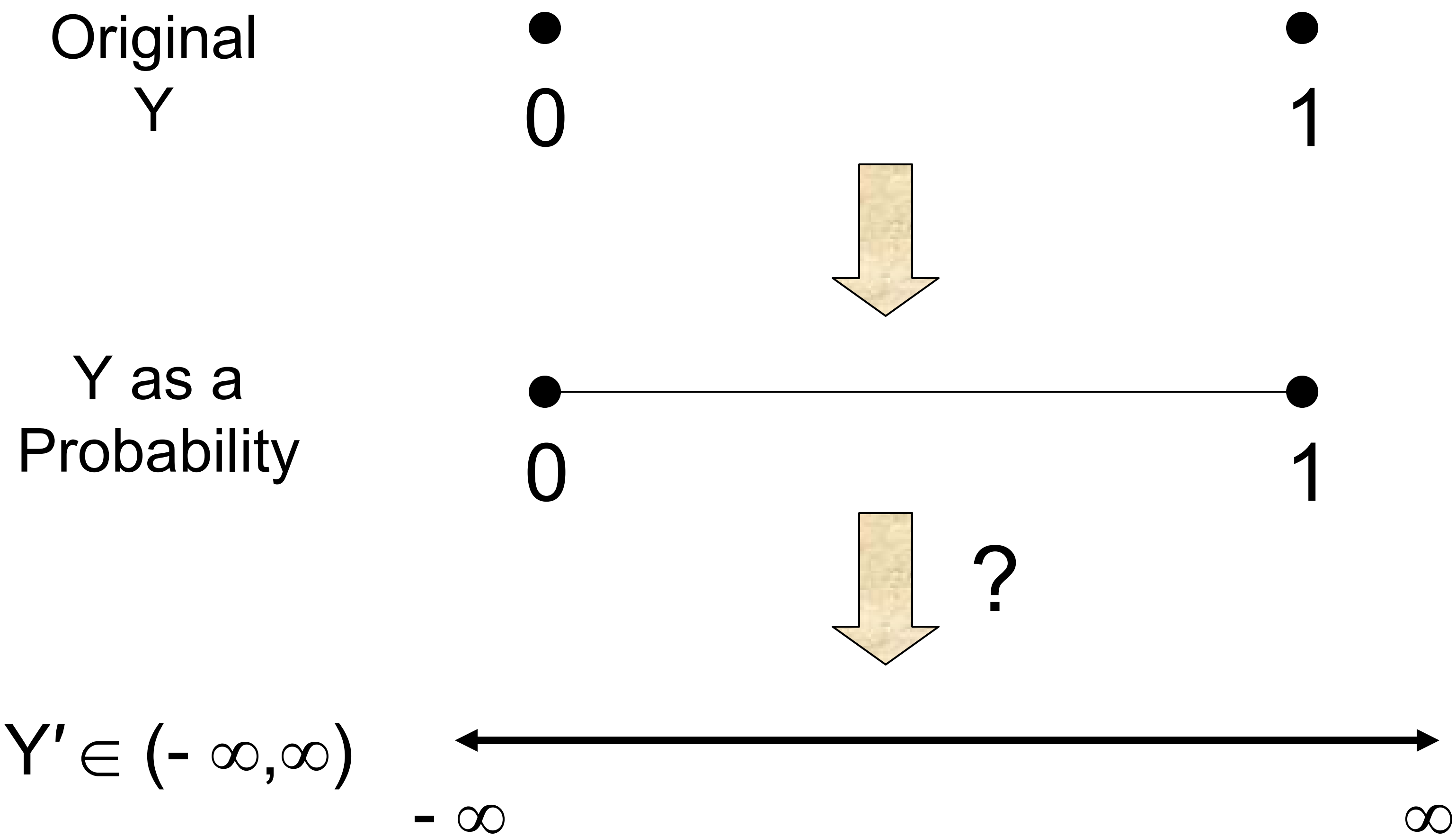
Redefining the Dependent Var.

Original Y	• 0	• 1

Redefining the Dependent Var.



Redefining the Dependent Var.



Redefining the Dependent Var.

- What function F(Y) goes from the [0,1] interval to the real line?
- Well, we know at least one function that goes the other way around.
  - That is, given any real value it produces a number (probability) between 0 and 1.
- This is the…

Redefining the Dependent Var.

- What function F(Y) goes from the [0,1] interval to the real line?
- Well, we know at least one function that goes the other way around.
  - That is, given any real value it produces a number (probability) between 0 and 1.
- This is the cumulative normal distribution Φ
  - That is, given any Z-score, $\Phi(Z) \in [0,1]$

Redefining the Dependent Var.

- So we would say that

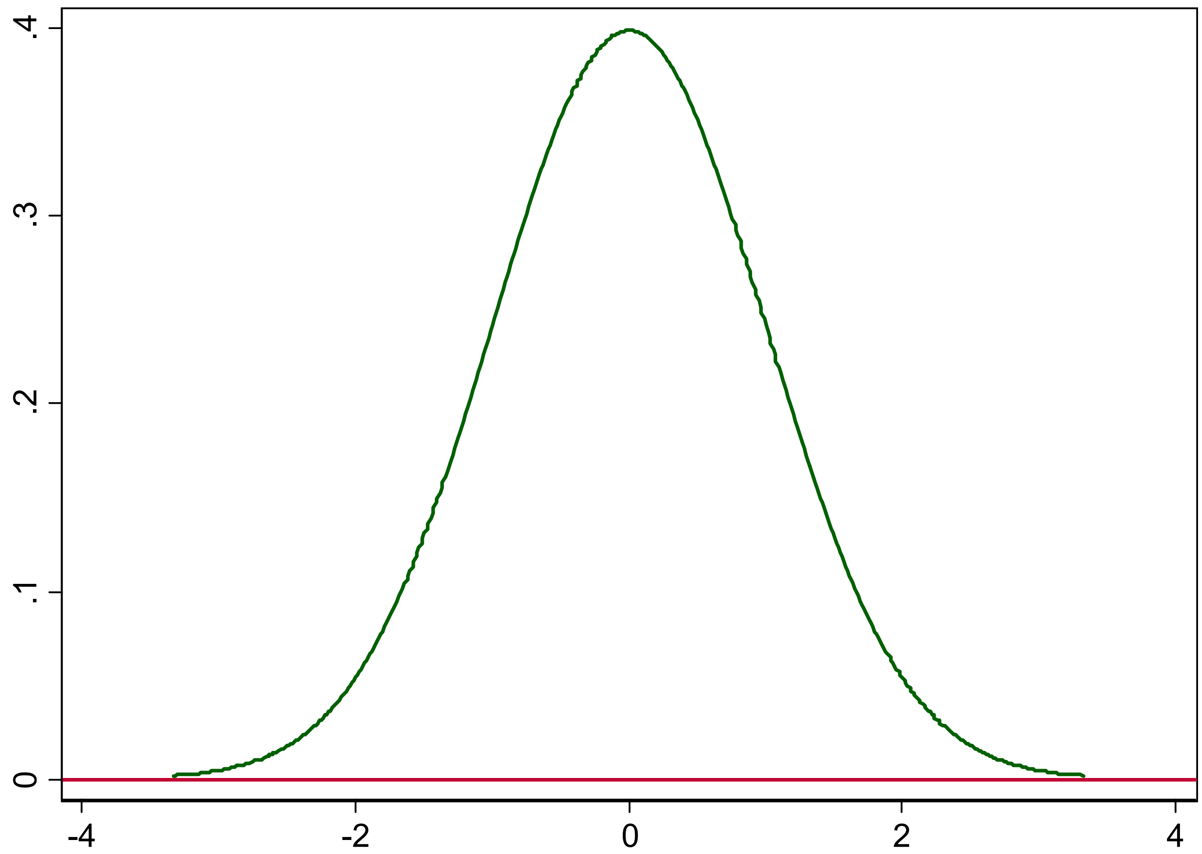
$$Y = \Phi(\mathbf{X}\beta + \varepsilon)$$

$$\Phi^{-1}(Y) = \mathbf{X}\beta + \varepsilon$$

$$Y' = \mathbf{X}\beta + \varepsilon$$

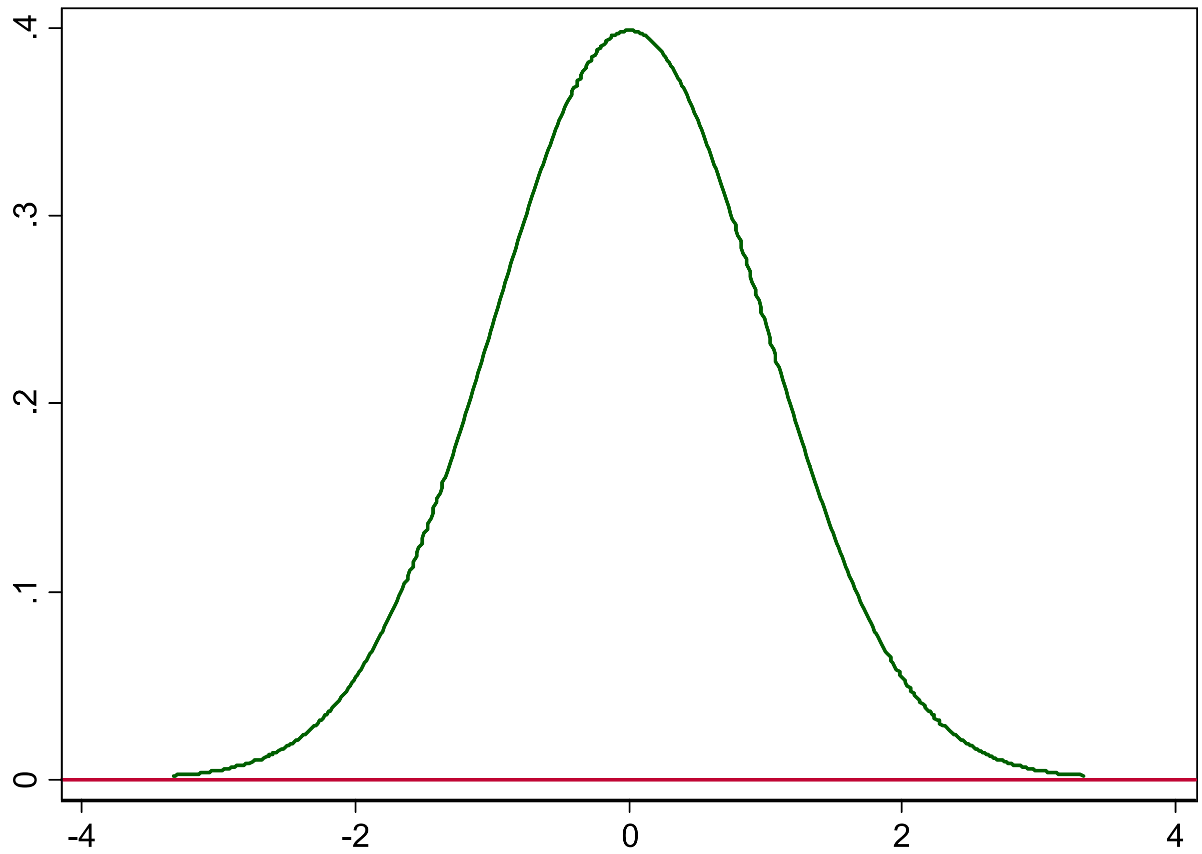
- Then our link function is $F(Y) = \Phi^{-1}(Y)$
- This link function is known as the <u>Probit</u> link
  - This term was coined in the 1930's by biologists studying the dosage-cure rate link
  - It is short for "probability unit"

Probit Estimation



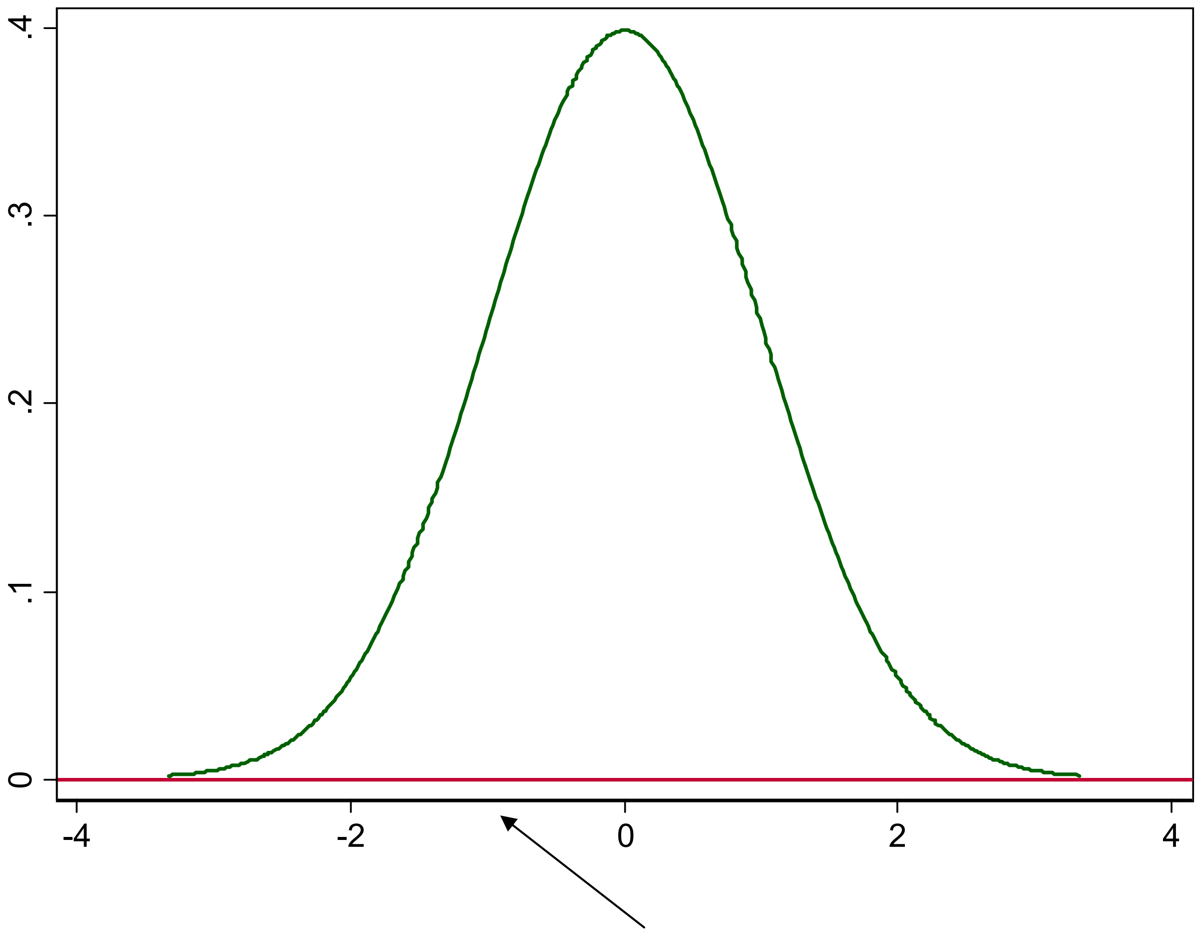
After estimation, you can back out probabilities using the standard normal dist.

Probit Estimation



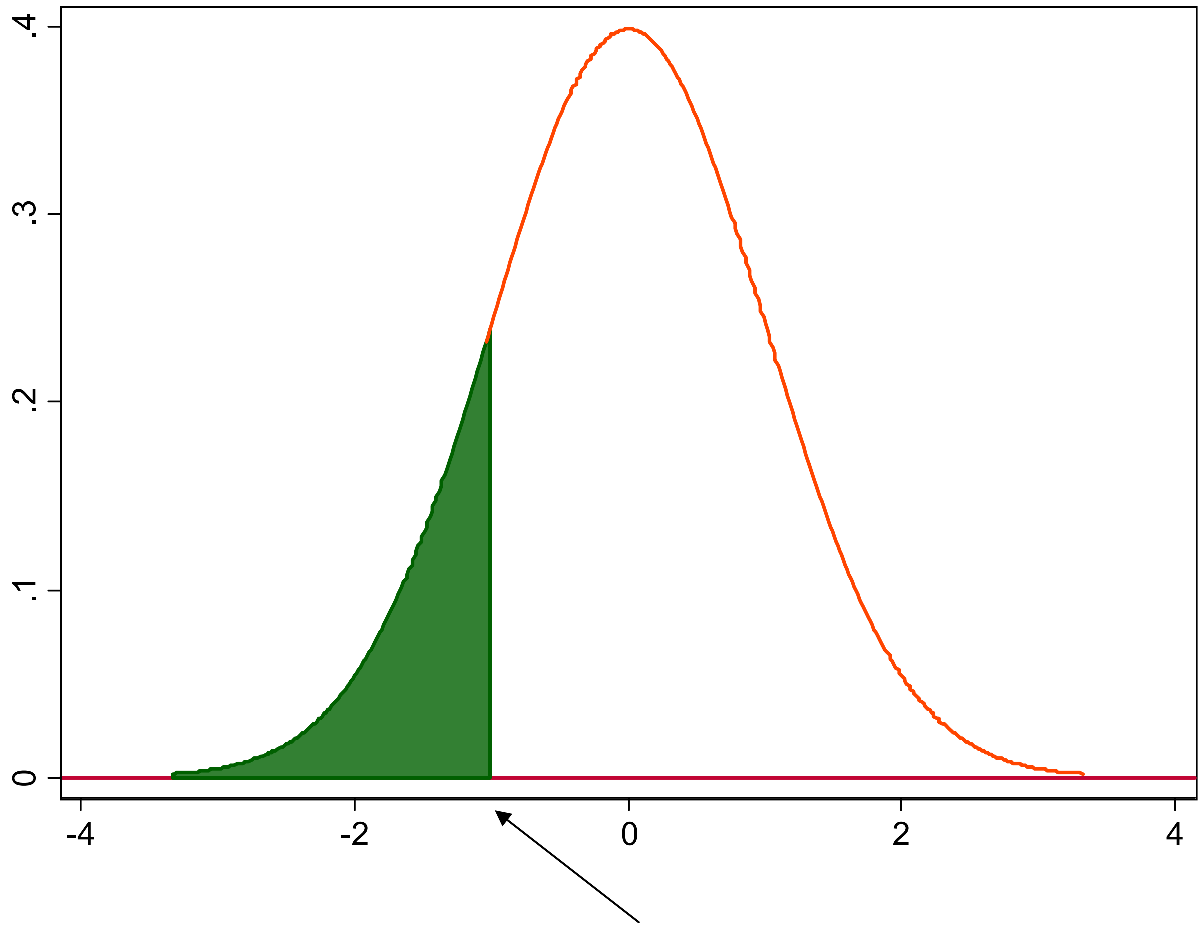
Say that for a given observation, $\mathbf{X}\beta = -1$

Probit Estimation



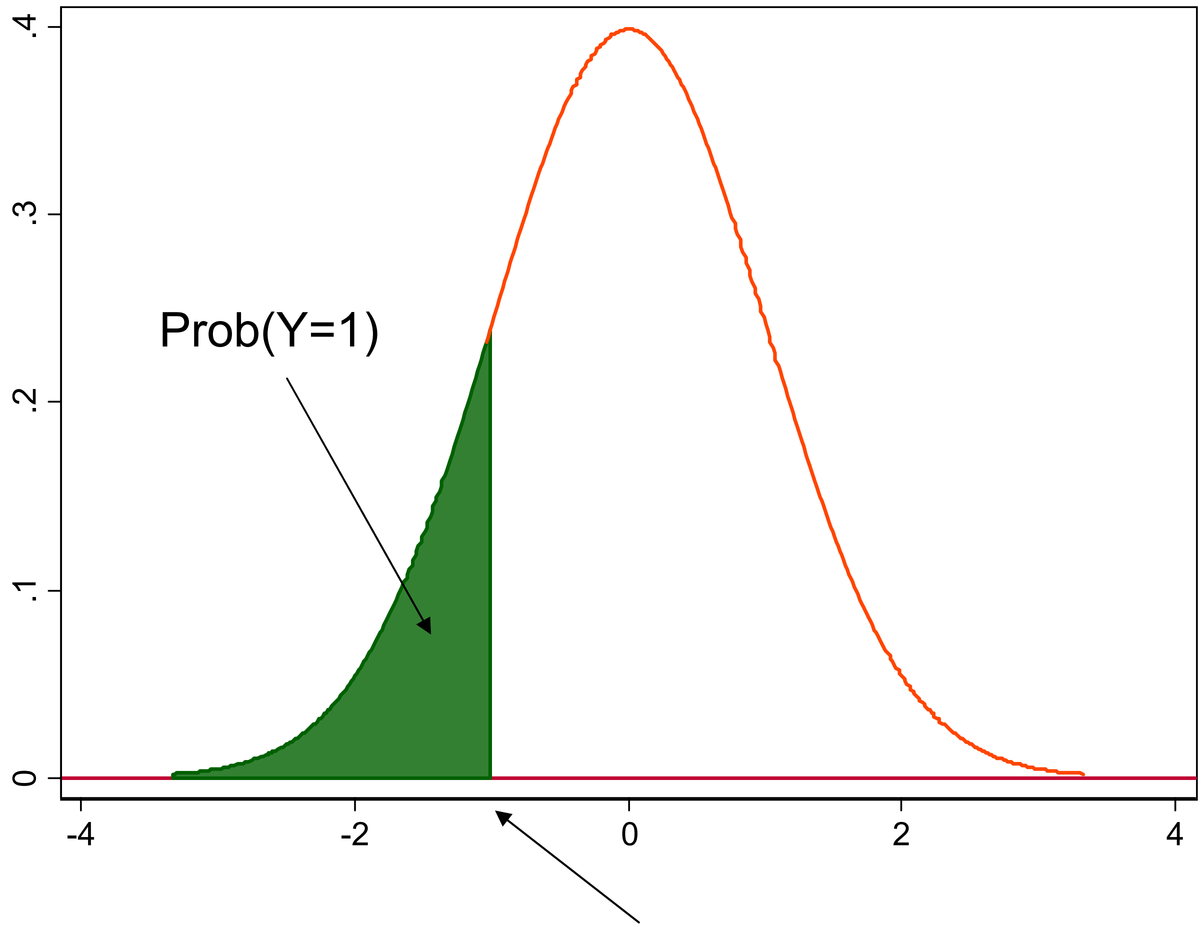
Say that for a given observation, $\mathbf{X}\beta = -1$

Probit Estimation



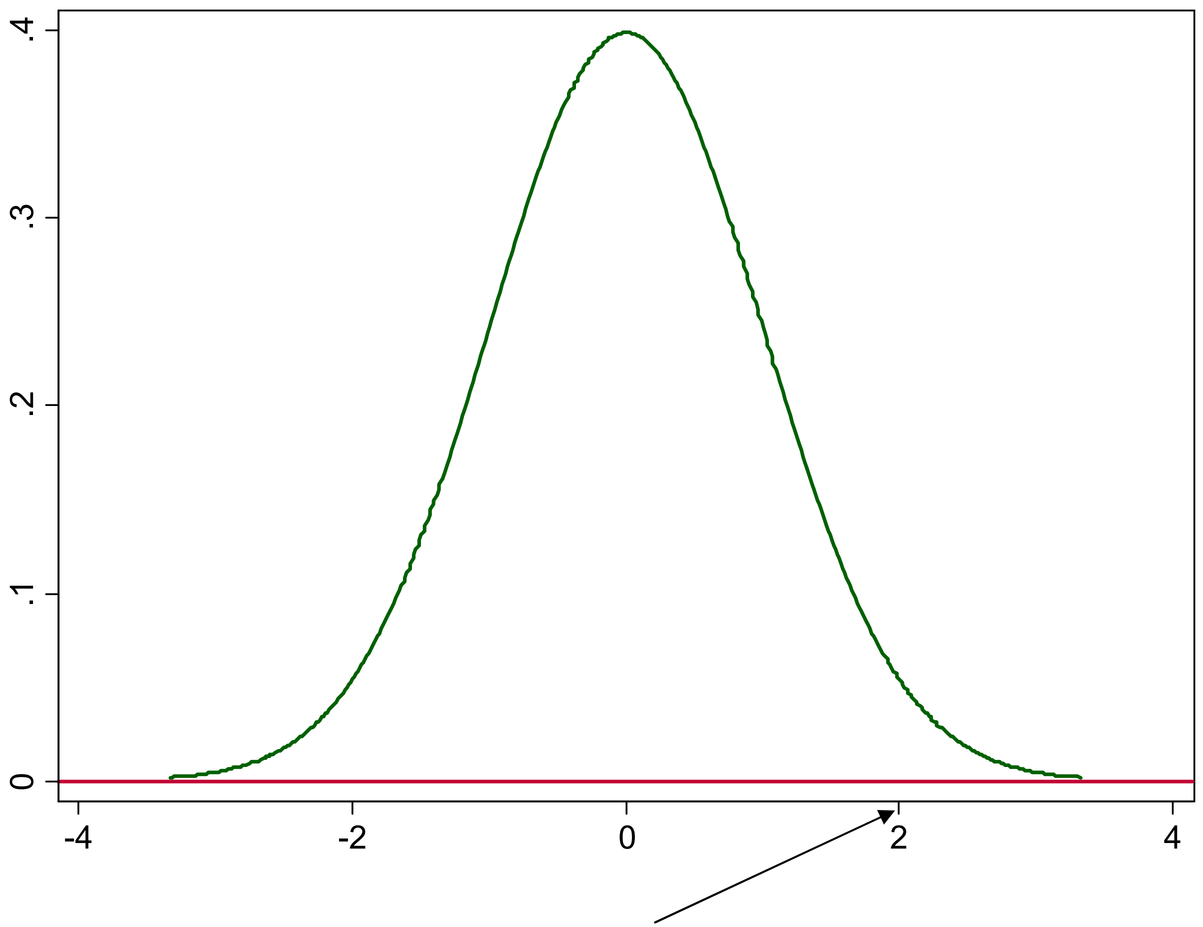
Say that for a given observation, $\mathbf{X}\beta = -1$

Probit Estimation



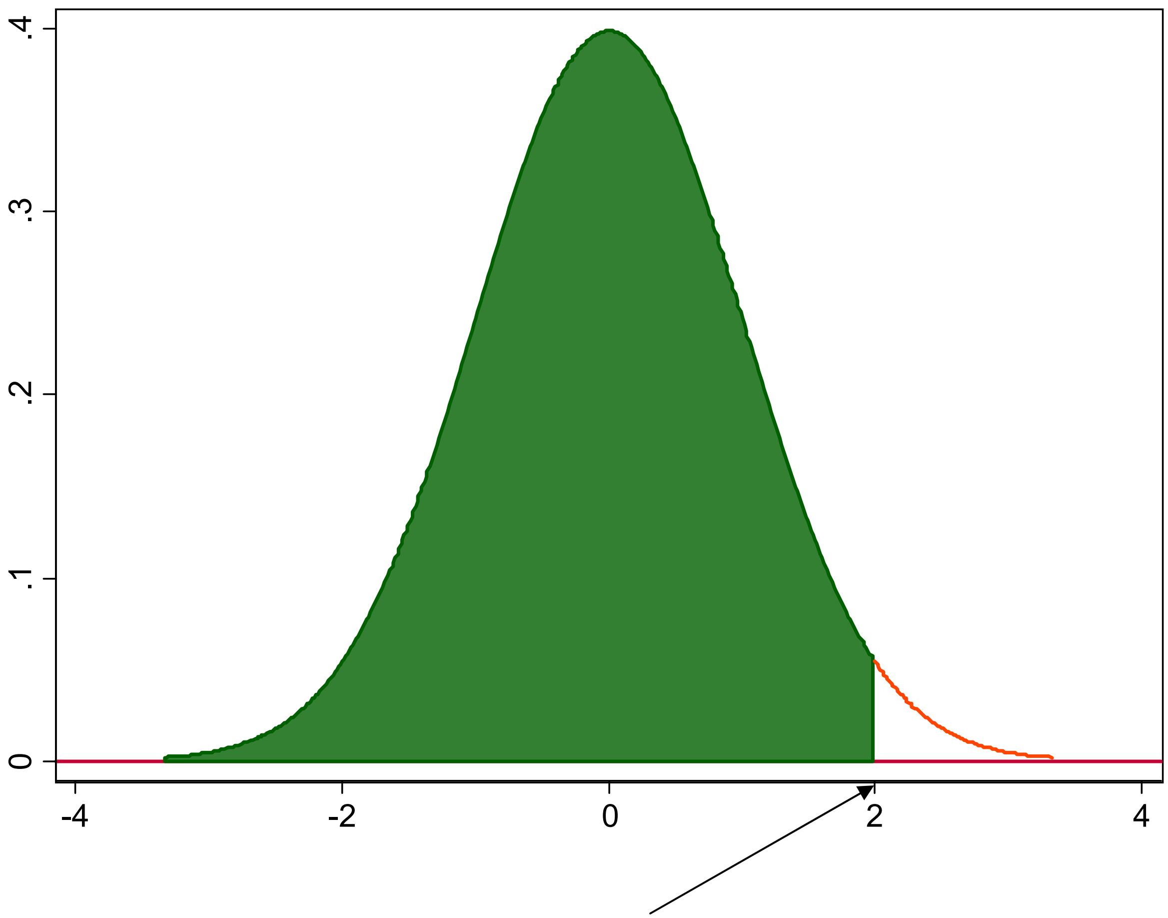
Say that for a given observation, **X**β = -1

Probit Estimation



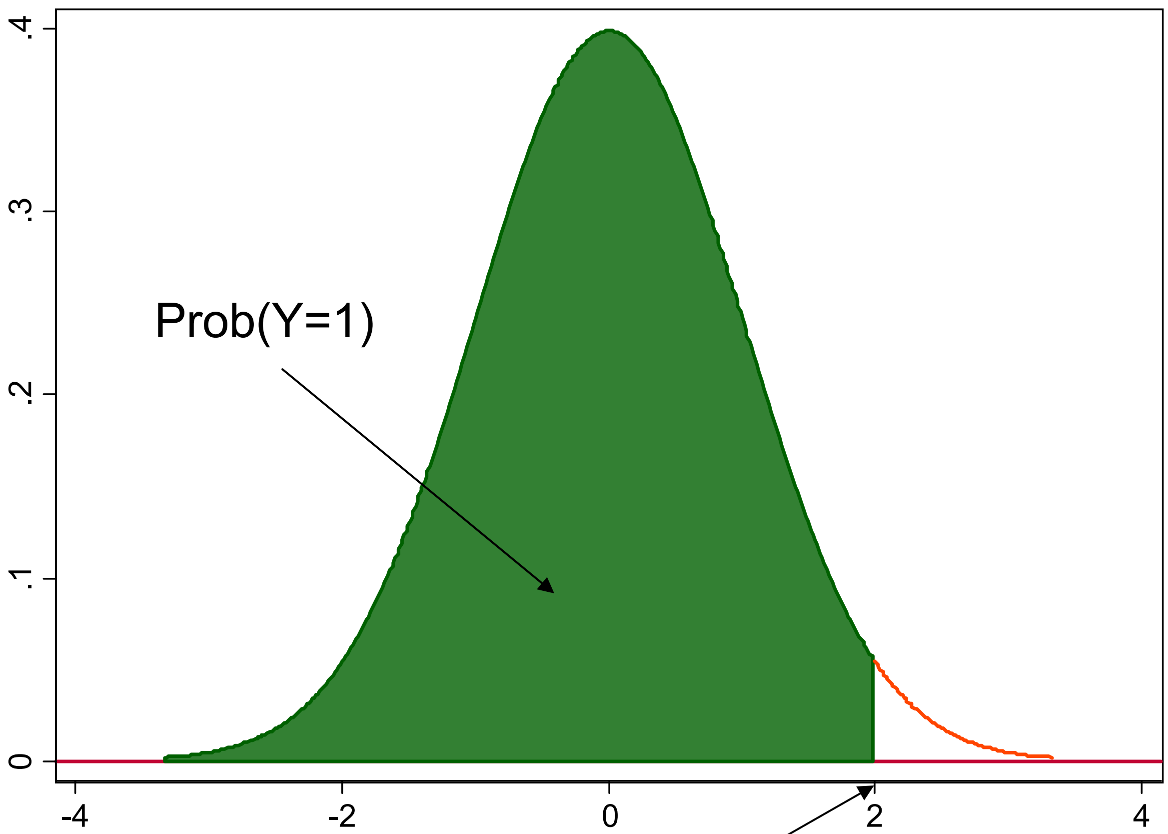
Say that for a given observation, **X**β = 2

Probit Estimation



Say that for a given observation, $\mathbf{X}\beta = 2$

Probit Estimation

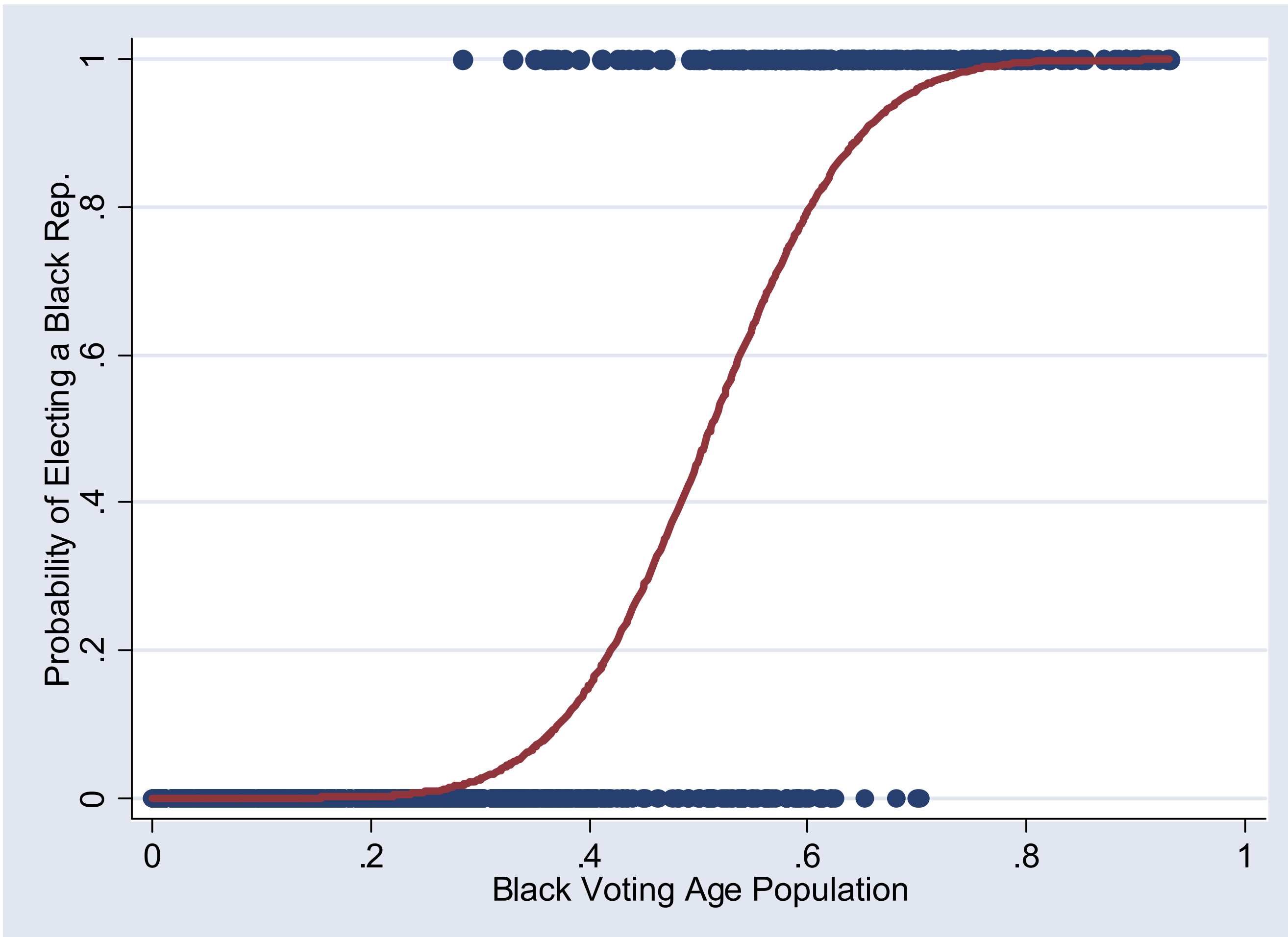


Say that for a given observation, **X**β = 2

Probit Estimation

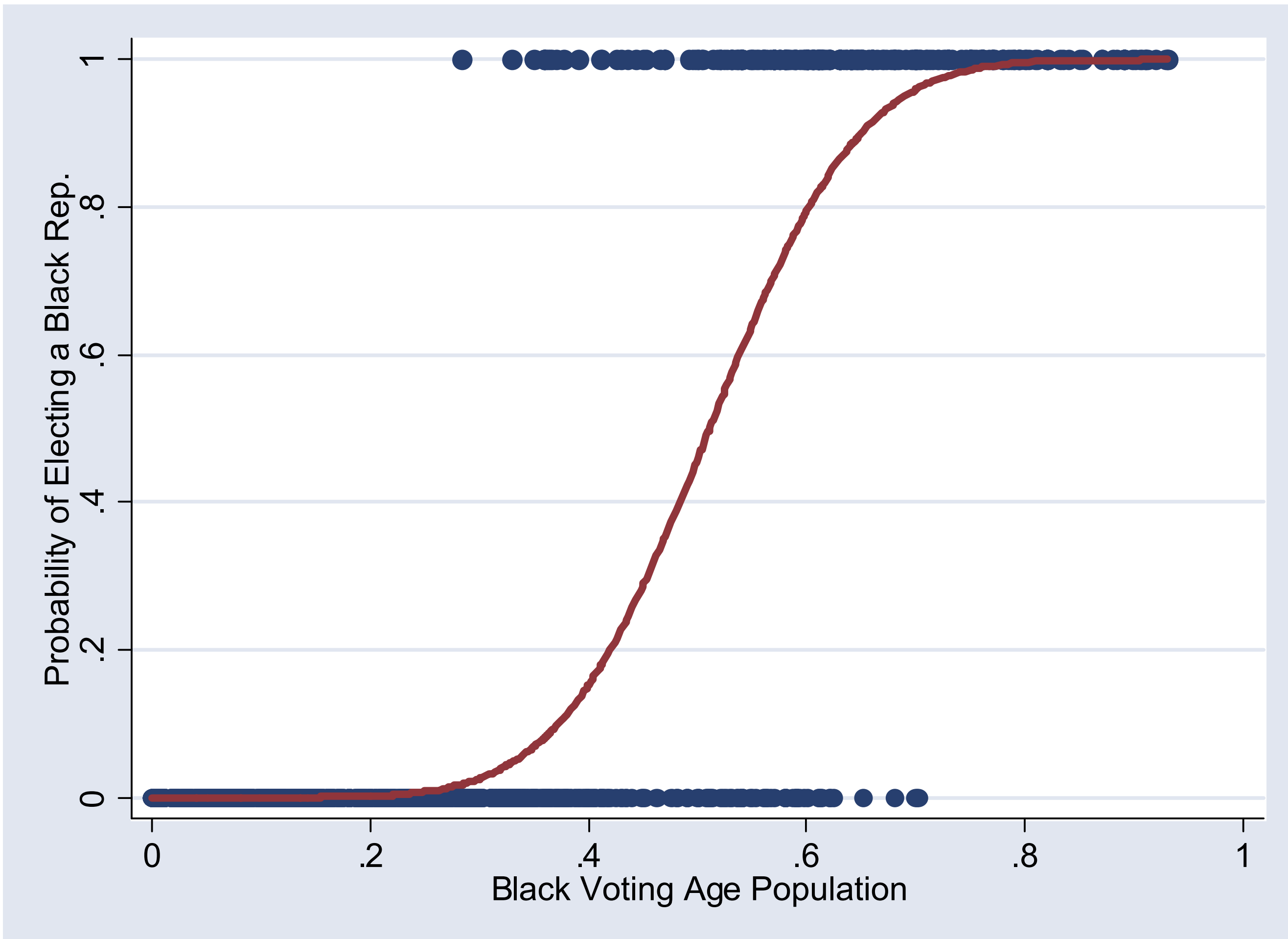
- In a probit model, the value of **X**β is taken to be the z-value of a normal distribution
  - Higher values of **X**β mean that the event is more likely to happen
- Have to be careful about the interpretation of estimation results here
  - A one unit change in X_i leads to a β_i change in the z-score of Y (more on this later…)
- The estimated curve is an S-shaped cumulative normal distribution

Probit Estimation



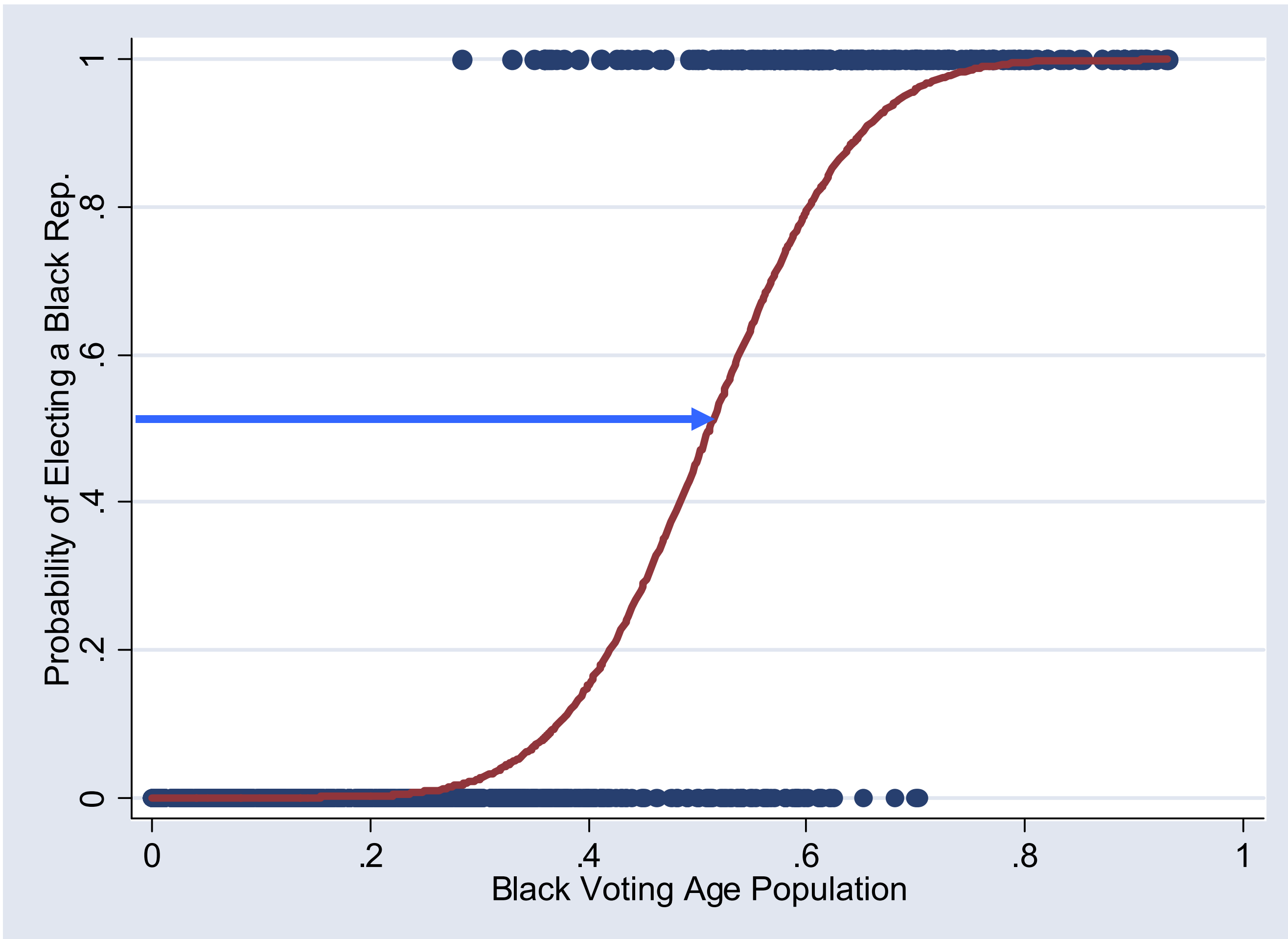
- This fits the data much better than the linear estimation
- Always lies between 0 and 1

Probit Estimation



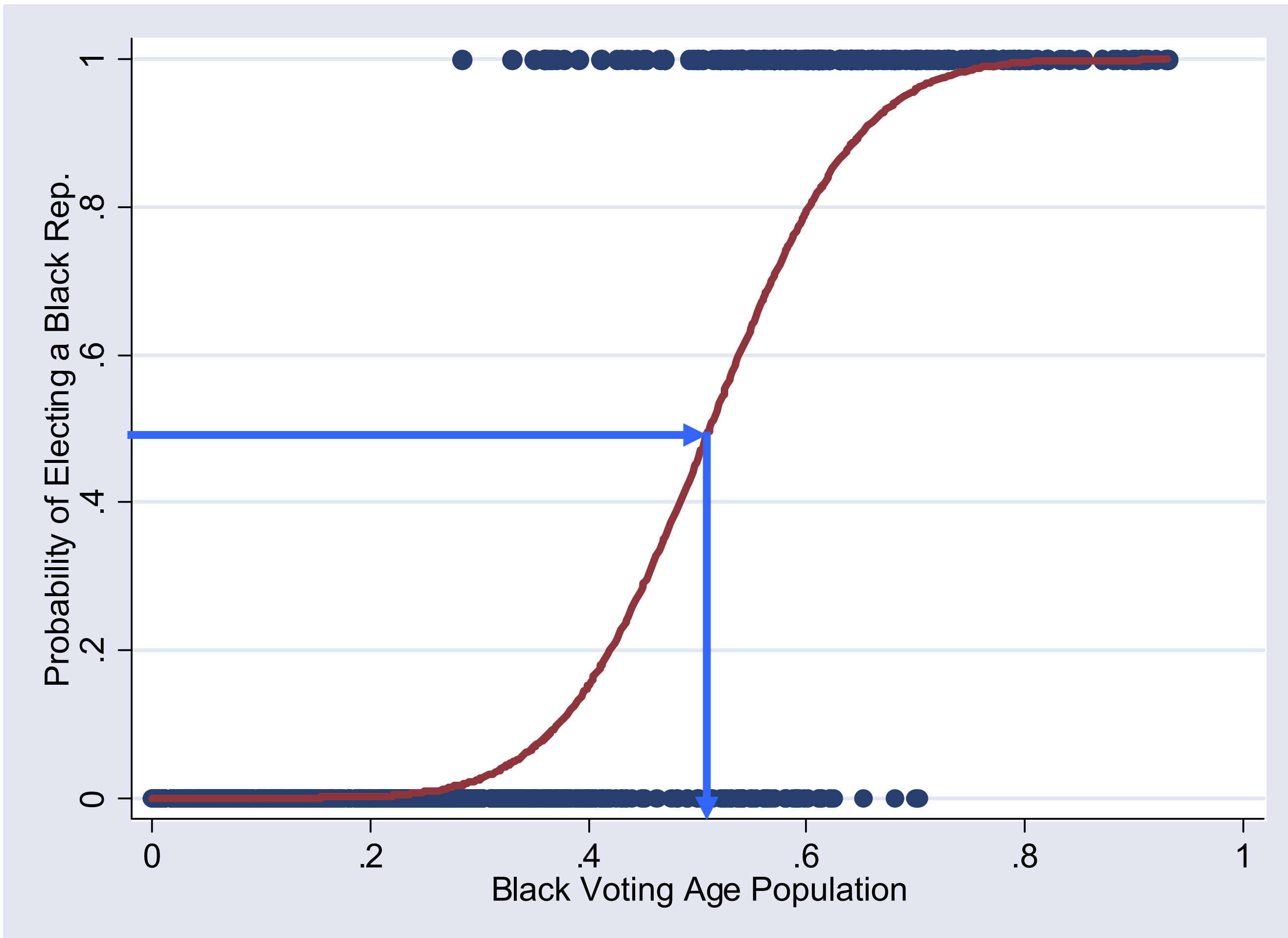
- Can estimate, for instance, the BVAP at which Pr(Y=1) = 50%
- This is the "point of equal opportunity"

Probit Estimation



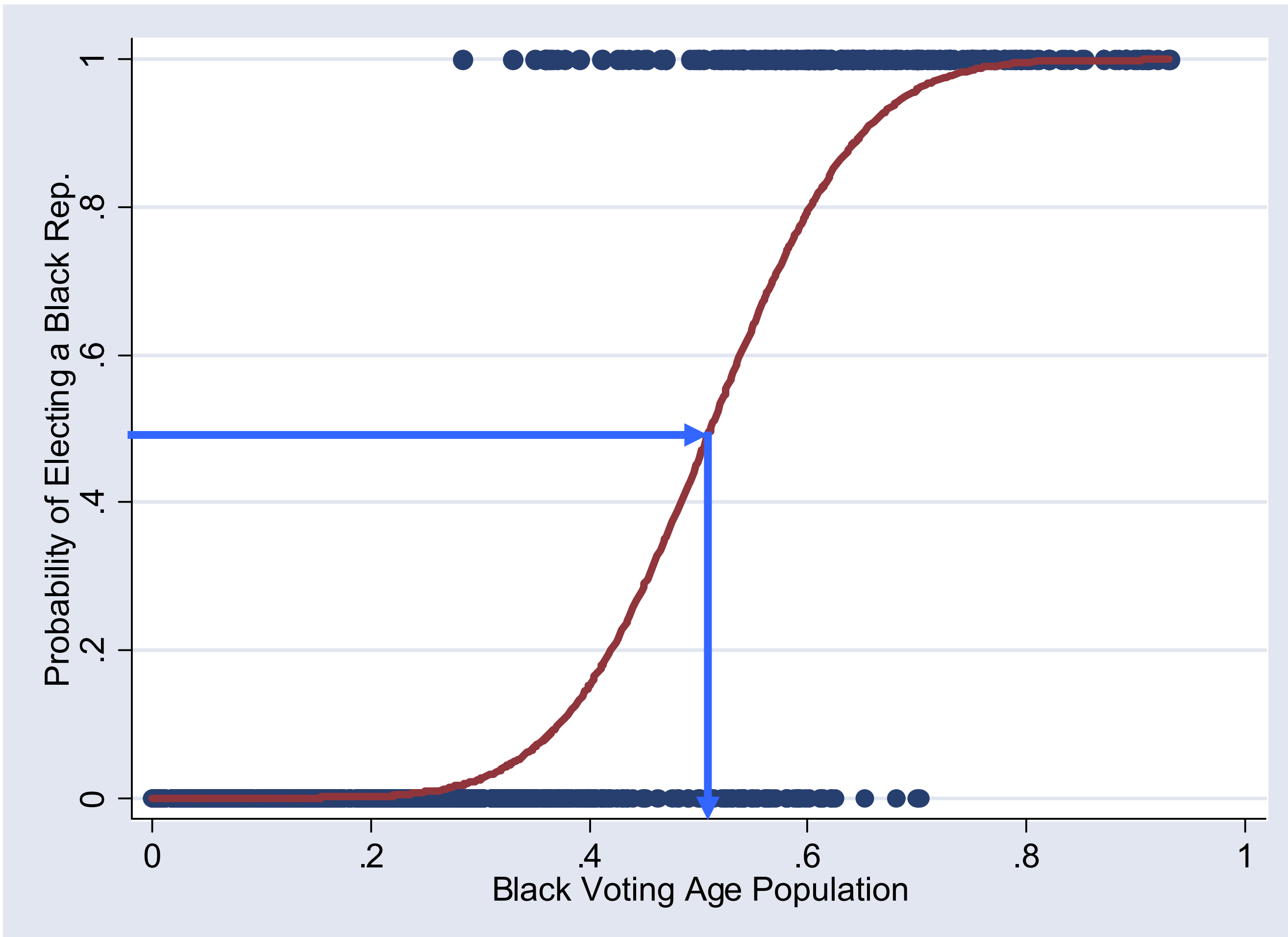
- Can estimate, for instance, the BVAP at which Pr(Y=1) = 50%
- This is the "point of equal opportunity"

Probit Estimation



- Can estimate, for instance, the BVAP at which Pr(Y=1) = 50%
- This is the "point of equal opportunity"

Probit Estimation

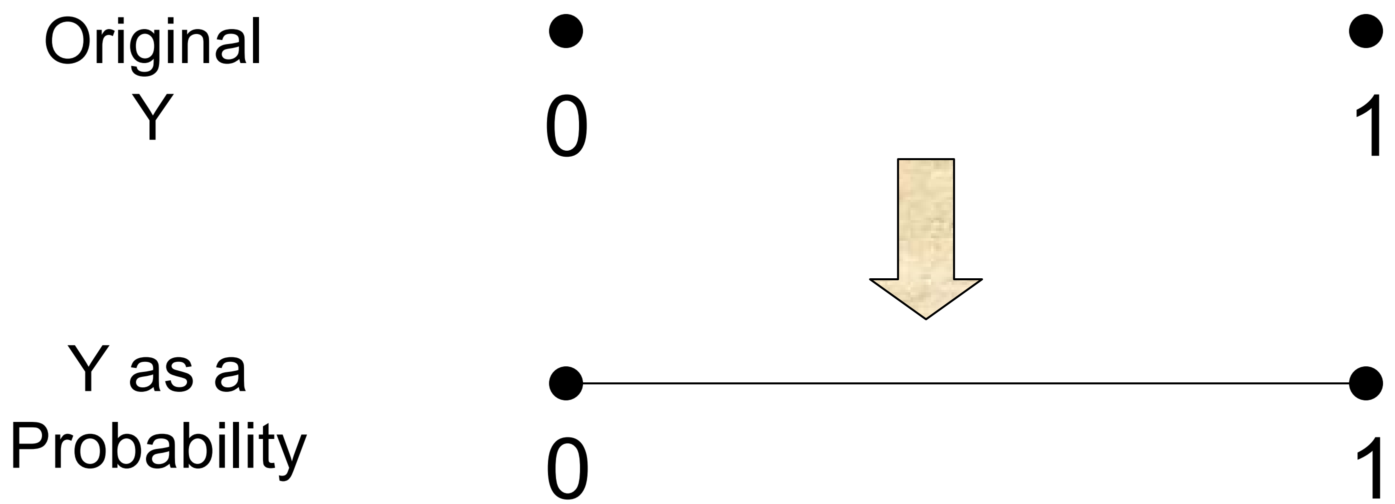


- This occurs at about 48% BVAP

Redefining the Dependent Var.

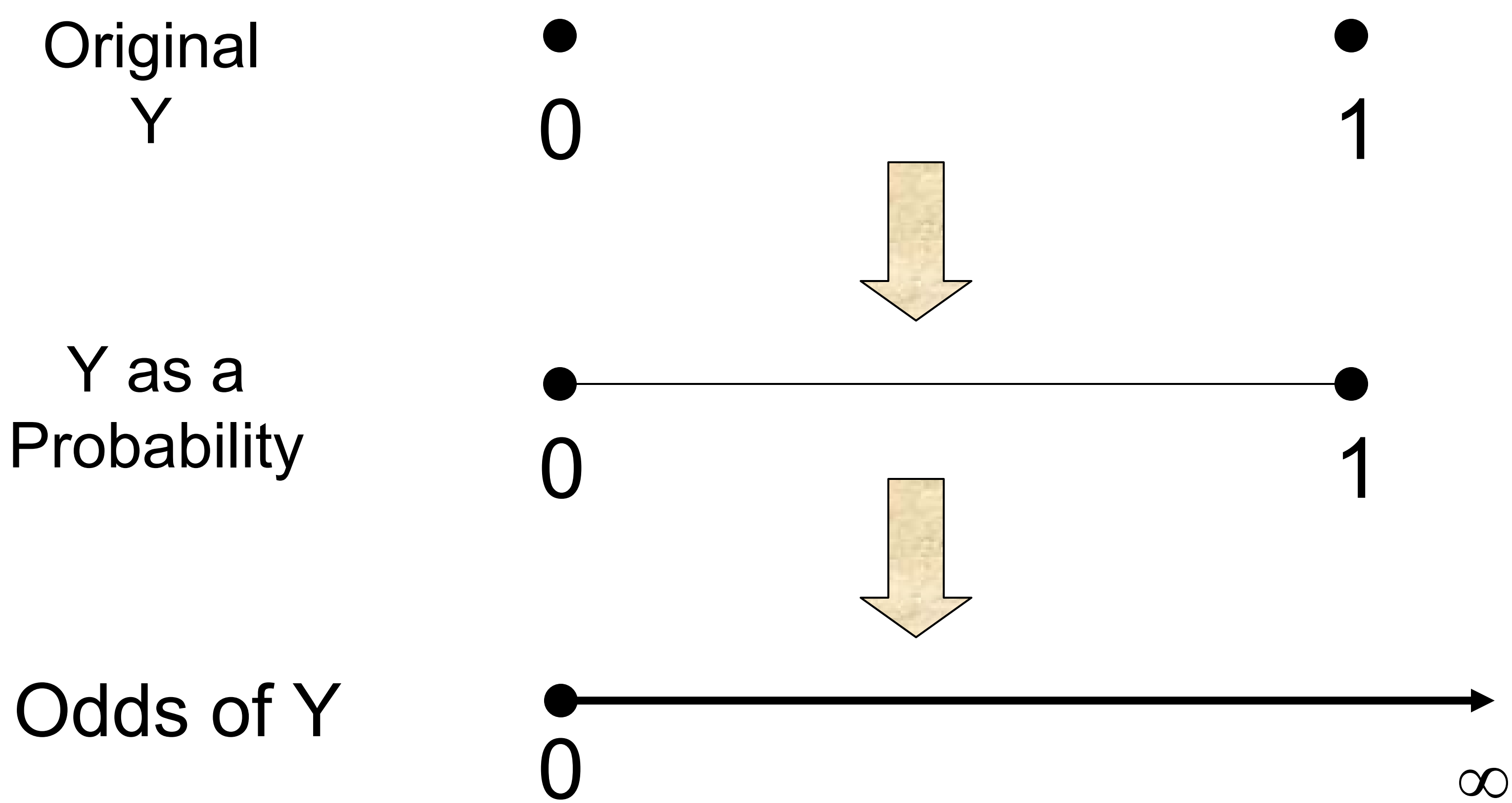
- Let's return to the problem of transforming Y from {0,1} to the real line
- We'll look at an alternative approach based on the odds ratio
- If some event occurs with probability p, then the odds of it happening are $O(p) = p/(1-p)$
  - $p = 0 \rightarrow O(p) = 0$
  - $p = 1/4 \rightarrow O(p) = 1/3$ ("Odds are 1-to-3 against")
  - $p = 1/2 \rightarrow O(p) = 1$ ("Even odds")
  - $p = 3/4 \rightarrow O(p) = 3$ ("Odds are 3-to-1 in favor")
  - $p = 1 \rightarrow O(p) = \infty$

Redefining the Dependent Var.



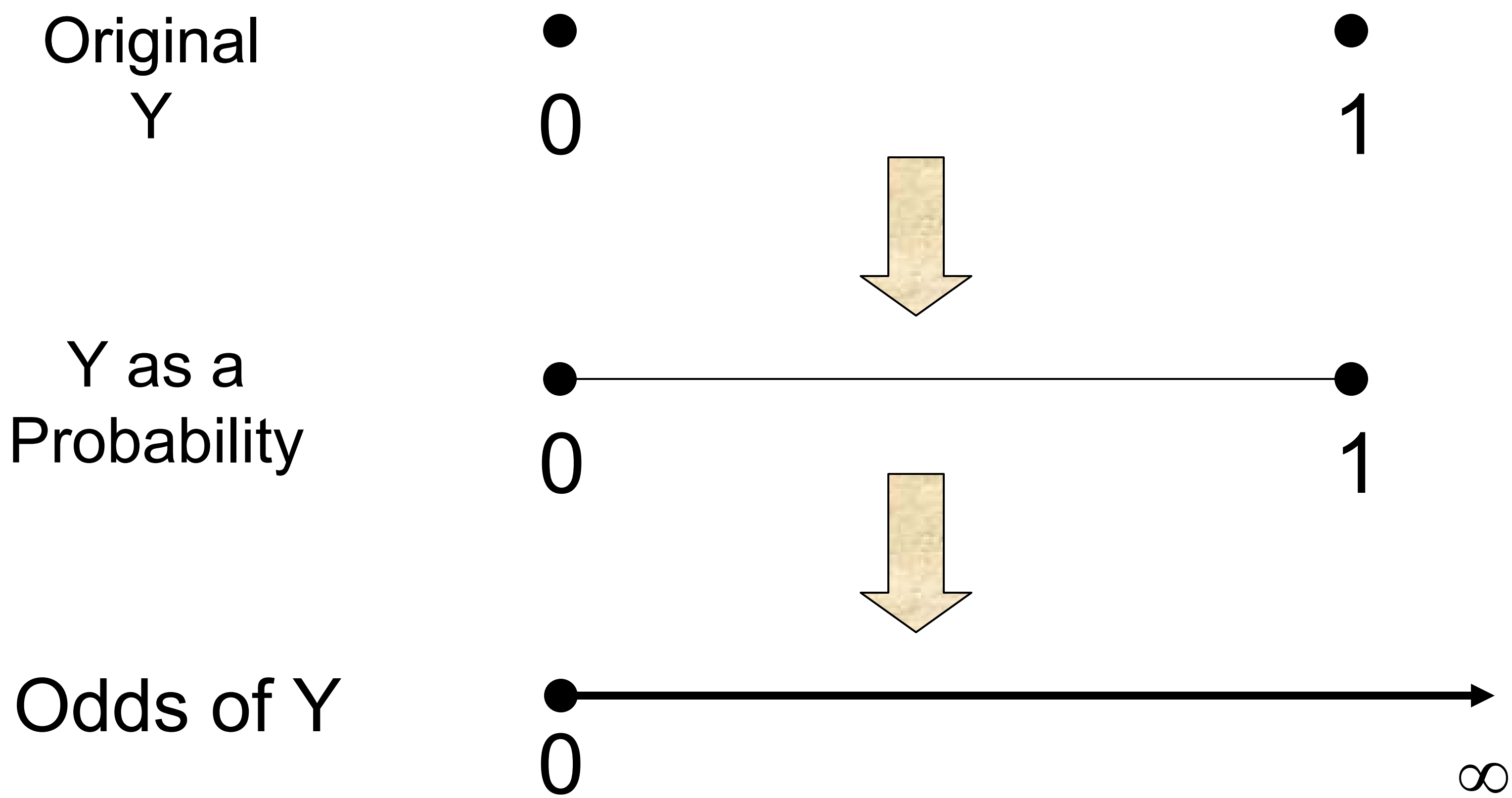
- So taking the odds of Y occuring moves us from the [0,1] interval…

Redefining the Dependent Var.



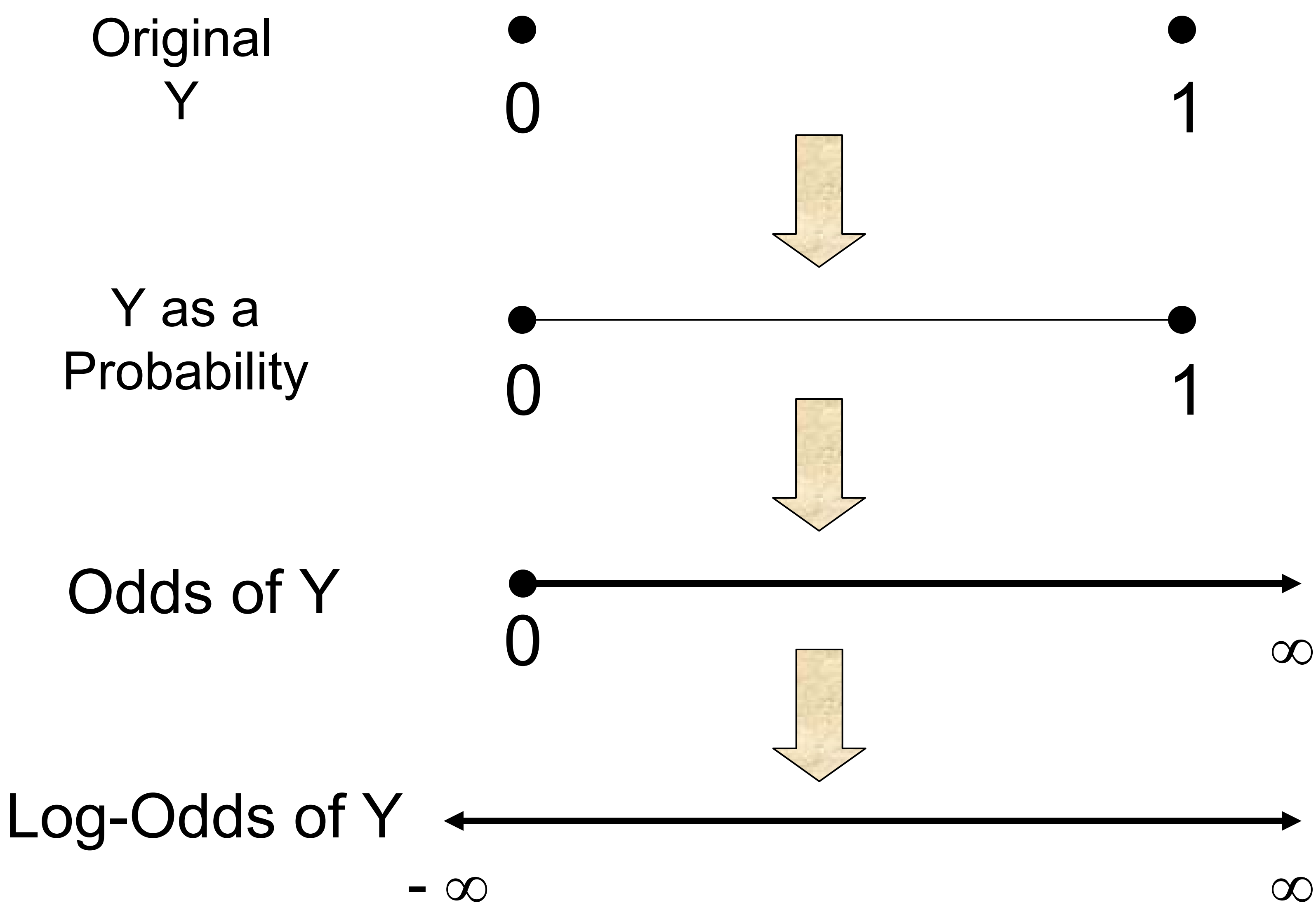
- So taking the odds of Y occuring moves us from the [0,1] interval to the half-line [0, ∞)

Redefining the Dependent Var.



- The odds ratio is always non-negative
- As a final step, then, take the log of the odds ratio

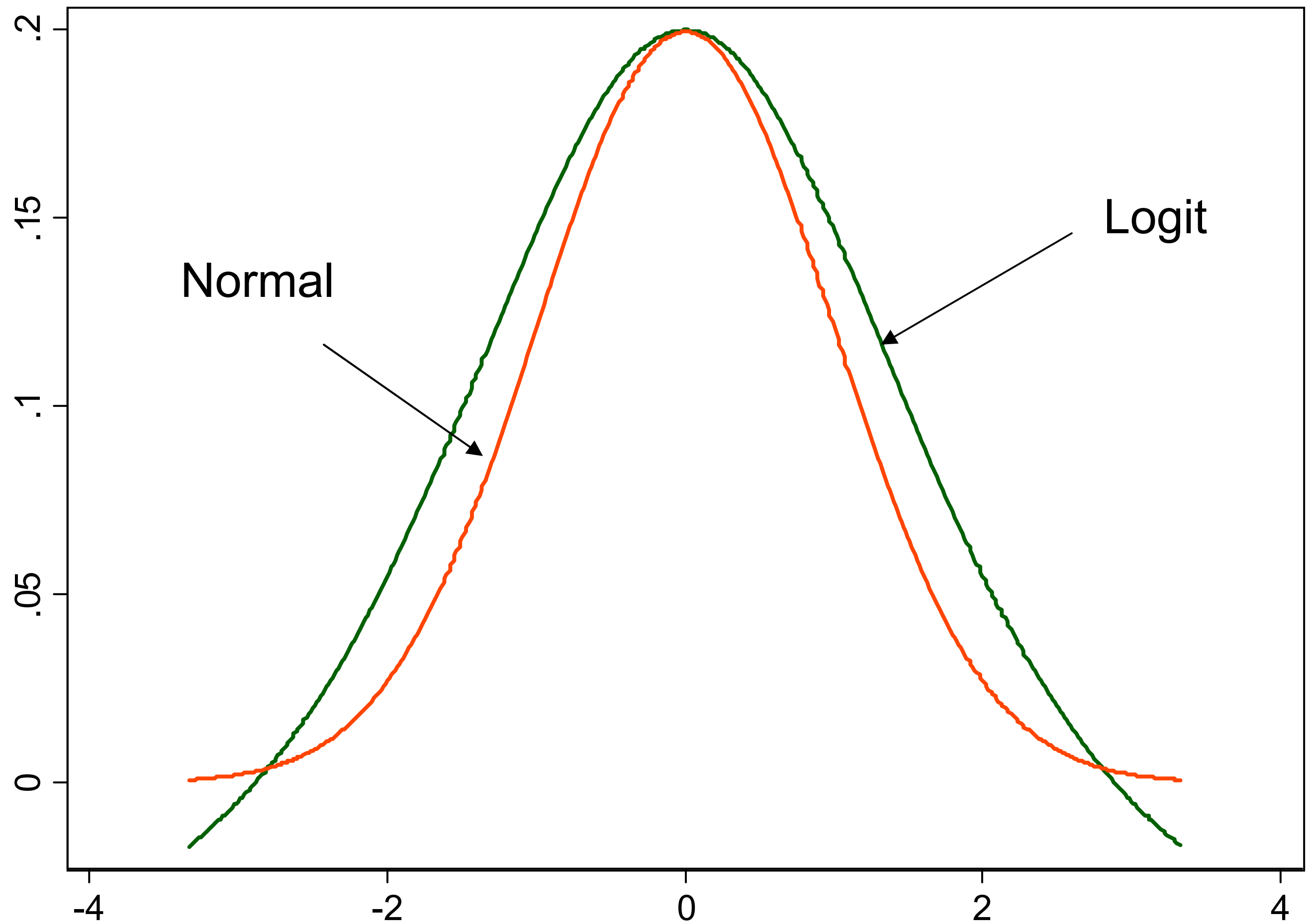
Redefining the Dependent Var.



Logit Function

- This is called the logit function
  - logit(Y) = log[O(Y)] = log[y/(1-y)]
- Why would we want to do this?
  - At first, this was computationally easier than working with normal distributions
  - Now, it still has some nice properties that we'll investigate next time with multinomial dep. vars.
- The density function associated with it is very close to a standard normal distribution

Logit vs. Probit



The logit function is similar, but has thinner tails than the normal distribution

Logit Function

- This translates back to the original Y as:

$$\log\left(\frac{Y}{1-Y}\right) = \mathbf{X}\beta$$

$$\frac{Y}{1-Y} = e^{\mathbf{X}\beta}$$

$$Y = (1-Y)e^{\mathbf{X}\beta}$$

$$Y = e^{\mathbf{X}\beta} - e^{\mathbf{X}\beta}Y$$

$$Y + e^{\mathbf{X}\beta}Y = e^{\mathbf{X}\beta}$$

$$\left(1+e^{\mathbf{X}\beta}\right)Y = e^{\mathbf{X}\beta}$$

$$Y = \frac{e^{\mathbf{X}\beta}}{1+e^{\mathbf{X}\beta}}$$

Latent Variables

- For the rest of the lecture we'll talk in terms of probits, but everything holds for logits too
- One way to state what's going on is to assume that there is a latent variable Y* such that

$$Y^* = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^2\right)$$

Latent Variable Formulation

- For the rest of the lecture we'll talk in terms of probits, but everything holds for logits too
- One way to state what's going on is to assume that there is a latent variable Y* such that

$$Y^{*} = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^{2}\right)$$ ⟵ Normal = Probit

Latent Variables

- For the rest of the lecture we'll talk in terms of probits, but everything holds for logits too
- One way to state what's going on is to assume that there is a latent variable Y* such that

$$Y^* = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^2\right)$$ ← Normal = Probit

- In a linear regression we would observe Y* directly
- In probits, we observe only

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

Latent Variables

- For the rest of the lecture we'll talk in terms of probits, but everything holds for logits too
- One way to state what's going on is to assume that there is a latent variable Y* such that

$$Y^* = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^2\right)$$ ← Normal = Probit

- In a linear regression we would observe Y* directly
- In probits, we observe only

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

These could be any constant. Later we'll set them to ½.

Latent Variables

- This translates to possible values for the error term:

$$y_i^* > 0 \Rightarrow \beta'\mathbf{x}_i + \varepsilon_i > 0 \Rightarrow \varepsilon_i > -\beta'\mathbf{x}_i$$

$$\Pr\left(y_i^* > 0 \mid \mathbf{x}_i\right) = \Pr\left(y_i = 1 \mid \mathbf{x}_i\right) = \Pr\left(\varepsilon_i > -\beta'\mathbf{x}_i\right)$$

$$= \Pr\left(\frac{\varepsilon_i}{\sigma} > \frac{-\beta'\mathbf{x}_i}{\sigma}\right)$$

$$= \Phi\left(\frac{-\beta'\mathbf{x}_i}{\sigma}\right)$$

- Similarly,

$$\Pr\left(y_i = 0 \mid \mathbf{x}_i\right) = 1 - \Phi\left(\frac{-\beta'\mathbf{x}_i}{\sigma}\right)$$

Latent Variables

- Look again at the expression for $\Pr(Y_i=1)$:

$$\Pr\left(y_i = 1 \mid \mathbf{x}_i\right) = \Phi\left(\frac{-\beta'\mathbf{x}_i}{\sigma}\right)$$

- We can't estimate both β and σ, since they enter the equation as a ratio
- So we set $\sigma=1$, making the distribution on ε a standard normal density.
- One (big) question left: how do we actually estimate the values of the b coefficients here?
  - (Other than just issuing the "probit" command in Stata!)

Maximum Likelihood Estimation

- Say we're estimating $Y=\mathbf{X}\beta+\varepsilon$ as a probit
  - And say we're given some trial coefficients β'.
- Then for each observation y_i, we can plug in $\mathbf{x}_i$ and β' to get $Pr(y_i=1)=\Phi(\mathbf{x}_i\,\beta')$.
  - For example, let's say $Pr(y_i=1) = 0.8$
- Then if the actual observation was $y_i=1$, we can say its likelihood (given β') is 0.8
- But if $y_i=0$, then its likelihood was only 0.2
  - And conversely for $Pr(y_i=0)$

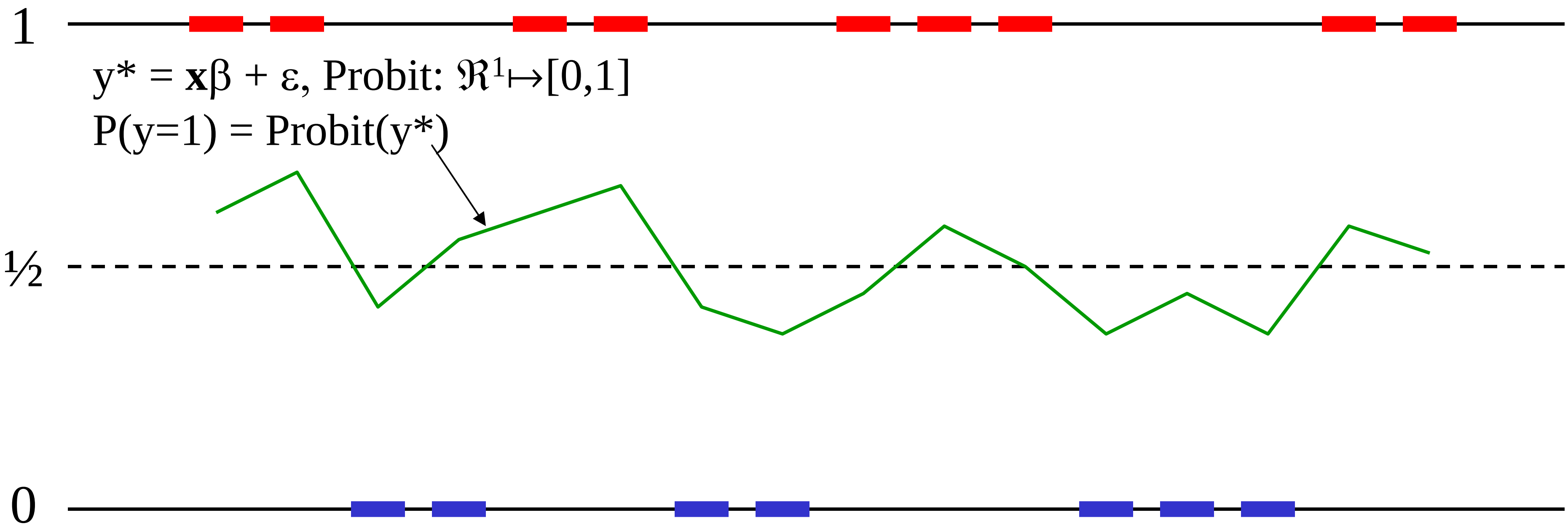
Maximum Likelihood Estimation

- Let $\mathcal{L}(y_i \mid \beta)$ be the likelihood of y_i given β
- For any given trial set of β' coefficients, we can calculate the likelihood of each y_i.
- Then the likelihood of the entire sample is:

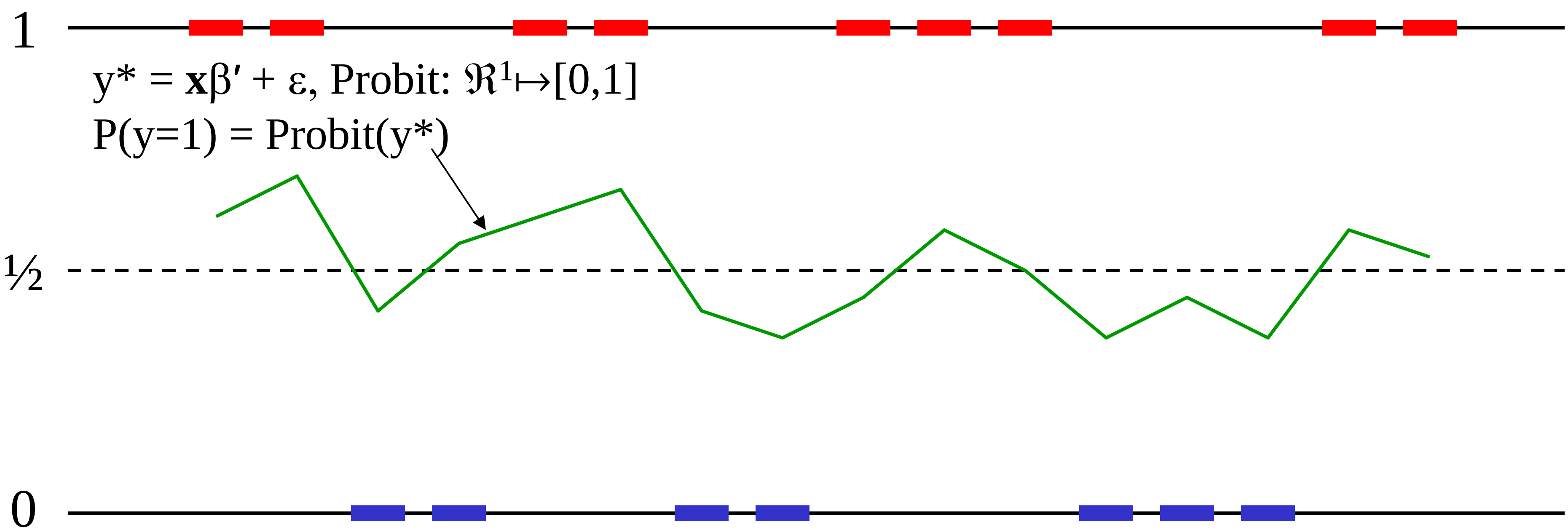
$$\mathcal{L}(y_1)\cdot\mathcal{L}(y_2)\cdot\mathcal{L}(y_3)\cdot\ldots\cdot\mathcal{L}(y_n) = \prod_{i=1}^{n}\mathcal{L}(y_i)$$

- Maximum likelihood estimation finds the β's that maximize this expression.
- Here's the same thing in visual form

Maximum Likelihood Estimation

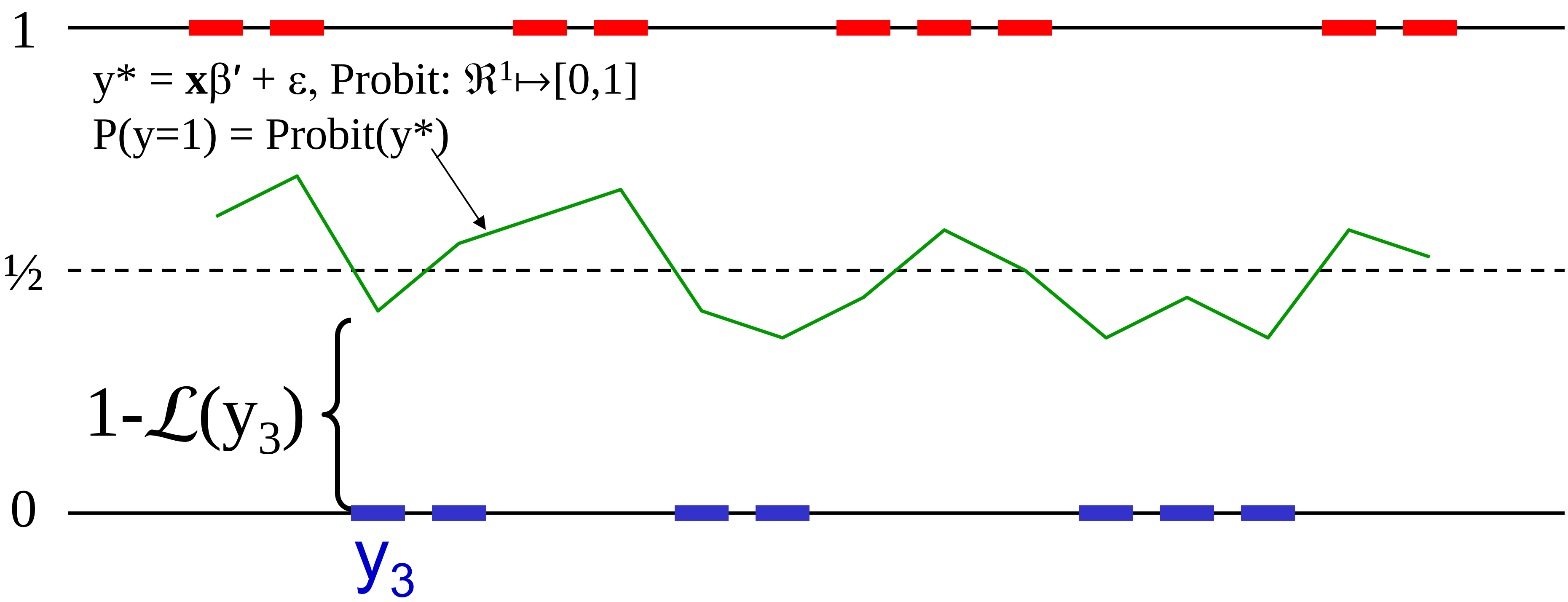


Maximum Likelihood Estimation



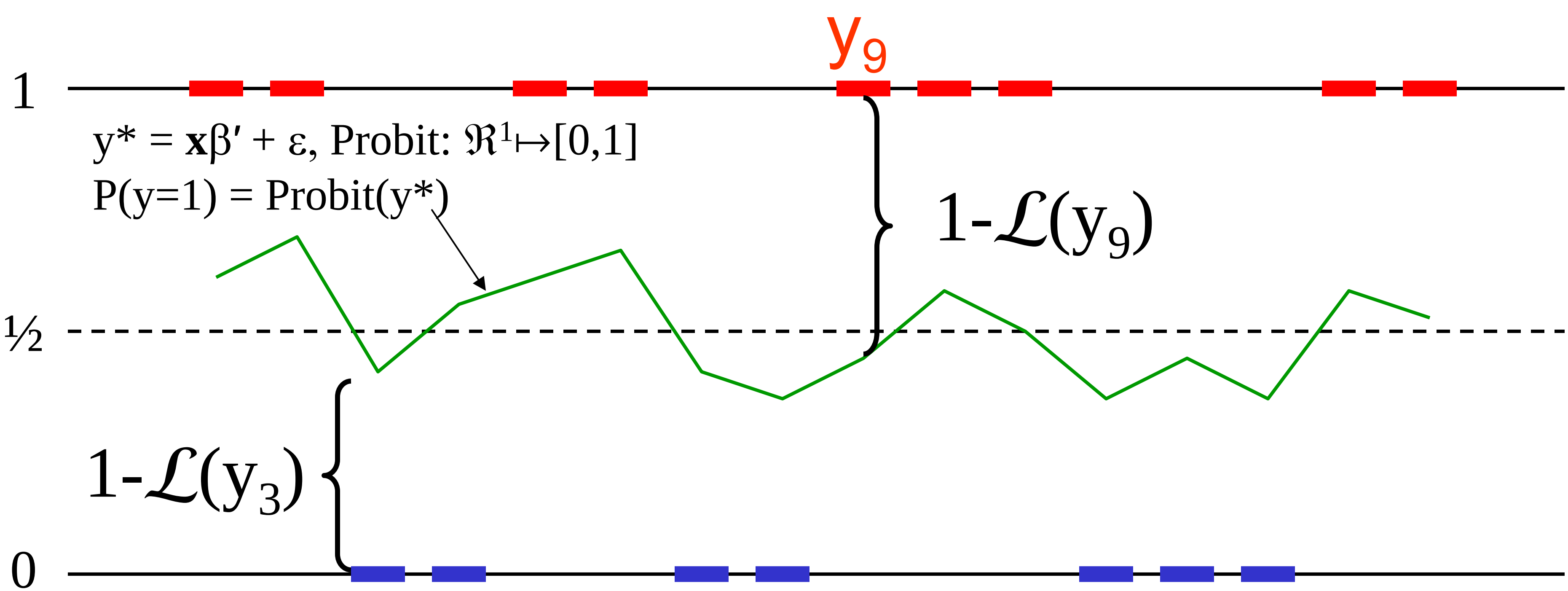
Given estimates β' of β, the distance from y_i to the line P(y=1) is $1-\mathcal{L}(y_i \mid \beta')$

Maximum Likelihood Estimation



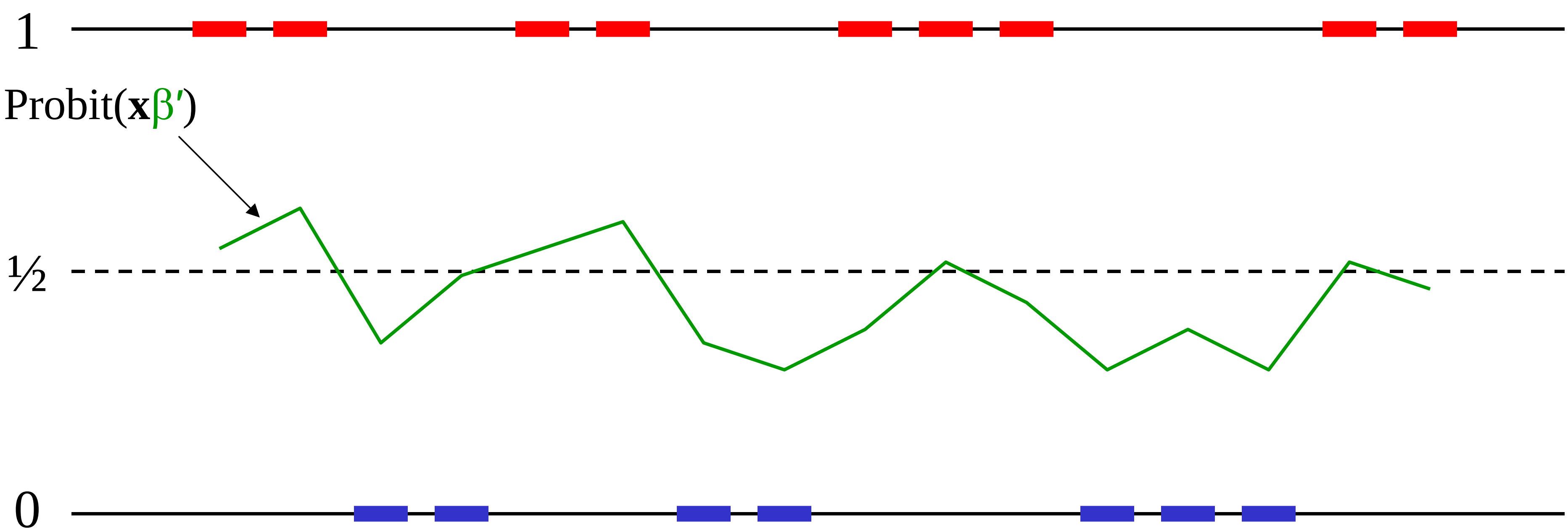
Given estimates β' of β, the distance from y_3 to the line P(y=1) is $1-\mathcal{L}(y_3 \mid \beta')$

Maximum Likelihood Estimation



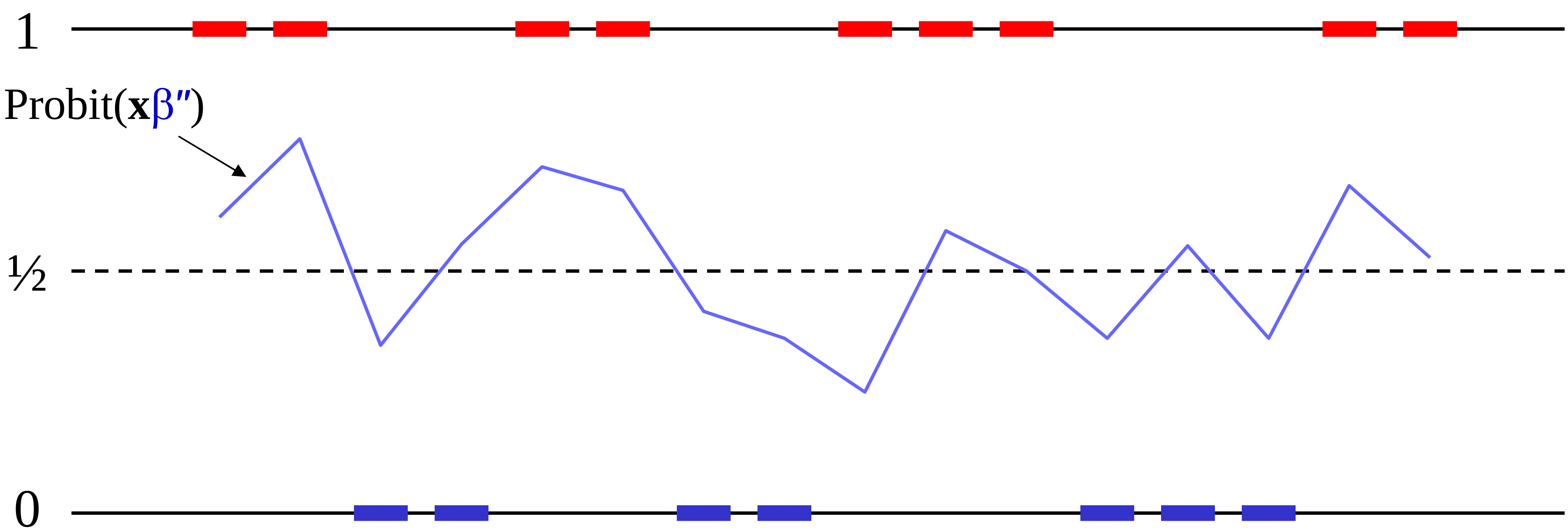
Given estimates β' of β, the distance from y_9 to the line P(y=1) is $1\text{-}\mathcal{L}(y_9 \mid \beta')$

Maximum Likelihood Estimation



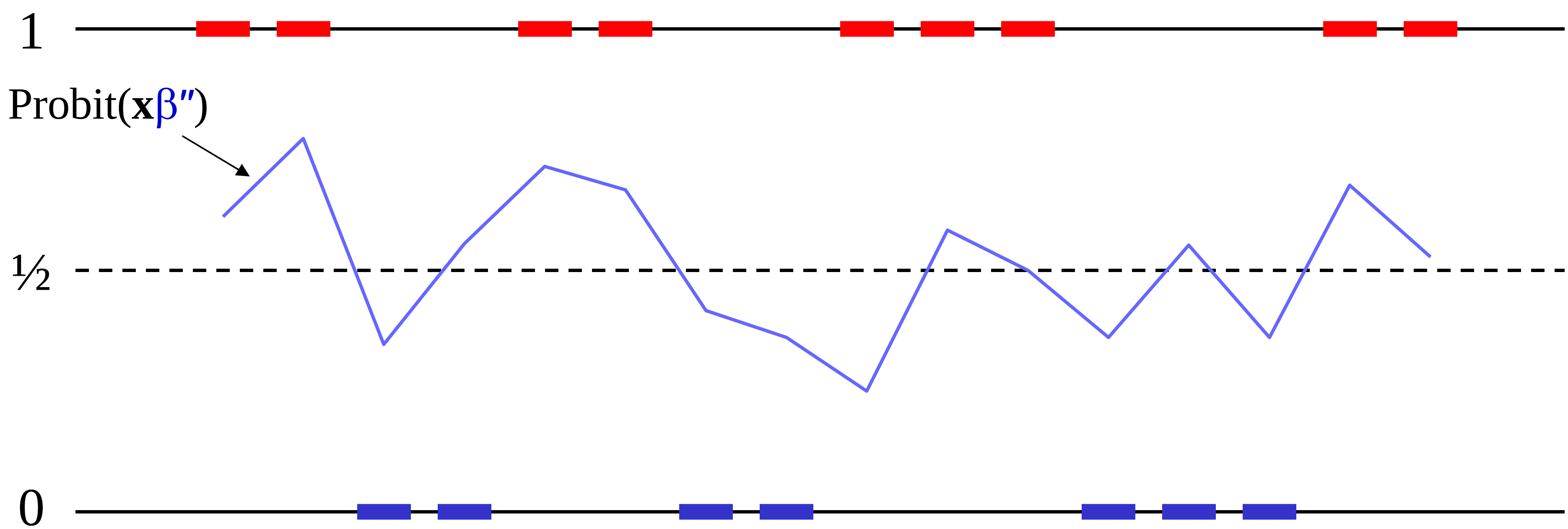
Impact of changing β'…

Maximum Likelihood Estimation



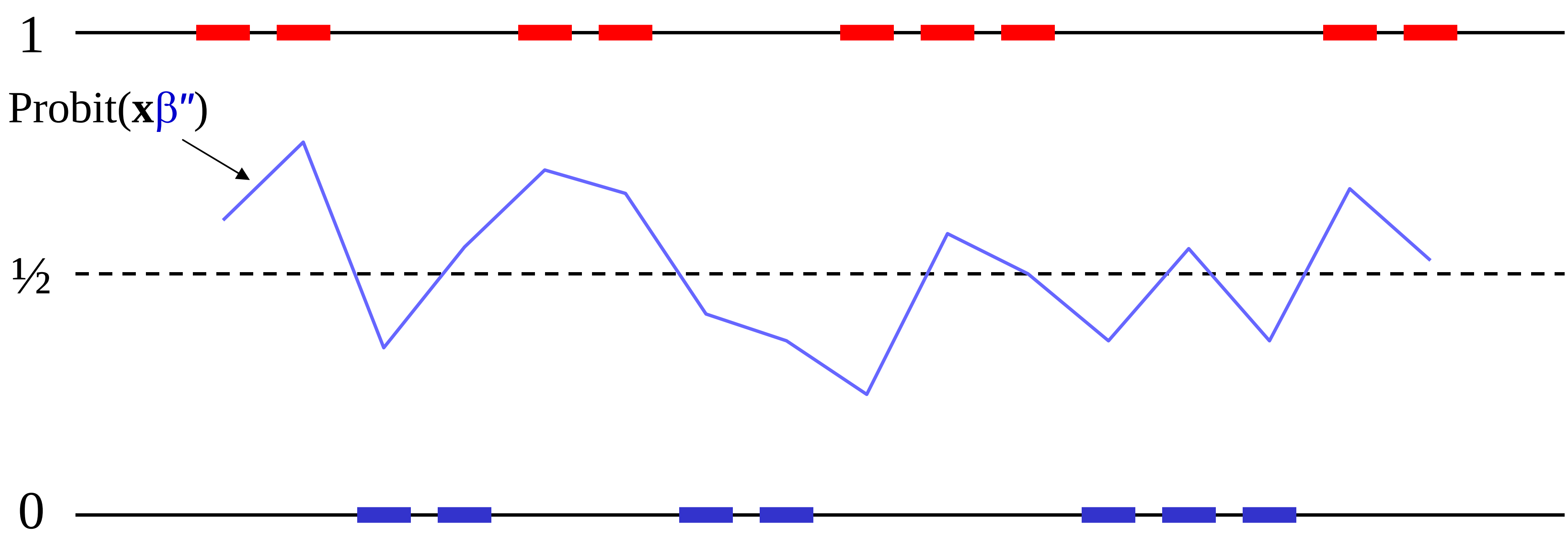
Impact of changing β′ to β″

Maximum Likelihood Estimation



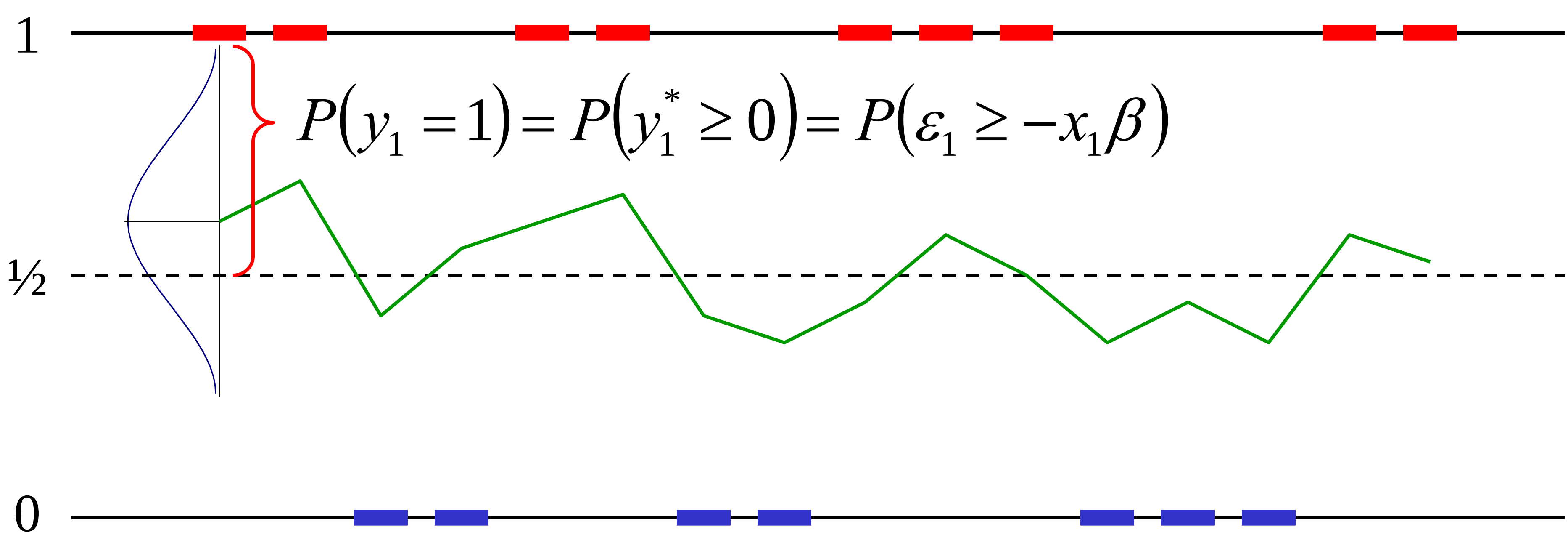
Remember, the object is to maximize the product of the likelihoods $\mathcal{L}(y_i \mid \beta)$

Maximum Likelihood Estimation



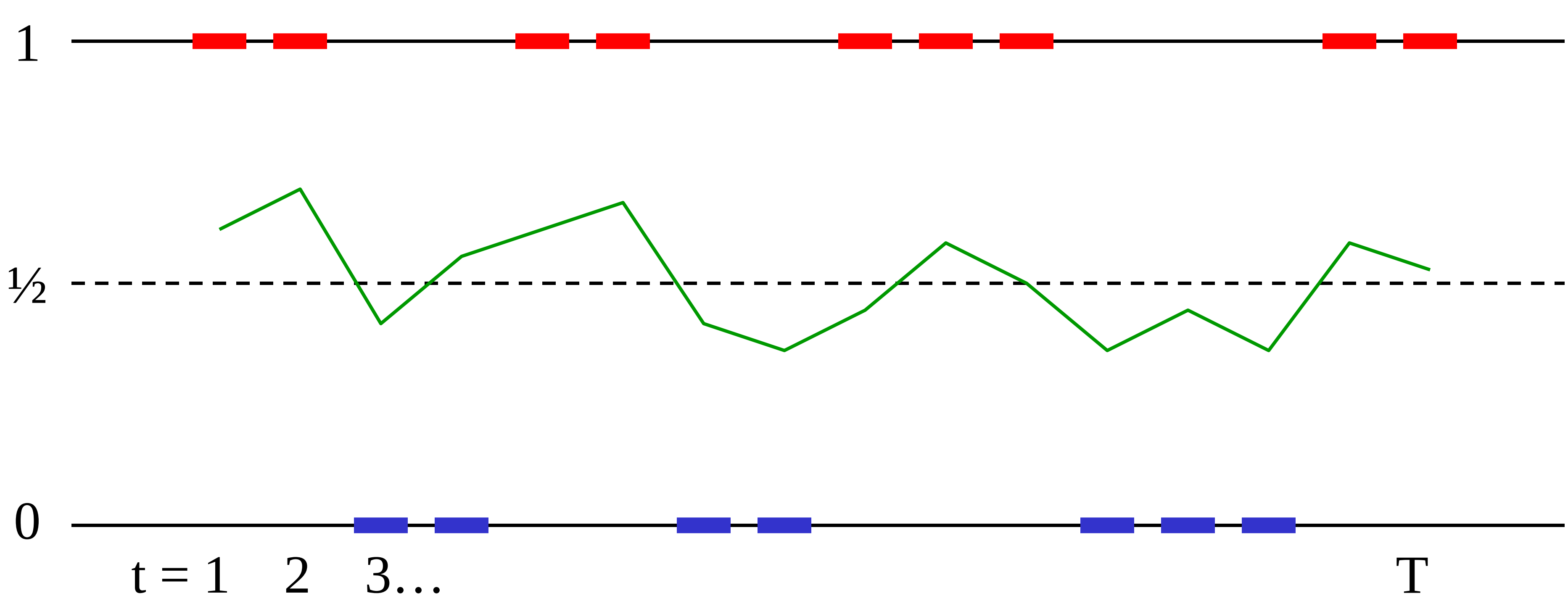
Using β″ may bring regression line closer to some observations, further from others

Maximum Likelihood Estimation



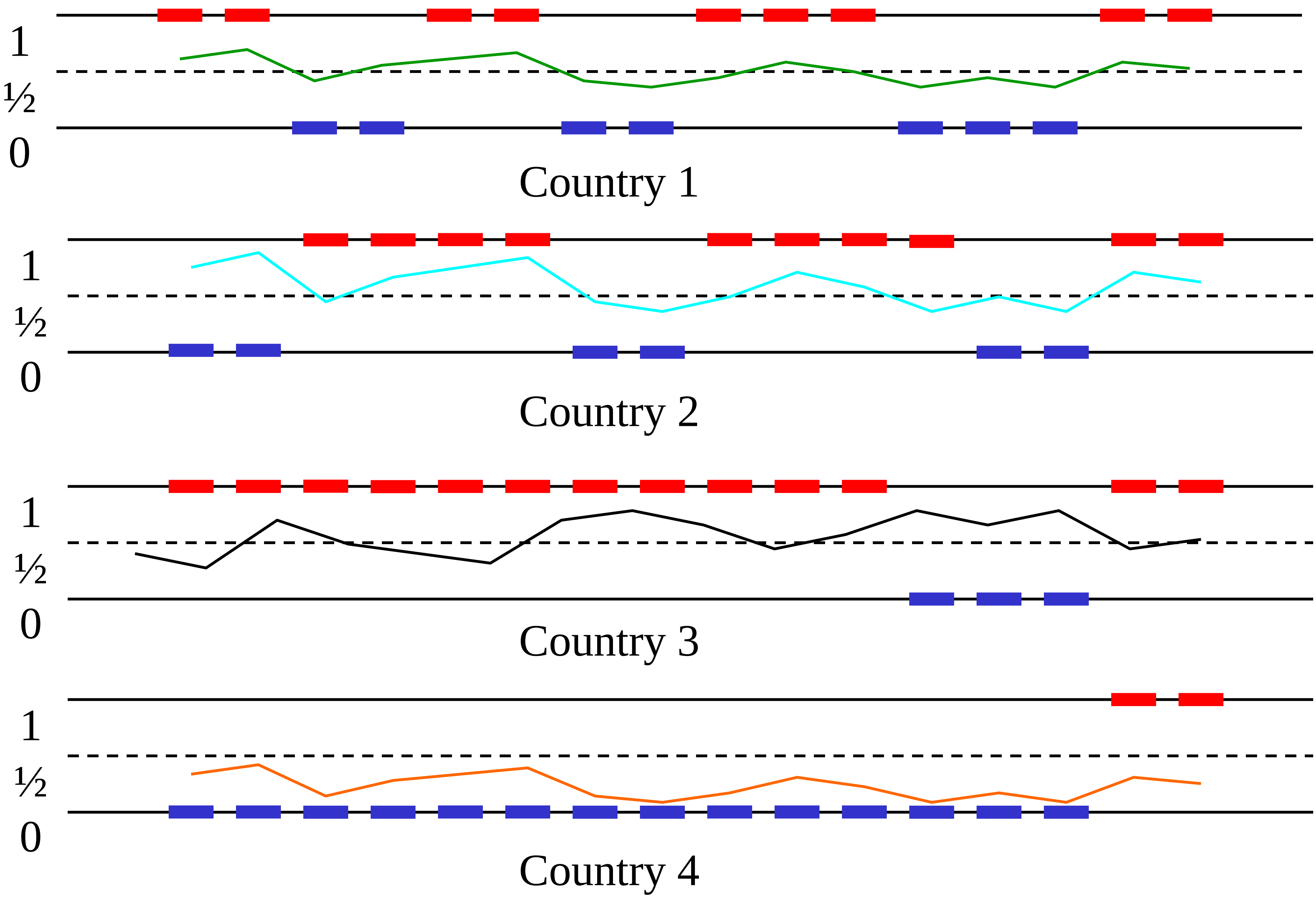
Error Terms for MLE

Maximum Likelihood Estimation



Time Series

Time Series Cross Section



Maximum Likelihood Estimation

- Recall that a likelihood function is:

$$\mathcal{L}(y_1)\cdot\mathcal{L}(y_2)\cdot\mathcal{L}(y_3)\cdot\ldots\cdot\mathcal{L}(y_n)=\prod_{i=1}^{n}\mathcal{L}(y_i)\equiv\mathcal{L}$$

- To maximize this, use the trick of taking the log first
  - Since maximizing the log($\mathcal{L}$) is the same as maximizing $\mathcal{L}$

$$\log(\mathcal{L})=\log\prod_{i=1}^{n}\mathcal{L}(y_i)$$

$$=\sum_{i=1}^{n}\log[\mathcal{L}(y_i)]$$

Maximum Likelihood Estimation

- Let's see how this works on some simple examples
- Take a coin flip, so that Y_i=0 for tails, Y_i=1 for heads
  - Say you toss the coin n times and get p heads
  - Then the proportion of heads is p/n
    - Since Y_i is 1 for heads and 0 for tails, p/n is also the sample mean
  - Intuitively, we'd think that the best estimate of p is p/n
- If the true probability of heads for this coin is ρ, then the likelihood of observation Y_i is:

$$\mathcal{L}(y_i) = \begin{cases} \rho \text{ if } y_i = 1 \\ 1-\rho \text{ if } y_i = 0 \end{cases}$$

$$= \rho^{y_i} \cdot (1-\rho)^{1-y_i}$$

Maximum Likelihood Estimation

- Maximizing the log-likelihood, we get

$$\max_{\rho} \sum_{i=1}^{n} \left[\log \mathcal{L}(y_i | \rho)\right] = \sum_{i=1}^{n} \log\left[\rho^{y_i} \cdot (1-\rho)^{1-y_i}\right]$$

$$= \sum_{i=1}^{n} y_i \log(\rho) + (1-y_i)\log(1-\rho)$$

- To maximize this, take the derivative with respect to ρ

$$\frac{\mathrm{d}\log \mathcal{L}}{\rho} = \frac{\mathrm{d}\left[\sum_{i=1}^{n} y_i \log(\rho) + (1-y_i)\log(1-\rho)\right]}{\rho}$$

$$= \sum_{i=1}^{n} y_i \frac{1}{\rho} - (1-y_i)\frac{1}{1-\rho}$$

Maximum Likelihood Estimation

- Finally, set this derivative to 0 and solve for ρ

$$\sum_{i=1}^{n}\left[\frac{y_i}{\rho}-\frac{(1-y_i)}{1-\rho}\right]=0$$

$$\frac{\sum_{i=1}^{n}\left[y_i(1-\rho)-(1-y_i)\rho\right]}{\rho(1-\rho)}=0$$

$$\sum_{i=1}^{n}\left[y_i-y_i\rho-\rho+(1-y_i)\rho\right]=0$$

$$n\rho=\sum_{i=1}^{n}y_i$$

$$\rho=\frac{\sum_{i=1}^{n}y_i}{n}$$

Maximum Likelihood Estimation

- Finally, set this derivative to 0 and solve for ρ

$$\sum_{i=1}^{n}\left[\frac{y_i}{\rho}-\frac{(1-y_i)}{1-\rho}\right]=0$$

$$\frac{\sum_{i=1}^{n}\left[y_i(1-\rho)-(1-y_i)\rho\right]}{\rho(1-\rho)}=0$$

$$\sum_{i=1}^{n}\left[y_i-y_i\rho-\rho+(1-y_i)\rho\right]=0$$

$$n\rho=\sum_{i=1}^{n}y_i$$

$$\rho=\frac{\sum_{i=1}^{n}y_i}{n}$$

Magically, the value of ρ that maximizes the likelihood function is the sample mean, just as we thought.

Maximum Likelihood Estimation

- Can do the same exercise for OLS regression
  - The set of β coefficients that maximize the likelihood would then minimize the sum of squared residuals, as before
- This works for logit/probit as well
- In fact, it works for <u>any</u> estimation equation
  - Just look at the likelihood function $\mathcal{L}$ you're trying to maximize and the parameters β you can change
  - Then search for the values of β that maximize $\mathcal{L}$
  - (We'll skip the details of how this is done.)
- Maximizing $\mathcal{L}$ can be computationally intense, but with today's computers it's usually not a big problem

Maximum Likelihood Estimation

- This is what Stata does when you run a probit:

```
. probit black bvap

Iteration 0:   log likelihood = -735.15352
Iteration 1:   log likelihood = -292.89815
Iteration 2:   log likelihood = -221.90782
Iteration 3:   log likelihood = -202.46671
Iteration 4:   log likelihood = -198.94506
Iteration 5:   log likelihood = -198.78048
Iteration 6:   log likelihood = -198.78004

Probit estimates                                  Number of obs   =       1507
                                                  LR chi2(1)      =    1072.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -198.78004                       Pseudo R2       =     0.7296

------------------------------------------------------------------------------
       black |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bvap |   0.092316   .5446756    16.95   0.000     0.081641    0.102992
       _cons |  -0.047147   0.027917   -16.89   0.000    -0.052619   -0.041676
------------------------------------------------------------------------------
```

Maximum Likelihood Estimation

- This is what Stata does when you run a probit:

```
. probit black bvap

Iteration 0:   log likelihood = -735.15352
Iteration 1:   log likelihood = -292.89815
Iteration 2:   log likelihood = -221.90782
Iteration 3:   log likelihood = -202.46671
Iteration 4:   log likelihood = -198.94506
Iteration 5:   log likelihood = -198.78048
Iteration 6:   log likelihood = -198.78004

Probit estimates                                  Number of obs   =       1507
                                                  LR chi2(1)      =    1072.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -198.78004                       Pseudo R2       =     0.7296

------------------------------------------------------------------------------
       black |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bvap |   0.092316   .5446756    16.95   0.000     0.081641    0.102992
       _cons |  -0.047147   0.027917   -16.89   0.000    -0.052619   -0.041676
------------------------------------------------------------------------------
```

Maximizing the log-likelihood function!

Maximum Likelihood Estimation

- This is what Stata does when you run a probit:

```
. probit black bvap

Iteration 0:   log likelihood = -735.15352
Iteration 1:   log likelihood = -292.89815
Iteration 2:   log likelihood = -221.90782
Iteration 3:   log likelihood = -202.46671
Iteration 4:   log likelihood = -198.94506
Iteration 5:   log likelihood = -198.78048
Iteration 6:   log likelihood = -198.78004

Probit estimates                                  Number of obs   =       1507
                                                  LR chi2(1)      =    1072.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -198.78004                       Pseudo R2       =     0.7296

------------------------------------------------------------------------------
       black |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bvap |   0.092316   .5446756    16.95    0.000     0.081641    0.102992
       _cons |  -0.047147   0.027917   -16.89    0.000    -0.052619   -0.041676
------------------------------------------------------------------------------
```

Maximizing the log-likelihood function!

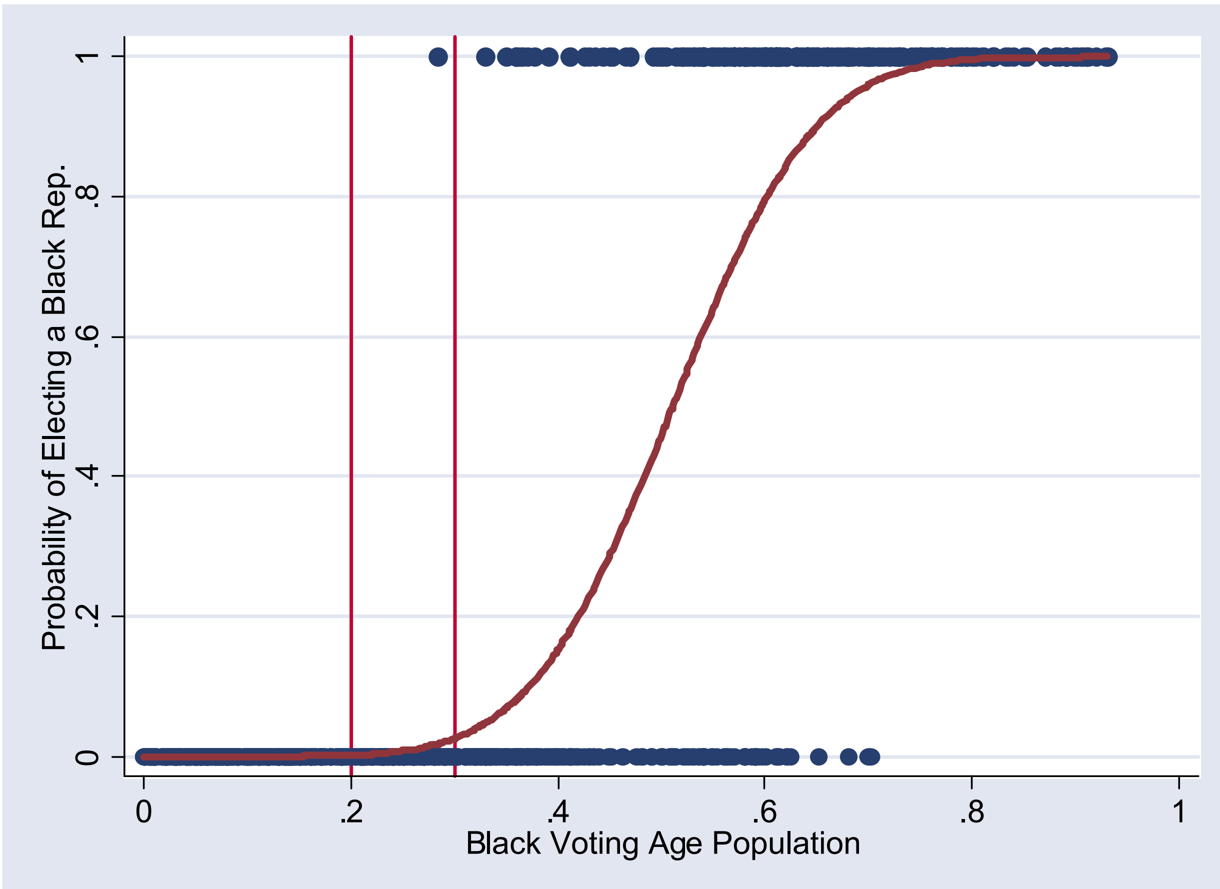
Coefficients are significant

Marginal Effects in Probit

- In linear regression, if the coefficient on x is β, then a 1-unit increase in x increases Y by β.
- But what exactly does it mean in probit that the coefficient on BVAP is 0.0923 and significant?
  - It means that a 1% increase in BVAP will raise the z-score of Pr(Y=1) by 0.0923.
  - And this coefficient is different from 0 at the 5% level.
- So raising BVAP has a constant effect on Y′.
- But this doesn’t translate into a constant effect on the original Y.
  - This depends on your starting point.

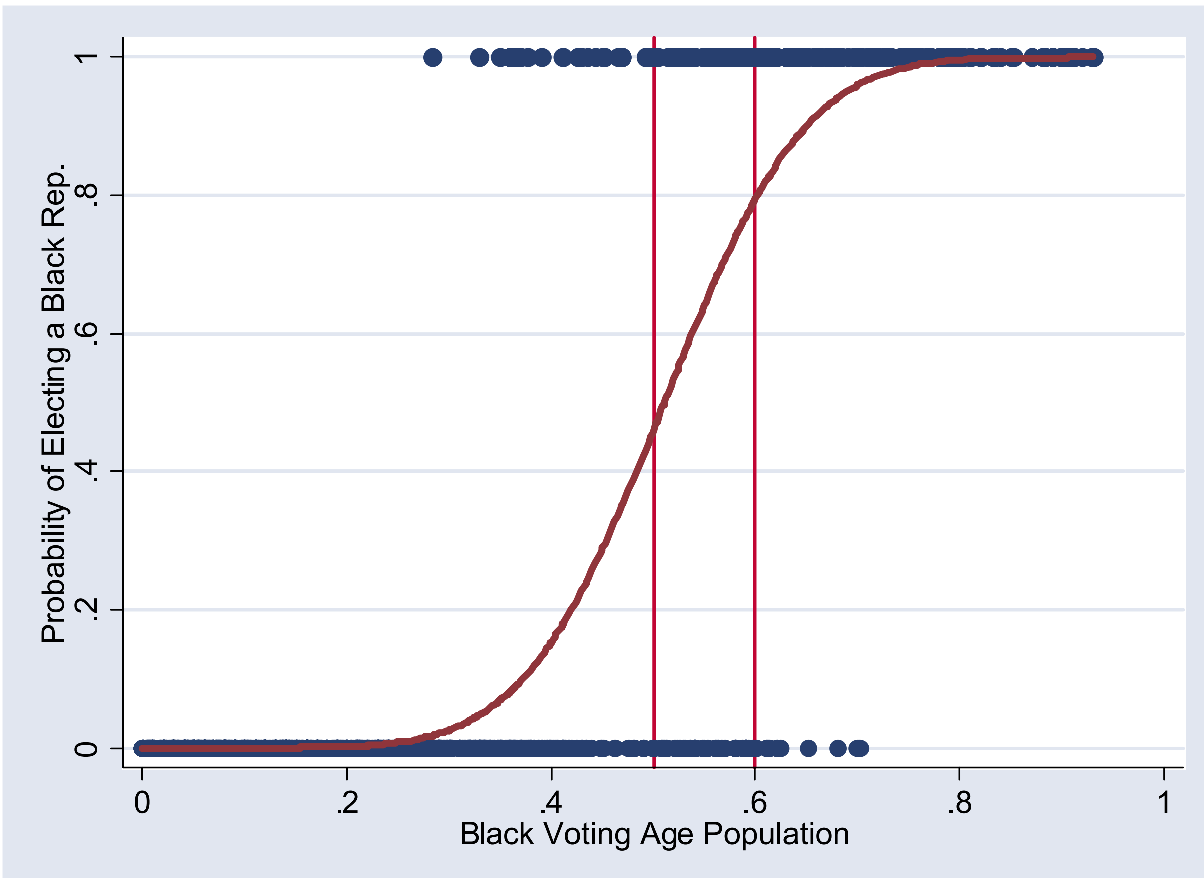
Marginal Effects in Probit

- For instance, raising BVAP from .2 to .3 has little appreciable impact on Pr(Black Elected)



Marginal Effects in Probit

- But increasing BVAP from .5 to .6 <u>does</u> have a big impact on the probability



Marginal Effects in Probit

- So lesson 1 is that the marginal impact of changing a variable is not constant.
- Another way of saying the same thing is that in the linear model

$$Y = \beta_0 + \beta_1 x_1 + \beta_1 x_1 + \ldots + \beta_n x_n, \text{ so}$$

$$\frac{\partial Y}{\partial x_i} = \beta_i$$

- In the probit model

$$Y = \Phi\left(\beta_0 + \beta_1 x_1 + \beta_1 x_1 + \ldots + \beta_n x_n\right), \text{ so}$$

$$\frac{\partial Y}{\partial x_i} = \beta_i \phi\left(\beta_0 + \beta_1 x_1 + \beta_1 x_1 + \ldots + \beta_n x_n\right)$$

Marginal Effects in Probit

- This expression depends on not just β_i, but on the value of x_i and <u>all other variables</u> in the equation
- So to even calculate the impact of x_i on Y you have to choose values for all other variables x_j.
  - Typical options are to set all variables to their means or their medians
- Another approach is to fix the x_j and let x_i vary from its minimum to maximum values
  - Then you can plot how the marginal effect of x_i changes across its observed range of values

Example: Vote Choice

- Model voting for/against incumbent as

$$\text{Probit}(Y) = \mathbf{X}\beta + \varepsilon \text{ , where}$$

$$x_{1i} = \text{Constant}$$

$$x_{2i} = \text{Party ID } s\text{ame as incumbent}$$

$$x_{3i} = \text{National economic conditions}$$

$$x_{4i} = \text{Personal financial situation}$$

$$x_{5i} = \text{Can recall incumbent's name}$$

$$x_{6i} = \text{Can recall challenger's name}$$

$$x_{7i} = \text{Quality challenger}$$

Example: Vote Choice

Table 6.1: Probability of Voting for the Incumbent Member of Congress

variable	Probit MLEs
Intercept	.184 (.058)
Party identification	1.35 (.056)
National economic performance (Retrospective Judgment)	–.114 (.069)
Personal financial situation (Retrospective Judgment)	.095 (.068)
Recall incumbent's name	.324 (.0808)
Recall challenger's name	–.677 (.109)
Quality of challenger	–.339 (.073)

Notes: Standard errors in parentheses. $N = 3341$. $-2 \ln L = 760.629$ Percent correctly predicted = 78.5%

Example: Vote Choice

Table 6.1: Probability of Voting for the Incumbent Member of Congress

variable	Probit MLEs
Intercept	.184
	(.058)
Party identification	1.35
	(.056)
National economic performance (Retrospective Judgment)	–.114
	(.069)
Personal financial situation (Retrospective Judgment)	.095
	(.068)
Recall incumbent's name	.324
	(.0808)
Recall challenger's name	–.677
	(.109)
Quality of challenger	–.339
	(.073)

Notes: Standard errors in parentheses. $N = 3341$. $-2 \ln L = 760.629$ Percent correctly predicted = 78.5%

Significant Coefficients

Example: Vote Choice

Table 6.2: Marginal Effects on Probability of Voting for the Incumbent Member of Congress

variable	$\hat{\beta}_j \phi(\hat{\boldsymbol{\beta}}' \mathbf{x}_i)$
Party identification	.251
National economic performance (Retrospective Judgment)	–.021
Personal financial situation (Retrospective Judgment)	.018
Recall incumbent's name	.060
Recall challenger's name	–.126
Quality of challenger	–.063

Notes: Explanatory variables are set equal to their medians in the sample.

This backs out the marginal impact of a 1-unit change in the variable on the probability of voting for the incumbent.

Example: Vote Choice

Table 6.2: Marginal Effects on Probability of Voting for the Incumbent Member of Congress

variable	$\hat{\beta}_j \phi(\hat{\boldsymbol{\beta}}' \mathbf{x}_i)$
Party identification	.251
National economic performance (Retrospective Judgment)	–.021
Personal financial situation (Retrospective Judgment)	.018
Recall incumbent's name	.060
Recall challenger's name	–.126
Quality of challenger	–.063

Notes: Explanatory variables are set equal to their medians in the sample.

big

big

This backs out the marginal impact of a 1-unit change in the variable on the probability of voting for the incumbent.

Example: Vote Choice

- Or, calculate the impact of facing a quality challenger by hand, keeping all other variables at their median.

$$
\begin{aligned}
\Pr(y_i = 1 | x_{7i} = 0) =& \Phi(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i}) \\
=& \Phi(.184 + 1.355 \times 1 - .114 \times .5 + .095 \times .5 \\
& + .324 \times 0 - .677 \times 0 - .339 \times 0) \\
=& .936
\end{aligned}
$$

Example: Vote Choice

- Or, calculate the impact of facing a quality challenger by hand, keeping all other variables at their median.

$$\begin{aligned}
\Pr(y_i = 1 | x_{7i} = 0) =& \Phi(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i}) \\
=& \Phi(.184 + 1.355 \times 1 - .114 \times .5 + .095 \times .5 \\
& + .324 \times 0 - .677 \times 0 - .339 \times 0) \\
=& .936
\end{aligned}$$

$\Phi(1.52)$

Example: Vote Choice

- Or, calculate the impact of facing a quality challenger by hand, keeping all other variables at their median.

$$
\begin{aligned}
\Pr(y_i = 1 | x_{7i} = 0) =& \Phi(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i}) \\
=& \Phi(.184 + 1.355 \times 1 - .114 \times .5 + .095 \times .5 \\
& + .324 \times 0 - .677 \times 0 - .339 \times 0) \\
=& .936
\end{aligned}
$$

$\Phi(1.52)$

From standard normal table

Example: Vote Choice

- Or, calculate the impact of facing a quality challenger by hand, keeping all other variables at their median.

$$
\begin{aligned}
\Pr(y_i = 1 | x_{7i} = 0) =& \Phi(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i}) \\
=& \Phi(.184 + 1.355 \times 1 - .114 \times .5 + .095 \times .5 \\
& + .324 \times 0 - .677 \times 0 - .339 \times 0) \\
=& .936
\end{aligned}
$$

$\Phi(1.52)$

From standard normal table

$$\Pr(y_i = 1 | x_{7i} = 1) = .881$$

$\Phi(1.52-.339)$

So there's an increase of .936 - .881 = 5.5% votes in favor of incumbents who avoid a quality challengers.

Example: Senate Obstruction

- Model the probability that a bill is passed in the Senate (over a filibuster) based on:
  - The coalition size preferring the bill be passed
  - An interactive term: size of coalition X end of session

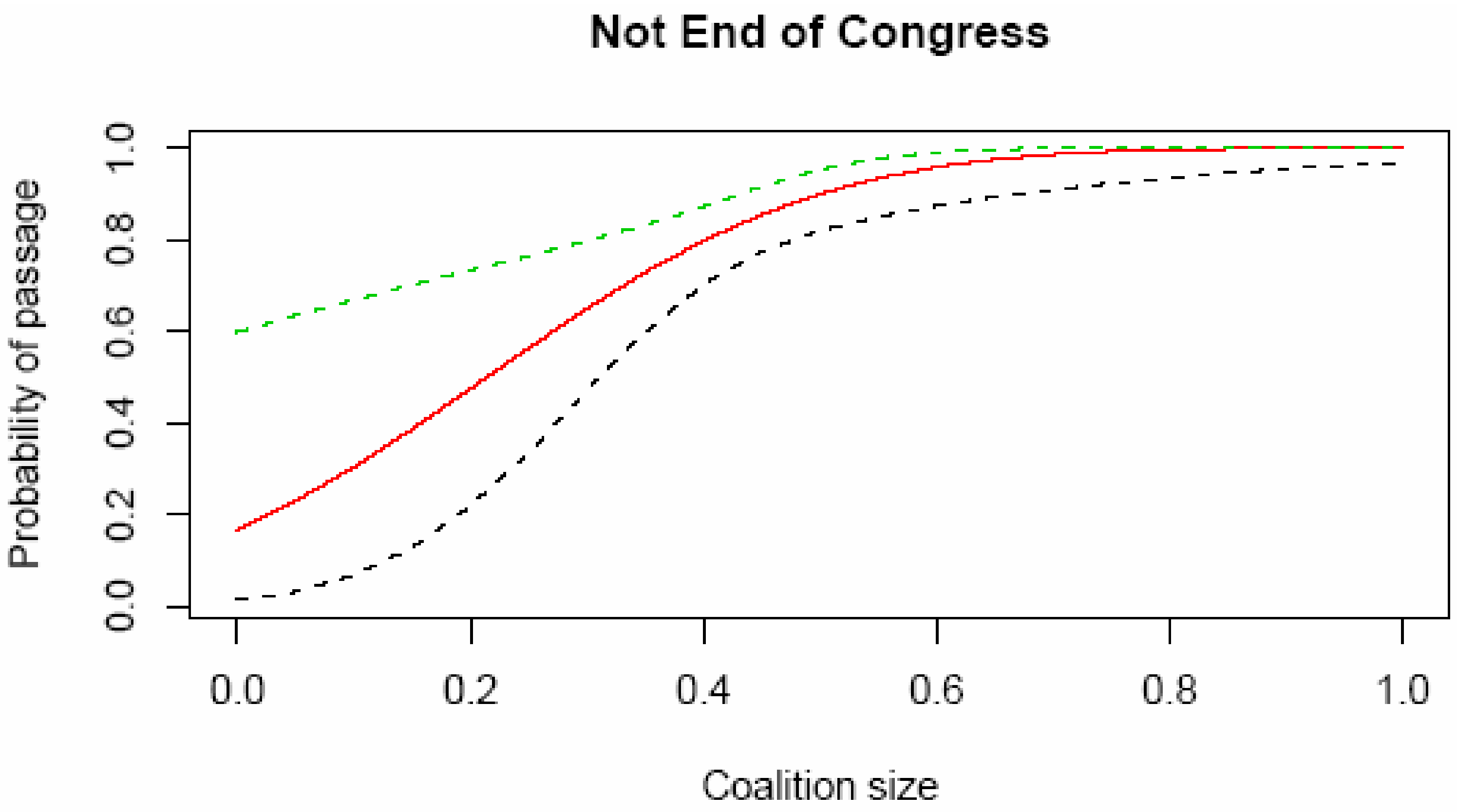
Table 6.3: Probit analysis of passage of obstructed measures, 1st–64th Congresses

Variable	Coefficient	Std. Err.
Constant	-1.671	0.962
Coalition size	6.155	2.224
Coalition size $\times$ end of session	-1.944	0.690
Likelihood ratio test	12.84	$(p = 0.002)$
% correctly predicted	72	

Note: $N = 114$.

Example: Senate Obstruction

- Graph the results for end of session = 0



Example: Senate Obstruction

- Graph the results for end of session = 1