

# APPLYING FRAILTY MODELS TO ANALYZE THE DURATION OF UNEMPLOYMENT IN TURKEY

**Nihal Ata Tutkun**  
Hacettepe University  
Faculty of Science  
Department of Statistics  
06800, Beytepe, Ankara, TURKEY.  
e-mail: nihalata@hacettepe.edu.tr

**Durdu Karasoy**  
Hacettepe University  
Faculty of Science  
Department of Statistics  
06800, Beytepe, Ankara, TURKEY.  
e-mail: durdu@hacettepe.edu.tr

## 1. Introduction

In recent years, unemployment in developing countries has received great attention. Like in all other countries, unemployment is one of the major problems of Turkey because of its negative socioeconomic effects. Turkey experienced a series of financial and economic crises in its history and one of the leading economic crises is experienced recently in 2001. This crisis affects the unemployment rate; it reached 10.3% in 2002 whereas the unemployment rate was 6.5% in 2000. Since then, unemployment duration becomes a serious problem.

According to the 2012 Turkish Household Labour Force Survey (HLFS), the population of 15 year and over is 54 million 724 thousand persons. Number of unemployed persons decreased by 97 thousand persons in 2012 and has reached to 2 million 518 thousand persons in Turkey. Unemployment rate realized as 9.2% with 0.6 percentage points decrease. Youth (15-24 age group) unemployment rate was realized as 17.5% with 0.9 percentage points decrease. Number of employed persons increased by 711 thousand persons compared to the previous year and has reached to 24 million 821 thousand persons in 2012. Employment rate increased from 45% to 45.4% with 0.4 percentage point compared to the previous year. Labor force participation rate (LFPR) realized as 50% with 0.1 percentage points increase compared to the previous year for Turkey in 2012. LFPR was realized as 71% with 0.7 percentage points decrease for male and realized as 29.5% with 0.7 percentage points increase for female compared to previous year (Diriöz, 2013; TurkStat, 2012).

For most of the developed countries like France (van den Berg and van Ours, 1999), Portugal (Addison and Portugal, 2003), Russia (Foley, 1997 and Grogan and van den Berg, 2001) and Slovakia (Lubyova and van Ours, 1997) it is possible to find studies on unemployment duration. However the studies for developing countries are scarce; the studies are for Egypt (Tunali and Asaad, 1992), for Ethiopia (Serneels, 2004), for Turkey (Tansel and Taşçı, 2004) and for Ukraine (Kupets, 2006).

In this study, the assessment of current unemployment in Turkey is discussed. Also, a real Turkish data set consists of 471368 unemployed person is analyzed. The factors that effect the unemployment duration in Turkey are tried to be determined by duration models.

## 2. Duration analysis

Survival analysis is a class of statistical methods for studying the occurrence and timing of events. Survival analysis is known as “event history analysis” in sociology, “duration analysis” in political science or economics, “reliability analysis (failure time analysis) in engineering, and “survival (hazard) analysis ” in biostatistics or epidemiology. Duration analysis include various methods such as life tables, Kaplan-Meier estimates, Cox (1972) regression model (CM), parametric regression models, frailty models, competing risks models and discrete-time models which are sometimes complementary. Duration models can handle the censored and time-dependent data which is difficult to deal with traditional statistical models.

Duration models depend mostly on hazard function,  $h(t)$ , that shows the probability that interested event (failure) occurs in an interval, given that no failure occurs before time  $t$ . The interested event can be onset of disease, equipment failures, earthquakes, automobile accidents, stock market crashes, revolutions, job terminations, births, marriages, divorces, promotions, retirements and arrests (Allison, 1995). In unemployment studies, the interested event occurs when unemployed finds a job and becomes employed and  $T$  shows the unemployment duration. Unemployment duration can be modelled by non-parametric, semi-parametric or parametric models. Although the parametric models give much more information than the non-parametric ones, results may be invalid if the data does not fit a parametric distribution. In these cases, the semi-parametric models can be used because they give more information than non-parametric models and they impose fewer assumptions than parametric models. Although Cox's semi-parametric model is the most frequently employed regression tool for survival data, fully parametric models may have some advantages. Efron (1977) and Oakes (1977) showed that, under certain circumstances, parametric models lead to more efficient parameter estimates than Cox's model. When empirical information is sufficient, parametric models can provide some insight into the shape of the baseline hazard (Nardi&Schemper, 2003). Two approaches to the modelling of covariate effects on survival have become popular in the statistical literature. The first approach is similar to the classical linear regression approach (so called accelerated failure-time model) and the second is proportional hazards approach. Although the accelerated failure time model provides a direct extension of the classical linear model's construction for explanatory variables under conventional data, its usage is restricted by the error distributions for survival data. The major approach to modelling the effects of covariates on survival is to model the conditional hazard rate as a function of covariates. It is difficult to use a formal statistical test to discriminate between parametric models because the models are not nested. One way of selecting an appropriate parametric model is to use the Akaike's Information Criterion (AIC). For determining the appropriateness of parametric models, some graphical diagnostics shall be used.

Frailty models are extensions of the proportional hazards model. Normally, in most clinical applications, survival analysis implicitly assumes a homogenous population to be studied. This means that all individuals sampled into that study are subject in principle under the same risk (e.g., risk of death, risk of disease recurrence). In many applications, the study population can not be assumed to be homogeneous but must be considered as a heterogeneous sample, i.e. a mixture of individuals with different hazards. For example, in many cases it is impossible to measure all relevant covariates related to the disease of interest, sometimes because of economical reasons, sometimes the importance of some covariates is still unknown. The frailty approach is a statistical modelling concept which aims to account for heterogeneity, caused by unmeasured covariates (Wienke, 2003). The notion of frailty provides a convenient way to introduce random effects, association and unobserved heterogeneity into models for survival data. In its simplest form, a frailty is an unobserved random proportionality factor that modifies the hazard function of an individual, or of related individuals (Yashin et.al, 2001). A frailty model in the univariate case introduces an unobservable multiplicative effect  $\alpha$  on the hazard, so that conditional on the frailty

$$h(t/\alpha) = \alpha h(t)$$

where  $\alpha$  is some random positive quantity assumed to have mean one (for purposes of model identifiability) and variance  $\theta$ . Many distributions can be chosen for the frailty, but the most common frailty distribution is the gamma distribution.

### 3. Application

In unemployment duration studies, it is common to analyse the effects of unemployment insurance on unemployment duration. The unemployment insurance is a new social insurance system for Turkey and implemented in 2000. First benefit is paid in 2002. Therefore, the studies over this topic are rare.

During the 2001 crisis, the per capita Gross National Product (GNP) declined by 9.6 percent which was the largest contraction ever in the history of the Republic. Unemployment rates increased during those crises and remained high since then. The ratio of unemployment stood at about 9.2 % in 2012. Therefore, in Turkey, unemployment remains as a serious problem in the agenda of the policy makers. Turkish Employment Agency established for aiding activities of protecting, improving, generalizing of employment and preventing unemployment. It becomes a respected and leader organisation that plays an effective role in service quality and diversity and also in tackling unemployment via adaptation to changes and innovations. Also, it plays a role in executing unemployment insurance services and implements active and passive labour force policies alongside its classical services of finding jobs and employees within an extended mission area. In accordance with the needs of labour market; providing the service of job and employee agency effectively; enabling, maintaining, developing, diversifying the employment and providing a temporal income support to people who have lost their jobs via implementation of active labour market policies aimed at enhancing the employability of workforce.

In this study, we used the data of a survey consist of unemployed people in 2009 in Turkey. The survey includes 471368 people who have worked before, leave their jobs for various reasons and qualify for unemployment benefits are taken into the study. Finding a new job is the endpoint of interest. This variable is measured in months. There was about a 2-year-follow-up period. People who were still unemployed at the end of the follow-up period (January, 2011) were treated as censored observations. Out of 471368 unemployed people, 34194 of them (7.3%) found a job and the rest 437174 people (92.7%) stayed as unemployed. Mean unemployment duration is calculated as 23 months.

In the concept of analysis, factors which affect the survival time of unemployed people are tried to be determined. In the study, covariates such as gender ( $x_1$ ), marital status ( $x_2$ ), age( $x_3$ ), education level ( $x_4$ ), vocational course ( $x_5$ ) and the city of application ( $x_6$ ) are used. In the non-parametric approach to duration analysis we provide the estimates of Kaplan-Meier survival function and also the log-rank test allows for testing the equality of survival functions different groups. For all covariates, we observe that the equality of the survival functions are rejected at a 95% significance level. The survival function shows the proportion of people who become employed as time proceeds. Women has smaller unemployment durations than men. Single people have smaller unemployment durations than married ones. The duration of unemployment decreases as the education level increases. The people who takes a vocational course have distinctly smaller unemployment duration than the others. Hence, we can say that the vocational course is helpful and suggest people to take this course. The duration of unemployment increases as the age increases.

In this study, proportional hazards assumption is assessed by a statistical test. This test is accomplished by finding the correlation between the Schoenfeld residuals for a particular covariate and the ranking of individual failure times (Therneau and Grambsch, 2000). For unemployment duration, it is found that covariates taken into consideration in this study do not hold the proportional hazards assumption ( $p$ -value=0). Since the proportional hazards assumption does not satisfied, extensions of CM (stratified CM, CM with time-dependent covariates), parametric regression models with and without unobserved heterogeneity are estimated as an alternative to CM. One has to be careful in using the extensions of CM. CM with time-dependent covariates needs additional parameters and the choose of time function is arbitrary; and stratified CM needs a determination of a strata variable. In this study, it is difficult to strata the data since all covariates have nonproportional hazards.

The analysis suggests that parametric regression models (with/without heterogeneity) are suitable for our data than the others. Kolmogorov-Smirnov and Anderson Darling tests showed that the log-logistic distribution could provide a better fit than the other common distributions for duration time.

Furthermore, Cox-Snell residuals plots and AIC suggest that log-logistic regression model is a suitable model for our unemployment data. We have also estimated the parametric regression models with and without unobserved heterogeneity and tested for the inclusion of unobserved heterogeneity with likelihood ratio tests since the models are nested. When we include the effect of the unobserved heterogeneity in the model by gamma distribution or inverse Gaussian distribution, the likelihood ratio test suggests the inclusion of unobserved heterogeneity. Therefore we can say that, at 5% significance level the heterogeneity is significant for the models and we should use the log-logistic model with unobserved heterogeneity for our data.

The duration of the unemployment is high, because job seekers do not want to work until they find a job in the qualifications they want to work. However, unemployment salary is still a new situation in Turkey. Revealing what influences the duration of finding a job will be useful for job seekers. Besides, the characteristics of the applicants who apply to get unemployment salary can be determined by duration models.

## References

- Addison, J.T., Portugal, P. (2003), Unemployment Duration: Competing and Defective Risks, *Journal of Human Resources*, Vol.38, pp.156-191.
- Allison, P.D. (1995), *Survival Analysis Using SAS: A Practical Guide*, SAS Institute Inc., Cary, NC, USA.
- Bulut, V. (2011), *Analysis of Factors Influencing Time of Unemployment Using Survival Analysis in Turkey*, Unpublished M.Sc. Thesis, Hacettepe University, Institute of Graduate Studies in Science, Statistics,
- Cox, D.R. (1972), Regression models and life-tables, *Journal of the Royal Statistical Society Series B*, Vol.34, pp. 187-220.
- Diriöz, S. Ç. (2013), *Employment Assessment Report of 2012*, Republic of Turkey Ministry of Development.
- Efron, B. (1977), The efficiency of Cox's likelihood function for censored data, *Journal of the American Statistical Association*, Vol.72, pp.557-565.
- Foley, Mark C. (1997), *Determinants of Unemployment Duration in Russia*, Economic Growth Center: Center Discussion Paper no. 779, New Haven: Yale University.
- Grogan, L. and van den Berg, G.J. (2001), The Duration of Unemployment in Russia, *Journal of Population Economics*, Vol. 14, pp. 549-568.
- Kupets, O. (2006), 'Determinants of Unemployment Duration in Ukraine', *Journal of Comparative Economics*, Vol.34, pp. 228-247.
- Lubyova, M. and van Ours J. C. (1997), *Work Incentives and the Probability of Leaving Unemployment in the Slovak Republic*, Tinbergen Institute Discussion Papers 97-071/3, Tinbergen Institute.
- Nardi, A., Schemper, M. (2003) Comparing Cox and Parametric Models in Clinical Studies, *Statistics in Medicine*, Vol.22, pp.3597-3610.
- Oakes, D. (1977), The asymptotic information in censored survival data, *Biometrika*, Vol.64, pp.441-448.
- Serneels, P. (2004), *The Nature of Unemployment in Urban Ethiopia*, Centre for the Study of African Economies, Working Paper 04-01, Oxford University.
- Tansel, A., Tasci, H.M. (2004), *Determinants of Unemployment Duration for Men and Women in Turkey*, ERC Working Paper in Economic 04/04.
- Therneau, T.M. and Grambsch, P.M. (2000), *Modelling Survival Data: Extending the Cox Model*, Springer, New York.
- Tunali, I. and Asaad, R. (1992), Market Structure and Spells of Employment and Unemployment: Evidence from the Construction Sector in Egypt, *Journal of Applied Econometrics*, Vol.7, pp. 339-367.
- Van den Berg, G.J. and Van Ours J.C. (1999), Duration dependence and heterogeneity in French youth unemployment durations, *Journal of Population Economics*, Vol. 12, pp.273-28
- Wienke, A. (2003), *Frailty Models*, Max Planck Institute for Demographic Research, MPIDR Working Paper WP 2003-032.
- Yashin, A.I., Iachine, I.A., Begun, A.Z., Vaupel, J.W. (2001), *Hidden Frailty: Myths and Reality*, Research report, Odense University, Department of Statistics and Demography.