Lapai Journal of Economics; Volume 2, No.2; 2018
Print ISSN: 2659-028X
Online ISSN: 2659-0271
Published by Department of Economics, IBB University Lapai, Niger State, Nigeria

Ordered Logit Model, Proportional Odds Assumption and Marginal Effects: Demysfying the Derivational steps

Muktar Bala

Department of Economics
IBB University Lapai
Email: muktarbala@yahoo.co.uk

Abstract
Ordered models, such as the Ordered Logistic regression model, are used when the dependent variable of a model is categorical and ordinal. However, the literatures on Ordered models often skip the derivational steps of the model, which may make researchers apply the model as a dogma without knowing how the output of the model is expected to be. This study provides a detailed breakdown of the derivational steps of the model; the Proportional Odds assumption; the marginal effects and some practical examples, with a view to helping researchers have a better understanding of the output of the model when used in their studies.

Keywords: Ordered model, derivational steps, researchers
JEL Classification: C51, C52, C61

1. Introduction

Econometric models such as Ordered Logit and Ordered Probit models are used when the dependent variable is categorical (with more than two outcomes) and has ranking (Long & Freese, 2006). For instance, in a model whose dependent variable has three outcome categories, such that category 1 is less than category 2, and category 2 is less than category 3, then the parameters of such type of a model can be estimated using an Ordered Logit or Probit model (Long, 1997; Orme & Combs-Orme, 2009). Although both Ordered Logit and Probit models give similar results, some researchers often prefer Ordered Logit model due to the simplicity of its mathematical procedures (Gujarati & Porter, 2009), and hence the focus on it in this study.

The size, not the sign, of the output of an Ordered Logit model is normally a crude estimate, which is not interpretable in its crude form unless it is standardized by transforming it into marginal effects, odd ratios, or predicted probabilities (Long & Freese, 2006). According to Leeper (2017), marginal effect provides the most unified and intuitive way of interpreting discrete outcome models, such as the Ordered Logit model. However, before the output of the model is accepted as being valid, the data must meet the Proportional Odds assumption or Parallel line regression assumption

(McCullaph, 1980). Therefore, in order to fulfill the conditions of the model (Ordered Logit), and to be able to analyze and interpret the output of the model, researchers need to have a clear understanding of the derivational steps of the model (Long, 1997). They also need to have a clear understanding of the Proportional Odds assumption and the different standardization methods, such as the marginal effect (Long, 1997; McCullaph, 1980).

In deriving Ordered Logit model, most of the literatures either skip or fail to explain some ancillary steps (e.g Long, 1997; Cameroon & Trivedi, 2005; Gujarati & Porter, 2009; Long & Freese, 2006), thereby leaving some readers with no option than to apply the model as a dogma. However, this may not assist them in understanding the output of the model and in interpreting the result of their findings. In this study, effort is being made to break down the derivational steps of the Ordered Logit model, Proportional Odds assumption and marginal effects in the simplest form. This is with a view to assisting researchers have full understanding of the model and guide them in interpreting their findings. The article is presented in seven sections. Section one is the introduction, while literature was reviewed in section two. Section three provides the derivational steps of the Ordered Logit model and in section four, the derivation of the Proportional Odss assumption was discussed. Section five is devoted to deriving the marginal effects of the Ordered Logit model, while section six gives practical examples of the items discussed in the study. The conclusion is drawn subsequently in section seven.

2. Review of the Literature

Models are simplified algebraic representations of a scientific theory (Meyer, 1982). It means that models ought to represent theories in a simplified manner, given details of each step and how it was arrived at for a clear understanding of the theory it represents. However, the description of Ordered Logit models given by most econometric text books did not show the derivational steps. For instance, in Long and Freese (2006) the derivational steps of the Ordered Logit model were explained in the following statement: "….substituting……and using some algebra leads to the standard formula for the predicted probability in the Ordered Logit model". Similarly, in explaining the marginal effect of the model, Long and Freese (2006) only show the formula for the marginal effect without detailed explanation of its derivational steps.

In Cameron and Trivedi (2005), though the derivational steps were explained, the derivational steps of the extreme outcomes were not clearly explained. The derivational steps of the marginal effects were also lumped and not explained in detail. Gujarati and Porta (2009) provided only the introductory treatment of the model, without mathematical explanation. On the other hand, Long (1997) provided one of the most detailed explanations of the derivational steps of the model, however, effort was not made to explain clearly, how the cumulative distribution function (CDF) of the lowest outcome of the model was translated to zero and the highest outcome translated to one. Some simplifications using the lowest common multiple

(LCM) were also applied in the derivational process of the model, which were not reported in the text. Similarly, the derivational steps of the marginal effect of the model were not explained in detail.

3. Derivation of the Ordered Logit Model

In this section, the derivational steps of an Ordered Logit model are discussed. In order to understand the derivation, children's health status will be used as an example throughout this study. The children's health status is taken as the dependent variable of the Ordered Logit model. According to the World Health Organization [WHO] (O'donnell et al. 2008), children's health status can be categorized based on their Z-score values. Assuming that the Z-score values for some group of children were obtained, then WHO considers all children in the range of Z < -3 as severely malnourished; those in the range of -3 ≤ Z < -2 as moderately malnourished and those in the range of -2 ≤ Z < +2 as adequately nourished (Webb & Bhatia, 2005). However, in this study, let's assume that adequately nourished children are in the range of Z ≥ –2. This means that all children in the range of Z-score of -3.1 or less, are severely malnourished; those in the range of Z-score between -3 to -2.1 are moderately malnourished and those in the range of -2 or more are considered as adequately nourished. An example of the Z-score values for some children is shown in Example 1.

Example 1: -0.41, -1.31, -0.28, -0.81, -0.98, -0.45, -0.94, -0.44, -1.26, -0.72, -0.82,-0.23,-0.34, -4.3, -3.39, -3.02, -4.08, -5.11, -3.73, -7.88, -3.05, -3.12, -4.88, -3.00, -2.51, -2.62, -2.3, -2.89, 1.97, 1.59.

Assuming that the Z-score value of each child, as shown in Example 1, is represented by y^*, then y^* becomes the latent dependent variable in an Ordered Logit model. The y^* is a latent dependent variable because it would be hidden in the estimation process when estimating the parameters (betas) of the explanatory variables of the model.

However, Example 2 shows the categorized version of the children's Z-scores that was presented in Example 1.

Example 2:

$$\left.\begin{array}{c} -4.3, -3.39, -3.02, -4.08, -5.11 \\ -3.73, -7.88, -3.05, -3.12, -4.88, \end{array}\right\} = \text{Severely malnourished } = 1$$

$$\left.-3.0, -2.51, -2.62, -2.33, -2.89\right\} = \text{Moderately malnourished} = 2$$

$$\left.\begin{array}{c} -0.41, -1.31, -0.28, -0.81, -0.98 \\ -0.45, -0.94, -0.44, -.126, -0.72 \\ -0.82, -0.23, -0.34, 1.59, 1.97 \end{array}\right\} = \text{Adequately nourished } = 3$$

It is shown in Example 2 that all children with a Z-score value within the range of -3.1 or less, are severely malnourished and are coded as 1. Those in the range of Z-score between -3 to -2.1 are moderately malnourished and are coded as 2, while those in the range of Z-score of -2 or more are considered as adequately nourished and are coded as 3. Therefore, the coding of the latent dependent variable y^* into Z-score groups ranked as 1, 2 and 3, gives a dependent variable that is actually observed (not hidden) in the estimation process. This dependent variable is called the "observed dependent variable" and is represented by y. Consequently, the measurement equation for the physical health of child i (y_i) using an Ordered Logit model is presented in Equation (1).

y_i = M if $\tau_{M-1} \leq y_i^* < \tau_M$, for M = 1, 2, … j………………......……………….……....1

where y_i^* is the latent physical health status of child i. The latent physical health values are the measured Z-scores for each child, which are not observed, but are coded as 1, 2 and 3 in order to derive the observed physical health outcomes for each child (i.e y_i). The τ's are the cutoff points or the thresholds. The cutoff points define where each outcome begins. M is the total number of outcomes or categories of the dependent variable of the Ordered model. The total number of cutoff points (τ's) in any Ordered model is always the total number of outcomes (M) minus one (i.e $M -$ 1).

Figure 1 shows the probability distribution and the Cumulative Distribution Functions (CDF's) of the children's Z-scores. From the figure, it is shown that all children with Z-score values of less that -3 (i.e from -∞ to -3.1) are grouped under outcome 1. Also, all children with Z-score values of -3 to -2.1 are grouped under outcome 2, while all children with Z-score values of -2 to positive infinity (+ ∞) are grouped under outcome 3. Therefore, since there are three outcomes (M=3), the number of cutoff points (τ's) in this Ordered Logit model will be two (i.e $M - 1 = 3 - 1 = 2$). The cutoff points, as shown in Figure 1, are points c and e (i.e τ_1 and τ_2), which also correspond to -3 for τ_1 and -2 for τ_2.

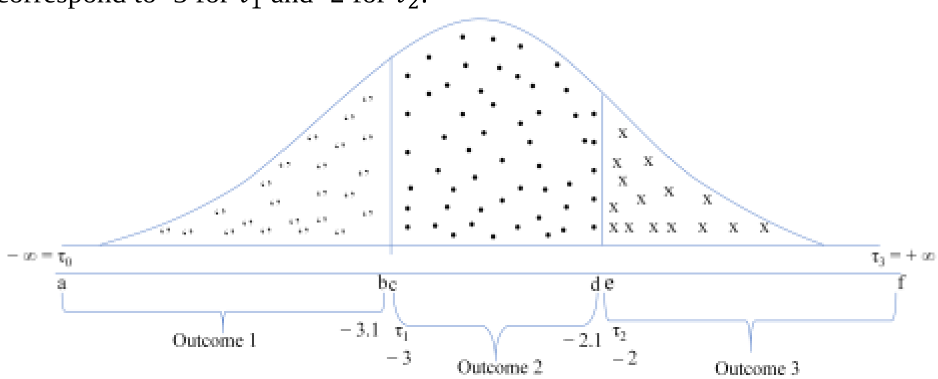

Figure 1: Probability distribution and the CDF's of the children's Z-scores.

The measurement model of Equation (1) is therefore shown in Equation (2).

$$y_i = \begin{cases} 1 \text{ if} -\infty = \tau_0 \leq y^* < \tau_1 \\ 2 \quad \text{if } \tau_1 \leq y^* < \tau_2 \\ 3 \text{ if } \tau_2 \leq y^* < \tau_3 = +\infty \end{cases} \quad \text{……………………………..……………………..…2}$$

In other words, Equation (2) is represented as Equation (3).

$$y_i = \begin{cases} 1 \text{ if } -\infty \leq \ y^* < -3 \\ 2 \text{ if } -3 \leq \ y^* \ < -2 \\ 3 \text{ if } -2 \leq \ y^* \ < +\infty \end{cases} \quad \ldots\ldots 3$$

However, following Long (1997), for child i, and for some set of independent variables, the structural Ordered Logit model is specified in Equation (4).

$$y_i^* = X_i \, \beta + \varepsilon_i \quad \ldots\ldots 4$$

where X_i is the vector of the regressors in the model, β is the vector of the parameters to be estimated while the Logistic random variable in the Ordered Logit model is represented by ε_i.

Therefore, the probability of observing outcome 1, which is the severely malnourished health status of the children, is given by Equation (5).

$$\Pr(y_i = 1 | \, X_i) = \Pr(-\infty = \tau_0 \leq y_i^* \ < \ \tau_1 | X_i) \ldots\ldots 5$$

Equation (5) shows that outcome 1 of the children's health status would be observed when y_i^* is greater than or equal to τ_0 (i.e -∞) but less than τ_1 (i.e less than -3) given X_i. The X_i is the set of explanatory variables that affect the health status of child i. Assuming y_i^* in Equation (4) is substituted into Equation (5), Equation (6) is derived:

$$\Pr(y_i = 1 | \, X_i) = \Pr(-\infty \leq \ X_i \, \beta \ + \ \varepsilon_i < \tau_1 | X_i) \ldots\ldots 6$$

By subtracting $X_i \, \beta$ within the inequalities, Equations (7) and (8) are derived:

$$\Pr(y_i = 1 | \, X_i) = \Pr(-\infty - X_i \, \beta \ \leq \ X_i \, \beta \ - X_i \, \beta + \ \varepsilon_i \ < \tau_1 - X_i \, \beta \, | X_i) \ldots\ldots 7$$

$$\Pr(y_i = 1 | \, X_i) = \Pr(-\infty - X_i \, \beta \ \leq \ \varepsilon_i \ < \tau_1 - X_i \, \beta \, | X_i) \ldots\ldots 8$$

According to Long (1997), CDF expresses the probability of a variable being less than some value. Therefore, from Equation (8), the CDF evaluated between −∞ and τ_1 (i.e between negative infinity and point b in Figure 1) yields the probability that a random variable X_i lies between the two values (−∞ and τ_1), as shown in Equations (9) and (10). Literally, this means that when all the Z-score values from negative infinity to a point immediately below the τ_1 threshold (i.e immediately below point *c* in Figure 1) are added together, they yield the probability that a particular explanatory variable (X_i) would determine the likelihood of child i being in the physical health outcome 1 (i.e severely malnourished health category).

$$\Pr(y_i = 1 | \, X_i) = \Pr\,(\varepsilon_i < \tau_1 - X_i \, \beta \, | X_i) - \Pr(\varepsilon_i \leq -\infty - X_i \, \beta\,) \ldots\ldots 9$$

Therefore, the probability that a child would fall in outcome 1 is the CDF of all the latent Z-score values of the children that are below the τ_1 threshold (i.e below point *c* or -3 in Figure 1), which is shown in the first part [$\Phi(\tau_1 - X_i \, \beta)$] of Equation (10). Where Φ is the symbol of the CDF.

$$\Pr(y_i = 1 | \, X_i) = \Phi\,(\tau_1 - X_i \, \beta) - \Phi\,(-\infty - X_i \, \beta) \ldots\ldots 10$$

However, the second part [$\Phi(-\infty - X_i \, \beta)$] of Equation (10) disappears in the sense that the CDF of a Logistic function is given as:

$$\text{Logistic CDF } (\Phi) = \frac{1}{1+e^{-z}} \ldots\ldots 10a$$

where $Z = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$. In the second part of Equation (9), $Z = -\infty - X_i \, \beta$. Therefore:

Logistic CDF $(\Phi) = \frac{1}{1+e^{-(-\infty - X_i \beta)}}$ ……………………………………..……………10b

Logistic CDF $(\Phi) = \frac{1}{1+e^{\infty + X_i \beta}}$ ……………………………………..……….......…10c

In Equation (10c), one of the components of the divisor, that is the exponent of infinity (e^{∞}) alone, is sufficient enough to make the outcome of the division almost equal to zero. Therefore, the second part of the Equation (10) disappears, leaving only the first part, as shown by Equation (11). Equation (11) gives the probability that the physical health of child i would be in the outcome 1 category, and is represented by points a to b in Figure 1.

$\Pr(y_i = 1 | X_i) = \Phi(\tau_1 - X_i \beta)$ ……………………..……………………..…………...…11

Similarly, the probability of selecting outcome 2, which is the moderately malnourished health status of the children, given some set of explanatory variables (X_i) is indicated in Equation (12):

$\Pr(y_i = 2 | X_i) = \Pr(\tau_1 \leq y_i^* < \tau_2 | X_i)$…………………………………………..............12

Equation (12) shows that outcome 2 of the children's health status would be observed, when y_i^* is greater than or equal to τ_1, but less than τ_2 (i.e -3 to -2.1 in Figure 1), given X_i. Assuming y_i^* in Equation (4) is substituted in Equation (12), we have Equation (13):

$\Pr(y_i = 2 | X_i) = \Pr(\tau_1 \leq X_i \beta + \varepsilon_i < \tau_2 | X_i)$……………......................…………...13

By subtracting $X_i \beta$, within the inequalities Equations (14) and (15) are obtained:

$\Pr(y_i = 2 | X_i) = \Pr(\tau_1 - X_i \beta \leq X_i \beta - X_i \beta + \varepsilon_i < \tau_2 - X_i \beta | X_i)$.........................14

$\Pr(y_i = 2 | X_i) = \Pr(\tau_1 - X_i \beta \leq \varepsilon_i < \tau_2 - X_i \beta | X_i)$…….…………………..……...15

Thus, the CDF evaluated between τ_1 and τ_2 gives the probability that a random variable X_i lies between the two values as shown in Equation (16) and (17):

$\Pr(y_i = 2 | X_i) = \Pr(\varepsilon_i < \tau_2 - X_i \beta | X_i) - \Pr(\varepsilon_i \leq \tau_1 - X_i \beta)$…..….………………..….16

$\Pr(y_i = 2 | X_i) = \Phi(\tau_2 - X_i \beta) - \Phi(\tau_1 - X_i \beta)$…………………………………..…17

Equation (17) gives the probability that the physical health of child i would be in the outcome 2 category. The first part $[\Phi(\tau_2 - X_i \beta)]$ of Equation (17) is represented by points a to d of Figure 1. The second part $[\Phi(\tau_1 - X_i \beta)]$ of Equation (17) is represented by points a to b. Therefore, when a to b is subtracted from a to d, the balance is c to d, which is the outcome 2 of the children's health status as shown by Figure 1.

Moreover, the probability of observing outcome 3, which is the adequately nourished health status, given some set of explanatory variables (X_i), is shown by Equation (18):

$\Pr(y_i = 3 | X_i) = \Pr(\tau_2 \leq y_i^* < \tau_3 | X_i)$………………...……18

when Equation (4) is substituted into Equation (18), we have Equation (19):

$\Pr(y_i = 3 | X_i) = \Pr(\tau_2 \leq X_i \beta + \varepsilon_i < \tau_3 | X_i)$…………………………………........…19

By subtracting $X_i \beta$, within the inequality and simplifying, we have Equation (20) and (21):

$\Pr(y_i = 3 | X_i) = \Pr(\tau_2 - X_i \beta \leq X_i \beta - X_i \beta + \varepsilon_i < \tau_3 - X_i \beta | X_i)$……………..….20

$\Pr(y_i = 3 | X_i) = \Pr(\tau_2 - X_i \beta \leq \varepsilon_i < \tau_3 - X_i \beta | X_i)$………....................................…21

Since the probability that a random variable is between two values is the difference between the CDF evaluated at the two values (Long, 1997), Equation (22) is derived from Equation (21), and Equation (23) is obtained by simplifying Equation (22):

$\Pr(y_i = 3 \mid X_i) = \Pr(\varepsilon_i < \tau_3 - X_i\,\beta \mid X_i) - \Pr(\varepsilon_i \leq \tau_2 - X_i\,\beta \mid X_i)$……….………..….22

$\Pr(y_i = 3 \mid X_i) = \Phi(\tau_3 - X_i\,\beta) - \Phi(\tau_2 - X_i\,\beta)$……………...…………..……….…..23

Equation (23) gives the probability that the physical health of child i would be in the outcome 3 category. The first part $[\Phi(\tau_3 - X_i\,\beta)]$ of Equation (23) is represented by points a to f in Figure 1, which is the probability distribution of the three outcomes. Therefore, the first part of Equation (23) is equal to 1 since the sum of probabilities in a probability distribution is equal to 1. The second part $[-\Phi(\tau_2 - X_i\,\beta)]$ of Equation (23) is represented by points a to d. Therefore, when a to d is subtracted from a to$f$, the balance is e to f, which is the outcome 3 of the children's health status. By simplifying Equation (23), we have Equation (24), which gives the probability that the physical health of child i would be in the outcome 3 category.

$\Pr(y_i = 3 \mid X_i) = 1 - \Phi(\tau_2 - X_i\,\beta)$……………...………………….……..…..…….24

Consequently, in order to estimate the effect of some set of explanatory variables (Xs) on the the probability of child i being in severely malnourished health category (outcome 1), moderately malnourished health category (outcome 2) or adequately nourished health status (outcome 3), Equations (11), (17) and (24) can be estimated.

Generally, therefore, Equation (1) can be applied to any M number of Ordered Logit outcome model. However, the lowest category $[\Pr(y_i = 1 \mid X_i)]$ is always the CDF from negative infinity to a value immediately before the first threshold [i.e $\Phi(\tau_1 - X_i\,\beta)$]. Also, the highest category is always the CDF of the entire probability distribution less the CDF from negative infinity to a value immediately before the $M - 1$ threshold [i.e $1 - \Phi(\tau_{m-1} - X_i\,\beta)$].

4. The Proportional Odds Assumption of the Ordered Models

The parameters, β's, as well as the threshold parameters, τ's in Equations (11), (17) and (24) can only be estimated using an Ordered Logit model when the data meet the parallel line assumption or proportional odds assumptions (McCullaph, 1980). The parallel line regression assumption states that given a particular set of probabilities of the outcome categories, then the log of the odds must form an arithmetic series for it to be estimated using an Ordered Logit model. According to Fullerton (2009), the first step is that the dependent variable with M categories is split into $M - 1$ binary Logit equations of log odds ratios, and then estimate simultaneously using the higher outcomes as the reference category.

In our own example, let us assume that the probability of severely malnourished health status is denoted by y_1, moderately malnourished health status is denoted by y_2, and adequately nourished health status is denoted by y_3. Therefore, since there are three outcomes ($M = 3$), the will be two Logit equations ($3 - 1 = 2$) of log odds ratios. These two Logit equations, according to Fullerton (2009), will be obtained by finding the ratio of outcome 1 versus 2 and 3, then outcome 1 and 2 versus 3. However, before getting the log odds ratios, there is need to find the odd ratios. The first Logit equation of the odds ratio is shown by Equation (25).

Odds of severely malnourished health: $\frac{y_1}{y_2+y_3}$ ……………………………..……………25

The odds (probability) of severely malnourished health (y_1), is the ratio of y_1 to the rest outcomes (i.e $y_2 + y_3$). In a binary Logit, assuming the probability of $y_1 = 1$, then the probability of the rest outcomes will be $1 - (y_2 + y_3)$. However, the probability of observing outcome1 (i.e $y_1 = 1$), is a function of the distribution of the error term. According to Gujarati and Porter (2009), the distribution of the error term in a Logitistic model is given by the function shown in Equation (25a):

$$\Pr(y_1 = 1) = f\left(\frac{1}{1+e^{-z}}\right) = \frac{1}{1+e^{-z}} = \frac{1}{1+\frac{1}{e^z}} \qquad \text{25a}$$

Since the LCM of $1+\frac{1}{e^z} = \frac{1+e^z}{e^z}$, therefore, by substituting the denominator of Equation (25a) with $\frac{1+e^z}{e^z}$ and operationalizing, we have Equation (25b):

$$\Pr(y_1 = 1) = \frac{1}{\frac{1+e^z}{e^z}} = f\left(\frac{e^z}{1+e^z}\right) \qquad \text{25b}$$

Consequently, by substituting y_1 in Equation (25) with $\frac{e^z}{1+e^z}$, the denominator of Equation (25) will be $1-\frac{e^z}{1+e^z}$. But the LCM of $1-\frac{e^z}{1+e^z}$ is $\frac{1}{1+e^z}$. Therefore, the odds of severely malnourished health (y_1) is computed as shown by Equation (26):

Odds of severely malnourished health: $$\frac{\frac{e^z}{1+e^z}}{\frac{1}{1+e^z}} = \frac{e^z}{1+e^z} \text{ x } \frac{1+e^z}{1} = e^z \qquad \text{26}$$

where $z = \tau_1 - X'\beta$.

Similarly, the second Logit equation of the odds ratio is given by Equation (27):

Odds of severely and moderately malnourished health: $\frac{y_1+y_2}{y_3}$ ………………………27

In a binary Logit, assuming the probability of $y_1 + y_2 = 1$, then the probability of the rest outcomes will be $1 - y_3$. However, the probability of observing outcome1 (i.e $y_1 + y_2 = 1$), is a function of the distribution of the error term, which is given by Equation (28) [the steps were skipped since have already been discussed]:

$$\Pr(y_1 + y_2 = 1) = f\left(\frac{e^z}{1+e^z}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 27a$$

Consequently, by substituting $y_1 + y_2$ in Equation (27) with $\frac{e^z}{1+e^z}$, the denominator of Equation (27) will be $1-\frac{e^z}{1+e^z}$. But the LCM of $1-\frac{e^z}{1+e^z}$ is $\frac{1}{1+e^z}$. Therefore, the odds of severely and moderately malnourished health (y_1+y_2) is computed as shown by Equation (29):

Odds of severely and moderately malnourished health:

$$\frac{\frac{e^z}{1+e^z}}{\frac{1}{1+e^z}} = \frac{e^z}{1+e^z} \text{ x } \frac{1+e^z}{1} = e^z \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 28$$

where $z = \tau_2 - X'\beta$.

The log odds ratios are therefore given in Equation (30) and (31):

$$\log\left(\frac{y_1}{y_2+y_3}\right) = \log(e^{z_1}) = z_1 = \tau_1 - (X_i\,\beta) = \tau_1 - (\beta_{11}X_{11} + \beta_{21}X_{21} + \ldots\ldots \beta_{k1}X_{k1}) \ldots 30$$

$$\log\left(\frac{y_1+y_2}{y_3}\right) = \log(e^{z_2}) = z_2 = \tau_2 - (X_i\,\beta) = \tau_2 - (\beta_{12}X_{12} + \beta_{22}X_{22} + \ldots\ldots \beta_{k2}X_{k2}) .. 31$$

where z_1 and z_2 are the logistic CDF's, while $\tau_1 < \tau_2$.

When the parallel line regression assumption holds, therefore, $\beta_{11} = \beta_{12}$, $\beta_{21} = \beta_{22}, \ldots\ldots \beta_{k1} = \beta_{k2}$. This means that the value of the coefficient of a particular explanatory variable, say X_{11} (i.e β_{11}) in Equation (30) should be almost the same with the value of the coefficient of the same explanatory variable (i.e β_{12}) in Equation (31). Therefore, what should differentiate Equation (30) and (31) should be the τ_i. The size of the τ_2 in Equation (31), which represents higher order category of the physical health (moderately malnourished), should be bigger than the size of τ_1 in equation (30), which represents the lower order category of the physical health status (severely malnourished). Hence, the log odds ratio forms a sequence, from lower order categories (outcome 1 versus outcome 2 and 3) to higher order category (outcome 1 and 2 versus outcome 3).

5. Derivation of the Marginal Effects in an Ordered Logit Model

Like other nonlinear models, the magnitude of the coefficients of an Ordered Logit model cannot be interpreted in its crude form, unless it is standardized. This is because it expresses the influence of each of the explanatory variable on the latent variable, y^*, not the observed discrete outcome variable (Long & Freese, 2006). Therefore, in order to get the effect of each one of the explanatory variables on the probability of a child being in any one of the physical health outcomes (outcomes 1, 2 and 3), the marginal effects is obtained as shown in Equations (11), (17) and (24). However, getting the marginal effects of Equations (11), (17) and (24) means that we obtain the first derivatives of the equations with respect to X_i. Recall that:

Equation (11) = $\Pr(y_i = 1| X_i) = \Phi\ (\tau_1 - X_i\ \beta)$.

Equation (17) = $\Pr(y_i = 2| X_i) = \Phi(\tau_2 - X_i\ \beta) - \Phi(\tau_1 - X_i\ \beta)$

Equation (24) = $\Pr(y_i = 3| X_i) = 1\ - \Phi(\tau_2 - X_i\ \beta)$

By inspection, Equation (11) is a product function. The first function is $\Phi()$, and the second is $\tau_1\ -\ X_i\ \beta$. Therefore, the derivative of Equation (11) with respect to X_i (marginal effect) can be obtained using the Chain rule. The Chain rule requires that:

> *The derivative of outside function be obtained, leaving the inside function alone and then multiply by the derivative of the inside function.*

The derivative of the outside function in Equation (11) is $\Phi()^{1-1} = \Phi()^0 = \Phi(1) = \Phi$. According to the rule, we leave the inside function alone, which means that we have $\Phi(\tau_1\ -\ X_i\ \beta)$ after differentiating the outside function, then we multiply by the derivative of the inside function again. The derivative of the inside function with respect to X_i is $-\beta$. Therefore, the derivative of Equation (11) with respect to X_i is given by Equation (32):

$$\frac{\partial \Pr(y_i = 1| X_i)}{\partial X_i} = \phi\ (\tau_1 - X_i\ \beta)\ x\ -\beta_i = -\beta_i[\phi\ (\tau_1 - X_i\ \beta)] \quad \ldots\ldots\ldots\ldots..\ldots\ldots\ldots..\ldots\ldots\ldots...32$$

Similarly, Equations (17) and (24) are also product functions. Therefore, in order to get the marginal effects of outcome 2 and 3, the derivatives of Equations (17) and (24) with respect to X_i were obtained, using Chain rule, as shown by Equations (33) through (36):

$$\frac{\partial \Pr(y_i = 2| X_i)}{\partial X_i} = [\phi\ (\tau_2 - X_i\ \beta)\ x - \beta] - [\phi\ (\tau_1 - X_i\ \beta)x - \beta] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.........33$$

$$\frac{\partial \Pr(y_i = 2| X_i)}{\partial X_i} = \beta_i[\phi\ (\tau_2 - X_i\ \beta) - \phi\ (\tau_1 - X_i\ \beta)] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots...\ldots\ldots\ldots...34$$

$$\frac{\partial \Pr(y_i = 3| X_i)}{\partial X} = -\ \phi(\tau_2 - X_i\ \beta)x - \beta \quad ..\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.....\ldots...............\ldots35$$

$$\frac{\partial \Pr(y_i = 3| X_i)}{\partial X} = \beta_i\phi(\tau_2 - X_i\ \beta) \quad ..\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..\ldots...............\ldots36$$

It is noticed that in Equation (32), which is the marginal effect of each explanatory variable on the probability of a child being severely malnourished ($y_i = 1| \mathrm{X_i}$), the sign of the beta outside the bracket ($-\beta$) is opposite to the sign of the coefficient of $\mathrm{X_i}$ (i.e β). The negative sign of the marginal effect of the lowest category also implies that a unit increase in the explanatory variable decreases the probability of a child belonging to the lower outcome (outcome 1) of the physical health. On the other hand, in Equation (36), which is the marginal effect of each explanatory variable on the probability of a child being adequately nourished ($y_i = 3| \mathrm{X_i}$), it is observed that the sign of the beta outside the bracket (β) is the same with the sign of the coefficient of $\mathrm{X_i}$ (i.e β). This implies that a unit increase in the explanatory variable increases the probability of a child being in the highest category of the physical health. However, the signs of the betas of other outcomes, such as outcome 2 are ambiguous.

6. Practical Examples

Assuming that there is a model whose dependent variable is the children's physical health status (*phealth*), which has three categories that are ordered as 1, 2 and 3 for the severely malnourished, moderately malnourished and adequately nourished health, as in our previous example. Assuming also that the model has some set of predictors, which are: *cagiver2, cagiver3, ecgiver, lrem, phelthe, clction, age, gender*. Assuming we are to estimate this model using an Ordered Logit model in STATA, then the following command is typed in the command window:

ologit phealth cagiver2 cagiver3 ecgiver lrem phelthe clction age gender

The result in Table 1 is the output of the model (*ologit phealth cagiver2 cagiver3 ecgiver lrem phelthe clction age gender*). The output in Table 1 is a crude estimate of the model in the sense that it indicates the effects of each of the explanatory variables on the latent physical health of the children (as shown by Equation 4), not the observed physical health [i.e not the ones represented by Equations (11), (17) and (24)]. From the result in Table 1, it could be seen that since the dependent variable of the model has three outcomes ($M = 3$), only two cutoff points (*cut1* and *cut2*) were estimated as shown by point (A) in Table 1. Therefore, the rule that $\tau = M - 1$, is observed.

Table 1:

```
Ordered logistic regression                     Number of obs     =      1,127
                                                Wald chi2(9)      =     133.64
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -148.48986               Pseudo R2         =     0.4357
```

phealth	B Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
cagiver2	-4.639935	.5987202	-7.75	0.000	-5.813405	-3.466465
cagiver3	.3898169	.4499802	0.87	0.386	-.4921281	1.271762
ecgiver	.2099331	.0689098	3.05	0.002	.0748724	.3449938
C lrem	1.706483	.6808263	2.51	0.012	.3720883	3.040878
phelthe	-5.133709	1.074398	-4.78	0.000	-7.23949	-3.027928
clction	-.0650005	.0881287	-0.74	0.461	-.2377295	.1077285
fincom	2.21e-06	.000049	0.05	0.964	-.0000939	.0000983
age	1.26122	.2460799	5.13	0.000	.7789123	1.743528
gender	-.4742114	.3500913	-1.35	0.176	-1.160378	.2119549
A /cut1	21.49194	6.575963			8.603293	34.38059
/cut2	23.46994	6.632449			10.47058	36.4693

It is also shown in Table 1 that although the dependent variable of the model has three outcomes, each of the predictors has only one coefficient (e.g the coefficient of *cagiver2* is only -4.639935; *cagiver3* is only 0.3898169 and *ecgiver* is only 0.2099331). This implies that the coefficient of each of the predictors in the model is assumed to have the same effect on each of the three outcomes of the dependent variable of the model (Proportional Odds Assumption). The effects should only differ across the ranks of the outcomes by the magnitude of the size of the coefficient of the threshold parameters (i.e cutoff points), not by the size of the coefficients of the predictors themselves. However, in order to investigate whether this assumption holds, "brant" test is normally conducted (Brant, 1990). In order to carry out this test, "*brant*" command is typed in the command window of STATA immediately after *ologit* command, as follows:

ologit phealth cagiver2 cagiver3 ecgiver lrem phelthe clction age gender

brant

The result in Table 2 shows that the p-value of "*All"* (point D in Table 2) is not statistically significant (0.616), which implies that all the variables in the model are jointly not statistically significant. Therefore, it means that there is NO evidence of violation of the Proportional Odds assumption (Brant, 1990).

Table 2

Brant test of parallel regression assumption

	chi2	p>chi2	df
All	7.21	0.616	9
cagiver2	0.00	0.994	1
cagiver3	0.00	0.995	1
ecgiver	1.48	0.224	1
lrem	1.53	0.216	1
phelthe	-0.00	-999.000	1
clction	1.88	0.171	1
fincom	0.14	0.705	1
age	1.60	0.206	1
gender	0.03	0.854	1

A significant test statistic provides evidence that the parallel regression assumption has been violated.

Haven confirmed that the Proportional Odds assumption is not violated, it is ideal to estimate the marginal effect and find out the size of the effect of each of the predictors on the probability of having outcome 1, 2 and 3 of the children's physical health statuses by typing "*margins, dydx(_all)"* command after *ologit* command in STATA command window, as follows:

ologit phealth cagiver2 cagiver3 ecgiver lrem phelthe clction fincom age gender

margins, dydx(_all)

The output of the above command is presented in Table 3. The coefficient of *cagiver2* for example, is 0.0432 under outcome 1 (severely malnourished); 0.09 under outcome 2 (moderately malnourished) and 0.133 under outcome 3 (adequately nourished). This means that, a one unit increase in *cagiver2* increases the probability of a child being in the severely malnourished health status by 4.32 percentage points (i.e 0.0432*100 = 4.32%); increases the child's likelihood of being in the moderately malnourished health by 9 percentage points (i.e 0.09*100 = 9%) and decrease their probability of being in the adequately nourished health category by 13.3 percentage points (i.e -0.133*100 = 13.3%), all things being equal.

Table 3: Marginal effects of the Ordered Logit model of the children health status

Variables	(1) severely malnourished	(2) Moderately Malnourished	(3) Adequately nourished
cagiver2	0.0432***	0.0900***	-0.133***
	(0.00527)	(0.0167)	(0.0187)
cagiver3	-0.00363	-0.00756	0.0112
	(0.00413)	(0.00900)	(0.0131)
ecgiver	-0.00195***	-0.00407***	0.00603***
	(0.000668)	(0.00143)	(0.00203)
lrem	-0.0159**	-0.0331**	0.0490**
	(0.00652)	(0.0145)	(0.0205)
phelthe	0.0478***	0.0996***	-0.147***
	(0.0121)	(0.0232)	(0.0326)
elction	0.000605	0.00126	-0.00187
	(0.000817)	(0.00175)	(0.00256)
fincom	-2.06e-08	-4.28e-08	6.34e-08
	(4.57e-07)	(9.51e-07)	(1.41e-06)
age	-0.0117***	-0.0245***	0.0362***
	(0.00237)	(0.00569)	(0.00735)
gender	0.00441	0.00920	-0.0136
	(0.00329)	(0.00705)	(0.0103)

One thing that is worthy of note is that the sign of a particular coefficient under outcome 1 of the marginal effect (Column 1 of Table 3) is always opposite to the sign of the same coefficient of the crude estimate in Table 1, while the sign of the highest outcome (for instance, outcome 3 in Column 3 of Table 3) is always the same with the sign of the same coefficient in the crude estimate in Table 1. For example, the sign of c*agiver2* in Table 1 is negative (*cagiver2* = -4.639935) as shown by point (B) of Table 1, but in Column 1 of Table 3, the sign of *cagiver2* is positive (0.0432) as shown by point (E) of Table 3. However the sign of the same variable (*cagiver2*) is negative (-0.133) in the highest outcome (outcome 3 in Column 3 of Table 3) as shown by point (G) of Table 3. Similarly, the sign of *ecgiver* in Table 1 is positive (*ecgiver* = 0.2099331) as shown by point (C) of Table 1, but in Column 1 of Table 3, the sign of *ecgiver* is negative (-0.00195) as shown by point (F) of Table 3. However the sign of the same variable (*ecgiver*) is positive (0.00603) in the highest outcome (outcome 3 in Column 3 of Table 3) as shown by point (H) of Table 3. These examples correspond to Equations (32) and (36) as explained in the derivation section.

7. Conclusion

This study has provided a detailed explanation and the derivational steps of the Ordered Logistic regression model, Proportional Odds assumption and the marginal effects in Ordered models. Practical example has also been provided in order to discern how the output of the model from STATA software tallies with the derivational steps discussed. The study thus sheds light on how the model works, thereby helping researches have better understanding of the output of the model when used in their studies.

References

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, *46*(4), 1171-1178.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge: Cambridge University press.

Fullerton, A. S. (2009). A conceptual framework for ordered logistic regression models. *Sociological Methods & Research*, *38*(2), 306-347.

Gujarati, D. N., & Porter, D. (2009). *Basic Econometrics*. New York: McGraw-Hill International Edition.

Leeper, T. J. (2017). Interpreting Regression Results using Average Marginal Effects with R's margins. Retrieved from https://rdrr.io/cran/margins/f/inst/doc/TechnicalDetails.pdf

Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata*. College Station, Texas: Stata press.

Long, J.S. (1997). *Regression models for categorical and limited dependent variables. Regression models for categorical and limited dependent variables: Advanced quantitative techniques in the social sciences, 7*. London: Sage Publications.

Long, S. J., Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata*. Stata press.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, *42*(2), 109-142.

Meyer, A. R. (1982). What is a model of the lambda calculus?. *Information and control*, *52*(1), 87-122.

O'donnell, O., Van Doorslaer, E., Wagstaff, A., & Lindelow, M. (2008). *Analyzing health equity using household survey data.* Washington, DC: World Bank

Orme, J. G., & Combs-Orme, T. (2009). *Multiple regression with discrete dependent variables*. Oxford University Press.

Webb, P., & Bhatia, R. A. (2005). *Manual: Measuring and Interpreting Malnutrition and Mortality*. Rome: Nutritioni Servce, World Food Programme.