Intermediate Econometrics

Limited Dependent Variable Models

Text: Wooldridge, 17

July 28, 2011

Logit and Probit Models for Binary Dependent Variables

- We are interested in modeling the **response probability**

$$\mathsf{P}(y = 1|x_1, x_2, \ldots, x_k) = \mathsf{P}(y = 1|\boldsymbol{x}) = p(\boldsymbol{x}_i),$$

  where,

$$\boldsymbol{x} = [1, x_1, x_2, \ldots, x_k] \tag{1}$$

  denotes the set of independent variables, *including the constant*.[1]

- To overcome the problems of the linear probability model, consider a specification of the form

$$\mathsf{P}(y = 1|x_1, x_2, \ldots, x_k) = G\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\right) = G(\boldsymbol{x\beta}), \tag{2}$$

  where $G(z)$ is a function such that

$$0 < G(z) < 1 \quad \text{for all real } z. \tag{3}$$

[1] Including the constant simplifies the notation somewhat.

- For example, in the linear probability model, G in (2) is

$$G(z) = z,$$

which obviously violates (3).

- Various nonlinear functions are better suited.

- The two most common functions are:

1. the cumulative distribution function (cdf) of the logistic distribution,

$$G(z) = \frac{e^z}{1+e^z} =: \Lambda(z), \tag{4}$$

which gives rise to the **logit model**. The response probability is then

$$\mathsf{P}(y=1|x_1, x_2, \ldots, x_k) = \frac{\exp\{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\}}{1+\exp\{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\}}. \tag{5}$$

2. The standard normal cdf, leading to the **probit model**,

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(\xi) d\xi, \tag{6}$$

where $\phi(z)$ is the standard normal density,

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right). \tag{7}$$

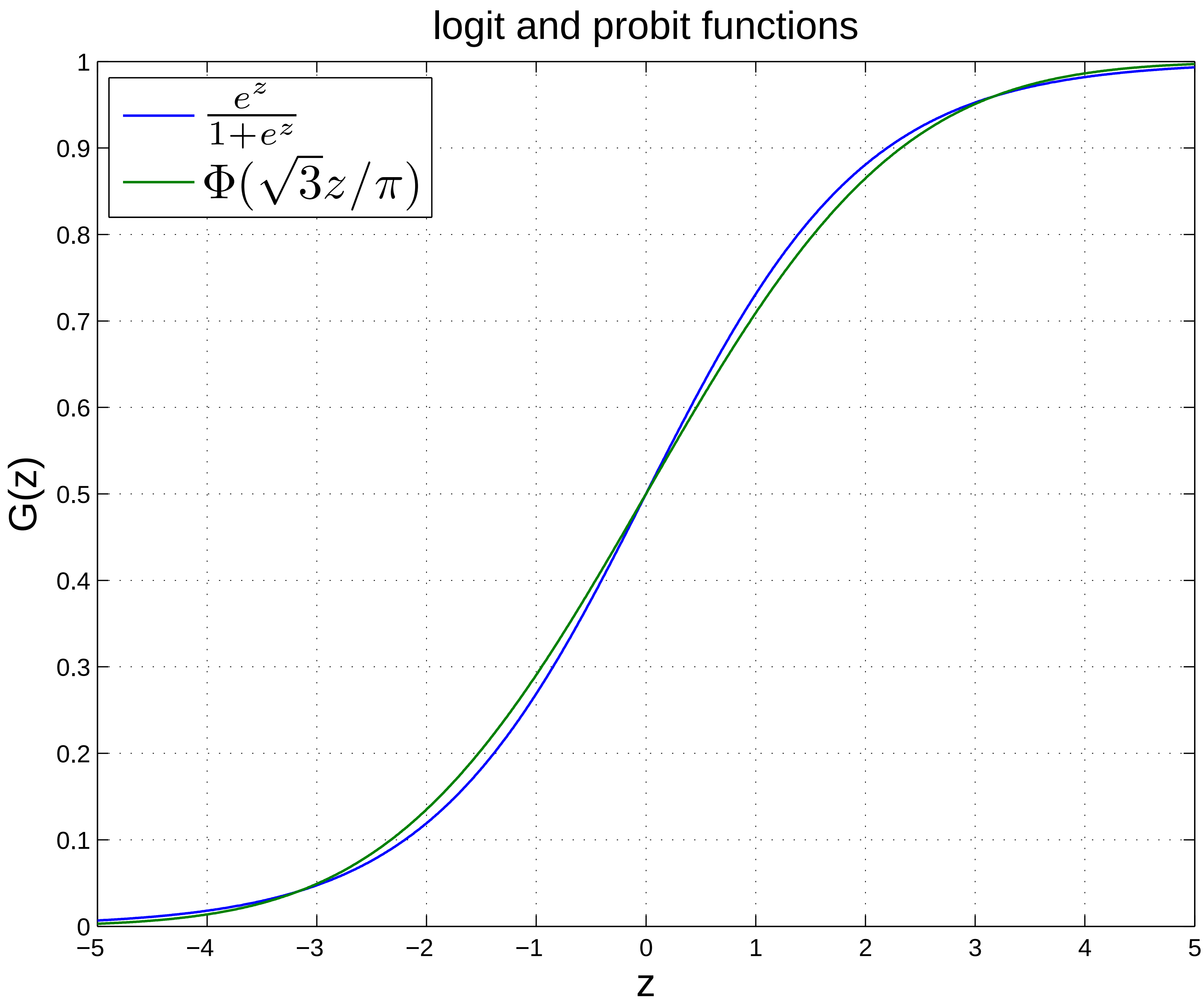
The response probability is then

$$\mathsf{P}(y = 1 | x_1, x_2, \ldots, x_k) = \Phi(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k). \tag{8}$$

- Both functions have a very similar shape.

- To illustrate this graphically, we have to take into account that the distributions (4) and (6) have different standard deviations.

- Namely, the standard deviation of a logistic variable with cdf (4) is $\pi/\sqrt{3} \approx 1.8138$, so we plot the cdf of a normal variable with the same standard deviation, i.e., the function $\Phi(\sqrt{3}z/\pi)$.


logit and probit functions
$\frac{e^z}{1+e^z}$
$\Phi(\sqrt{3}z/\pi)$
G(z)
z
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−5
−4
−3
−2
−1
0
1
2
3
4
5

- That is, for many practical purposes,

$$\Lambda(z) \approx \Phi(\sqrt{3}z/\pi). \tag{9}$$

- Thus also

$$\Lambda(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) \approx \Phi(\sqrt{3}/\pi(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)). \tag{10}$$

- Hence, if the two functions, up to standardization, would be *identical*, then we would observe that logit parameter estimates are about 1.8 times bigger than the probit estimates.

- However, since the distribution functions are clearly not identical, it turns out that they tend to be about 1.6 times bigger.[2]

- According to Greene (*Econometric Analysis*), "[i]n most applications, the choice between these two seems not to make much difference."

[2] Cf. P. Kennedy (2008): *A Guide to Econometrics*, 6e, p. 248.

Partial Effects

- What is the partial effect of a quantitative variable on the response probability $p(\boldsymbol{x})$?

- We calculate

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\boldsymbol{x}\boldsymbol{\beta}) \cdot \beta_j, \tag{11}$$

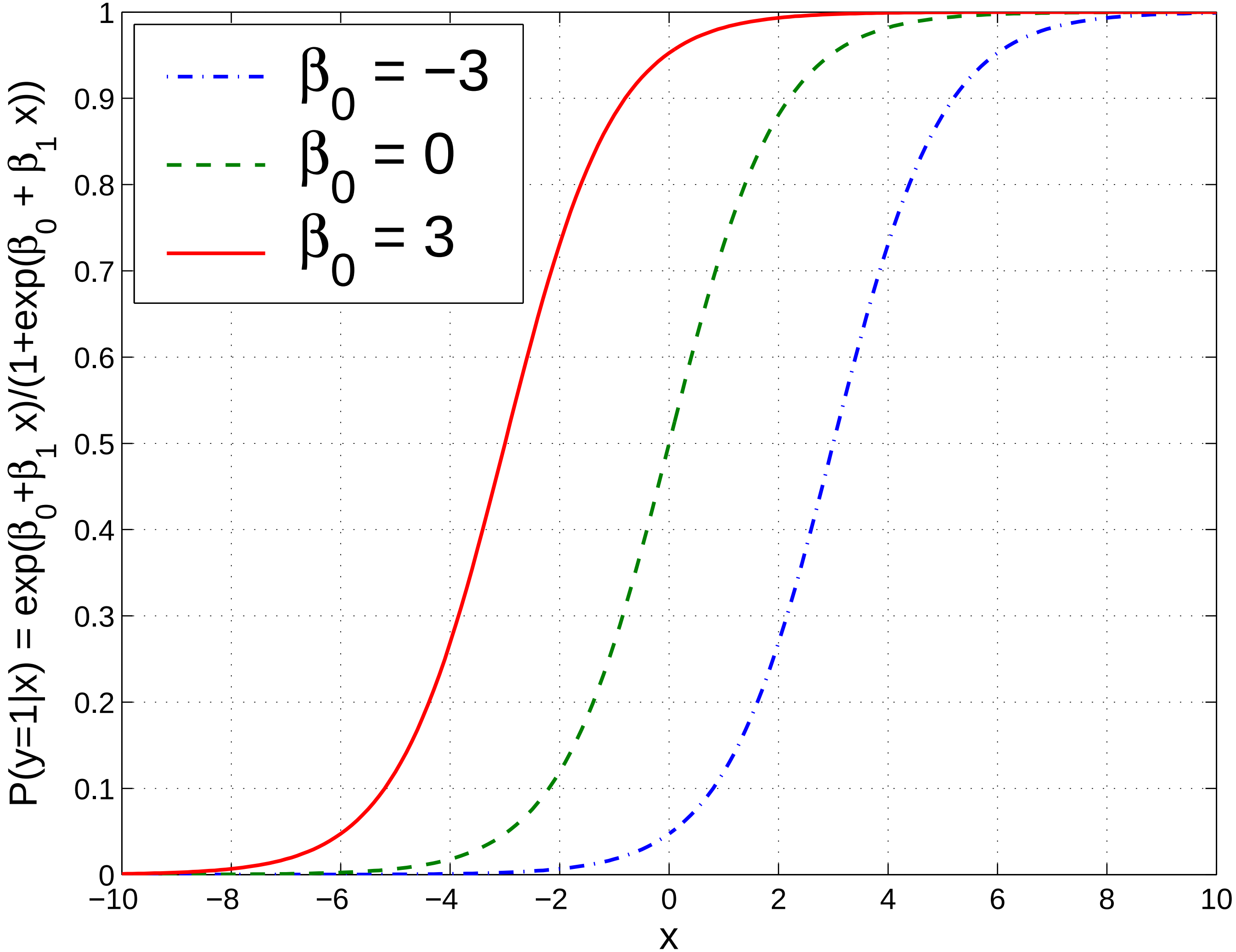
  where

$$g(z) = \frac{dG(z)}{dz} = \begin{cases} \frac{e^z}{(1+e^z)^2} = p(\boldsymbol{x})(1 - p(\boldsymbol{x})) & \text{(logit)} \\ \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} & \text{(probit)} \end{cases}$$

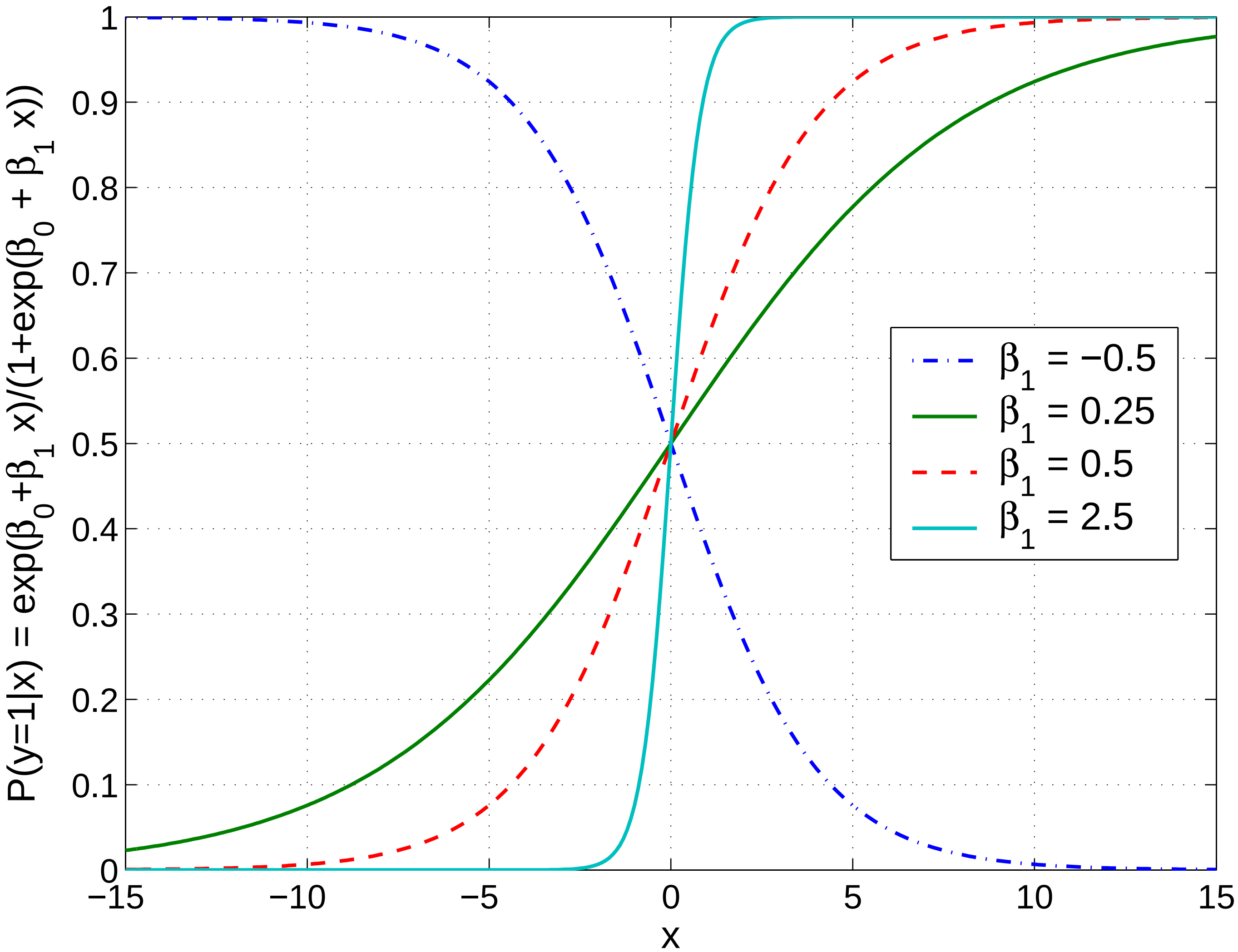
  is the density of either the logistic (logit) or normal (probit) distribution, respectively.

- Since g(z) is positive, the partial effect has the same sign as β_j.

$$P(y=1|x) = \frac{\exp\{\beta_0 + \beta_1 x\}}{1 + \exp\{\beta_0 + \beta_1 x\}}, \quad \beta_1 = 1.$$



$$P(y=1|x) = \frac{\exp\{\beta_0 + \beta_1 x\}}{1 + \exp\{\beta_0 + \beta_1 x\}}, \quad \beta_0 = 0.$$



- Moreover, the *relative* effects of any two quantitative explanatory variables do not depend on $\boldsymbol{x}$.

- In contrast to the linear probability model, however, the magnitude of the effect depends on $\boldsymbol{x}$, i.e., the current values of all the explanatory variables.

- In applications, we may then wish to calculate (11) for various combinations of interest of the independent variables.

- For reporting purposes, to summarize the magnitudes of the partial effects, it may often be desirable to have a single scale factor that can be used to multiply each estimate $\widehat{\beta}_j$ in (11).

- For example, with estimates $\widehat{\boldsymbol{\beta}}$,

$$g(\bar{\boldsymbol{x}}\widehat{\boldsymbol{\beta}}) = g\left(\widehat{\beta}_0 + \sum_{j=1}^{k} \widehat{\beta}_j \bar{x}_j\right) \tag{12}$$

can be used to this end.

- The idea behind using (12) is that it reflects the partial effects for the "average person" in the sample.

- Expression (12) is difficult to interpret when some of the independent variables are binary, e.g., $x_1 = 1$ for females and zero otherwise.

- In such situations, it is preferable to report the "typical" marginal effect separately both for females and males.

- Alternatively, the **average partial effect** may be used, given by the scale factor

$$\frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{x}_i\widehat{\boldsymbol{\beta}}) \tag{13}$$

in (11).

- For binary (or other discrete) variables, the calculation of partial effects based on (11) is not tenable.

- Instead we can directly calculate the change in probability induced by a change in the variable.

- For example, if x_1 is a binary variable, then the (estimated) partial effect from changing x_1 from zero to one, with all the other variables fixed at their mean values, is

$$G\left(\widehat{\beta}_0 + \widehat{\beta}_1 + \widehat{\beta}_2\overline{x}_2 + \cdots + \widehat{\beta}_k\overline{x}_k\right) - G\left(\widehat{\beta}_0 + \widehat{\beta}_2\overline{x}_2 + \cdots + \widehat{\beta}_k\overline{x}_k\right), \tag{14}$$

or, when (13) is adopted,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{G\left(\widehat{\beta}_0 + \widehat{\beta}_1 + \widehat{\beta}_2 x_{i2} + \cdots + \widehat{\beta}_k x_{ik}\right) - G\left(\widehat{\beta}_0 + \widehat{\beta}_2 x_{i2} + \cdots + \widehat{\beta}_k x_{ik}\right)\right\}.$$

- Clearly the method of comparing the implied probabilities for different values of the independent variables is applicable for quantitative variables also.

- Summary measures for quantitative variables can also be calculated in this spirit. E.g., for a quantitative and roughly continuous variable x_1, we may calculate the *centered unit change* for x_1 as

$$G\left(\widehat{\beta}_0+\widehat{\beta}_1\left(\overline{x}_1+\frac{1}{2}\right)+\widehat{\beta}_2\overline{x}_2+\cdots+\widehat{\beta}_k\overline{x}_k\right)$$
$$-G\left(\widehat{\beta}_0+\widehat{\beta}_1\left(\overline{x}_1-\frac{1}{2}\right)+\widehat{\beta}_2\overline{x}_2+\cdots+\widehat{\beta}_k\overline{x}_k\right)$$

or

$$\frac{1}{n}\sum_{i=1}^{n}\left\{G\left(\widehat{\beta}_0+\widehat{\beta}_1\left(x_{i1}+\frac{1}{2}\right)+\widehat{\beta}_2x_{i2}+\cdots+\widehat{\beta}_kx_{ik}\right)\right.$$
$$\left.-G\left(\widehat{\beta}_0+\widehat{\beta}_1\left(x_{i1}-\frac{1}{2}\right)+\widehat{\beta}_2x_{i2}+\cdots+\widehat{\beta}_kx_{ik}\right)\right\}.$$

as analogues to (12) and (13), respectively.

- In summary, "[c]ompared with the linear probability model, the cost of using probit and logit models is that the partial effects [...] are harder to summarize" (Wooldridge, *Introductory Econometrics*, Chapter 17).

- With estimates $\widehat{\boldsymbol{\beta}}$ and for small Δx_j, the change in probability of x_j is sometimes approximated by

$$\Delta\widehat{\mathsf{P}}(y = 1|\boldsymbol{x}) \approx \{g(\boldsymbol{x}\widehat{\boldsymbol{\beta}})\widehat{\beta}_j\}\Delta x_j \tag{15}$$

for given values of $\boldsymbol{x}$ of interest.

- However, in some situations, approximation (15) may give rise to quite misleading results.[3]

- Consider the example

$$p(x_1) = \frac{e^{1.5 \cdot x_1}}{1 + e^{1.5 \cdot x_1}}, \quad \frac{\partial p(x_1)}{\partial x_1} = 1.5 \frac{e^{1.5 \cdot x_1}}{(1 + e^{1.5 \cdot x_1})^2}.$$

[3] E.g., S. B. Caudill and J. D. Jackson (1989): Measuring Marginal Effects in Limited Dependent Variable Models, *The Statistician* 38, 203–206.

- For $x = 0.6$, $p(x_1) = 0.7109$, and (15) is

$$1.5 \times 0.7109 \times (1 - 0.7109) = 0.3083,$$

which cannot be the increase in probability due to $\Delta x_1 = 1$, since $0.7109 + 0.3083 > 1$.

- The exact value of a unit increase in x_1 is

$$G(\beta_1(x_1+1)) - G(\beta_1 x_1) = \frac{e^{1.5\cdot(0.6+1)}}{1+e^{1.5\cdot(0.6+1)}} - \frac{e^{1.5\cdot 0.6}}{1+e^{1.5\cdot 0.6}} = 0.2059. \quad (16)$$

Estimation

- These models are most efficiently estimated by means of maximum likelihood.

- For given $\boldsymbol{x}_i$, y_i is Bernoulli with probability $p(\boldsymbol{x}_i)$, i.e., its density

$$f(y_i|\boldsymbol{x}_i;\boldsymbol{\beta}) = p(\boldsymbol{x}_i)^{y_i}(1-p(\boldsymbol{x}_i))^{1-y_i}, \quad y_i \in \{0,1\}.$$

- The density of the entire sample (of size n) is

$$f(y_1,\ldots,y_n|\boldsymbol{X};\boldsymbol{\beta}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i)^{y_i}(1-p(\boldsymbol{x}_i))^{1-y_i}. \tag{17}$$

- When we view (17) as a function of the unknown parameters, for a given sample, we obtain the likelihood function.

- Thus, the log–likelihood function is given by

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \log p(\boldsymbol{x}_i) + (1 - y_i) \log(1 - p(\boldsymbol{x}_i)) \right]. \tag{18}$$

- The maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$ is the value of $\boldsymbol{\beta}$ that maximizes (18).

- Consider the logit model.

- For this model, with

$$p(\boldsymbol{x}) = \frac{e^{\boldsymbol{x}\boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}\boldsymbol{\beta}}} = \frac{\exp\{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\}},$$

we have

$$\begin{aligned} \frac{\partial p(\boldsymbol{x}_i)}{\partial \beta_j} &= \frac{x_{ij} e^{\boldsymbol{x}_i \boldsymbol{\beta}} (1 + e^{\boldsymbol{x}_i \boldsymbol{\beta}}) - x_{ij} e^{2\boldsymbol{x}_i \boldsymbol{\beta}}}{\left(1 + e^{\boldsymbol{x}_i \boldsymbol{\beta}}\right)^2} = \frac{x_{ij} e^{\boldsymbol{x}_i \boldsymbol{\beta}}}{\left(1 + e^{\boldsymbol{x}_i \boldsymbol{\beta}}\right)^2} \\ &= x_{ij} p(\boldsymbol{x}_i)(1 - p(\boldsymbol{x}_i)). \end{aligned}$$

- Hence, we get the likelihood equations

$$\frac{\partial \log L(\widehat{\boldsymbol{\beta}})}{\partial \widehat{\beta}_j} = \sum_{i=1}^{n} \left[y_i \frac{1}{\widehat{p}(\boldsymbol{x}_i)} \frac{\partial \widehat{p}(\boldsymbol{x}_i)}{\partial \widehat{\beta}_j} - (1 - y_i) \frac{1}{1 - \widehat{p}(\boldsymbol{x}_i)} \frac{\partial \widehat{p}(\boldsymbol{x}_i)}{\partial \widehat{\beta}_j} \right]$$
$$= \sum_{i=1}^{n} \left[\frac{y_i - \widehat{p}(\boldsymbol{x}_i)}{\widehat{p}(\boldsymbol{x}_i)(1 - \widehat{p}(\boldsymbol{x}_i))} \right] \frac{\partial \widehat{p}(\boldsymbol{x}_i)}{\partial \widehat{\beta}_j}$$
$$= \sum_{i=1}^{n} (y_i - \widehat{p}(\boldsymbol{x}_i)) x_{ij} = 0, \quad j = 0, \ldots, k. \tag{19}$$

- Note that, since a constant is included, (19) implies (for $j = 0$ in (19))

$$\frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} \widehat{p}(\boldsymbol{x}_i), \tag{20}$$

so the *average* estimated probability of $y = 1$ is equal to the fraction of times that $y = 1$ was actually observed.

- The likelihood equations (19) have no closed–form solution and have to be solved numerically.

- The maximum likelihood estimator is consistent and asymptotically normally distributed, with asymptotic standard errors deriving from likelihood theory.

- Thus asymptotic t tests can be conducted and asymptotic confidence intervals calculated in the usual manner.

- **Multiple Hypotheses** can be tested by means of **likelihood ratio tests** (LRT).

- To do so, we calculate the value of the maximized log–likelihood function both

  – under the null $(\log L_0)$
  – and the alternative hypothesis $(\log L_1)$,

  where the former and the latter correspond to the restricted and unrestricted models, respectively.

- Then the likelihood ratio test statistic,

$$\mathsf{LRT} = -2(\log L_0 - \log L_1) \overset{asy}{\sim} \chi^2(q),$$

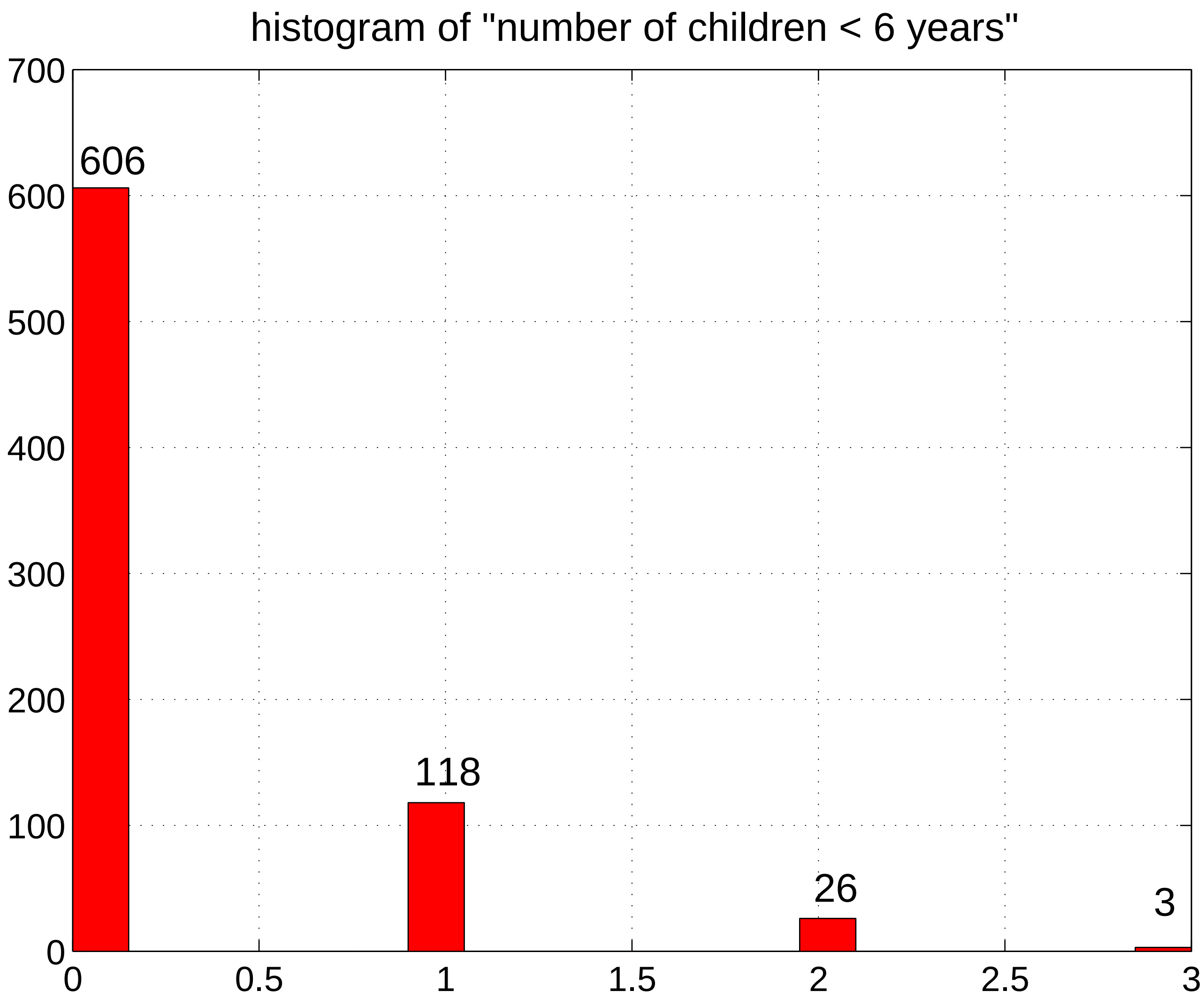
where q is the number of restrictions under the null hypothesis, e.g., q exclusion restrictions.

Labor Market Participation Estimates

Table 1: labor market participation of married woman ($n = 753$)

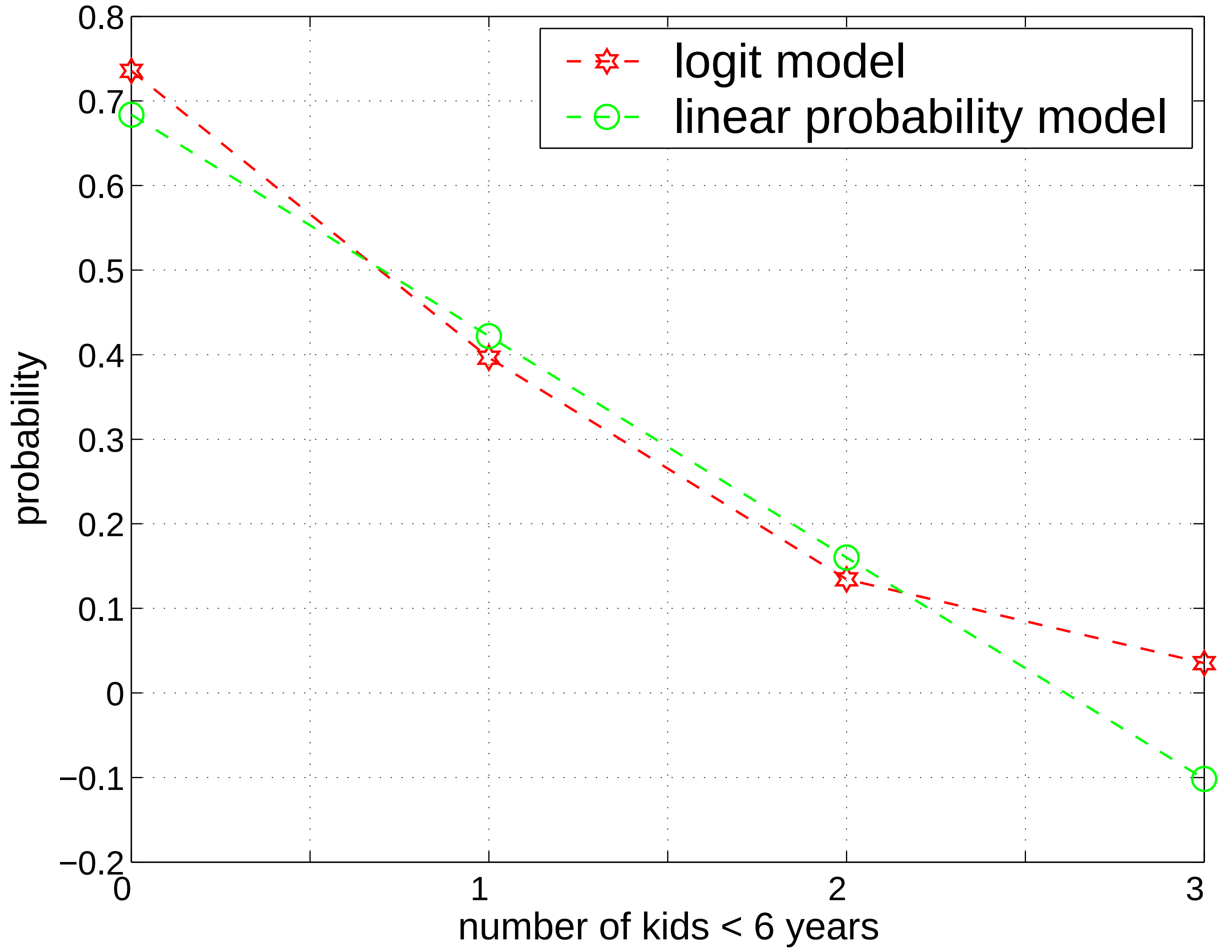
variable	LPM (OLS)	Logit	Probit	Logit/Probit
constant	0.5855	0.4255	0.2701	1.5753
	(0.1514)	(0.8604)	(0.5086)	
other income	-0.0034	-0.0213	-0.0120	1.7753
	(0.0015)	(0.0084)	(0.0048)	
education	0.0380	0.2212	0.1309	1.6896
	(0.0072)	(0.0434)	(0.0253)	
experience	0.0395	0.2059	0.1233	1.6690
	(0.0058)	(0.0321)	(0.0187)	
experience^2	-0.0006	-0.0032	-0.0019	1.6714
	(0.0002)	(0.0010)	(0.0006)	
age	-0.0161	-0.0880	-0.0529	1.6655
	(0.0024)	(0.0146)	(0.0085)	
# kids under 6 years	-0.2618	-1.4434	-0.8683	1.6622
	(0.0316)	(0.2036)	(0.1185)	
# kids 6–18	0.0130	0.0601	0.0360	1.6696
	(0.0135)	(0.0748)	(0.0435)	
$g(\bar{\boldsymbol{x}}\widehat{\boldsymbol{\beta}})$	1	0.2431	0.3906	$\frac{1}{1.6063}$
$\frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{x}_i\widehat{\boldsymbol{\beta}})$	1	0.1786	0.3008	$\frac{1}{1.6842}$
$\overline{cp}$	0.7344	0.7357	0.7344	–
$\tilde{cp}$ (25)	1.4424	1.4477	1.4439	–

- For example, consider a 30 year old women with average other sources of household income, average education, 5 years of experience, and no kids between 6 and 18.

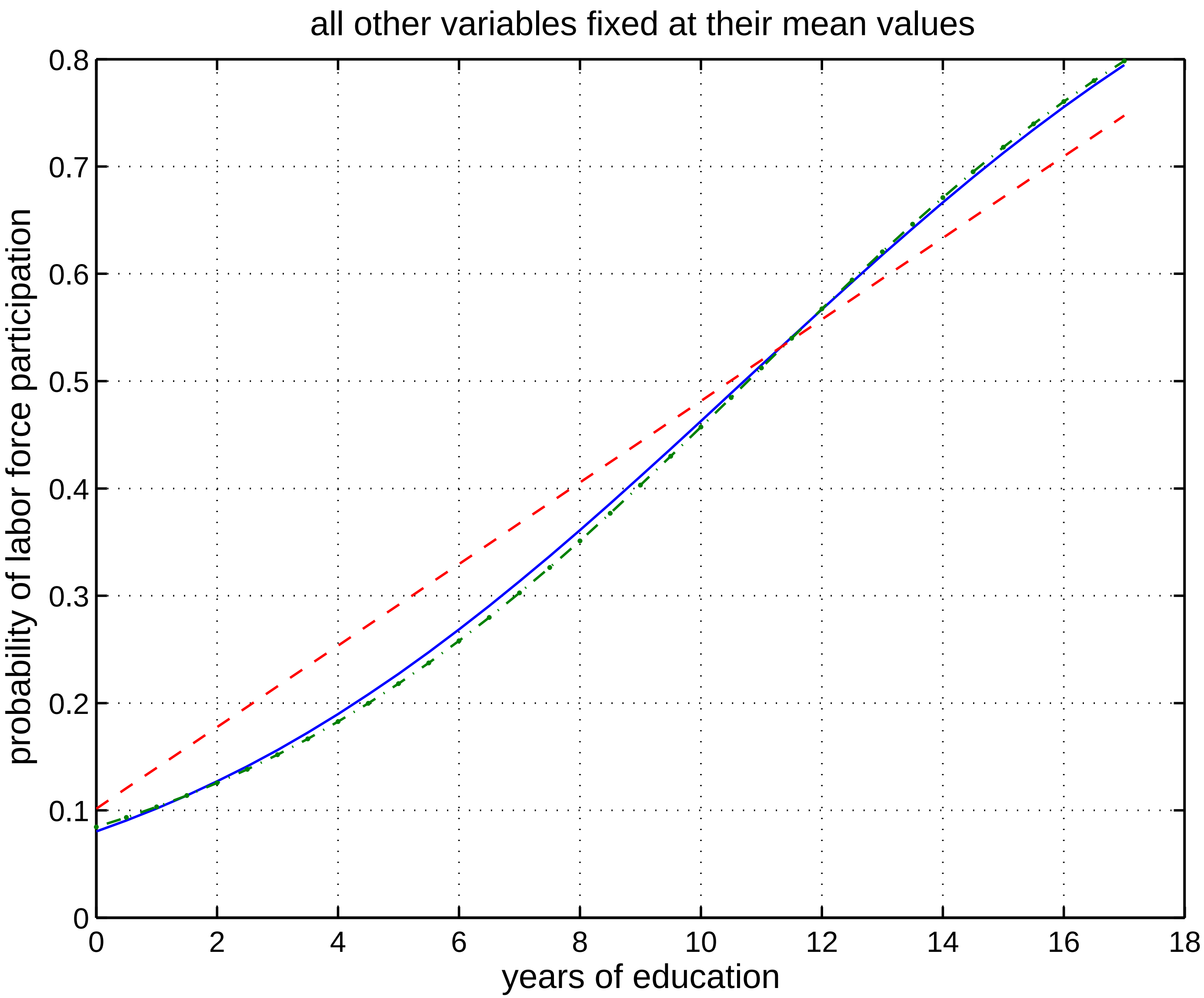


- There is a diminishing marginal effect of the third young child (whereas in the LPM, the marginal effect is constant).

- But for the relevant range (cf. the numbers in the histogram on previous slide) the probabilities are not that different.



red (dashed): linear; blue (solid): probit; green (dash–dotted): logit



Likelihood Ratio Tests (LRT)

- To illustrate the use of the LRT, we estimate several logit models via maximum likelihood:

Table 2: ✓ indicates whether a variable is in the model

variable	Model0	Model1	Model2	Model3	Model4
constant (β_0)	✓	✓	✓	✓	✓
other income (β_1)		✓	✓	✓	✓
education (β_2)		✓	✓	✓	✓
experience (β_3)		✓	✓	✓	
experience^2 (β_4)		✓	✓	✓	
age (β_5)		✓	✓	✓	✓
# kids under 6 years (β_6)		✓	✓		✓
# kids 6–18 (β_7)		✓			✓
$\log L$	-514.87	-401.76	-402.09	-432.78	-454.18

- We first do a test for the overall significance of the (logistic) regression, i.e., test Model0 against Model1, where the null hypothesis is

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0. \tag{21}$$

- The likelihood ratio test statistic is

$$2 \times \{-401.76 - (-514.87)\} = 226.2161.$$

- Comparing this with the χ^2 distribution with 7 degrees of freedom, we observe that this is significant at any reasonable level. For example, the critical value at the 5% level is 14.0671.

- As the t statistic of kids between 6 and 18 is not significant, we expect to obtain the same result from an LRT.

- The test statistic for Model2 against Model1 ($H_0 : \beta_7 = 0$) is

$$2 \times \{-401.76 - (-402.09)\} = 0.6480,$$

which is indeed not significant given the asymptotically valid χ^2 statistic with one degree of freedom.

- If we were interested in whether children (older or younger) have an impact at all, we test Model3 against Model1 ($H_0 : \beta_6 = \beta_7 = 0$), which gives rise to two degrees of freedom, with test statistic

$$2 \times \{-401.76 - (-432.78)\} = 62.0225,$$

which is highly significant at essentially any significance level.

- Same is true for *experience* (Model4 against Model1), again using a $\chi^2(2)$ distribution.

Table 3: Quantiles of the χ^2 distribution (ν denotes degrees of freedom)

ν	0.9	0.95	0.975	0.99
1	2.7055	3.8415	5.0239	6.6349
2	4.6052	5.9915	7.3778	9.2103
3	6.2514	7.8147	9.3484	11.3449
4	7.7794	9.4877	11.1433	13.2767
5	9.2364	11.0705	12.8325	15.0863
6	10.6446	12.5916	14.4494	16.8119
7	12.0170	14.0671	16.0128	18.4753
8	13.3616	15.5073	17.5345	20.0902
9	14.6837	16.9190	19.0228	21.6660
10	15.9872	18.3070	20.4832	23.2093
11	17.2750	19.6751	21.9200	24.7250
12	18.5493	21.0261	23.3367	26.2170
13	19.8119	22.3620	24.7356	27.6882
14	21.0641	23.6848	26.1189	29.1412
15	22.3071	24.9958	27.4884	30.5779

Goodness–of–Fit

- McFadden **pseudo** R^2,

$$R^2_{pseudo} = 1 - \frac{\log L(\text{estimated model})}{\log L(\beta_1 = \beta_2 = \cdots = \beta_k = 0)}, \tag{22}$$

  i.e., the maximized log–likelihood of a given model is compared with that of a model without explaining variables (constant probability).

- Note that, since the log–likelihood for these models is always negative, a better fit is indicated by it being closer to zero in magnitude.

- Quantity (22) is zero if the independent variables have no explanatory power.

- It is unity for a "perfect fit" in the sense that $p(\boldsymbol{x}_i) = 1$ for $y_i = 1$ and $p(\boldsymbol{x}_i) = 0$ for $y_i = 0$, since then the log–likelihood of the model is zero.

- But this will never happen with logit and probit (except in pathological cases, when it signals that something is wrong with the model).

- For example, if the sign of x_1 is a perfect predictor of y, then $\beta_1 \to \pm\infty$ leads to a "perfect fit".

- This may happen, e.g., if one mistakenly includes a dummy variable that is nearly identical to the dependent variable.

- Thus, a "perfect fit" just indicates a flaw in the model.

- Note that the restricted likelihood in (22) is just

$$\log L(\beta_1 = \beta_2 = \cdots = \beta_k = 0) = n_1 \log\left(\frac{n_1}{n}\right) + (n - n_1)\log\left(\frac{n - n_1}{n}\right),$$

where

$$n_1 = \sum_{i=1}^{n} y_i.$$

- This is because, with a constant probability $\mathsf{P}(y = 1)$, $\widehat{\beta}_0$ will be determined so that

$$\mathsf{P}(y = 1) = \frac{e^{\widehat{\beta}_0}}{1 + e^{\widehat{\beta}_0}} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{n_1}{n},$$

see Equation (20).

- Clearly $0 \leq R^2_{pseudo} < 1$.

- An alternative way to assess goodness–of–fit is by considering the **percent correctly predicted**.

- For example, our predictor may be

$$\widehat{y}_i = 1 \text{ if } p(\boldsymbol{x}_i) > 0.5 \text{ and } \widehat{y}_i = 0 \text{ otherwise.} \tag{23}$$

- This can be done for the linear probability model even in the case that some probabilities are not in $[0, 1]$.

- The percent correctly predicted of a model is the percentage of times for which $\widehat{y}_i = y_i$, which can be compared with that of a "naive" model (intercept only), where

$$\widehat{y}_i = \begin{cases} 1 & \text{for all } i \text{ if } n_1/n > 0.5 \\ 0 & \text{for all } i \text{ otherwise.} \end{cases} \tag{24}$$

- The percent correctly predicted for each of the outcomes separately is also of interest.

- In particular, if n_1/n is rather large, then the naive model will have a high percent correctly predicted, but the challenge is then to predict the (unconditionally) low–probability events when $y = 0$.

- Let n_{ij}, $i, j = 0, 1$, be the number of times that $y_i = i$ has been predicted and $y_i = j$ has been observed.

Table 4: Predicted and observed outcomes

	$y = 0$	$y = 1$
$\widehat{y} = 0$	n_{00}	n_{01}
$\widehat{y} = 1$	n_{10}	n_{11}
$\sum$	n_0	n_1

- The overall percent correctly predicted (cp) is

$$\overline{cp} = \frac{n_{00} + n_{11}}{n} = \frac{n_{00} + n_{11}}{n_{00} + n_{10} + n_{01} + n_{11}},$$

and those for the outcomes $y = 0$ and $y = 1$ are

$$cp_0 = \frac{n_{00}}{n_0} = \frac{n_{00}}{n_{00} + n_{10}}, \quad \text{and} \quad cp_1 = \frac{n_{11}}{n_1} = \frac{n_{11}}{n_{01} + n_{11}},$$

respectively.

- Measure

$$\tilde{cp} := cp_0 + cp_1 \tag{25}$$

should exceed unity if the model predicts better than the naive specification (for which it is always equal to one).

- These measures can also be used to evaluate out–of–sample forecasts, which are of predominant interest in some applications.

- Henriksson and Merton (1981) develop a test based on (25) for assessing statistical significance of market timing skills of investment managers.[4]

[4]On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills, *Journal of Business*, 54, 513–533.

- As an example, consider the logit model for the labor market participation of married women.

Table 5: Predicted and observed outcomes for the labor force participation example

	$y=0$	$y=1$
$\widehat{y}=0$	207	81
$\widehat{y}=1$	118	347
$\sum$	325	428

- Thus, the percent correctly predicted is

$$\overline{pc} = \frac{n_{00}+n_{11}}{n_{01}+n_{10}+n_{00}+n_{10}} = \frac{207+347}{753} = 0.7357. \tag{26}$$

- The overall frequency of ones is

$$\frac{1}{753}\sum_{i=1}^{753} y_i = \frac{n_{11}+n_{01}}{753} = \frac{347+81}{753} = \frac{428}{753} = 0.5684, \tag{27}$$

which is the percent correctly predicted of a model with an intercept only.

- Moreover

$$cp_0 = \frac{n_{00}}{n_0} = \frac{207}{325} = 0.6369, \quad \text{and} \quad cp_1 = \frac{n_{11}}{n_1} = \frac{347}{428} = 0.8107. \quad (28)$$

so

$$\tilde{pc} = cp_0 + cp_1 = 1.4477. \quad (29)$$

- The intercept–only model would correctly predict all of the ones but none of the zeros (because the overall frequency (27) > 0.5), so for this "model" always $\tilde{pc} = 1$ in (29).