Estimating the Number of Unemployed Months for Individuals in Turkey with the Poisson and Negative Binomial Regression Models

Elvan HAYAT* Afet SÖZEN ÖZDEN**

ABSTRACT

Unemployment is one of the greatest economic and social problems in Turkey, as well as it is in many other countries in the world. Unemployment is often explained by macroeconomic factors. However, demographic and individual characteristics also have an effect on the unemployment duration of individuals, in addition to the macroeconomic factors. The present study aims to find the factors that have an effect on the duration of unemployment of individuals in Turkey with count data regression models. Therefore, the present study examined Poisson Regression (PR) and Negative Binomial Regression (NBR) models, which are used in cases that the dependent variable is count data. The study also aims to determine the model with the best fit to the dataset among the estimated models. In the study, the number of months in which individuals were unemployed was modeled, using the data obtained from the Survey of Income and Living Conditions (SILC) micro dataset of the Turkish Statistical Institute (TURKSTAT) in 2019. 62713 people aged 15 and over participated in the SILC, of which 5889 reported that they were unemployed for one month or more. A model with the best fit and with the independent variables of marital status, education status, and general health status was determined among the seven models determined by the forward selection method. It has been determined that the model that best fits the dataset among the predicted models is the NBR model according to the Akaike Information Criterion (AIC).

Key Words: *Count data, Poisson Regression Model, Negative Binomial Regression Model*

JEL Classification: *C10, C46, D30*

Poisson ve Negatif Binom Regresyon Modelleri ile Türkiye'de Bireylerin İşsiz Kaldığı Ay Sayısının Tahmini

ÖZ

İşsizlik dünyanın pek çok ülkesinde olduğu gibi, Türkiye'de de ekonomik ve sosyal sorunların en başında yer almaktadır. İşsizlik genellikle makro iktisadi faktörlerle açıklanmaya çalışılmaktadır. Bireylerin işsizlik süresi ise makro iktisadi faktörlerin yanı sıra, demografik ve bireysel özelliklerle de ilgilidir. Bu çalışma, Türkiye'de bireylerin işsiz kalma sürelerini etkileyen faktörleri sayma veri regresyon modelleri ile belirlemeyi amaçlamaktadır. Bu amaçla, bağımlı değişkenin sayma verisi olduğu durumlarda kullanılan Poisson Regresyon (PR) ve Negatif Binom Regresyon (NBR) modelleri ele alınmıştır. Tahmin edilen modellerden veri setine en iyi uyum sağlayan modelin belirlenmesi hedeflenmiştir. Çalışmada, Türkiye İstatistik Kurumu (TÜİK) tarafından 2019 yılında yapılan Gelir ve Yaşam Koşulları Araştırması (GYKA) mikro veri setinden elde edilen veriler kullanılarak bireylerin işsiz kaldığı ay sayısı modellenmiştir. GYKA'ya 15 yaş ve üzeri 62713 kişi katılmış, 5889'u işsiz kaldığı ay sayısının sıfırdan farklı olduğunu belirtmiştir Bağımsız değişkenlerin seçimi için ileri seçim yöntemi kullanılmıştır. İleri seçim yöntemi ile

*Doç. Dr., Aydın Adnan Menderes Üniversitesi Aydın İktisat Fakültesi, Ekonometri Bölümü. elvan.hayat@adu.edu.tr, ORCID Bilgisi: 0000-0001-8200-8046

** Uzman Ekonometrist. afetsozen@gmail.com, ORCID Bilgisi: 0000-0001-9821-7029

(Makale Gönderim Tarihi: 03.02.2022 / Yayına Kabul Tarihi:09.03.2023)

Doi Number: 10.18657/yonveek.1067907

Makale Türü: Araştırma Makalesi

belirlenen yedi modelden en uygun olanının medeni durum, eğitim durumu ve genel sağlık durumu bağımsız değişkenlerinin olduğu model olduğu tespit edilmiştir. Akaike Bilgi Kriteri'ne (AIC) göre, tahmin edilen modeller içerisinde veri setine en iyi uyum sağlayan modelin NBR modeli olduğu belirlenmiştir.

Anahtar Kelimeler: *Sayma verisi, Poisson Regresyon Modeli, Negatif Binom Regresyon Modeli*

JEL Sınıflandırması: *C10, C46, D30*

INTRODUCTION

Unemployment is defined as "being unemployed, being unable to find a job" in the dictionary of the Turkish Language Association (TLA). Unemployment is one of the most important economic and social problems for societies. The internationally recognized standard definition of the term "unemployment" is based on three criteria: not having a job, being ready to start working, and looking for a job. In order for an individual to be considered as unemployed, they must have all three criteria simultaneously (Boztepe, 2007). The term unemployed is defined by the Turkish Statistical Institute (TURKSTAT) as all non-institutional persons of working age who have used at least one of the active job search channels in the last four weeks, are able to start working within 2 weeks, and are not employed during the reference period (TURKSTAT, 2021).

Previous studies generally examine unemployment in terms of macroeconomic factors (Akhtar and Shahnaz, 2005; Anyanwu 2013; Arslan and Zaman, 2014; Günaydın and Çetin, 2015; Abugamea, 2018; Ayhan, 2019; Kabaloğlu et al., 2022). However, the factors affecting unemployment can also be at an individual level. In addition to macroeconomic factors, the demographic and unique characteristics of individuals can also play a role in being unemployed. Unemployment and prolongation of unemployment have a significant impact on the lives of individuals. Moreover, the unemployment period is a problem that has special importance both in individual and social dimensions. In fact, the longer the unemployment period, the more likely it is that unemployment can turn from an individual problem to a social one. As the duration of unemployment increases, people's self-confidence may decrease, along with an increase in family problems. Some psychological problems can also occur in individuals who experience unemployment for a long time. Individuals can shorten the unemployment period by improving their qualifications while they are in an unemployed period. This is a choice within the control of individuals. However, macro-level factors cannot be changed at the individual level. These factors are beyond the control of individual (Arslan and Şentürk, 2018: 114).

There are also many studies in the literature in which the factors affecting unemployment or youth unemployment are investigated using micro data. However, in these studies, unemployment status (as unemployed or working) was considered as a two-category variable and analyzes were made with logit and probit regression models. In studies on unemployment or youth unemployment by microdata in Turkey and in the world, it has been found that the main factors are generally gender, marital status, education level and geographical conditions (Msigwa and Kipesha, 2013; Green, Loon and Mangan, 2000; Terzo and Giaconia, 2018; Emeç et al., 2021). In the study by Green, et al. (2000), it is aimed to determine the youth employment situation in Queensland (Australia). The data used in the research have obtained from the Queensland Youth Survey. Variables such as education level, gender, language, experience and age were used in the study. The logit model was used to examine the effect of these variables on youth unemployment. As a result of the research, it has been determined that the factor that affects youth unemployment the most is the level of education and that the young people who are unsuccessful in education are in a bad situation in terms of both finding a job and entering employment. In the study conducted by Msigwa and Kipesha (2013), it was aimed to examine the factors that determine youth unemployment in Tanzania. The Multinomial Logistics Model was used to analyze the determinants of unemployment in Tanzania in the study. As a result of the research, it was determined that the variables of gender, geographical location, education, skills and marital status were important factors in explaining the difference in employment status of young people in Tanzania. In the study conducted by Tunçsiper and Rençber (2020), data from the Household Labor Force Survey (2019) of TURKSTAT were used. Micro factors affecting youth unemployment (as education, gender, age, position in the household, household size and type of settlement) were modeled with the Binary Logit model. As a result of the research, it has been determined that the variables of education, position in the household, household size, gender and type of settlement are effective factors on youth unemployment in Turkey.

In the studies in the literature in which unemployment time is considered as the dependent variable, it has been determined that they generally use the Cox regression method by considering the unemployment time as censored data. In the study of Taşçı and Özdemir (2006), it was aimed to determine the factors affecting long-term unemployment in Turkey by using the data of the Household Labor Force Survey of TURKSTAT. Factors such as settlements, gender, marriage, region, education, occupation, age and job type were examined with the Two-Stage Probit model. As a result of the research, it has been determined that gender, age, education, settlements, region, status in the household and occupation factors are effective on long-term unemployment in Turkey. Hunt (1995) aimed to examine unemployment compensation for unemployed individuals without children and unemployment insurance for unemployed individuals aged 41 and over in West Germany. In this study, the unemployment duration of unemployed individuals who received unemployment compensation and unemployment insurance were analyzed in terms of demographic characteristics, with the Cox regression model. The results of the analysis showed that the gender factor has no effect on unemployment insurance or on the unemployment duration of individuals receiving unemployment compensation. The study found that the unemployment duration length of individuals is more affected by their personal behavior than by their demographic characteristics. Grogan and Berg have been used the Russian Longitudinal Monitoring Survey (RLMS) to assess factors affecting the duration of unemployment and underemployment in Russia between 1994 and 1996. In the study, duration models were estimated using non-parametric and parametric estimation techniques. As a result of the research, it has been determined that those who have completed higher education have relatively short periods of unemployment and underemployment, regardless of their previous job type. Furthermore, it was concluded that married women have relatively long periods of unemployment. The study by Denisova (2002) investigated the effect of individual characteristics on unemployment durations of registered unemployed persons collected by the Federal Employment Service between 1996 and 2000 in Voronezh, Russia. The study showed that women had a shorter time of unemployment than men, and the more educated individuals had shorter time as an unemployed.

In their study, Borsic and Kavkler (2008) estimated the effect of the demographic characteristics of individuals on their unemployment durations for Slovenia between the years 2002-2005 using the Cox regression model. In the study, it was concluded that there are significant differences in the duration of unemployment between age groups. It was found that the region of residence also had an effect on the period of being unemployed. It has been found that as the education level increases, the duration of finding a job shortens, and it is concluded that this, in turn, shortens the period of being unemployed. The study found that women were unemployed for longer periods of time than men. Danacica (2008) aimed to examine the effect of selected variables on unemployment duration in Romania. Using data from the Romanian National Employment Agency, duration data models such as Kaplan Meier method, Cox proportional risk model and time dependent covariate Cox regression were estimated. The probability of men being unemployed was found to be higher than that of women. It has been concluded that age is a disadvantage in finding a job. It has also been determined that as the level of education increases, the period of being unemployed decreases. Karasoy et al. (2015) aimed to analyze the factors affecting the duration of unemployment in Turkey with survival analysis. The analysis results showed that the gamma regression model had the best fit. In addition, the study showed that gender, marital status, education level, course, province and age factors had an effect on the duration length of unemployment.

In the study conducted by Sözen Özden and Hayat (2022), it was aimed to determine the factors affecting the number of months in which unemployed individuals were unemployed. In the study, the data of TURKSTAT 2019 SILC was used. In the study, the effects of marital status, education status, general health and chronic disease variables on the duration of unemployment were analyzed with count data regression models. It was determined that the most suitable method for the model determined in the study was the Zero Inflated Negative Binomial (ZINB) Regression model. As a result of the research, it was determined that the variables of marital status, education status, general health status and chronic disease status were effective factors on the duration of unemployment.

The present study aims to determine the factors that affect the number of months in which individuals are unemployed, using data from the 2019 SILC conducted by TURKSTAT. The study aims to find the most appropriate estimation model by using count data regression models for estimating months unemployed. Properties of PR and NBR methods are described and independent variable selection methods are explained in the following section of the study. Afterwards, empirical analysis is presented. In this section, both the dataset and the model are explained in detail, and descriptive statistics and model estimation results are presented. In the conclusion and recommendations section, the analysis findings are summarized and a general discussion is presented.

I. METHOD

A. Poisson Regression Model

The PR Model is a model that forms the basis of count data models. It is a nonlinear regression model. In the PR model, the distribution of the data of the dependent variable has a Poisson distribution. The Poisson distribution is used for determining the number of events that occur frequently in real life within a certain time interval. In Equation 1, the probability density function of the Poisson distribution is given with the discrete random variable Y and the μ parameter μ > 0 (Cameron and Trivedi, 1998: 3):

$$f(Y = y) = \frac{\mu^y e^{-\mu}}{y!}, \qquad y = 0,1,2,3, \dots \tag{1}$$

The mean and variance of the Poisson distribution are the same, as shown in Equation 2:

$$E(Y) = Var(Y) = \mu \tag{2}$$

This is expressed as equidispersion in the PR model, where the conditional mean and variance are equal. In practice, this equality generally cannot be achieved, and when not, overdispersion or under dispersion is often observed.

B. Negative Binomial Regression Model

The excess of zero values or higher values in the dataset obtained based on the counting violates the assumption of equal dispersion. If the dependent variable variance is greater than the conditional mean, overdispersion occurs. In cases of overdispersion, the PR model is not sufficient, although consistent. This may cause biased standard errors and misinterpretation of the coefficients by the researcher. An alternative method used for cases where the dependent variable shows overdispersion is the NBR Model (Beaujean and Morgan, 2016: 4). The probability density function of the NBR model is shown in Equation 3 (Lawless, 1987, p.210):

$$f(y_i) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\Gamma(\alpha^{-1})} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}} , \quad y_i = 0,1,2, \dots \quad \alpha > 0 \tag{3}$$

Gamma function is represented by Γ(.) in Equation 3. α represents the distribution parameter. In case that α>0, it refers to the presence of overdispersion. The NBR model is a special case of the PR model. Parameter estimates for the NBR model are obtained by the Maximum Likelihood method and iterative (repetitive) methods.

C. Selection of the Independent Variables

In regression analysis, some independent variables in the model may not be statistically significant. In this case, independent variables that have no effect on the dependent variable should be excluded from the model. Determining the independent variables that best represent the dependent variable in the regression analysis is of vital importance. Different selection methods have been developed to be used for selecting the independent variables to be included in a model. The choice of independent variables is important in cases where there are three or more variables. Forward selection, backward elimination, stepwise selection and all possible subsets selection methods are frequently used in the selection of variables that should be included into the model.

The forward selection method starts with the estimation of the model, where merely the constant term is included. Afterwards, among the independent variables, the variable with the highest correlation value with the dependent variable is added to the model. The process continues by adding a new highly correlated variable without removing the variable added to the model. The goodness-of-fit criteria of all estimated models are compared and the most appropriate model is selected (Tarı, 2018: 337).

In the backward elimination method, all the variables that are considered as independent variables are added to the model, then the model is estimated. In the estimated model, the weakest independent variable is removed from the model and the model is estimated once more. If the variable which is removed from the model is statistically significant and the model becomes significantly weaker upon removal, it is decided not to exclude this dependent variable from the model. The process continues until only the independent variables that contribute significantly to the model remain in the model (Tarı, 2018: 336).

In the stepwise selection method, model estimations are evaluated after each variable is added to the model sequentially. An added variable remains only if it contributes to the model. All of the other variables in the model are tested for their contribution to the model once more. Variables that do not contribute significantly are excluded from the model (Küçüksille, 2014: 260).

In the all possible subsets method, an n number of variables that are likely to have an effect on the dependent variable are determined. The number of models to be estimated is calculated using the 2^n-1 formula. Once all the determined models have been estimated, the model comparison is performed. By this way, the model with the independent variables that best explains the dependent variable is determined. The all possible subset is an alternative variable selection method that requires more computation in comparison with the other variable selection methods.

II. EMPIRICAL ANALYSIS

A. Data and Model

In the present study, questionnaire data of 2019 SILC obtained from the TURKSTAT micro database were used. It was aimed to determine the socio-demographic factors that have an impact on the months unemployed in the 2019 SILC data. The question "How many months was the individual unemployed in 2018?" was answered by 62713 people aged 15 and older. 5889 out of 62713 people stated that the number of months they had been unemployed was different from zero. Stata 15.0 statistical package program was used for the analysis. By combining categories for the analysis, the "education" variable was reduced from 11 categories to 6 categories, and the "general health" variable was reduced from 5 to 3 categories. Table 1 shows the independent variables used in the analysis.

Table 1. Independent variables

Independent Variables	Description
X_1: Marital Status	1: Married 2: Never married 3: Spouse passed away 4: Divorced
X_2: Education	0: Illiterate 1: Less than high school 2: High school 3: College 4: Bachelor's 5: Postgraduate
X_3: General state of health	1: Good 2: Average 3: Bad
X_4: Chronic Disease	1: Yes 2: No
X_5: Sex	1: Male 2: Female
X_6: Age	Age of the individual

The variables given in Table 1 show all the variables obtained from the 2019 TURKSTAT SILC that may affect the number of months of unemployment.

Among these variables, it is thought that the variable of marital status may have an effect on the number of months of unemployment depending on the responsibility of supporting for the family. Due to the high number of dependents, it is assumed that the unemployment duration of married individuals is shorter than that of single or divorced individuals. Another variable that is thought to be effective on the duration of unemployment is the variable of educational status (Msigwa and Kipesha, 2013; Terzo and Giaconia, 2018). Having a high level of education may provide an advantage in getting a job, but this may cause individuals to act selectively in their job preferences (Sözen Özden and Hayat, 2022: 12).

It is thought that the general health status of individuals is among the determining factors for the duration of unemployed. Since the poor general health status of individuals may affect the length of time they are unemployed, it is included in the model assuming that it is an effective factor on the duration of unemployment. (Bayrak, 2020: 53; Emeç et al., 2021; Sözen Özden and Hayat, 2022: 12). The variable of chronic disease status has been included in the model as an independent variable since it is thought that the chronic diseases of individuals may cause a prolongation of the unemployment period. Another variable that is thought to be effective on the duration of unemployment is the gender variable. (Perugini and Signorelli, 2010; Emeç et al., 2021). Since the patriarchal society understanding prevails in our country, men are thought to be unemployed for a shorter period of time than women (Bayrak, 2020: 52). Another variable that is thought to be effective on the duration of unemployment is the age variable. Young people are thought to have more to find a job opportunities than middle-age groups have. Therefore, the age variable was included in the model with the assumption that the time for young people to be unemployed would be shorter. In addition, it is assumed that as the age of the individual increases, the experience increases and the experienced individuals are unemployed for a shorter period of time. (Bayrak, 2020: 52).

The distribution of unemployed individuals aged 15 and over participating in the 2019 TURKSTAT SILC according to the Turkey Statistical Regional Units Classification (NUTS) is given in Chart 1.

Chart 1. Distribution of the individuals participating in the research according to the NUTS

When Chart 1 is examined, it is concluded that the unemployed individuals aged 15 and over participating in the research are mostly in the TRC (South East Anatolia) region with 13.38%, and at least in the TR9 (Eastern Black Sea) region with 4.89%.

Independent variables to be used for unemployed month numbers of individuals were determined by the forward selection method. 7 models were created by adding the independent variables sequentially. The models established with the forward selection method are given in Table 2.

Table 2. Models Established for Appropriate Independent Variable Selection

Model Number	Established Model
Model 1	β_0
Model 2	$\beta_0 + \beta_1 X_1$
Model 3	$\beta_0 + \beta_1 X_1 + \beta_2 X_2$
Model 4	$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$
Model 5	$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$
Model 6	$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$
Model 7	$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$

The PR and NBR models were estimated for the models shown in Table 2. The results of the model selection can be seen in Table 3 below.

Table 3. Estimation Results for Model Selection

Poisson Regression				
Model	Deviation	Deviation Degrees of Freedom	AIC	α (Distribution Parameter)
1	12119.49	5888	33307.52	2.058
2	11749.26	5885	32943.29	1.996
3	11686.8	5881	32888.83	1.987
4	11643.46	5879	32849.49	1.981
5	11643.45	5878	32851.49	1.981
6	11642.02	5877	32852.05	1.981
7	11642.02	5876	32854.05	1.981
Negative Binomial Regression				
Model	Deviation	Deviation Degrees of Freedom	AIC	α (Distribution Parameter)
1	31587.394	5887	31591.394	5.366
2	31405.300	5884	31415.300	5.337
3	31373.090	5880	31391.090	5.336
4*	31351.222	5878	31373.222	5.334
5	31351.216	5877	31375.216	5.335
6	31350.480	5876	31376.480	5.335
7	31350.479	5875	31378.479	5.336

As it can be seen in Table 3, the distribution parameter is greater than 1 in all models, thus it is concluded that overdispersion is present in all models. The most appropriate model for analysis is the one with the smallest Akaike Information Criterion (AIC). In this regard, Model 4 has the smallest AIC. However, this model is not suitable for PR analysis because of overdispersion. Therefore, the resulting most appropriate method for Model 4 is the NBR model. The independent variables used in Model 4 are marital status, education, and general state of health.

B. Descriptive Statistics

The "Months Unemployed" variable used in the analysis is a discrete variable, while the "Marital status", "Education", "General state of health" variables are categorical variables. The frequency table of the "Months Unemployed" variable is shown in Table 4.

Table 4. Frequencies of the dependent variable

Months Unemployed	Frequency	Percentage	Cumulative
1	365	6.20	6.20
2	455	7.73	13.92
3	460	7.81	21.74
4	515	8.75	30.48
5	438	7.44	37.92
6	1004	17.05	54.97
7	397	6.74	61.71
8	365	6.20	67.91
9	308	5.23	73.14
10	206	3.50	76.63
11	105	1.78	78.42
12	1271	21.58	100.00

Table 4 shows that 1271 people were unemployed for 12 months and 1004 people were unemployed for 6 months. 365 of the participants stated that they were unemployed for 1 month and 105 of them were unemployed for 11 months. Individuals who have been unemployed for 12 months constitute 21.58% of unemployed individuals who participated in the study. This is followed by individuals who have been unemployed for 6 months with 17.05%. 6.20% of the individuals participating in the research stated that they were unemployed for 1 month in the last year, 7.73% for 2 months in the last year, 7.81% for 3 months in the last year, 8.75% for 4 months in a year, 7.44% for 5 months in the last year, 6.74% for 7 months in the last year, 6.20% for 8 months in a year. In addition, 1.78% of the participants stated that they were unemployed for 11 months in the last year. Table 5 shows the descriptive statistics of the dependent variable.

Table 5. Descriptive Statistics of the Dependent Variable

Dependent Variable	Number of Observations	Mean	Standard Deviation	Min-Max	Q_1	Q_2	Q_3
Months Unemployed	5889	6.77	3.59	1-12	4	6	10

As seen in Table 5, 5889 people attended SILC in 2019 and stated that they were unemployed for at least 1 month in 2018. The average months unemployed is approximately 6.8 months. The standard deviation of the dependent variable was found to be 3.59, and consequently, the variance of the dependent variable was 12.89. It was concluded that the variance of the dependent variable was larger than the mean, which is one of the indicators of overdispersion. Q_1, Q_2, Q_3 seen in Table 5 represent the quartiles. 1st quartile (Q_1) is calculated as 4, 2nd quartile (Q_2) as 6 and 3rd quartile (Q_3) as 10. The second quartile also shows the median value. Since the values of the calculated quartiles fulfill the formula Q3 - Q2 > Q2 - Q1, it is concluded that the distribution of the dependent variable is skewed to the right. Based on these results and due to the structure and distribution of the dependent variable, it can be said that one of the count data regression models may be appropriate in the model estimation. The frequency table of the categorical independent variables is shown in Table 6.

Table 6. Frequencies of independent variables

Independent Variables	Frequency	Percentage	Cumulative
Marital Status			
1: Married	2.769	47.03	47.03
2: Never Married	2.874	48.80	95.82
3: Spouse passed away	19	0.32	96.15
4: Divorced	227	3.85	100.00
Education			
0: Illiterate	121	2.06	2.06
1: Less than high school	3.197	54.29	56.35
2: High school	1.208	20.51	76.86
3: College	489	8.30	85.16
4: Bachelor’s	828	14.06	99.22
5: Postgraduate	46	0.78	100.00
General state of health			
1: Good	4617	78.40	78.40
2: Average	995	16.90	95.30
3: Bad	277	4.70	100.00

In Table 6, the marital status variable consists of four categories as married, never married, spouse passed away and divorced. Educational status variable consists of 6 categories as illiterate, less than high school, high school graduate, college graduate, Bachelor's degree and postgraduate degree. General state of health variable consists of 3 categories of good, average and bad. Among unemployed SILC participants, 47.03% are married, 48.8% have never been married, 0.32% are widowed, and 3.85% are divorced individuals. 2.06% of the unemployed individuals participating in the research are illiterate, 54.29% have a less than high school education level, 20.51% are high school graduates, 8.3% are college graduates, 14%, 06 are Bachelor's graduates and 0.78% are postgraduate graduates. 78.4% of unemployed individuals stated that their general state of health status was good, 16.9% stated that it was average and 4.7% stated that it was bad.

C. Model Estimation

In the study, Model 4 was estimated with the NBR model and it was found by the forward selection method that it would be the most suitable model for independent variable selection. The results of the estimated NBR model are shown in Table 7.

Table 7. NBR Model Estimation Results

Log likelihood = -1567.242 Number of Observations = 5889		LR $\chi^2(10)$ = 236,91	Prob. > χ^2 = 0,000	
Independent Variables	Coefficient (β)	Standard Error	IRR	Probability
Constant	2.089	0.0382	8.074	0.000***
Marital Status (Reference: Married)				
Never married	0.198	0.016	1.219	0.000***
Spouse passed away	-0.110	0.132	0.896	0.406
Divorced	0.176	0.037	1.193	0.000***
Education (Reference: Postgraduate)				
Illiterate	-0.095	0.092	0.909	0.300
Less than high school	-0.182	0.078	0.834	0.019**
High school	-0.170	0.079	0.844	0.031**
College	-0.180	0.081	0.835	0.026**
Bachelor's	-0.068	0.079	0.934	0.390
General health status (Reference: Bad)				
Good	-0.141	0.034	0.869	0.000***
Average	-0.088	0.036	0.916	0.016**
α(Distribution parameter)	5.334			
LR test of alpha=0: χ^2 = 1477.50 Prob. >= χ^2 = 0.000				
*** Statistically significant with a 1% margin of error. ** Statistically significant with a 5% margin of error.				

The coefficient, standard error, incidence rate ratio (IRR) and probability values of the NBR model are given in Table 7. The model is significant in a general sense because the chi-square test probability value of the estimated model is $p < 0.01$. In the estimated model, the "Married" category is the reference category for the marital status variable. In terms of marital status variable, the probability value of the "Never married" and "Divorced" categories is $p<0.01$, therefore the estimated coefficient and IRR values for these categories are statistically significant. The reference category for the educational status variable is "Postgraduate". In terms of the education variable, since the probability values of the "Less than high school", "High school" and "College" categories are $p<0.05$, the estimated coefficient and IRR values for these categories are statistically significant. In the general state of health, the reference is the category titled "Bad". In the categories belonging to the general state of health variable, the probability value of the "Good" category is $p<0.01$ and the probability value of the "Medium" category is $p<0.05$, therefore the coefficient and IRR values of the categories are statistically significant. The interpretations of the statistically significant IRR values of the estimated NBR model are as follows:

- **Never married:** The number of unemployed months for an individual who has never been married is approximately 1.22 times more than the number of unemployed months for a married individual.
- **Divorced:** The number of unemployed months of a divorced individual is approximately 1.193 times more than the number of unemployed months of a married individual.
- **Less than high school:** The number of unemployed months for an individual with an education level that is less than high school is approximately 0.83 times less than the number of unemployed months for an individual with a postgraduate degree.

- **High school:** The number of unemployed months of a high school graduate is approximately 0.84 times less than the number of unemployed months of an individual with a postgraduate degree.
- **College:** The number of unemployed months of a college graduate is approximately 0.84 times less than the number of unemployed months of an individual with a postgraduate degree.
- **Good:** The number of unemployed months of an individual with good general state of health is approximately 0.87 times less than the number of unemployed months of an individual with bad general state of health.
- **Average:** The unemployed months of an individual with an average general state of health is approximately 0.92 times less than the number of unemployed months of an individual with a bad general state of health.

As expected before the analysis, individuals who have never been married or have divorced are unemployed longer than married individuals. This result show that the married individuals are not likely to act selectively in their job prefences due to of providing for their families. As expected, it was concluded that the increase in education level increases the duration of unemployment.The reason for this is that as the level of education increases, individuals become more selective in their job preferences. As the general health condition of the individual improves, the duration of unemployed decreases, as expected before the analysis. A good health status of the individual will reduce the probability of being unemployed. As the health status of the individual gets better, the duration of being unemployed will decrease.

III. CONCLUSION AND RECOMMENDATIONS

Previous studies in the literature have generally modeled unemployment with macroeconomic factors. However, the factors affecting unemployment can also be at an individual level. In addition to macroeconomic factors, unemployment may also be related to the demographic characteristics of individuals. There are many studies in which unemployment is also investigated using micro data. However, in these studies, unemployment was generally considered as a two-category variable. In addition, there are studies in the literature that use the Cox regression method by considering the unemployment duration as censored data.

Being unemployed and the prolongation of the period of unemployed significantly affect the lives of individuals. For these reasons, the present study aimed to determine the individual factors affecting unemployment. In the study, the answers given by individuals aged 15 years and older who participated in the 2019 Income and Living Conditions Survey to the question of "How many months did the individual spend unemployed in 2018?" were determined as dependent variable. 56824 people who stated the number of months they were unemployed as zero were not included in the analysis. The "number of unemployed months" data determined as the dependent variable of the present study is a count data.

Generally, count data regression models are used when the dependent variable is obtained based on counting. In the study, PR and NBR models from count data regression models were examined. The PR model is an equally dispersed model based on the assumption of mean and variance equality. NBR, on the other hand, is an over dispersion model based on the assumption that the variance is greater than the mean. In the present study, it was found that the dataset showed over dispersion and the NBR model was used in the analysis. The dependent variable was accepted as the number of unemployed months of individuals. The independent variables were determined as marital status, education and general state of health variables with the forward selection method.

A general evaluation of the estimation results showed that marital status, education, general state of health variables has an impact on the number of months unemployed. It has been concluded that the number of unemployed individuals who have never been married is higher in 2018 than married individuals. Divorced individuals were also unemployed for more months than married individuals in 2018. These results regarding marital status show that married individuals have fewer unemployed months in a year. It was found that the number of unemployed months of individuals with a postgraduate degree was higher than that of individuals with other education levels. In addition, the worse the general state of health of the individual, the higher the number of months that they are unemployed.

This study establishes a guidance for future studies to determine the factors affecting the number of unemployed months for individuals. Conducting similar studies at the micro level and conducting replication studies on a regular basis will provide useful information to policy makers and decision makers.

Araştırma ve Yayın Etiği Beyanı

Makalenin tüm süreçlerinde Yönetim ve Ekonomi Dergisi'nin araştırma ve yayın etiği ilkelerine uygun olarak hareket edilmiştir.

Yazarların Makaleye Katkı Oranları

Yazarlar çalışmaya eşit oranda katkı sağlamıştır.

Çıkar Beyanı

Yazarların herhangi bir kişi ya da kuruluş ile çıkar çatışması yoktur.

REFERENCES

Abugamea, G. H. (2018). Determinants of Unemployment: Empirical Evidence from Palestine. Munich Personal RePEc Archive. https://mpra.ub.uni-muenchen.de/89424/

Akhtar, S. and Shahnaz, L. (2005). Understanding the Youth Unemployment Conundrum in Pakistan: Preliminary Empirical Macro-Micro Analysis. *Indian Journal of Labour Economics*, Vol. 49, Iss. 2, 233-248.

Anyanwu, J. C. (2013). Characteristics and Macroeconomic Determinants of Youth Employment in Africa. *African Development Review*, Vol. 25, Iss. 2, 107-129.

Arslan, H., and Şentürk, İ. (2018). Türkiye'de İşsizlik Süresinin Bireysel Belirleyicileri. *Eskişehir Osmangazi Üniversitesi İİBF Dergisi,* Vol. 13, Iss. 1, 113-128.

Arslan, M. and Zaman, R. (2014). Unemployment and Its Determinants:A Study of Pakistan Economy (1999-2010). *Journal of Economics and Sustainable Development*, Vol. 5, Iss. 13, 20-25.

Ayhan, F. (2019). Türkiye Ekonomisinde İşsizliğin Belirleyicisi Olan Temel Makroekonomik Değişkenlerin Tespitine İlişkin Bir Uygulama. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, Iss. 25, 235-252.

Bayrak, Y. (2020). Türkiye'de Genç İşsizliği ve Yükseköğretim Öğrencileri Arasında İşsizlik Kaygısı: Bir Alan Araştırma Örneği. Yüksek Lisans Tezi. Bursa: Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü.

Boztepe, Y. (2007). Türkiye'de İşsizlik Kavramı ve İşsizliğin Ortadan Kaldırılması İle İlgili Bir Model Oluşturulması. *Yüksek Lisans Tezi*. İstanbul: Yıldız Teknik Üniversitesi, Sosyal Bilimler Enstitüsü.

Beaujean, A. A. and Grant, M. B. (2016). Tutorial on Using Regression Models with Count Outcomes Using R. *Practical Assessment, Research, and Evaluation,* Vol.21, Iss. 2.

Borsic, D., and Kavkler A. (2008). Modeling unemployment suration in Slovenia using Cox regression models. *Springer Transit Stud Rev.,* Vol. 16, 145-156.

Cameron, A. C. and Trivedi, P. K. (1998). Regression Analysis of Count Data. New York: Cambridge University Press.

Danacica, D. E. (2008). The Analysis of Unemployment in Romania Through Duration Models. International Conference in Economics & Finance.

Denisova, I. A. (2002). Staying Longer in Unemployment Registry in Russia: Lack of Education, Bad Luck or Something Else?, *Center for Economic and Finanacial Research and New Economic School.*

Emeç, H.; Üçdoğruk Birecikli, Ş. and Acar Balaylar, N. (2021). Türkiye'de Genç İşsizliğinin Profili: Panel Logit Model Tahmini. *Yönetim ve Çalışma Dergisi*. Vol. 5, Iss. 2, 137-154.

Green, C., Loon, A. and Mangan, J. (2000). Youth Labour Markets, Education and Employment Destination: Results From The Queensland Survey of Youth. *Labour Market Research Unit Department of Employment and Training*.

Grogan, L. and Berg, G. J. (1999). The duration of unemployment in Russia. *Serie research memoranda,* Iss. 15, 1-44.

Günaydın, D. and Çetin, M. (2015). Genç İşsizliğin Temel Makroekonomik Belirleyicileri: Ampirik Bir Analiz. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,* Iss. 22, 17-34.

Hunt, J. (1995). The Effect of Unemployment Compensation on Unemployment Duration in Germany. *Journal of Labor Economics,* Vol. 13, Iss.1, *88-120.*

Kabaloğlu, G., Konya, S. and Küçüksucu, M. (2022). Gelişmiş ve Gelişmekte Olan Ülkelerde Genç İşsizliğinin Ekonomik Belirleyicileri: Ampirik Bir Analiz. *MANAS Sosyal Araştırmalar Dergisi*, Vol. 11, Iss. 4, 1430-1448

Karasoy, D., Ata Tutkun, N. and Bulut, V. (2015). Türkiye'de İşsizlik Süresini Etkileyen Faktörler. *Uluslararası Yönetim İktisat ve İşletme,* Vol. 11, Iss. 24*, 57-76.*

Küçüksille, E. (2014). Çoklu Doğrusal Regresyon Modeli. Ş. Kalaycı içinde, SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri (p. 259-266). Ankara: Asil Yayın Dağıtım Ltd. Şti.

Lawless, J. F. (1987). Negative Binomial and Mixed Poisson Regression. *The Canadian Journal of Statistics,* Vol.15, Iss. 3*, 209-225.*

Msigwa, R. and Kipesha, E. F. (2013). Determinants of Youth unemployment in Developing Countries: Evidences from Tanzania. *Journal of Economics and Sustainable Development*, Vol. 4, Iss. 14, 67-76.

Perugini, C. and Signorelli, M. (2010). Youth Labour Market Performance in European Regions. *Economic Change and Restructuring*, Vol. 43, 151-185.

Sözen Özden, A., and Hayat, E. (2022). Sayma Verilerinin Modellenmesi ve Bireylerin İşsiz Kalma Süresi Üzerine Bir Uygulama. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* Iss. 50, 1-18. Erişim Linki: https://doi.org/10.30794/pausbed.1055797

Taşçı, H. M. and Özdemir, A. R. (2006). Trends in Long-Term Unemployment and Determinants of Incidence of Long-Term Unemployment in Turkey. *Journal of Economic and Social Research*, Vol. 7, Iss. 2, 1-33.

Terzo, G. and Giaconia, C. (2018). The Socio-Demographic Determinants of Youth Unemployment in Italy: Evidences from National Labour Survey. 23-25 January, Seville, Spain

Tunçsiper, B. and Rençber, E. Z. (2020). Türkiye'de genç işsizlik sorunu ve belirleyicileri üzerine logit model analizi. *International Journal of Social Sciences and Education Research*, Vol. 6, Iss. 4, 514-530.

Tarı, R. (2018). Ekonometri. 13. Baskı, Kocaeli: Umuttepe Yayınları.

TURKSTAT (2021). Türkiye İstatistik Kurumu, 2021. Erişim Linki: https://www.tuik.gov.tr/