Alternative Link Functions

Andrew Zieffler
Department of Educational Psychology

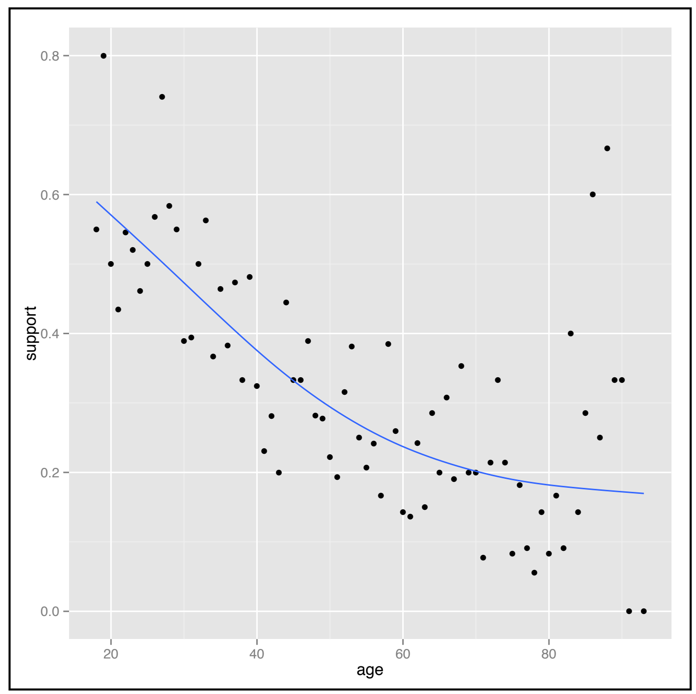
Reading and Examining the Data

```
# Read in the data
> gay = read.csv(file = "http://www.tc.umn.edu/~zief0002/Data/Gay-Marriage.csv")

# Create outcome
> gay$support = ifelse(gay$marriage == 1, 1, 0)
```

```
# Plot the relationship between age and proportion of support
> ggplot(data = gay, aes(x = age, y = support)) +
    stat_summary(fun.y = mean, geom = "point") +
    geom_smooth(se = FALSE)
```

Use `stat_summary()`

It seems like there could be a quadratic relationship between age and proportion of support for gay marriage.



Goal: Non-Linear Mapping of a Function to the [0, 1] Space

Goal of our function is to:

- Produce conditional mean y-values between 0 and 1
- Provide a nonlinear, monotonic relationship between x and the conditional means
- Not constrain the coefficient estimates (i.e., no matter what value the coefficients take, the conditional mean of y will be between 0 and 1)

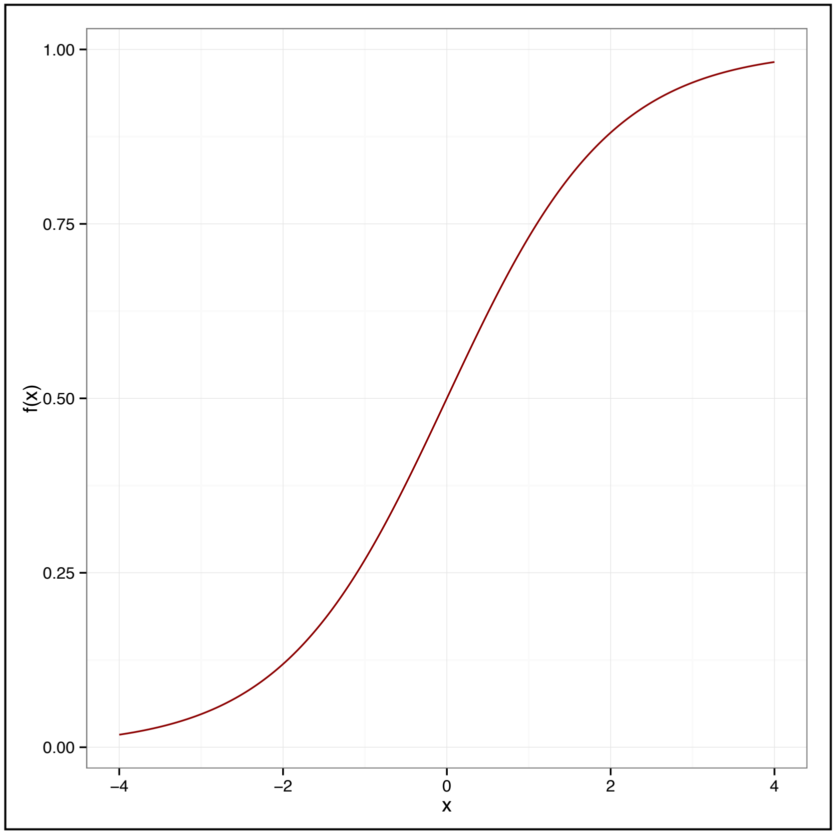
Logistic function

Using a continuous covariate, x, the probability of a binary outcome being = 1 is,

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$

We linearize this **via the link**, in this case the logit (or logistic) function

$$g(x) = \ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \beta_0 + \beta_1 x$$



Plot of the logistic function.

```
> glm.a <- glm(support ~ age + I(age^2), data = gay,
     family = binomial(link = "logit"))
> summary(glm.a)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.7012522  0.3817714   4.456 8.34e-06 ***
age         -0.0731393  0.0166302  -4.398 1.09e-05 ***
I(age^2)     0.0004209  0.0001674   2.514   0.0119 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2239.0  on 1745  degrees of freedom
Residual deviance: 2124.2  on 1743  degrees of freedom
AIC: 2130.2

Number of Fisher Scoring iterations: 4
```

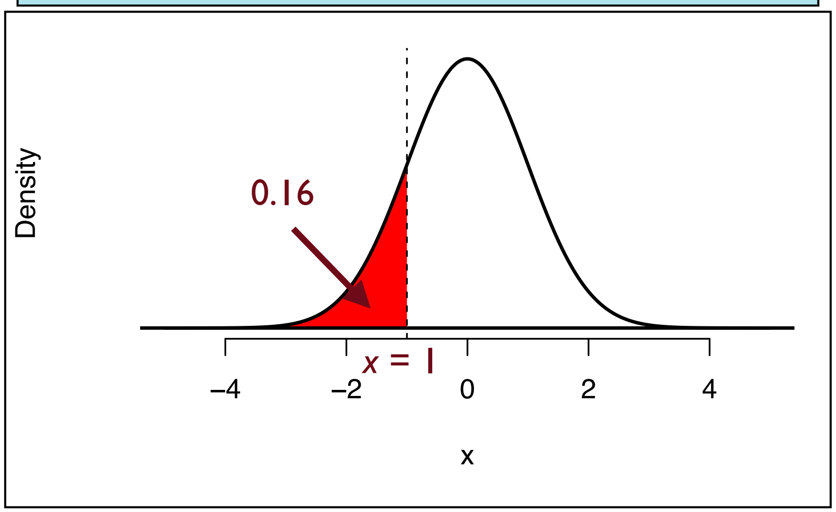
Alternative Link #1: Probit

Common nonlinear function that meets these specifications is the **cumulative density function** (CDF)

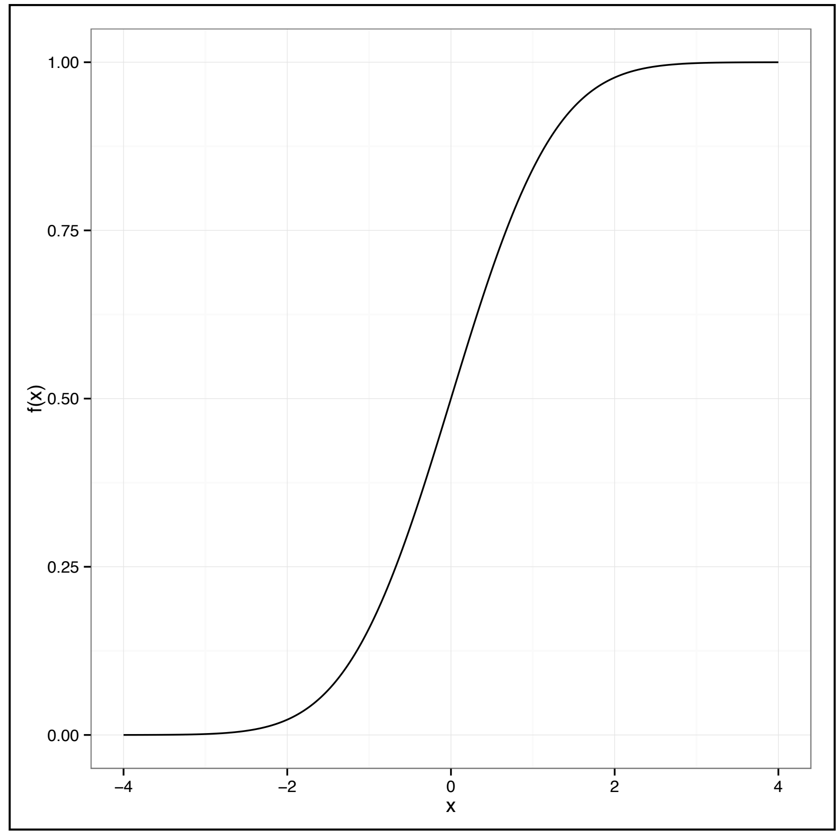
Although any CDF would work, here we use the unit normal distribution's CDF

$$f(x) = \frac{1}{2\pi} \int_{-\inf}^{x} e^{\left(-\frac{1}{2}x^2\right)} dx$$

The CDF for the unit normal distributions provides the cumulative density for some value of *x*.



This **CDF** takes a value of *x* and maps it to the cumulative density, between [0, 1].



Plot of the Unit Normal CDF.

$$\pi(x) = \Phi\left(\beta_{P0} + \beta_{P1}x\right)$$

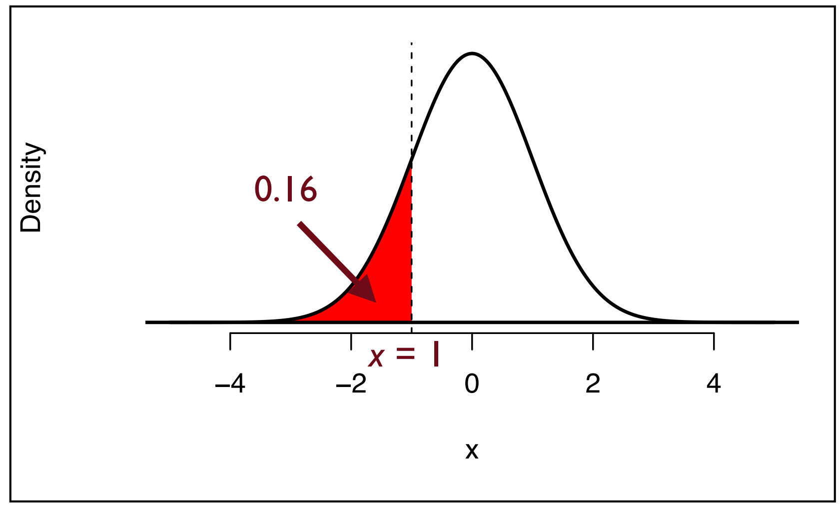
Φ denotes the use of the standard normal distribution.

The P stands for probit.

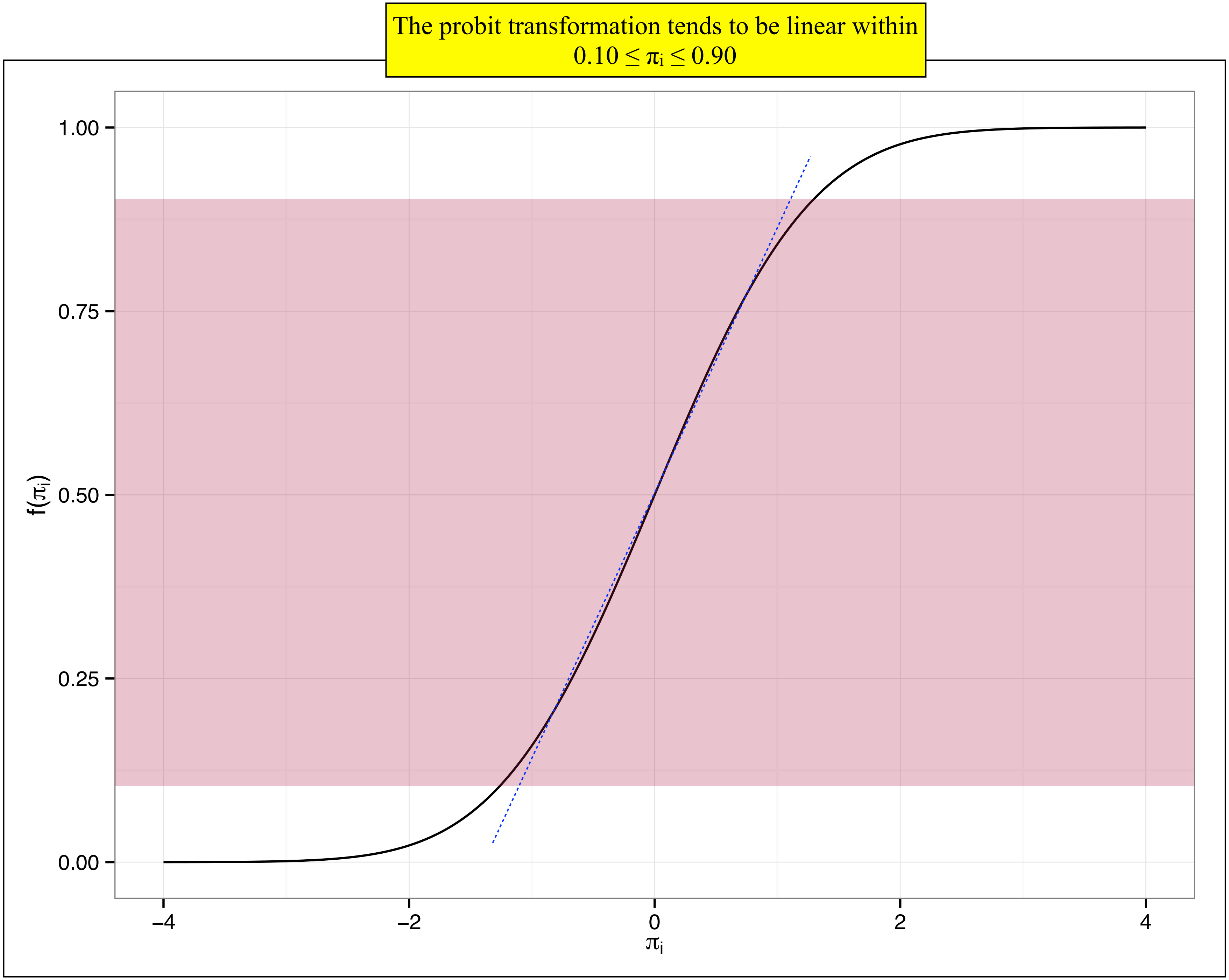
We linearize this via the inverse of the CDF function.

$$g(x) = \Phi^{-1}\Big(\pi(x)\Big) = \Phi^{-1}\left(\beta_0 + \beta_1 x\right)$$

This is called the **probit** transformation.



The probit transformation of 0.16 is 1.


The probit transformation tends to be linear within
$0.10 \le \pi_i \le 0.90$
1.00
0.75
0.50
0.25
0.00
$f(\pi_i)$
−4
−2
0
2
4
π_i

```
> glm.b <- glm(support ~ age + I(age^2), data = gay,
     family = binomial(link = "probit"))
> summary(glm.b)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.068e+00  2.316e-01   4.612 3.98e-06 ***
age         -4.620e-02  9.965e-03  -4.636 3.55e-06 ***
I(age^2)     2.732e-04  9.918e-05   2.754  0.00588 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2239.0  on 1745  degrees of freedom
Residual deviance: 2124.1  on 1743  degrees of freedom
AIC: 2130.1

Number of Fisher Scoring iterations: 4
```

To fit the probit mapping, use `link="probit"`

	Logit			Probit		
Predictor	B	SE	*p*	B	SE	*p*
Age	–0.073	0.02	< 0.001	–0.046	0.30	< 0.001
Age^2	0.0004	0.0002	0.012	0.00027	0.010	0.006
(Intercept)	1.701	0.38	< 0.001	1.068	0.23	< 0.001

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

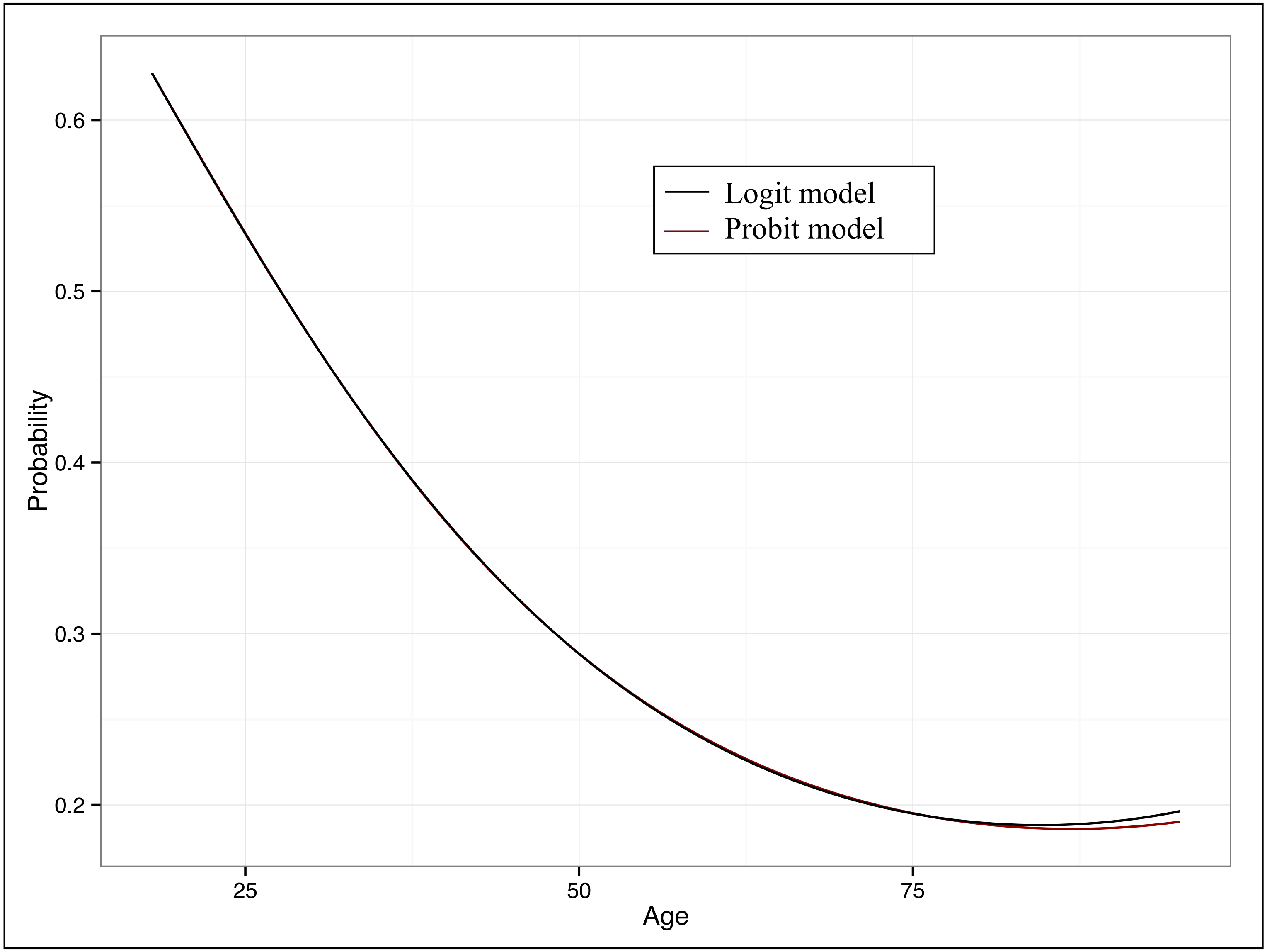
Model evaluation	Logit	Probit
Deviance	2124.2	2124.1
AIC	2130.2	2130.1
BIC	2146.6	2146.5
Pseudo R-squared		
Cox and Snell's	0.064	0.064
Efron's	0.066	0.066
McFadden's	0.051	0.051
Nagelkerke	0.088	0.088

Logit vs. Probit models

- Different coefficient and standard error values
- Similar findings about the need for a quadratic effect of age (albeit at different significance values)
- Similar values for goodness-of-fit

Logit vs. Probit models (cntd.)

- Predicted probabilities are essentially the same…
- ...this is because the predicted probabilities are between 0.2 and 0.6



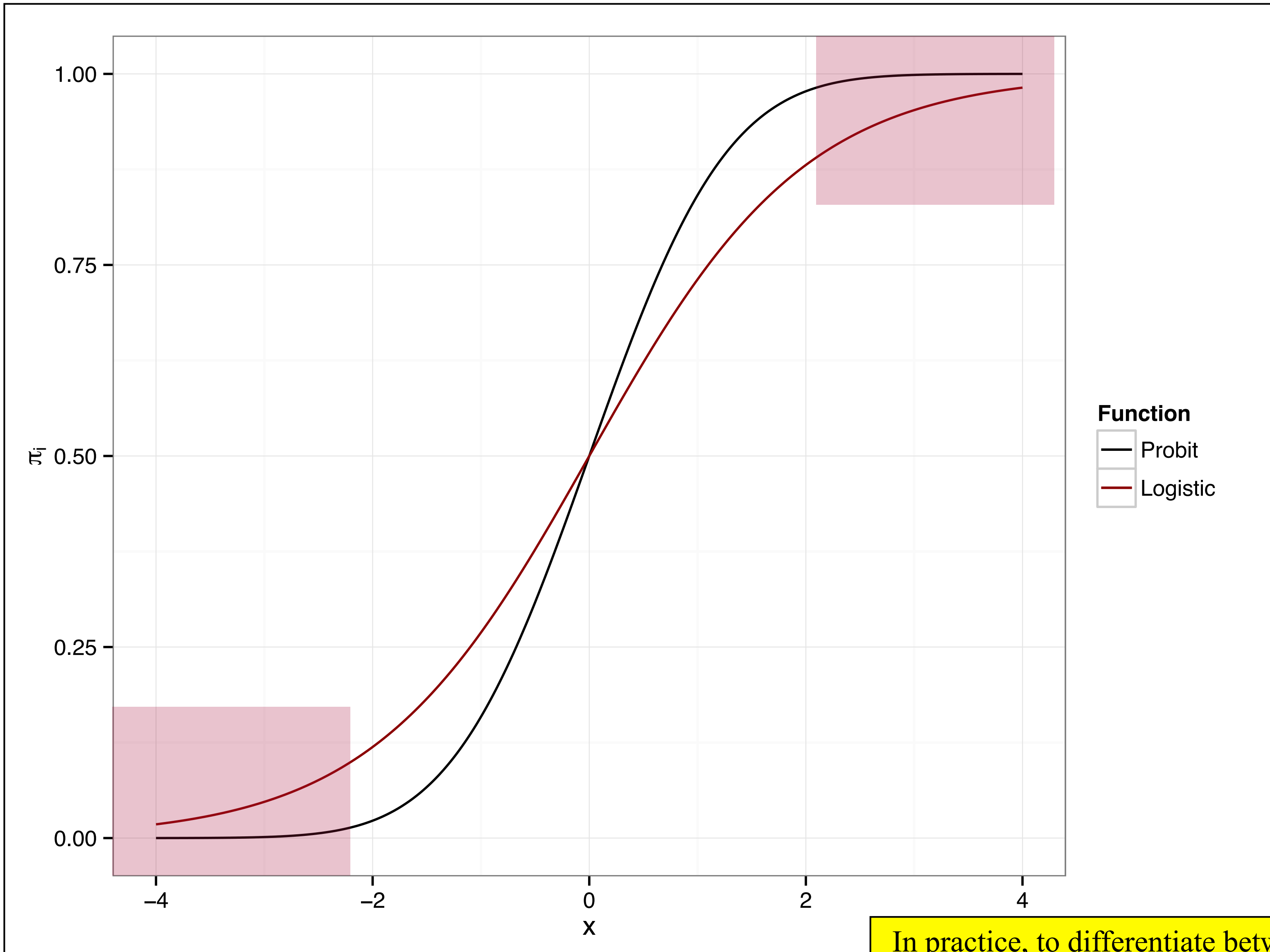
- In general, measures of fit and *p*-values will be quite similar for logit and probit models (although not identical)

- Parameter estimates and SEs tend to be quite a bit lower for the probit model than for the logit model
  - ✓ Typically coefficients in the logit model are about 80% higher than the probit model coefficients
  - ✓ The difference merely reflects the wider range covered by the logit transformation...it is not a substantive difference

	Logit			Probit		
Predictor	B	SE	*p*	B	SE	*p*
Age	–0.073	0.02	< 0.001	–0.046	0.30	< 0.001
Age^2	0.0004	0.0002	0.012	0.00027	0.010	0.006
(Intercept)	1.701	0.38	< 0.001	1.068	0.23	< 0.001

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

For really extreme values of *x* the logistic model tends to better differentiate the predicted *y*-value (note how flat the probit model is).

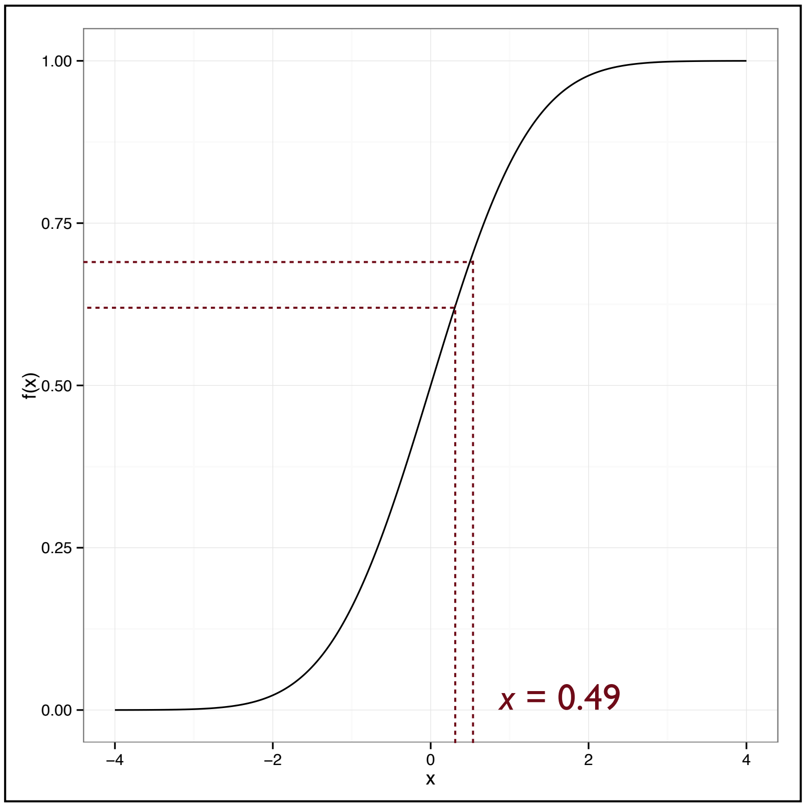


In practice, to differentiate between the predicted probabilities produced from a logit and probit model takes extremely large samples!

Interpreting Coefficients from the Probit Fitted Model

	Logit	Probit
Predictor	B	B
Age	–0.03	–0.02
(Intercept)	0.84	0.49

Note. *p < .05, **p < .01, ***p < .001



To make this easier, we will use the coefficients from a model without the quadratic effect of age

Probit model

$$\hat{y} = q = 0.49 - 0.02(\text{Age})$$

The predicted values from a probit model represent the quantile from the standard normal distribution.

Age = 0

$$\hat{y} = 0.49$$

Age = 1

$$\hat{y} = 0.49 - 0.02(1) = 0.47$$

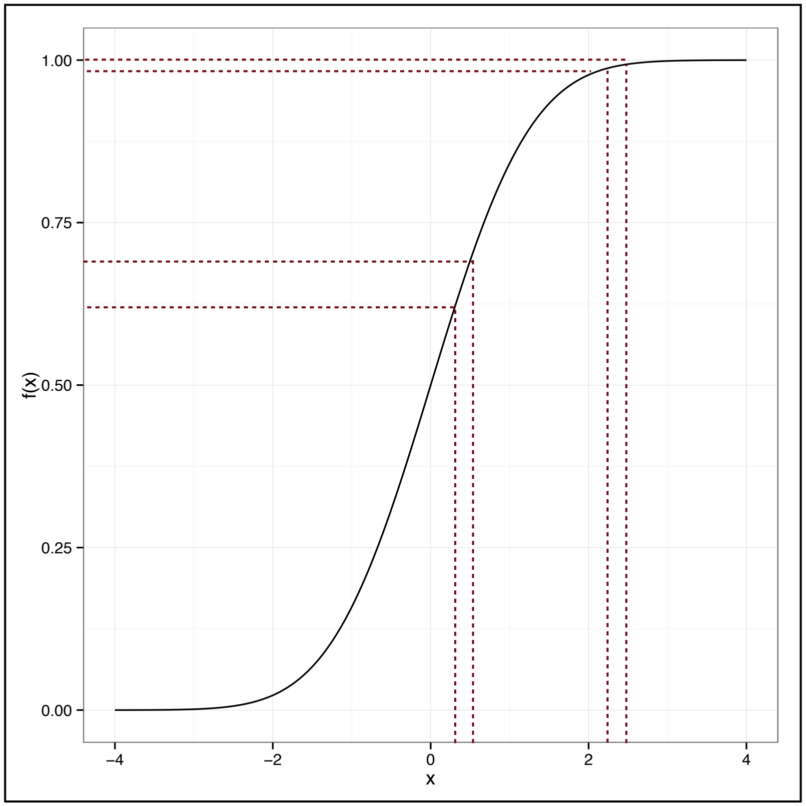
To get a sensible interpretation, of the effect we need to transform this to probabilities using the inverse function

$$\hat{\pi}(x) = \Phi^{-1}\Big(0.49 - 0.02(\text{Age})\Big)$$

Caution: Interpretation of the coefficients in probit regression is not as straightforward as the interpretations of coefficients in linear regression or logit regression.

Two primary problems

- There is no closed form solution for the inverse of the coefficients
- The change in probability attributed to a one-unit change in a given predictor *is dependent on the starting value of the predictor*.



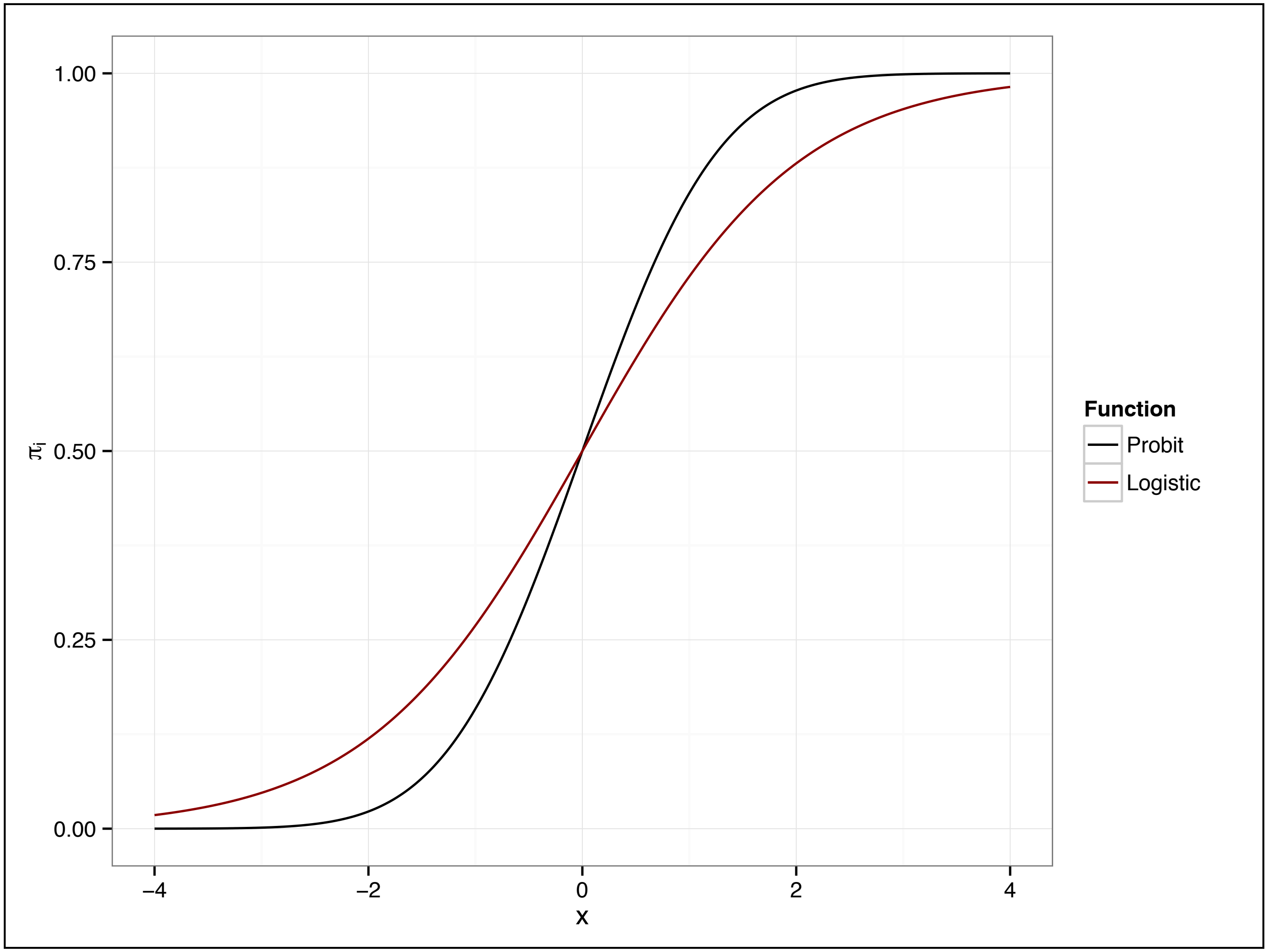
Age = 0	Age = 1
$\hat{y} = 0.49$	$\hat{y} = 0.47$

```
> pnorm(0.49)

[1] 0.6879331
```

```
> pnorm(0.47)

[1] 0.6808225
```

We can still somewhat interpret the coefficients from a probit model.

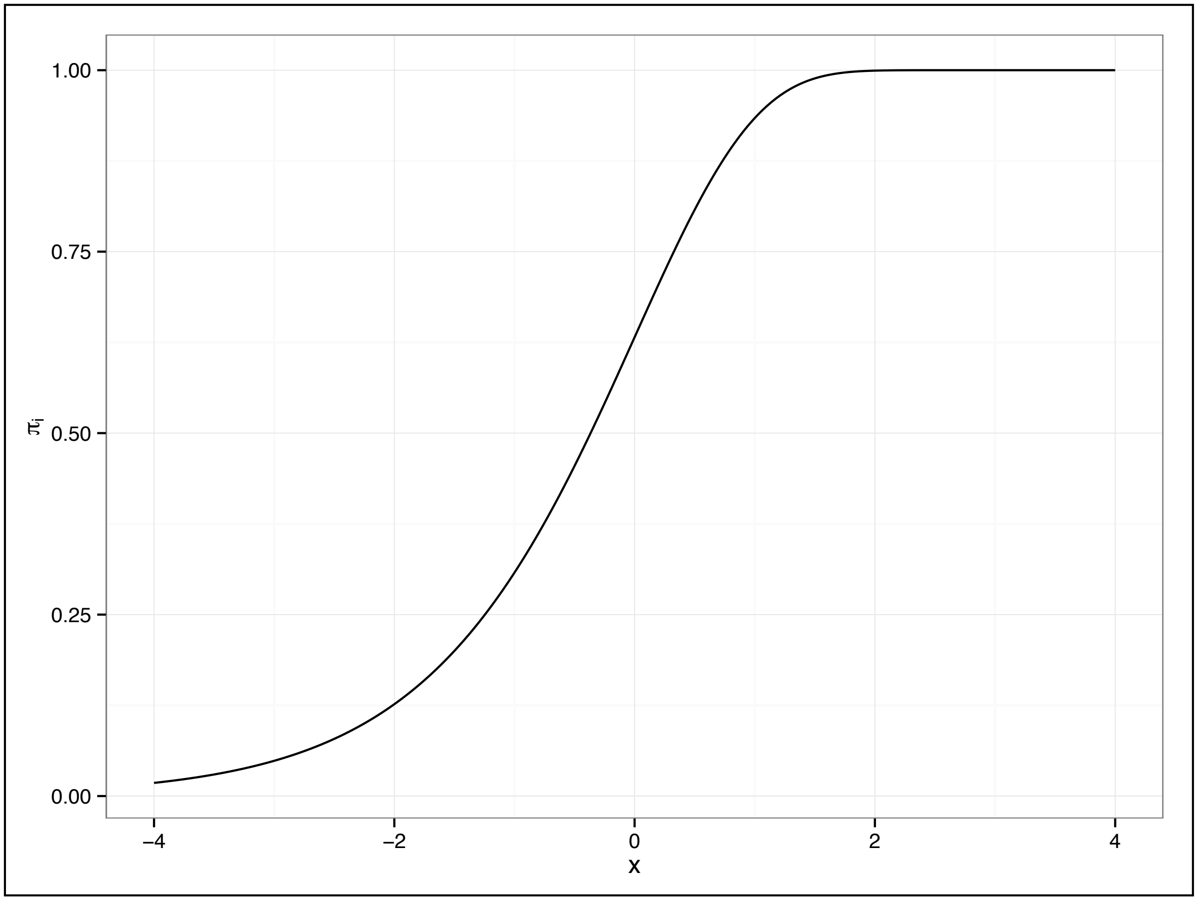
- A positive coefficient means that an increase in the predictor corresponds to an increase in the predicted probability of the outcome.
- A negative coefficient means that an increase in the predictor corresponds to a decrease in the predicted probability of the outcome.



Both the probit model and logit model are symmetric around $\pi = 0.5$.

The probabilities approach 0 and 1 at the same rate.

Alternate Link #2: Complementary Log–Log



The complementary log–log model is asymmetric around $\pi = 0.5$.

The probabilities approach 0 more slowly and approach 1 more quickly...look at $x = 1$ vs. $x = -1$

The conditional probability of the outcome is expressed as

$$\pi(x) = 1 - \exp\left[-\exp\left(\beta_0 + \beta_1 x\right)\right]$$

To linearize this, we compute the inverse

$$g(x) = \ln\left[-\boxed{\ln\left[1 - \pi(x)\right]}\right] = \beta_0 + \beta_1 x$$

ln[1 – π] is always a negative number

```
> glm.c <- glm(support ~ age + I(age^2), data = gay,
     family = binomial(link = "cloglog"))
> summary(glm.c)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.8430710  0.2859232   2.949  0.00319 **
age         -0.0514063  0.0130689  -3.933 8.37e-05 ***
I(age^2)     0.0002632  0.0001363   1.931  0.05354 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2239.0  on 1745  degrees of freedom
Residual deviance: 2124.9  on 1743  degrees of freedom
AIC: 2130.9

Number of Fisher Scoring iterations: 5
```

To fit the complementary log–log mapping, use `link="cloglog"`

Predictor	Logit			Probit			Complementary Log–Log		
	B	SE	*p*	B	SE	*p*	B	SE	*p*
Age	–0.073	0.02	< 0.001	–0.046	0.30	< 0.001	–0.051	0.01	< 0.001
Age^2	0.0004	0.0002	0.012	0.00027	0.010	0.006	0.0003	0.0001	0.054
(Intercept)	1.701	0.38	< 0.001	1.068	0.23	< 0.001	0.843	0.29	0.003

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Model evaluation	Logit	Probit	Complementary Log–Log
Deviance	2124.2	2124.1	2124.9
AIC	2130.2	2130.1	2130.9
BIC	2146.6	2146.5	2147.3
Pseudo R-squared			
Cox and Snell's	0.064	0.064	0.064
Efron's	0.066	0.066	0.065
McFadden's	0.051	0.051	0.051
Nagelkerke	0.088	0.088	0.088

Interpreting Coefficients from the Fitted Complementary Log–Log Model

The complementary log–log model is closely related to the *Cox proportional hazards model* that is used to model **time to event occurrence**.

Time to event occurrence are known as **survival models** (a.k.a., event history models, hazards models, Cox proportional hazards models, Cox regression models, duration models, and failure time models)

A time to event variable reflects **the time until a participant has an event of interest** (e.g., heart attack, goes into cancer remission, death, supports gay marriage, etc.).

In survival analysis, "survival" refers to remaining "free" (0) of the event over time

In **survival analysis**, the goal is to answer questions such as, What is the probability that a participant survives 5 years? Are there differences in survival between groups? How do certain personal, behavioral or clinical characteristics affect participants' chances of survival?

In a Cox proportional hazards regression model, the measure of effect is the **hazard rate,** which is the *risk of failure* (i.e., the risk or probability of suffering the event of interest), *given that the participant has survived up to a specific time*.

For example, in our analysis, the hazard rate would measure the probability supporting gay marriage ("failure") given that the person has not supported gay marriage ("survived") up to a specific time

To make this easier, we will again use the coefficients from a model without the quadratic effect of age

$$\hat{y} = 0.34 - 0.03(\text{Age})$$

We interpret the coefficients from the complementary log–log model in a similar manner to how we would interpret coefficients from a proportional hazards model.

Note: The coefficients obtained from the Cox proportional hazards model and the complementary log–log model will **not** be the same.

Nonetheless, they are estimating the same thing, so we interpret them in a similar manner.

We exponentiate the coefficients to obtain a hazard ratio….

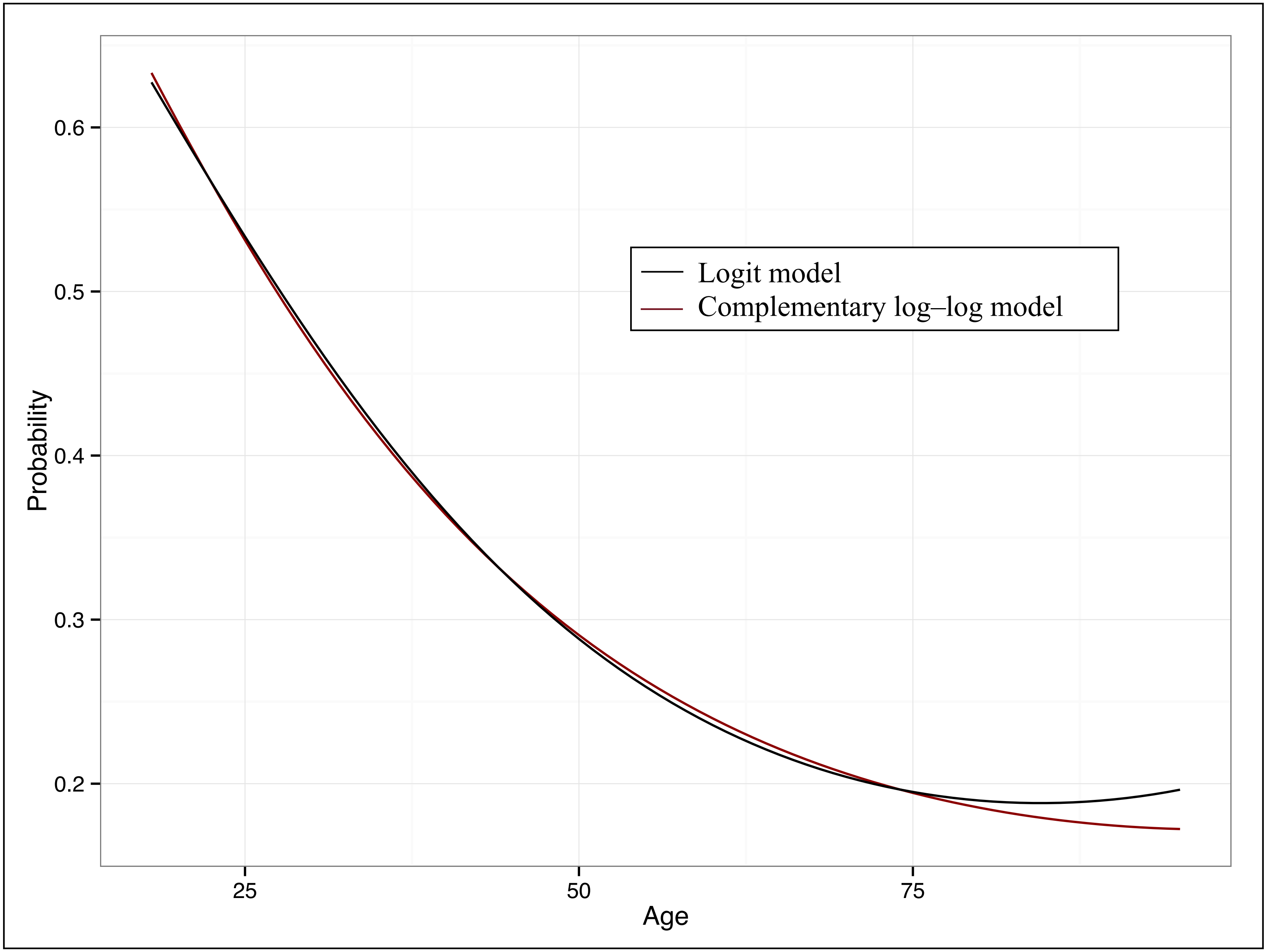
```
> exp(-0.03)

[1] 0.9704455
```

The estimated hazard of supporting gay marriage changes by a factor of 0.97 for a one year difference in age…

OR

The hazard of supporting gay marriage decreases by 3% for each additional year of age…

given that that person has not supported gay marriage previously.


Logit model
Complementary log–log model
Probability
Age
0.6
0.5
0.4
0.3
0.2
25
50
75

Asymmetry in the Link: Another View

In coding the variable `support`

- 1 = support gay marriage
- 0 = do not support gay marriage

Predictor	Logit B	Logit SE	Probit B	Probit SE
Age	-0.073^{***}	0.02	-0.046^{***}	0.30
Age^2	0.0004^{*}	0.0002	0.00027^{**}	0.010
(Intercept)	1.701^{***}	0.38	1.068^{***}	0.23

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

What if these were coded in the other direction?

In coding the variable `support`

- 1 = do not support gay marriage
- 0 = support gay marriage

Predictor	Logit B	Logit SE	Probit B	Probit SE
Age	0.073^{***}	0.02	0.046^{***}	0.30
Age^2	-0.0004^{*}	0.0002	-0.00027^{**}	0.010
(Intercept)	-1.701^{***}	0.38	-1.068^{***}	0.23

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Symmetry in the link function gives you the same coefficient estimates (except the sign is reversed) and standard errors. This suggests that the model doesn't rely on which level of the outcome is coded 1 and which is coded 0.

Predictor	Complementary Log–Log B	SE	Complementary Log–Log (Reverse coded) B	SE
Age	–0.051***	0.01	0.052***	0.01
Age^2	0.0003	0.0001	–0.0003***	0.0001
(Intercept)	0.843**	0.29	–1.568***	0.251

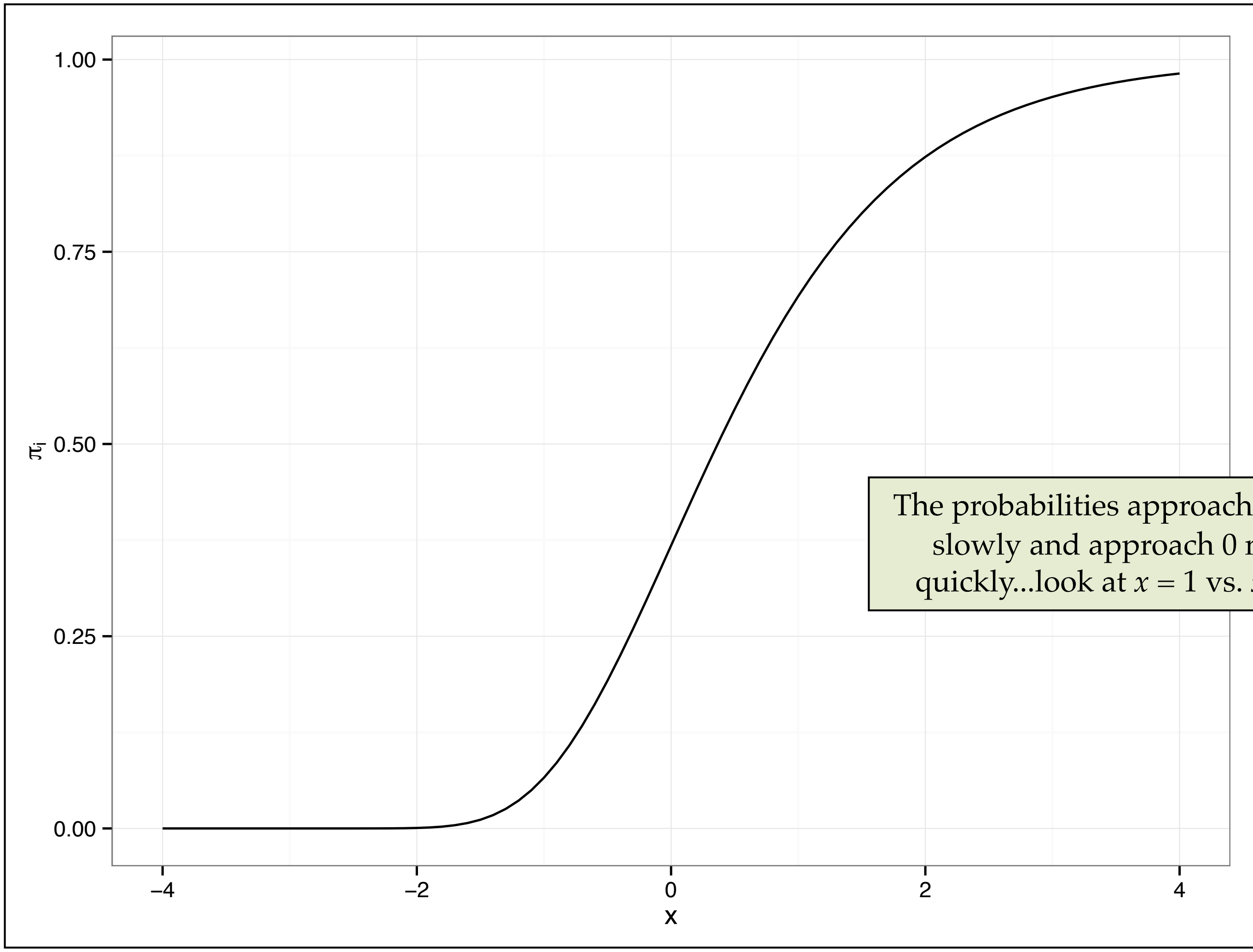
Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Asymmetry in the link function gives you different coefficient estimates and standard errors. This suggests that the model absolutely relies on which level of the outcome is coded 1 and which is coded 0.

Alternate Link #3: Log–Log

Because of the asymmetry in the complementary log–log link function, there is another related link function referred to as the log–log link.

The log–log model is also asymmetric around $\pi = 0.5$.



The probabilities approach 1 more slowly and approach 0 more quickly...look at $x = 1$ vs. $x = -1$

The conditional probability of the outcome is expressed as

$$\pi(x) = \exp\left[-\exp\left(\beta_0 + \beta_1 x\right) \right]$$

To linearize this, we compute the inverse

$$g(x) = \ln\left[-\ln\left[\pi(x)\right] \right] = \beta_0 + \beta_1 x$$

For the **complementary log–log model** the conditional probability and linearization are:

$$\pi(x) = 1 - \exp\left[-\exp\left(\beta_0 + \beta_1 x\right) \right]$$

$$g(x) = \ln\left[-\ln\left[1 - \pi(x)\right] \right] = \beta_0 + \beta_1 x$$

It is the **complement** of the log–log link

To fit the log–log mapping, reverse code the outcome and use `link="cloglog"`

```
# Reverse code the outcome
> gay$noSupport = ifelse(gay$support == 0, 1, 0)

> glm.d <- glm(noSupport ~ age + I(age^2), data = gay,
     family = binomial(link = "cloglog"))
> summary(glm.d)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.568e+00  2.512e-01  -6.245 4.24e-10 ***
age          5.238e-02  1.025e-02   5.109 3.24e-07 ***
I(age^2)    -3.318e-04  9.814e-05  -3.381 0.000722 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2239.0  on 1745  degrees of freedom
Residual deviance: 2123.4  on 1743  degrees of freedom
AIC: 2129.4

Number of Fisher Scoring iterations: 5
```

Predictor	Logit			Probit			Complementary Log–Log			Log–Log		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Age	–0.073	0.02	< 0.001	–0.046	0.30	< 0.001	–0.051	0.01	< 0.001	0.0523	0.01	< 0.001
Age^2	0.0004	0.0002	0.012	0.00027	0.010	0.006	0.0003	0.0001	0.054	–0.003	0.0001	< 0.001
(Intercept)	1.701	0.38	< 0.001	1.068	0.23	< 0.001	0.843	0.29	0.003	–1.568	0.251	< 0.001

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Notice the effects of age and age^2 are close to symmetric for the log–log and complementary log–log models

Model evaluation	Logit	Probit	Complementary Log–Log	Log–Log
Deviance	2124.2	2124.1	2124.9	2123.4
AIC	2130.2	2130.1	2130.9	2129.4
BIC	2146.6	2146.5	2147.3	2145.8
Pseudo R-squared				
Cox and Snell's	0.064	0.064	0.064	0.064
Efron's	0.066	0.066	0.065	0.066
McFadden's	0.051	0.051	0.051	0.051
Nagelkerke	0.088	0.088	0.088	0.088

To make this easier, we will again use the coefficients from a model without the quadratic effect of age

$$\hat{y} = -0.79 + 0.02(\text{Age})$$

We interpret the coefficients from the log–log model in the same way we would interpret coefficients from the complementary log–log model.

We exponentiate the coefficients to obtain a hazard ratio....

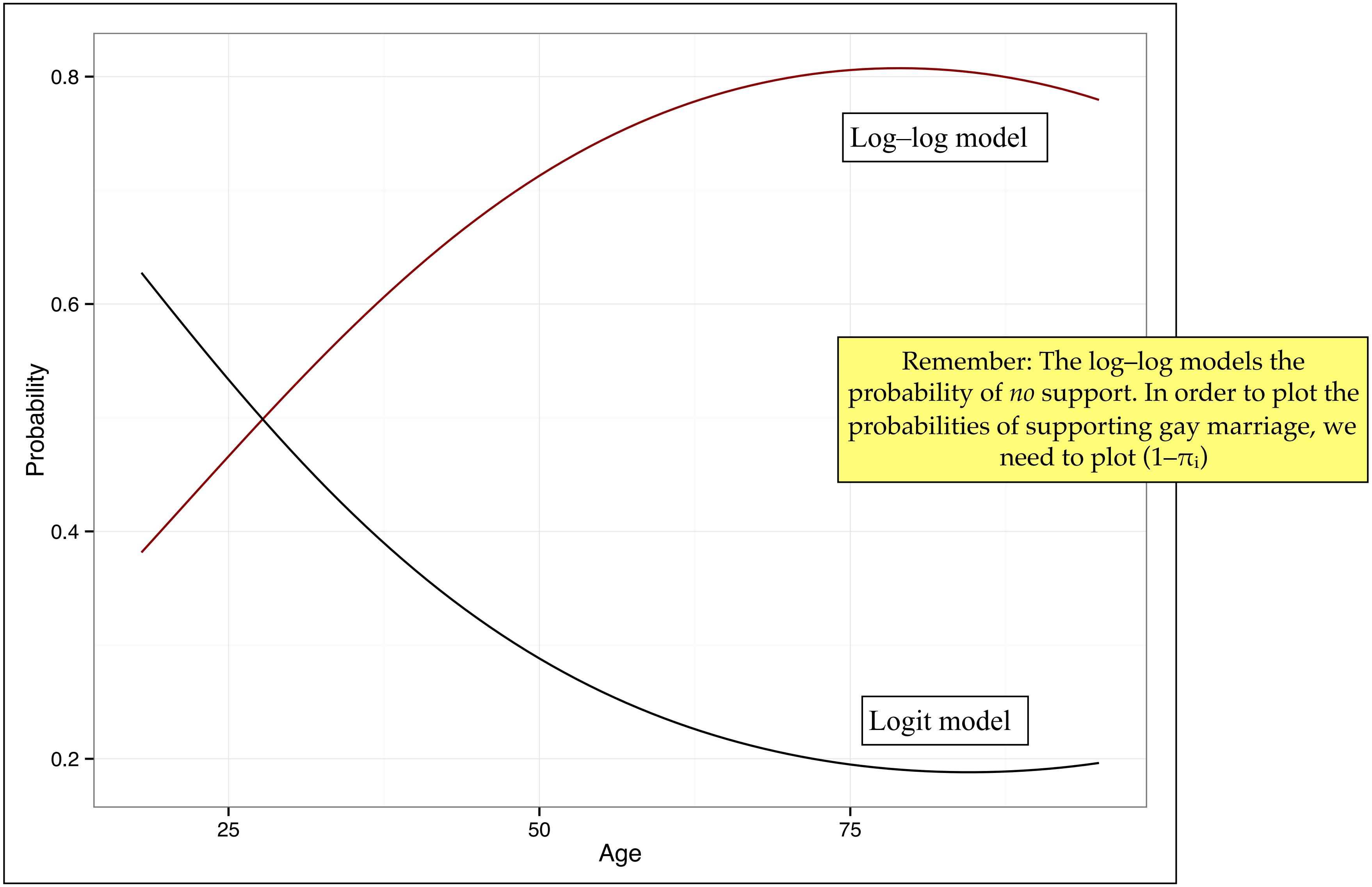
```
> exp(0.02)

[1] 1.020201
```

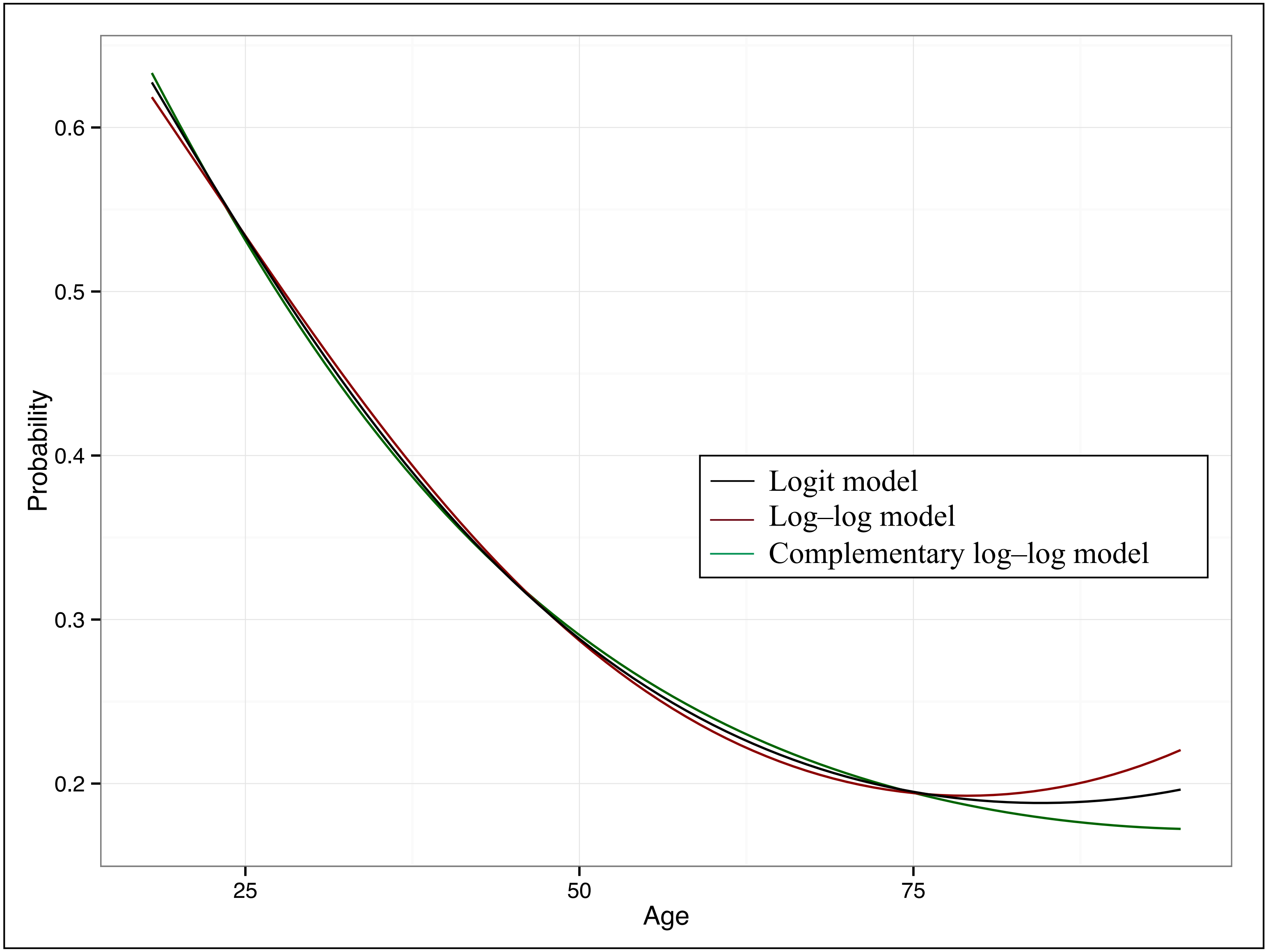
The estimated hazard of *not* supporting gay marriage changes by a factor of 1.02 for a one year difference in age...

OR

The hazard of *not* supporting gay marriage decreases by 2% for each additional year of age...

given that that person has supported gay marriage previously.


Log–log model
Logit model
Remember: The log–log models the probability of *no* support. In order to plot the probabilities of supporting gay marriage, we need to plot $(1-\pi_i)$
Probability
0.8
0.6
0.4
0.2
25
50
75
Age



0.6
0.5
0.4
0.3
0.2
Probability
25
50
75
Age
Logit model
Log–log model
Complementary log–log model