Comparing Effects with Logit Models

G. Rohwer

Version 1 **May 2011**

Contents

In social research one is often interested in investigating how a binary outcome variable depends on conditions. A widely used tool is the logit model which connects the probability distribution of a binary outcome variable via a nonlinear function with values of explanatory variables. An ongoing debate concerns the comparison of explanatory variables across nested models. The focus often is on model parameters (and log odd ratios which are directly connected with such parameters). The present paper argues that this focus can be easily misleading when comparing models and instead takes effects, defined as differences of conditional expectations, as the quantities of main interest.

The first section introduces the notation. The second section briefly criticizes the idea that parameters in reduced models should be viewed as 'biased estimates' of corresponding parameters in more comprehensive models. The third section discusses how to compare effects across models.

1 Defining Effect Sizes

1.1 A single explanatory variable

The most simple model connects just two variables and can be graphically depicted as $X \longrightarrow\!\!\!\!\!\!\rightarrow Y$. The double arrow heads are intended to indicate a stochastic relationship. As an example, one can think that the model concerns the dependence of children's success in school ($Y = 1$ if success, $Y = 0$ otherwise) on parents' educational level represented by X (e.g. 0 low, 1 high). The model assumes a probabilistic relationship, that is, the probability distribution of Y is assumed to depend on values of X. Since Y is binary, one can simply use the function

$$x \longrightarrow \Pr(Y\!=\!1 \,|\, X\!=\!x) = \mathrm{E}(Y \,|\, X\!=\!x) \tag{1}$$

This function shows how the expectation of Y depends on values of X. In the present paper, the interest concerns effects, that is, effects of *changes (differences) of values* of X on the distribution of Y. I use the notation

$$\Delta^s(Y; X[x', x'']) := \mathrm{E}(Y \,|\, X\!=\!x'') - \mathrm{E}(Y \,|\, X\!=\!x') \tag{2}$$

and refer to this as the stochastic effect of a change in the variable X from x' to x''. Notice that, in general, the relationship is not linear, implying that the effect not only depends on the amount of change, $(x'' - x')$, but also on x'. Except for the special case when X is binary, effects cannot be represented by single numbers.

I now consider a logit model as a parametric representation of the functional relationship (1). It is based on using a logistic link function

$$F(v) := \frac{\exp(v)}{1+\exp(v)} \tag{3}$$

to approximate (1), resulting in the model

$$\mathrm{E}(Y \mid X = x) \approx F(\alpha + x\beta_x) \tag{4}$$

(Using here an equality sign instead of $\approx$ would presuppose that the model is correctly specified. However, in particular when thinking of the possibility that further explanatory variables should be included, this cannot be assumed just from the beginning.) The effect defined in (2) is then approximated by

$$\Delta^a(Y; X[x', x'']) := F(\alpha + x''\beta_x) - F(\alpha + x'\beta_x) \tag{5}$$

where the 'a' is intended to indicate 'approximation'.

1.2 Adding another explanatory variable

I now consider the addition of another explanatory variable, say Z. To continue with the example, one can imagine that the child's success (Y) not only depends on the parents' educational level (X), but also on the school type (Z). Graphically depicted, the model then is $(X, Z) \longrightarrow\!\!\!\!\rightarrow Y$, and the corresponding functional relationship is

$$(x, z) \longrightarrow \Pr(Y = 1 \mid X = x, Z = z) = \mathrm{E}(Y \mid X = x, Z = z) \tag{6}$$

In contrast to the simple model (1), effects of X can now be defined only conditional on values of Z:

$$\Delta^s(Y; X[x', x'']; Z = z) := \mathrm{E}(Y \mid X = x'', Z = z) - \mathrm{E}(Y \mid X = x', Z = z) \tag{7}$$

Table 1 Fictitious data for the illustration.

x	z	y	cases
0	0	0	600
0	0	1	600
0	1	0	240
0	1	1	560
1	0	0	40
1	0	1	160
1	1	0	80
1	1	1	720

Again, one can use a logit model as a parametric approximation to (6). Including an interaction effect, the model is

$$\mathrm{E}(Y \mid X = x, Z = z) \approx F(\alpha^* + x\beta_x^* + z\beta_z^* + xz\beta_{xz}^*) \tag{8}$$

(Of course, since this model differs from (4), also the parameters must be distinguished.) The parameterized effect then is

$$\Delta^a(Y; X[x', x'']; Z = z) := F(\alpha^* + x''\beta_x^* + z\beta_z^* + x''z\beta_{xz}^*) - F(\alpha^* + x'\beta_x^* + z\beta_z^* + x'z\beta_{xz}^*) \tag{9}$$

To illustrate, I use the data shown in table 1. Y represents the child's success ($Y = 1$), X represents the parents' educational level (0 low, 1 high), and Z represents the school type (0 or 1). Nonparametric estimates can be derived directly from the observed frequencies as shown in the following table:

x	z	$\mathrm{E}(Y \mid X = x, Z = z)$
0	0	0.5
0	1	0.7
1	0	0.8
1	1	0.9

(10)

(Using the logit model (8) would result in identical estimates. Leaving out the interaction effect would lead to slightly different values.) One then finds the effects:

$$\Delta^s(Y; X[0,1]; Z=0) = 0.8 - 0.5 = 0.3 \tag{11}$$
$$\Delta^s(Y; X[0,1]; Z=1) = 0.9 - 0.7 = 0.2$$

showing how the effect of a difference in parents' educational level depends on the school type.

2 Parameters in Reduced Models

The parameters β_x and β_x^* cannot immediately be compared and must be considered as belonging to different models. In order to stress this point, I briefly criticize the idea that parameters in reduced models should be viewed as 'biased estimates' of corresponding parameters in more comprehensive models. To illustrate the argument, I use an example taken from Mood (2010: 71). The example assumes a correctly specified logit model

$$\mathrm{E}(Y \mid X=x, Z=z) = F(x\beta_x + z\beta_z) \tag{12}$$

Values of X and Z are taken from two independent standard normal distributions. Written with a latent variable, the model is

$$Y_l := x\beta_x + z\beta_z + L \tag{13}$$

where L is a random variable with a standard logistic distribution, defined by $\Pr(L \leq l) = F(l)$, implying that $Y_l \geq 0 \iff Y = 1$ (based on the symmetry of L).[1] Mood uses this model with $\beta_x = 1$ and three different values for β_z. I begin with assuming that also $\beta_z = 1$.

One can then consider a model which omits Z. Taken as a standard logit model, it can be written in terms of a latent variable as

$$Y_l^r := x\beta_x^r + L \tag{14}$$

Estimating this model with simulated data, Mood finds $\beta_x^r = 0.84$, which is obviously less than $\beta_x = 1$, and concludes that the estimate is 'clearly

biased towards zero' (p. 71).[2] However, this statement presupposes that (14) has the task to estimate β_x as defined by (12), and this is at least debatable.

Viewing (14) as a reduced version of (12), it provides estimates of probabilities which have a clear and sensible meaning: they approximate probabilities which are averages w.r.t. the (a presupposed) distribution of the omitted variable. In the example, $F(x\beta_x^r)$ approximates

$$\mathrm{E}_Z(\Pr(Y = 1 \mid X = x, Z)) := \int F(x\beta_x + z\beta_z)\,\phi(z)\,dz \tag{15}$$

where $\phi(z)$ denotes the standard normal density function. This shows that β_x^r is the correct parameter to be used when being interested in approximating the probabilities defined in (15). Instead intending to estimate β_x would not be sensible. In fact, knowing β_x without also knowing β_z would be almost useless because $F(x\beta_x)$ provides a correct estimate only for the special case where $z = 0$.

Note that the proposed interpretation of the reduced model (14) holds independently of the size of β_z. For example, assuming $\beta_z = 2$, Mood finds $\beta_x^r = 0.61$, even smaller than 0.84, but $F(x\beta_x^r)$ is still an (actually very good) approximation to the average w.r.t. the omitted variable as defined in (15).

3 Comparing Effects Across Models

I now consider the question how to compare the effects of X across the two models, (1) and (6).

3.1 Consideration of marginal effects

Obviously, an immediate comparison is not possible because in model (6) effects also depend on values of Z. One therefore needs to define marginal effects based on a reduced version of (6). This requires to think of Z as a variable that has an associated distribution. Taking into account that the distribution of Z could depend on values of X, one can start from the equation

$$\mathrm{E}(Y \,|\, X = x) = \sum\nolimits_z \mathrm{E}(Y \,|\, X = x, Z = z)\, \Pr(Z = z \,|\, X = x) \tag{16}$$

(Here I assume that Z is a discrete variable as it is the case in the school example; if Z is continuous, as it is the case in Mood's example, one would use an integral instead of the sum.) The effect defined in (2) can then be expressed as

$$\begin{aligned} \Delta^s(Y; X[x', x'']) = &\sum\nolimits_z \mathrm{E}(Y \,|\, X = x'', Z = z)\, \Pr(Z = z \,|\, X = x'') \\ - &\sum\nolimits_z \mathrm{E}(Y \,|\, X = x', Z = z)\, \Pr(Z = z \,|\, X = x') \end{aligned} \tag{17}$$

A simpler formulation is possible if Z is independent of X. The marginal effect is then an average of the conditional effects:

$$\Delta^s(Y; X[x', x'']) = \sum\nolimits_z \Delta^s(Y; X[x', x''], Z = z)\, \Pr(Z = z) \tag{18}$$

Note, however, that even in this case the effect of X depends on the distribution of Z. To illustrate, I use Mood's example where Z has a normal distribution independent of X. Corresponding to (18) one finds the approximation

$$\Delta^s(Y; X[x', x'']) \approx \int (F(x''\beta_x + z\beta_z) - F(x'\beta_x + z\beta_z))\, \phi(z)\, dz$$

showing how effects of X also depend on the distribution of Z. For example, assuming $Z \sim \mathcal{N}(0, 1)$, one finds $\Delta^s(Y; X[0, 1]) \approx 0.7 - 0.5 = 0.2$, but the effect will increase when the variance of Z becomes smaller and, conversely, will decrease when the variance becomes larger.

3.2 Correlated explanatory variables

In social research, explanatory variables are most often correlated, and the simple relationship (18) does not hold. A first problem then concerns how to think of correlations between observed explanatory variables. A further problem that will be deferred to a later section concerns possibly relevant omitted variables which, presumably, are correlated with already included explanatory variables.

How to take into account correlations between observed explanatory variables depends on the purpose of the model to be estimated. To facilitate the discussion, I now explicitly distinguish between exogenous and endogenous variables of a model. Endogenous variables are stochastic variables having conditional distributions which depend on values of other variables of a model; such variables will be marked by a single dot. In contrast, exogenous variables do not have a distribution and only serve to formulate conditions; they will be marked by two dots.[3]

One purpose of a model could be to describe the relationship between a dependent and several explanatory variables as found in a given data set (and assumed to exist in a correspondingly defined population). Given this purpose, one can ignore correlations between explanatory variables and, assuming two such variables, refer to a model as follows:

$$\begin{tikzcd} \ddot{X} \arrow[r] & \dot{Y} \\ \ddot{Z} \arrow[ur] & \end{tikzcd} \tag{19}$$

The model only concerns the dependency of the probability distribution of $\dot{Y}$ on values of the two explanatory variables and does not entail anything about relationships between these variables. In other words, the explanatory variables are treated as exogenous variables without associated distributions; and this entails that the model cannot be used to think about correlations between these variables. Of course, the model can be estimated also with data exhibiting correlations between the explanatory variables. Think for example of the data in table 1 where the statistical variables corresponding to $\ddot{X}$ and $\ddot{Z}$ are correlated.

Another purpose of a model could be to investigate effects of variables as defined in the first section. For example, one might be interested in the question how the expectation of $\dot{Y}$ (the child's success) depends on a change, or difference, in the variable $\ddot{X}$ (the parents' educational level). Obviously, the model (19) cannot be used to answer this question because the effect also depends on values of $\ddot{Z}$. The observation of correlations between explanatory variables then leads to an important question: Can values of $\ddot{Z}$ be fixed when referring to the effect of a change in the value of $\ddot{X}$?

Of course, given a function like (6), one can easily think of changes in values of $\ddot{X}$, and consequently of effects of $\ddot{X}$, while holding $\ddot{Z} = z$ fixed. However, in a more relevant understanding the question does not concern possibilities to manipulate formulas, but the behavior of the social processes which actually generate values of the variables represented in a model (see Rohwer 2010: 82ff). In this understanding, the question motivates to consider more comprehensive models which include assumptions about relationships between explanatory variables.

There are several different possibilities. Here I briefly consider two. The first one can be depicted as follows:

$$\begin{tikzcd} \ddot{X} \arrow[r, ->>] \arrow[d, ->>] & \dot{Y} \\ \dot{Z} \arrow[ur, ->>] & \end{tikzcd} \tag{20}$$

$\ddot{X}$ is still an exogenous variable, but $\ddot{Z}$ has now changed into an endogenous stochastic variable, $\dot{Z}$. In addition to the function (6), there is now another function

$$x \longrightarrow \Pr(\dot{Z}=z \mid \ddot{X}=x) \tag{21}$$

showing how the distribution of $\dot{Z}$ depends on values of $\ddot{X}$. In our example, based on the data in table 1, one finds $\Pr(\dot{Z} = 1 \mid \ddot{X} = 0) = 0.4$ and $\Pr(\dot{Z}=1 \mid \ddot{X}=1) = 0.8$, showing how the child's school type depends on the parents' educational level.

Given this model, a change in $\ddot{X}$ entails a change in the distribution of $\dot{Z}$. So it is not possible to fix $\dot{Z} = z$ when considering an effect of $\ddot{X}$, and this entails that effects of $\ddot{X}$ and $\dot{Z}$ cannot be separated.[4] It follows that one can only define a total effect of a change in $\ddot{X}$, and this total effect equals the effect (17) which is derived from a reduced model resulting from omitting $\dot{Z}$; in the example: $\Delta^s(\dot{Y}; \ddot{X}[0,1]) \approx 0.88 - 0.53 = 0.3$. In other words, assuming the model (20), marginalization w.r.t. $\dot{Z}$ is required in order to define the effect of interest.

Table 2 Modification of the data in table 1.

x	z	y	cases
0	0	0	257
0	0	1	257
0	1	0	360
0	1	1	840
1	0	0	17
1	0	1	69
1	1	0	120
1	1	1	1080

The situation is less clear when considering a model in which the explanatory variable of interest is endogenous, for example:

$$\begin{tikzcd} \dot{X} \arrow[r, ->>] & \dot{Y} \\ \ddot{Z} \arrow[u, ->>] \arrow[ur, ->>] & \end{tikzcd} \tag{22}$$

While the model can well be used to define an effect of $\ddot{Z}$, there is no straightforward answer to the question how to define an effect of a change in $\dot{X}$. One could fix $\ddot{Z} = z$ and nevertheless think of different values of $\dot{X}$ to be used for the calculation of an effect; but such effects are conditional on $\ddot{Z} = z$ and already available in the model (19). On the other hand, without a distribution for $\ddot{Z}$, one cannot derive a marginal model. Thinking instead of a variable $\dot{Z}$ that can be assumed to have a distribution, the marginal effect of $\dot{X}$ depends on the actual choice. For example, deriving the distribution of $\dot{Z}$ from the data in table 1, one finds the marginal effect 0.3. Using instead the data in table 2 (which entail the same functional relationships as specified in (22)), one finds 0.25. Given this model, it seems best not to attempt to attribute to $\dot{X}$ a definite (context-independent) effect.

3.3 Continuing with Mood's example

For further illustration of correlated explanatory variables I use a modification of Mood's example in which values of $\ddot{X}$ and $\ddot{Z}$ are taken from a bivariate normal distribution with correlation $\rho \neq 0$. One can again consider the reduced model (14). For example, assuming $\rho = 0.5$, one finds $\beta_x^r = 1.32$, now larger than $\beta_x = 1$ (this also shows that omitting a variable not always leads to an 'attenuated parameter'). As I have argued above, this is not a 'biased estimate' of β_x, but must be viewed as a parameter of the reduced model (14). In this understanding, β_x^r can be used to calculate a sensible approximation to the marginal expectation (16). In the example, $\mathrm{E}(\dot{Y}|\ddot{X}=0) \approx F(0) = 0.5$, and $\mathrm{E}(\dot{Y}|\ddot{X}=1) \approx F(1.32) = 0.79$.

These values could be used to calculate the effect $\Delta^s(\dot{Y};\ddot{X}[0,1]) \approx 0.79 - 0.5 = 0.29$, obviously larger than the value 0.2 that was calculated for Mood's original model with uncorrelated explanatory variables. In order to understand the difference, one needs an extended model that allows one to interpret the correlation between the two explanatory variables. I consider model (20) which is based on the assumption that the distribution of $\dot{Z}$ depends on values of $\ddot{X}$. In the example, the conditional density of $\dot{Z}$, given $\ddot{X} = x$, is a normal density $\phi(z;\mu,\sigma)$ with $\mu = x\rho$ and $\sigma = \sqrt{1-\rho^2}$, entailing that $\ddot{X}$ and $\dot{Z}$ are connected by a linear regression function.

This allows an easy interpretation of the effect. For example, if the value of $\ddot{X}$ changes from 0 to 1, this entails a change in the mean value of $\dot{Z}$ from 0 to ρ, and, if $\rho > 0$, the effect becomes larger compared with a situation where $\rho = 0$. In any case, assuming that $\dot{Z}$ depends on $\ddot{X}$ allows one to attribute the total effect to the change in $\ddot{X}$.

3.4 Comparing variables across models

Neither parameters nor effects can directly be compared across models. It is well possible, however, to compare the role played by explanatory variables. For example, one can compare the role played by $\ddot{X}$ across the models (4) and (8). One can begin with a look at the estimated parameters. Using the data in table 1, one finds $\hat{\beta}_x = 1.67$ and $\hat{\beta}_x^* = 1.39$.

Table 3 Modification of the data in table 1.

x	z	y	cases
0	0	0	400
0	0	1	600
0	1	0	100
0	1	1	400
1	0	0	200
1	0	1	300
1	1	0	200
1	1	1	800

This does not show, however, that $\ddot{X}$ is 'less important' when one 'controls for' values of $\ddot{Z}$. The total effect of $\ddot{X}$ is essentially identical in both models (differences only result from the parameterization of the models). Of course, the enlarged model provides an opportunity to think of this total effect in a more refined way.

Even if, by including a further variable, a parameter becomes zero one cannot conclude that the corresponding variable has no effect. To illustrate, I use the data in table 3. Using these data to estimate (4) and (8), one finds $\hat{\beta}_x = 0.32$ and $\hat{\beta}_x^* = 0$. This shows that the effect of $\ddot{X}$, conditional on values of $\ddot{Z}$, is zero. There nevertheless is a relevant total effect of $\ddot{X}$, namely $\Delta^s(\dot{Y};\ddot{X}[0,1]) \approx 0.73 - 0.67 = 0.06$.

How to interpret this effect depends on assumptions about the relationship between $\ddot{X}$ and $\ddot{Z}$. In our example, assuming that the choice of a school type depends on the parents' educational level, one would use model (20). The total effect of $\ddot{X}$ can then be explained by the difference in the probabilities $\Pr(\dot{Z}=1|\ddot{X}=0) = 1/3$ and $\Pr(\dot{Z}=1|\ddot{X}=1) = 2/3$.

3.5 Unobserved Heterogeneity

So far, I have assumed observed explanatory variables. Further questions concern 'unobserved heterogeneity'. I take this expression to mean that there are further unobserved explanatory variables that should be included in a model. So the question arises how the model would change if these additional variables would have been included. A reliable answer is obviously not possible, but a few remarks can be derived from the foregoing discussion.

As before I only consider logit models and begin with assuming that the interest concerns conditional expectations,

$$\mathrm{E}(\dot{Y} | \ddot{X} = x) \approx F(\alpha + x\beta_x) \tag{23}$$

When hypothetically adding a further explanatory variable, say $\ddot{Z}$, one gets a more comprehensive model. However, in order to think of (23) as a reduced version of that model, one needs to think of $\ddot{Z}$ as a variable $\dot{Z}$ that can be assumed to have a distribution (given, e.g., by values of $\ddot{Z}$ if such values could be observed). Equation (16) then shows that $\mathrm{E}(\dot{Y} | \ddot{X} = x)$ can be viewed as a mean value w.r.t. the distribution of $\dot{Z}$; and consequently $F(\alpha + x\beta_x)$ can be viewed as an approximation to this mean value. As an illustration remember Mood's example. Not having observed $\ddot{Z}$, one can estimate only the reduced model (14), but this model correctly provides an approximation to the expectation defined in (15). As shown by (16), this remains true when $\dot{Z}$ is correlated with $\ddot{X}$.

The situation is more complicated when the interest concerns effects as defined in (2). First assume that the hypothetically included unobserved variable $\dot{Z}$ is independent of the variable $\ddot{X}$. As shown by (18), the effect derived from the reduced model can then be viewed as a mean of effects which additionally condition on value of $\dot{Z}$. Of course, the not observed effects $\Delta^s(\dot{Y}; \ddot{X}[x', x''], \dot{Z} = z)$ can have quite different, even positive and negative, values. For example, one can easily modify the data in table 1 to get conditional expectations as follows:

x	z	$\mathrm{E}(\dot{Y} \mid \ddot{X} = x, \ddot{Z} = z)$
0	0	0.7
0	1	0.8
1	0	0.6
1	1	0.9

entailing effects $\Delta^s(\dot{Y}; \ddot{X}[0,1]; \ddot{Z} = 0) = -0.1$ and $\Delta^s(\dot{Y}; \ddot{X}[0,1]; \ddot{Z} = 1) = 0.1$. The observed effect is then positive if $\Pr(\dot{Z} = 1) > 0.5$ and negative otherwise.

When the omitted variable is correlated with observed explanatory variables a critical question concerns the sources of the correlation. To conceive of the observed effect of $\ddot{X}$ as a total effect requires the presupposition of a model in which the omitted variables functionally depends on $\ddot{X}$. Otherwise, as I have argued above, no easy interpretation of the observed effect seems possible.

Notes

[1]It is often said that the variance of the latent variable Y_l is 'not identified' (e.g., Allison 1999, Cramer 2007). This is true in the following sense: When starting from a regression model $Y_l = x\beta_x + z\beta_z + \epsilon$ with an arbitrary residual variable ϵ, and observations (values of Y) only provide information about the sign of Y_l, the variance of this variable cannot be estimated. The statement is misleading, however, when the latent variable is derived from a logit model. If Y_l is *defined* by (13), a variance of Y_l does exist only conditional on values of the explanatory variables, and is already known from the model's definition: $\mathrm{Var}(Y_l|X = x, Z = z) = \mathrm{Var}(L) = \pi^2/3$.

[2]For similar views see Allison (1999), Cramer (2007), Wooldridge (2002: 470).

[3]For a discussion of functional models based on this notation see Rohwer (2010).

[4]It is not even possible to clearly separate a direct and an indirect effect, see Rohwer (2010: 68f).

References

Allison, P. D. (1999). Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods & Research* 28, 186–208.

Cramer, J. S. (2007). Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specified Disturbances. *Oxford Bulletin of Economics and Statistics* 69, 545–555.

Mood, C. (2010). Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review* 26, 67–82.

Rohwer, G. (2010). *Models in Statistical Social Research*. London: Routledge.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press.