JIS Journal of Interdisciplinary Sciences
Volume 6, Issue 2; 21-41
November, 2022. © The Author(s)
ISSN: 2594-3405

Ordered Choice Models: Ordinal Logit and Ordinal Probit

Öznur İşçi Güneri[1]*, Burcu Durmuş[1] and Aynur İncekırık[2]

[1]Department of Statistics, Faculty of Science, Muğla Sıtkı Koçman University, Turkey
[2]Department of Econometrics, Faculty of Economics and Administrative Sciences, Manisa Celal Bayar University, Turkey
[1]Department of Statistics, Faculty of Science, Muğla Sıtkı Koçman University, Turkey, oznur.isci@mu.edu.tr*

Abstract: *Classical ordinal logit and probit models are used in studies where the dependent variable is categorical and ordinal. In order to use these models, the assumption of parallel slopes must be met. If this assumption is not met, the generalized ones of the ordinal logit and ordinal probit models, which are more flexible in terms of assumptions, or the multinomial logit model can be used. The aim of this study is to discuss in detail the ordered logit and ordered probit models, which are developed when the dependent variable category is more than two. For this purpose, a sample data set was taken and first of all, the assumption of parallel slopes was investigated. While the validity of the models was tested with the likelihood ratio test statistic, AIC and BIC and deviation statistics were used for the goodness of fit test. The results show that there are no significant differences between the models and there are no strict rules for choosing the probit model or the logit model.*

Keywords: Ordinal logit model; ordinal probit model; parallel slopes assumption

Introduction

In cases where the dependent variable is categorical in a multivariate model, the estimations obtained by the Ordinary Least Squares regression (OLS) method used in multiple regression analysis are insufficient. In a multivariate model, if the dependent variable has a nominal or categorical scale and the independent variables have the same type of scale, the most appropriate types of models for this data set are logit, probit, Linear Probability Model (LPM) or tobit models. Among them, logit and probit models are models developed as LPM alternatives. The Tobit model is used for censored data structures.

Logit models are a method that helps to perform classification and assignment process used to determine the relationship with independent variables in cases where the dependent variable is observed in binary, triple and multiple categories. For logit models, as in multiple regression analysis, there is no normality and continuity assumption prerequisite. There are three basic methods in logit models; binary logit, ordinal logit and nominal logit regression model. In the binary logit model, the dependent variable contains binary responses. It is applied when the dependent variable is ordinal in the ordered logit model. The ordinal scale dependent variable consists of at least 3 categories. When determining these categories, the answers should be in natural order. For example, if the severity of the disease is in question, categories as mild-moderate-severe can be determined. The code values of ordinal values must follow the same order of magnitude (1-2-3 etc). The nominal scale model is applied when the dependent variable is nominal scale. As with the ordinal scale, it should contain values observed in at least 3 categories. However, it is not necessary for the categories to follow a sequence in coding the observed values. For example, occupational groups such as engineering, banking, medicine, etc. can be nominally determined.

The dependent variable examined in most of the studies, especially in the field of social sciences and health, is categorical and ordinal. One of the mistakes that researchers frequently make in such studies is to calculate the categorical dependent variable probabilities with the help of Multinomial Logit analysis, forgetting that the categorical dependent variable is in an ordered structure (Finney, 1971). In logit models, the natural logarithm of the probabilities of the ordinal dependent variable is expressed as a linear function of the independent variables. Therefore, the logit model is a member of the "generalized linear models" family, and the logit transform, that is, the natural logarithm of the ratios of the independent variable, is used as a link function. In ordinal logit models, the latent variable approach is used and the dependent variable is estimated as a function of this unobservable variable (Jackman, 2000).

There are different Logit rendering formats used for dependent variable comparisons in ordinal logit models. Cumulative logit models are the easiest to interpret and the most frequently used. Cumulative logit models are divided into three; Proportional Odds Model (POM), Non-Proportional Odds Model (NPOM) and Partial Proportional Odds Model (PPOM) (Arı and Yıldız, 2014). In the POM, unlike the Multinomial Logit model, the cumulative logits created have a parallelism assumption, Parallel Slopes Assumption or a proportional risk assumption (Proportional Odds Assumption). For ordinal categories, estimation methods generally assume that the estimated coefficients of the independent variables do not vary between categories. This is called the parallel slopes (lines) assumption.

However, this assumption is mostly not met or it is seen that it is ignored by researchers. If this assumption is not met, the results of ordinal logit models cannot be trusted and alternative models are recommended. In cases where the assumption is not met, it is seen that the multinomial logit model is applied even in cases where there are models with dependent variables in an ordered structure. Applying an unordered model to a model whose dependent variable has an ordered structure will cause efficiency losses in the prediction results. On the contrary, when a model with an ordered structure is applied to a model with an unordered dependent variable, it will cause serious deviations in the estimation results (Amemiya, 1985).

Problems with the assumption of parallel slopes must be considered in the empirical analysis of categorical dependent variables. To deal with this, information is needed about the effects of independent variables on different categories. An analysis based on a basic theory that provides information about variables violating the parallel slopes assumption may be preferable. However, the assumption of parallelism must be satisfied in the classical ordered logit and ordered probit models. In cases where this assumption is not met, the use of the proportional risk model is incorrect (Hosmer and Lemeshow, 2000).

Another method that can be used when the assumption of parallelism is not met is the Multinomial Logit model. However, in the Multinomial Logit model, the ordinal structure of the dependent variable is ignored and is nominally included in the model. Therefore, a search is made for a model that both takes into account the ordinal structure and relieves the rather rigid proportional risk assumption. In recent years, the Generalized Ordinal Logit Model, also known as the Non-Proportional Odds Model, has been used in cases where the assumption of parallelism is not met and the Proportional Odds Model is insufficient (Bender and Grouven, 1998; Arı and Yıldız, 2014). Another method is the Generalized Probit Model. This model both considers the ordinal structure of the dependent variable and does not restrict the proportional odds assumption (Hardin and Hilbe, 2001).

The probit model is an alternative to the logit model. This model belongs to the family of generalized linear models. It is used when the dependent variable is two-category and multi-category, as in the logit model. It is seen that the ordinal probit model is widely used if the values of the dependent variable take more than two values and are in an ordered structure. In this model, as in the logit model, parallelism assumption is required. Both models give very similar results. However, the logit model is more popular than the probit model. One of the most important reasons for this is that while the logit model uses OR (Odds Ratio) values, which are easier to interpret while calculating the coefficients, the cumulative normal distribution is used in the probit model.

Research Method

Parallel Slopes Assumption

The basic assumption for the ordered logit and ordered probit models is the parallel slopes assumption. The parallel slopes assumption states that the categories of the dependent variable are parallel to each other. When this assumption is not met, no parallelism is achieved between the categories. According to this assumption, the parameters should not change for different categories. In other words, the correlation between the independent variable and the dependent variable does not change for the categories of the dependent variable, and the parameter estimates for the cut-off points do not change (Ananth and Kleinbaum, 1997). Figure 1 shows whether the parallel slopes assumption is valid or not (Peterson and Harrell, 1990).

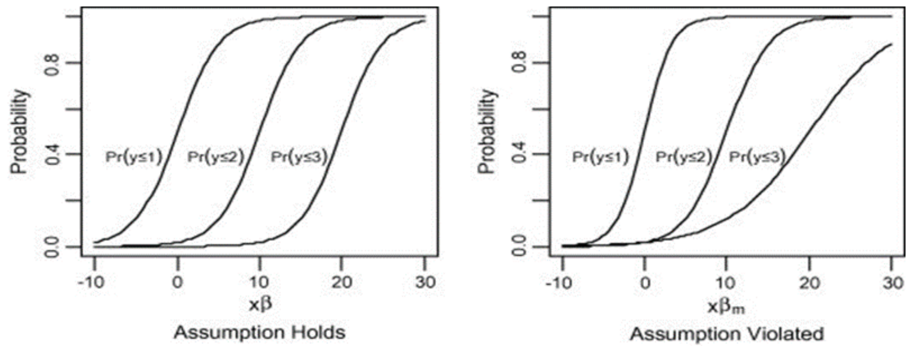

Figure 1. Conditions where the parallel slopes assumption is valid and not valid

Parallel slopes approach 1 in each of the 3 probabilities shown in Figure 1 with the increase in independent variables if the assumption is met. As seen in Figure 1, although the slopes of all 3 curves are the same, the threshold parameters differ. Whether the β coefficients of the independent variable are equal for each category is tested with the null hypothesis given in Equation 1:

$$H_0: \beta_{1j} = \beta_{2j} = \cdots = \beta_{(K-1)j} = \beta, \quad \text{j} = 1, 2, \dots, \text{J}) \tag{1}$$

To test whether the parallel slopes assumption is met, the Likelihood Ratio Test, Wald chi-square test and other related tests are used (Jackman, 2000; Liao, 1994). In the ordered logit model, these tests examine the equality of different categories and decide whether the assumption is valid. If the assumption of parallel slopes is not satisfied, interpretations of the results will be inaccurate. For this reason, alternative models such as generalized ordinal logit model and multinomial logit model are used instead of ordered logit and ordinal probit regression models to find correct results.

Ordinal Logit Model (OLM)

In some cases, the multi-category dependent variable may be in an ordered structure. Such situations are mostly encountered in Likert type scales in survey studies or in studies conducted to measure disease severity in the field of health. There is a clear orderable structure between the dependent variable categories. However, the distances between consecutive categories are not equal to each other (Bender and Grouven, 1998).

Since the dependent variable categories are not measured with a range scale and the distances between consecutive categories are not equal, this type of data cannot be easily modeled with classical regression. On the other hand, using the Multinomial logit model on such data ignores this structure of the ordinal dependent variable and is insufficient to use all the information in the dependent variable (Finney, 1971). For these reasons, ordinal logit models have found wide use in the analysis of this type of data (Timur and Akay, 2017). The ordinal logit model is a natural extension of the binary logit model and given as in Equation 2:

$$y^*_{OLM} = x^T\beta + \epsilon \tag{2}$$

Here y^* is the latent variable, β is the unknown parameters and ϵ is the error term. Considering that the dependent variable has J ordered categories, the relationship between the observed categories of τ cut off/threshold values can be given as in Equation 3:

$$y_i = \begin{cases} 0 & y^* \leq 0, \\ 1 & 0 < y^* \leq \tau_1, \\ 2 & \tau_1 < y^* \leq \tau_{2,} \\ & . \\ & . \\ & . \\ J & \tau_{j-1} < y^* \end{cases} \qquad i = 1,2,\ldots,N \tag{3}$$

Cumulative logit models are the most widely used method in ordinal logit models. Other logit models used in ordinal logit models, apart from cumulative logit models, are Adjacent Category Logit and Continuation Ratio Logit models. These three models differ according to which categories and how they are compared (Finney, 1971; Liao, 1994).

Cumulative Logit Models

Where the dependent variable category is ordinal, the category probabilities are often expressed in terms of cumulative probabilities. In general, the cumulative probability for category j is defined as in Equation 4:

$$P(y \leq j) = P(y = 1) + P(y = 2) + \cdots . + P(y = j) \tag{4}$$

Here, the dependent variable y consists of the j order categories. τ is the unknown threshold values separating ordinal categories. The expression in Equation 4 can be written more clearly as in Equation 5. Thus, the cumulative probability equation is reached (Jackman, 2000).

$$P(y \leq j) = F\left(\tau_j - \sum_{k=1}^{K} \beta_k x_k\right) = \frac{exp(\tau_j - \sum_{k=1}^{K} \beta_k x_k)}{1 + exp(\tau_j - \sum_{k=1}^{K} \beta_k x_k)}, j = 1,2,\ldots,j-1 \tag{5}$$

The cumulative logit model is created by using the cumulative probability P($y \leq j$) instead of the category probability P($y = j$) in the logit transformation. From the definition of the logit link function, the cumulative logit model is written as follows (Liao, 1994):

$$logit[P(y \leq j)] = log\left[\frac{P(y \leq j)}{1 - P(y \leq j)}\right], j = 1,2,\ldots,j-1 \tag{6}$$

The expression $\frac{P(y \leq j)}{1-P(y \leq j)}$ in Equation 6 given above is defined as the cumulative risk (Cumulative Odds) for the jth dependent variable category. Therefore, cumulative logit models use the ordinal structure and are obtained by dividing the probability of falling into a lower category by the probability of falling into a higher category for the jth dependent variable category. As with different models, a category is chosen as the reference category (usually the highest category). The $j-1$ cut off point is estimated in this way, and the estimates provide information about the cumulative probabilities for each consecutive category. In cumulative probability, the probability of being in the selected category or being in a subcategory is considered together (Fullerton and Xu, 2012).

Cumulative logit models are divided into three main model groups based on the assumption of parallelism. These models are proportional odds, non-proportional odds and partial proportional odds models (Butler and Moffitt, 1982; O'Connell, 2006; Long, 1997). These models are briefly described below.

i. Proportional Odds Model (POM)

The Proportional Odds Model was defined by McCullagh (1980) for ordinal logistic regression (O'Connell, 2006). The model is based on the cumulative distribution function. Proportional odds models are widely used when the dependent variable is ordinal and parallel slopes assumption is valid (Hosmer and Lemeshow, 2000; Barak, 2005). In the cumulative logit model, when the cumulative probability values are placed in the logistic distribution, the equations obtained for the proportional probability model are given in Equation 7 (Timur and Akay, 2017):

$$\text{logit}[P(y \leq j)] = \log\left[\frac{P(y \leq j)}{1-P(y \leq j)}\right] = \log\left[\frac{F(\tau_j - \sum_{k=1}^{K} \beta_k x_k)}{1-F(\tau_j - \sum_{k=1}^{K} \beta_k x_k)}\right]$$

$$\text{logit}[P(y \leq j)] = \log\left[\exp\left(\tau_j - \sum_{k=1}^{K} \beta_k x_k\right)\right] = \tau_j - \sum_{k=1}^{K} \beta_k x_k \,, j = 1,2, \dots, j-1 \quad (7)$$

In the proportional odds model given above, each cumulative logit has its own threshold value, represented by τ_i. When the dependent variable categories are denoted by $j = 1, 2, \dots, j-1$, it is seen that the β_k coefficients in the equation are independent of the dependent variable categories.

In the proportional odds model (Equation 7), there is a (-) sign in front of the β_k coefficients. The meaning of this negative sign is that the probability of falling into the lower category with a positive β_k coefficient is inversely proportional to $P(y \leq j)$. In other words, a positive β_k coefficient indicates that the probability of falling into a lower category decreases, while the probability of falling into a higher category increases if $P(y \geq j)$ (Finney, 1971). For the opposite situation, a negative β_k coefficient indicates that the probability of falling into the lower category $P(y \leq j)$ increases, and the probability of falling into the higher category $P(y \geq j)$ decreases. In the proportional risk model, the β_k coefficient gives the effect of a one-unit increase in the kth independent variable on the dependent variable, as in other linear models.

ii. Non-Proportional Odds Model (NPOM)

The proportional odds assumption may not be satisfied in some cases. In cases where this assumption is not met, generalized logit and Multinomial Logit models can be used. It should be noted that the ordinal structure of the dependent variable in the Multinomial Logit model is ignored and it is nominally included in the model (Hosmer and Lemeshow, 2000). Therefore, a search is made for a model that both considers the ordinal structure and relieves the rather rigid proportional odds assumption.

In the non-proportional odds model proposed by Fu (1998), contrary to the multinomial logit model, cumulative logits are used during logit creation, but the proportional odds assumption is not met (Butler and Moffitt, 1982). In other words, in this model, the effect of independent variables on the dependent variable odds (cumulative odds) is no longer equal and when the dependent variable category is denoted by $j = 1, 2, \dots, j-1$, the β_{jk} coefficients are different for each dependent variable category (Arı and Yıldız, 2014; Butler and Moffitt, 1982). The non-proportional odds model (the generalized ordinal logit model) can be written as in Equation 8:

$$logit[P(y \leq j)] = log\left[\frac{P(y\leq j)}{1-P(y\leq j)}\right] = log\left[\frac{F(\tau_j - \sum_{k=1}^{K}\beta_{jk}x_k)}{1-F(\tau_{ij} - \sum_{k=1}^{K}\beta_{jk}x_k)}\right] \quad (8)$$

$$logit[P(y \leq j)] = \log\left[\exp\left(\tau_j - \sum_{k=1}^{K}\beta_{jk}x_k\right)\right] = \tau_{ij} - \sum_{k=1}^{K}\beta_{jk}x_k\,, j = 1, 2, \dots, j-1$$

In the non-proportional risk model, each cumulative logit has its own threshold value, represented by τ_j. When the dependent variable categories are denoted by $j = 1, 2, \dots, j-1$, the coefficients of β_{jk} in the equation take different values in each dependent variable category. In the non-proportional odds model, the coefficient β_{jk} gives the effect of a one-unit increase in the kth independent variable on the cumulative logit.

iii. Partial Proportional Odds Model (PPOM)

The partial proportional odds model can be used when the parallel slopes assumption is valid or not. This model was first proposed by Peterson and Harrell (1990) (Long, 1997). The partial proportional odds model has the same characteristics as both the proportional odds model and the disproportionate odds model (Long, 1997; Agresti, 2002). PPOM has been defined in two ways; the constrained proportions model and the unconstrained proportions model.

a. Unconstrained Partial Proportional Odds Model (UPPOM)

Two different sets of coefficients are estimated in UPPOM; the first set holds the assumption of parallel slopes and the second set does not hold the assumption of parallel slopes. The general form of the model can be written as in Equation 9 (Agresti, 2002);

$$P(y \le j) = \frac{\exp(-\alpha_j - x^T\beta - t^T\gamma_j)}{1+\exp(-\alpha_j - x^T\beta - t^T\gamma_j)}, \quad j = 1, 2, \dots, k \tag{9}$$

In this model, if the value of γ_j is equal to 0, the assumption of parallel lopes is valid and the model takes the form of POM (Long, 1997; Agresti, 2002).

b. Constrained Partial Proportional Odds Model (CPPOM)

In the non-proportional risk model, constraints are defined for a group of disproportionate variables. The model becomes constrained when the coefficients at the changing breakpoints are multiplied by a predefined constant scalar. Since there will be parallelism between constrained variable coefficients, CPPOM will need fewer parameters than unrestricted PPOM and NPOM (Agresti, 2002). The general form of the model is given in Equation 10 (Long, 1997).

$$P(y \le j) = \frac{\exp(-\alpha_j - x^T\beta - t^T\gamma\Gamma_j)}{1+\exp(-\alpha_j - x^T\beta - t^T\gamma\Gamma_j)}, \quad j = 1, 2, \dots, k \tag{10}$$

In this model, Γ is a predefined constant scalar. y is a vector and does not depend on J.

Ordinal Probit Model (OPM)

The probit model is an alternative to the logit model. Analysis of ordered responses begins with the expansion made by Finney (1971) on Aitchison and Silvey (1957) (Snell, 1964; Fox, 1997). Another pioneering study is the parallel development of Snell's (1964) differential treatment of ordinal outcomes (Aitcheson and Silvey, 1957; Greene, 2012). The modern form of the ordinal probit model was proposed by McElvey and Zavoina (1975) for the analysis of ordered, categorical, non-quantitative choices, outcomes, and responses.

The ordinal probit model is used in many studies with ordinal structure. The places where this model is widely used include bond ratings, preferences in consumption, satisfaction and health status surveys. The model is used to describe the data generation process for a random result that takes one of a series of discrete, ordered results (Aitcheson and Silvey, 1957). For example, in clinical studies, when investigating the effect of a drug on a patient, while the dependent variable can be categorical variables such as complete cure (1), alleviation of symptoms (2), no effect (3), worsening (4), death (5), the independent variables may include variables such as age, gender, blood pressure, heart disease etc. In another example, the ordinal probit model in the Likert-type scale (strongly disagree, agree, strongly agree), which is widely used in survey research, makes no assumptions for the interval distances between the options, preserving the order of the options of the dependent variable, and adapts appropriately to this data.

The traditional ordinal probit model states that all variables are restrictive and satisfy the parallel slopes assumption. If this assumption is not met, one of the alternative methods is the generalized probit model. This model, on the other hand, uses a completely flexible approach and allows all coefficients to vary between categories, which is a very strong assumption. It is seen that the ordinal probit model is widely used if the values of the dependent variable take more than two values and are in an ordered structure. In general, the ordered probit model can be written as in Equation 11.

$$y^*_{OPM} = x^T\beta + \epsilon \tag{11}$$

Here y^* is the definite but unobserved dependent variable (order of levels); x is the vector of independent variables and β is the vector of regression coefficients that we want to predict. Also, even if y^* is not observed, it is assumed that the categories of the dependent variable can be observed (Fu, 1998), then it can be written as in Equation 12;

$$y_i = \begin{cases} 0 & y^* \leq 0, \\ 1 & 0 < y^* \leq \mu_1, \\ 2 & \mu_1 < y^* \leq \mu_{2,} \\ & . \\ & . \\ & . \\ J & \mu_{j-1} < y^* \end{cases}, \quad i = 1,2,\dots,N \tag{12}$$

Here, the μ refers to the unknown threshold values that separate the categories. With the assumption that the errors are normally distributed in the ordinal probit model, the likelihood function can be estimated using the Gauss-Hermite Quadrature approach developed by Butler and Moffitt (1982) (Greene and Hensher, 2010). The $j-1$ threshold parameter of the latent variable y^* is obtained. If the model is a constant term, $j-2$ threshold parameters are estimated. Since the first threshold parameter is zero, the threshold parameters are all positive. The probabilities of each ordinal outcome are as in Equation 13 (McCullagh, 1980):

$$\begin{aligned} P[y_i = 0] &= P[\, y^* \leq 0] = \phi(\mu_0 - x_i\beta) \\ P[y_i = 1] &= P[\, 0 < y^* \leq \mu_1] = \phi(\mu_1 - x_i\beta) - \phi(\mu_0 - x_i\beta) \\ P[y_i = 2] &= P[\, \mu_1 < y^* \leq \mu_{2,}] = \phi(\mu_2 - x_i\beta) - \phi(\mu_1 - x_i\beta) \\ &. \\ &. \\ &. \\ P[y_i = j] &= \phi(\mu_j - x_i\beta) - \phi(\mu_{j-1} - x_i\beta) \end{aligned} \tag{13}$$

The general format for $j = m$ (highest category) is reduced to Equation 14:

$$P[y_i = m] = \phi(\mu_m - x_i\beta) - \phi(\mu_{m-1} - x_i\beta) = 1 - \phi(\mu_{m-1} - x_i\beta) \quad (14)$$

We can use MLE to predict this model. For the MLE function, Z_{ij} is defined as the indicator variable, equal to 1 if $y_i = j$, equal to 0 otherwise. If $\phi_{ij} = \phi[\mu_j - x_i\beta]$ and $\phi_{i,j-1} = \phi[\mu_{j-1} - x_i\beta]$, the log likelihood function is given as in Equation 15 (McCullagh, 1980);

$$lnL = \sum_{i=1}^{N}\sum_{j=0}^{m} Z_{ij}\ln[\phi_{ij} - \phi_{i,j-1}] \quad (15)$$

Model Validity Tests and Goodness of Fit Indicators

Likelihood Ratio Test Statistics (LR)

The LR test is a common measure used to investigate the effect of the independent variable or variables in the model on the dependent variable. The LR test statistic can be generalized to test the significance of the independent variables included in the model. When the log-likelihood of the model with k independent variables is L_k and the log-likelihood of the model with k+p is represented by L_{k+p}, the test statistic of the Generalized Likelihood Ratio is as in Equation 16 (Brant, 1990):

$$G^2 = -2(\mathrm{L_k} - \mathrm{L_{k+p}}) \quad (16)$$

The likelihood ratio test statistic shows the chi-square distribution in p degrees of freedom ($G^2 \sim \chi_p^2$). It should be noted that the generalized likelihood ratio test statistic is used only for nested model comparisons.

Deviance (D)

In generalized linear models, the deviation measure is an indication of how much the actual values differ from the predicted values. With this aspect, the deviation measure is equivalent to the error sum of squares (SSE) in classical linear models (Finney, 1971). A complex model containing as many parameters as the number of observations is called a fully saturated model. When log-likelihood L_S obtained from the saturated model (Log-Likelihood of the data set) is shown with the log likelihood L_M (Log-Likelihood of the model) obtained from any sub-model of the saturated model created by the variables considered to be significant, the deviation measure is as in Equation 17:

$$D = -2(L_M - L_S) \quad (17)$$

A smaller deviation is an indication of a better fit of the model to the data, since the deviation measure will give the magnitude of the deviation from the true values.

Akaike Information Criteria (AIC)

With this criterion proposed by Akaike (1973), information can be obtained about which model fits the data better (Arı and Yıldız, 2014). The Akaike Information Criterion (AIC) is used for model comparisons that are either nonnested or calculated from different samples. When the log-likelihood of the model is L, the number of independent variables in the model is k, and the number of observations is n, AIC value is calculated as in Equation 18:

$$\mathrm{AIC} = \frac{-2\log(\mathrm{L})+2\mathrm{k}}{\mathrm{n}} \quad (18)$$

Smaller values of the AIC are an indication of a better fit of that model to the data. Therefore, among several models, the model that best fits the data is the one with the smallest AIC. In cases where the number of parameters is large compared to the sample size, AICc suggested by Hurvich and Tsai should be used instead of AIC.

Bayesian Information Criteria (BIC)

The Bayesian Information Criterion is based on Bayesian comparison of models. The Bayesian Information Criterion is used in model comparisons that are not nested or calculated from different samples, as in AIC. The Bayesian Information Criterion is calculated as in Equation 19:

$$BIC = -2log(\mathcal{L}) + klog(n) \quad (19)$$

Smaller values of the BIC are an indication of a better fit of that model to the data. Therefore, among several models, the model that best fits the data is the one with the smallest BIC.

McFadden's Adjusted Likelihood Ratio Index

When the log likelihood of the model with only constant term is L_0 and the log likelihood of the model with k independent variables is indicated by L_1, McFadden's adjusted Likelihood Ratio indicator is calculated as in Equation 20:

$$\overline{\mathrm{R}}^2_{\mathrm{McFadden}} = 1 - \frac{\mathrm{L}_1 - \mathrm{k}}{\mathrm{L}_0} \quad (20)$$

McFadden's corrected likelihood ratio indicator is referred to as Pseudo R^2 in most statistical packages. This indicator can be calculated for any model calculated with the MLE method. Among the models, the model with the largest R^2 is the model that best fits the data.

Application

The dataset was obtained from a survey conducted in 2018 to evaluate the experiences and opinions of patients and their family members on satisfaction with healthcare services in

Macedonia (Dimitrievska and Tomovska, 2020). The sample size consists of 500 people. Data were collected using a Likert scale between 1-5 points.

The dataset is available at https://www.kaggle.com/datasets/vdimitrievska/patient-satisfaction-dataset?select=datasetsatisfaction.csv. For more information, see the study of Dimitrievska and Tomovska, 2020 (Factors connected to patients' satisfaction in the health care system in North Macedonia). Stata 13 statistical package program was used for analysis.
Dependent variable: overall patient satisfaction (1: yes, 2: no, 3: partly)
Independent variables:

- Check-up appointment (Checkup_appointment)
- Time waiting (Time_waiting)
- Admin procedures (Admin_procedures)
- Hygiene and cleaning (Hygiene_cleaning)
- Time of appointment (Time_of_appointment)
- Quality/experience (Quality_experience_Dr)
- Doctor Specialists available (Specialists_avaliable_Dr)
- Communication with doctor (Communication_with _Dr)
- Exact diagnosis (Exact_diagnosis)
- Modern equipment (Modern_equipment)
- Friendly health care workers (Friendly_health_care_workers)
- Laboratory services (Lab_services)
- Availability of drugs (Availability_of_drugs)
- Waiting rooms (Waiting_rooms)
- Hospital rooms quality (Hospital_rooms_quality)
- Quality parking, playing rooms, cafe's (Parking_others)

The distribution of frequencies for the dependent variable is shown graphically in Figure 2.

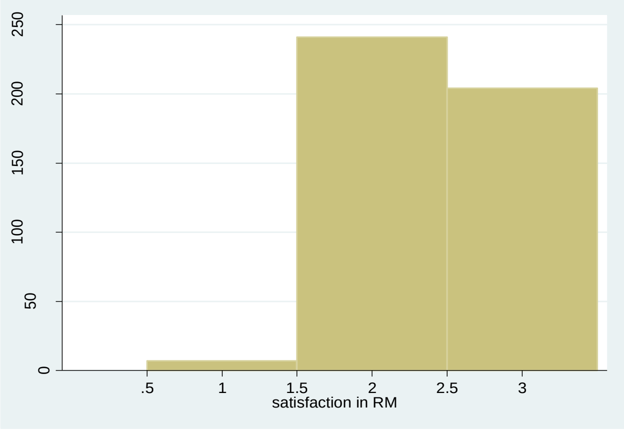

Figure 2. Distribution of the satisfaction variable
Frequency distribution across the categories of the satisfaction variable is given in Table 1.

Table 1. Frequency Distribution of Dependent Variable

Satisfaction	Freq.	Percent	Cum.
1	7	1.55	1.55
2	241	53.32	54.87
3	204	45.13	100.00
Total	452	100.00	

When the frequency distribution in the data set is examined, it is seen that it is concentrated in the 2th and 3th categories. According to Table 1, the percentages of category levels are 1.55%, 53.32% and 45.13%, respectively.

Ordinal logit model results

Due to the natural ordered structure of the dependent variable, the logit model was preferred first in the application. The results found for the ordinal logit model are given in Table 2.

Table 2. Ordinal Logit Model Results

Satisfaction	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
Checkup_appointment	.0006272	.0903114	0.01	0.994	-.1763799	.1776343
Time_waiting	-.0814159	.100526	-0.81	0.418	-.2784432	.1156114
Admin_procedures	-.0709189	.1038348	-0.68	0.495	-.2744313	.1325934
Hygiene_cleaning	-.0845556	.1152066	-0.73	0.463	-.3103564	.1412451
Time_of_appointment	.043907	.104093	0.42	0.673	-.1601115	.2479255
Quality_experience_Dr	.2159781	.094716	2.28	0.023	.0303381	.4016181
Specialists_avaliable_Dr	-.0318688	.1034881	-0.31	0.758	-.2347017	.1709641
Communication_with _Dr	-.0609047	.112261	-0.54	0.587	-.2809322	.1591228
Exact_diagnosis	.2602155	.1023569	2.54	0.011	.0595997	.4608313
Modern_equipment	-.2055631	.1059302	-1.94	0.052	-.4131826	.0020563
Friendly_health_care_workers	-.0769104	.1057555	-0.73	0.467	-.2841874	.1303666
Lab_services	-.004348	.0932292	-0.05	0.963	-.1870739	.178378
Availability_of_drugs	.0594051	.0910271	0.65	0.514	-.1190048	.2378149
Waiting_rooms	-.1065645	.1240415	-0.86	0.390	-.3496815	.1365525
Hospital_rooms_quality	.0322269	.1438462	0.22	0.823	-.2497065	.3141603
Parking_others	.0243604	.1328899	0.18	0.855	-.236099	.2848199
/cut1	-4.392466	.4754762			-5.324.382	-3.46055
/cut2	.089035	.2956981			-.4905226	.6685926
	LR $chi^2(16) = 30.81$		Prob > $chi^2 = 0.0142$		Pseudo $R^2 = 0.0449$	

According to Table 2, variables of quality experience Dr, exact diagnosis, and modern equipment were found to be significant ($p<0.05$). The cut-off points or thresholds used to distinguish between quality levels are shown at the bottom (cut 1, cut 2). The estimated threshold parameter for the 3-level dependent variable is 2. We can write the ordered logit model as in Equation 21.

$$y^*_{OLM} = .0006272\ \text{checkup\_appointment} - .0814159\ \text{time\_waiting} + \cdots + .0243604\ \text{parking\_others} \quad (21)$$

The representation of threshold values is as in Equation 22.

$$y = \begin{cases} 1 & -4.392466 < y^* \leq .4754762 \\ 2 & .089035 < y^* \leq .2956981 \end{cases} \quad (22)$$

According to the results obtained, when other variables are kept constant, the quality_experience Dr is at 1, 2 and 3 satisfaction level 1.24% [OR= exp(.2159781)=1.24] and the exact_diagnosis is 1.29% [OR=exp(.2602155)=1.29] increases. Other variables can be interpreted similarly.

The accuracy of the model was tested by calculating statistics on the threshold parameters ($p<0.05$). The coefficient, standard error and probability values for the threshold parameters are given in Table 3.

Table 3: Test of significance of threshold parameters

Satisfaction	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
cut 1- cut 2	4.481501	.385791	11.62	0.000	3.725.365	5.237.638

The ordinal logit (probit) model assumes that the distance between each category of outcome is proportional. In practice, violating this assumption may or may not change your material results. You need to test if this is the case. A Brant-developed Wald test or LR test can be used to test whether the proportional probabilities i.e. parallel slopes assumption is valid.

The validity of the parallel slopes assumption for the ordinal logit model was investigated with the Brant test. When Table 4 is examined, it is seen that the assumption is met since all the p values are greater than 0.05 ($p>0.05$). This indicates that the model is significant. Therefore, the classical logit model, that is, the proportional risk model, is suitable.

Table 4. Brant test of parallel regression assumption

	chi^2	$p>chi^2$	df
All	21.45	0.162	16
Checkup_appointment	0.04	0.836	1
Time_waiting	0.67	0.413	1

Admin_procedures	0.24	0.622	1
Hygiene_cleaning	5.19	0.023	1
Time_of_appointment	1.84	0.175	1
Quality_experience_Dr	0.08	0.781	1
Specialists_avaliable	0.14	0.704	1
Communication_with_Dr	0.00	0.954	1
Exact_diagnosis	0.10	0.752	1
Modern_equipment	0.18	0.673	1
Friendly_health_care_workers	0.16	0.690	1
Lab_services	0.14	0.704	1
Availability_of_drugs	0.06	0.800	1
Waiting_rooms	2.30	0.129	1
Hospital_rooms_quality	0.06	0.803	1
Parking_others	2.81	0.094	1

The parallel slopes test can help you evaluate whether it is reasonable to assume that the parameters are the same for all categories. This test compares the predicted model with a set of coefficients for all categories with a model that has a separate set of coefficients for each category. Poor model fit may also be due to the chosen ordering of the dependent variable categories. This can happen for many reasons, including using an incorrect link function or using the wrong model.

After the satisfaction of the assumption of parallel slopes, the probabilities of the ordinal logit model were calculated. Descriptive statistics for these probabilities are given in Table 5.

Table 5. Descriptive statistics of probabilities of OLM

Variable	**n**	**Mean**	**Std. Dev.**	**Min**	**Max**
p1ologit	452	.0155102	.0080081	.0030056	.0621422
p2ologit	452	.5320837	.1198791	.2073504	.7919829
p3ologit	452	.4524061	.1271395	.1458749	.7896441

The mean values of the ordinal logit model in Table 5 and the % values in Table 1 were found to be very close to each other. In Table 5 considering the mean values, the mean probability values for 1, 2 and 3 were found to be 1.55%, 53.2% and 45.24%, respectively. The model can be said to fit the data reasonably. Estimated probabilities are similar to real ones. Figure 3 shows the graph of the probabilities of the ordinal logit model.

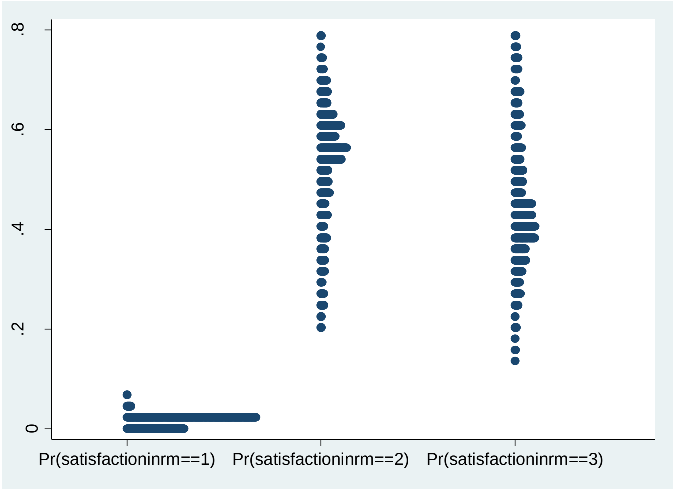

Figure 3. Graph of estimated probabilities for the ordinal logit model

According to Figure 3, satisfaction is below 0.20 at the 1st level between 0.20 and 0.80 at the 2nd and 3rd levels.

Ordinal probit model results

Ordinal probit was applied to the same data since the dependent variable was in an ordinal structure. The results found for the ordered probit model are given in Table 6.

Table 6. Ordinal Probit Model Results

Satisfaction	**Coef.**	**Std. Err.**	**z**	**P>z**	**[95% Conf.**	**Interval]**
Checkup_appointment	-.0047644	.0544886	-0.09	0.930	-.11156	.1020313
Time_waiting	-.060626	.0601714	-1.01	0.314	-.1785597	.0573078
Admin_procedures	-.03291	.062253	-0.53	0.597	-.1549237	.0891037
Hygiene_cleaning	-.0230793	.0686417	-0.34	0.737	-.1576145	.1114558
Time_of_appointment	.0066156	.0619377	0.11	0.915	-.1147801	.1280113
Quality_experience_Dr	.1240189	.0569111	2.18	0.029	.0124752	.2355625
Specialists-avaliable	-.016847	.0613708	-0.27	0.784	-.1371316	.1034377
Communication_with_Dr	-.0397645	.066984	-0.59	0.553	-.1710508	.0915219
Exact_diagnosis	.150589	.061192	2.46	0.014	.0306553	.2705241

	7	1				
Modern_equipment	-.1256695	.063257	-1.99	0.047	-.2496508	-.0016881
Friendly_health_care_workers	-.0438331	.0640225	-0.68	0.494	-.1693149	.0816487
Lab_services	.0034503	.0558779	0.06	0.951	-.1060685	.112969
Availability_of_drugs	.0399372	.0547702	0.73	0.466	-.0674104	.1472847
Waiting_rooms	-.08519	.073853	-1.15	0.249	-.2299393	.0595592
Hospital_rooms_quality	.0258864	.087083	0.30	0.766	-.1447932	.196566
Parking_others	.0372378	.0799058	0.47	0.641	-.1193747	.1938503
/cut1	-2.27381	.2241497	-2.713136	-1.834485		
/cut2	.082684	.1764375	-.2631272	.4284952		
	LR $chi^2(16)$ = 28.38		Prob > chi^2 = 0.0285		Pseudo R^2 = 0.0414	

Similar to the ordinal logit model, quality_experience_Dr, exact_diagnosis, and modern_equipment variables were found to be significant in this model, while the other variables were found to be insignificant (p<0.05). We can write the ordered probit model as in Equation 23.

$$y^*_{OLM} = -.0047644\ \text{checkup\_appointment} - .060626\ \text{time\_waiting} + \cdots +.\ .0372378\ \text{parking\_others} \quad (23)$$

The representation of the threshold values for probit model is as in Equation 24;

$$y = \begin{cases} 1 & -2.27381 < y^* \leq .2241497 \\ 2 & .082684 < y^* \leq .1764375 \end{cases} \quad (24)$$

As with binary probit models, the results here are z-scores. So, the interpretation here is no different from binary probit models, except that the sequencing is reflected in this case. For example, if an interpretation is made for the quality_experience_Dr variable, it shows that a unit increase in quality_experience_Dr will lead to an increase of 0.124 points in favour of satisfaction in the z-score. Z-scores are similar in sign and significance to log ratios (OR) (negative and positive variables). In addition, their slope is interpreted in the same way. As in the ordinal logit model, probabilities are calculated in the ordinal probit model. Descriptive statistics of these probabilities are given in Table 7.

Table 7. Descriptive statistics of OPM's probabilities

Variables	n	Mean	Std. Dev.	Min	Max
p1oprobit	452	.0162487	.0120205	.0008138	.0932951
p2oprobit	452	.5314265	.1091706	.2126567	.756548
p3oprobit	452	.4523248	.1194992	.1501569	.7865295

The mean values of the ordinal probit model in Table 7 and the % values in Table 2 were found to be very close to each other. Again, we can say that this model fits the data reasonably well. Thus, the estimated probabilities are similar to the real ones. Figure 4 shows the graph of the probabilities of the ordinal probit model.

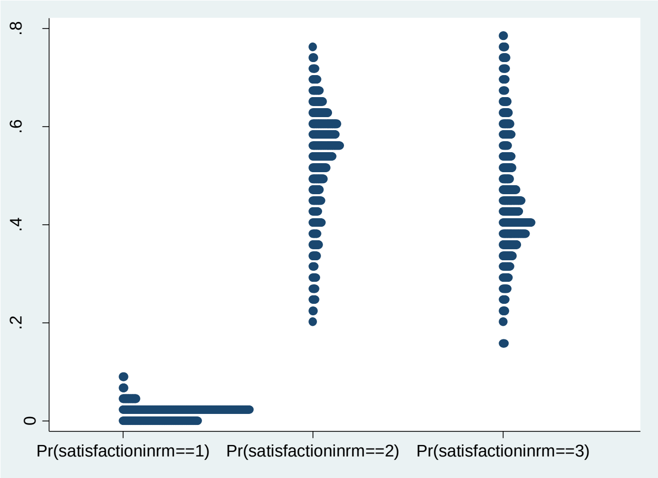

Figure 4. Graph of predicted probabilities for the ordinal probit model

When probabilities are examined according to the ordinal probit model in Figure 4, results close to the estimates obtained with the ordinal logit model given in Figure 3 are found.

In this study, ordinal logit and probit models were used to test the patients' satisfaction. First of all, the assumption of parallelism was checked for the data and it was seen that the assumption was met for the logit and probit models to be applied. Goodness-of-fit statistics for this data set in question are given in Table 8.

Table 8. Goodness of Fit Statistics

		OLM	OPM
Log-likelihood			
	Model	-327.624	-328.841
	Intercept-only	-343.030	-343.030
Chi-square			
	Deviance (df=434)	655.247	657.681
	LR (df=16)	30.814	28.380

p-value	0.014	0.028
R2		
McFadden	0.045	0.041
McFadden (adjusted)	-0.008	-0.011
McKelvey & Zavoina	0.083	0.090
Cox-Snell/ML	0.066	0.061
Cragg-Uhler/Nagelkerke	0.084	0.078
Count	0.613	0.611
Count (adjusted)	0.171	0.166
IC		
AIC	691.247	693.681
AIC divided by N	1.529	1.535
BIC (df=18)	765.294	767.728
Variance of		
e	3.290	1.000
y-star	3.587	1.099

When the goodness-of-fit statistics in Table 8 are examined, it is seen that both models are suitable for examining the patients' satisfaction, although there are generally smaller values for the ordinal logit model.

Conclusion

In ordinal choices models, the dependent variable has the feature of being ordinal in addition to its nominal feature. The ordinal logit and probit models are used when the categories of the dependent variable are more than two (at least 3) and are naturally ordered. This ordered style is mostly used in Likert type scales, especially in survey data. For example, in studies on consumer brand preferences, the ranking can be determined as "moderate" "good" "very good". In such cases, the categories (levels) of the dependent variable are determined by coding as the nominal scale or as 1, 2, 3, …. Thus, the categories have a natural ordering and do not have numerical superiority over each other. These models are also known as generalized linear models.

Newton Raphson algorithm is the most preferred method to obtain parameter estimates of ordinal dependent variable models. The threshold number for these regression models is 1 less than the category number of the dependent variable.

In order to apply the classical ordinal logit and ordinal probit models, the assumption of parallel slopes must be satisfied. When the dependent variable has more than two categories, multinomial logit models are also used. However, these model do not take into account the ordinal nature of the dependent variable. In other words, they neglect the order that exists between the levels of the dependent variable. If there is uncertainty about whether the dependent variable is ordinal or not, it can be estimated with multinomial models to determine whether the assumptions are met.

Multinomial analysis techniques used in ordinal or categorical dependent variable models ignore the ordinal nature of the dependent variable and assume that it has a nominal structure. For this reason, the use of multinomial analyses in cases where the assumption of parallel is not provided causes loss of information. At this point, another model has been proposed, which provides flexibility in the assumption of parallel slopes and takes into account the ordered structure. This model is a generalized ordinal logit model that uses cumulative logit models that do not satisfy the parallel assumption. It can also be used in the generalized probit model.

When the existing research is examined, it is seen that researchers generally prefer logit models. The main reason for this is thought to be the more widespread knowledge of the logit model. However, as can be seen from the results, no significant difference was found between the two models. There are no strict rules for choosing a probit model or a logit model. Sometimes it is desirable to deal with the tail part of the curve. In this case, the choice of logit or probit may be important because the tail thickness of the logit model is greater than that of the probit. In such cases, the choice of the appropriate model can be decided by looking at log probability or goodness-of-fit statistics. When the parallelism assumption is not met, other suitable models should be used as mentioned before.

References

Agresti, A. (2002). *Categorical Data Analysis*. Second Edition, Wiley, Hoboken, New Jersey.

Aitcheson, J & Silvey, S. (1957). The Generalization of Probit Analysis to the Case of Multiple Responses. *Biometrika, 44*, 131-140.

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, Harvard University Press.

Ananth, C. & Kleinbaum, D. (1997). Regression Models for Ordinal Responses: A Review of Methods and Applications. *International Journal of Epidemiology, 26*, 1232-1333.

Arı, E. & Yıldız, Z. (2014). Parallel Lines Assumptıon in Ordinal Logistic Regression and Analysis Approaches. *International Interdisciplinary Journal of Scientific Research, 1*(3), 8-23.

Barak, N. A. (2005). *Sıralı (Ordinal) ve Multinomial Logit Modeller Üzerine Bir Uygulama* (in Turkish), Unpublished Master's Thesis, Hacettepe University, Institute of Health Sciences, Ankara, 2005.

Bender, R. & Grouven, U. (1998). Using Binary Logistic Regression Models for Ordinal Data with Non-Proportional Odds. *Journal of Clinical Epidemiology, 51*(10), 809-816.

Brant, R. (1990). Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics, 46* (4), 1171-1178.

Butler, J. & Moffitt, R. (1982). A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model. *Econometica*,. *50*, 761-764.

Dimitrievska, V. & Tomovska, A.M. (2020). *Factors connected to patients' satisfaction in the health care system in North Macedonia*. Book of Abstracts of the Fifteenth Annual International Conference on European Integration – AICEI.

Finney, D. J. (1971). *Probit analysis*, 3rd Edition, Cambridge University Press, Cambridge.

Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Sage Publications, Thousand Oaks, CA, 1997.

Fullerton, A. S. & Xu, J. (2012). The Proportional Odds with Partial Proportionality Constraints Model for Ordinal Response Variables. *Social Science Research, 41*, 182-198.

Fu, V. (1998). Estimating Generalized Ordered Logit Models. *Stata Technical Bulletin, 8*(44), 160-164.

Greene, W. H. & Hensher, D. A. (2010). *Modeling Ordered Choices a Primer.* Cambridge University Press, New York, 2010.

Greene, W. H. (2012). Econometric Analysis, (7. ed.). Boston: Pearson Education. 827–831.

Hardin, J. & Hilbe, J. (2001). *Generalized Linear Models and Extensions*. Stata Press, College Station, Texas.

Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression*, Second Edition, Wiley, New York, 2000.

Jackman, S. (2000). *Models for Ordered Outcomes*. Political Science.

Liao, T. F. (1994). *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models.* Sage Publications, Thousand Oaks, CA, 1994.

Long, S. J. (1997*). Regression Models for Categorical and Limited Dependent Variables*. Sage Publications, Thousand Oaks, CA, 1997.

McElvey, R. & Zavoina, W. (1975). A Statistical Model for the Analysis of Ordered Level Dependent Variables. *Journal of Mathematical Sociology, 4,* 103-120.

McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society, Series B (Methodological), 42*(2), 109-142.

O'Connell, A. A. (2005). Logistic Regression Models for Ordinal Response Variables. *Quatitative Applications in the Social Sciences*, *146*, 105. Sage Publications

Petersonm, B. & Harrell, J. R. (1990). Partial Proportional Odds Models for Ordinal Response Variables. *Applied Statistics*, *39*(2), 205-217.

Snell, E. (1964). A Scaling Procedure for Ordered Categorical Data. *Biometrics*, *20*, 592-607.

Timur, B. & Akay, E. Ç. (2017). Kadın ve Erkeklerin Mutluluğunu Etkileyen Faktörlerin Genelleştirilmiş Sıralı Logit Modeli ile Analizi" (in Turkish), *Social Sciences Research Journal*, *6*(3), 88-105.

Paper Received March 1, 2022; Accepted June 23, 2022; Published November 2, 2022