USING PARAMETRIC AND NON-PARAMETRIC METHODS FOR ESTIMATING SCHOOL TO WORK TRANSITION

†**Ana Gorenca**
"Aleksandër Moisiu" University
a_gorenca@hotmail.com

†**Esmeralda Shehaj**
University of Tirana
a_gorenca@hotmail.com

Abstract

This paper investigates the effect of socio-economic factors on the unemployment duration of fresh graduates using parametric and non-parametric approaches. The 2012 Living Standards Measurement Survey of Albania is used to estimate the effects. We restrict the sample to only those aged from 19 to 35 years old to reduce the heterogeneity among unemployed graduates. The determinants of the school to work transition are derived from the job-search theory. We employ parametric methods, i.e., proportional hazard models (weibull and exponential models) to estimate the effect of gender, age, education, residence and occupation on the first period of unemployment. Non-parametric methods are used to visually present the school to work transition. Our findings indicate that the weibull model provides a better fit to the data compared to the exponential model. Furthermore, the estimates in the weibull specification are in line with the theoretical findings of the job-search theory. Education, age, gender and occupation are significant determinants of the first unemployment spell. Moreover, unemployment duration increases with age. Lastly, the unemployment spells are longer for those with low levels of education and low occupation profiles.

Key words: unemployment duration, job-search theory application, hazard rate, proportional hazard.

1. Introduction

Each and every year, a considerable number of graduates in Albania aim to join the labour market. The existence of a dynamic economic environment, capable to accommodate the increasing inflow of unemployed youth is necessary, but not sufficient. The transition from school to work represents a crucial stage in the life and future career of youth in Albania. In this paper, we examine the school to work transition of individuals aged from 19 to 35 years old. Given the socio-economic characteristics of the examined group, as well as the socio-economic environment in Albania, we expect a complex and non-smooth transition. However, policymaking has played a lead role in smoothing the process of exiting from unemployment. In specific, active labour market policies have been addressed mainly to

tackle unemployment among youngsters and to smooth the search and match process. In this paper we first aim to visually present the transition from school to work, and second to determine the effect of socio-economic factors on unemployment duration.

An increasing number of publications in transition analysis have paid special attention to school to work transition. Most of the studies suggest that graduates with less favorable socio-economic background are expected to have longer unemployment spells. Furthermore, schooling and the characteristics of the region where the unemployed graduate comes from play an important role in the probability of finding a job. For instance, Lassibille et al (2001) examine the impact of human capital on unemployment duration using a data set drawn from the Socio-Demographic Survey in Spain in 1991. They employ multinomial logistic models and find that males and those with high education profiles have shorter unemployment spells compared to females and those with low school attainment. In addition, their results indicate insignificance of family background on the probability of exiting from unemployment. Jaunky and Khadaroo (2007) study the unemployment duration of graduates from universities in Mauritania from 1995 to 2000 using several proportional hazard (PH) and accelerated failure time models. The main findings suggest that age affects negatively the first duration of unemployment. Further, the father's education is positively correlated with the graduate's unemployment spell. Canals and Diebolt (2002) employs a survival model to determine the effect of education background on the unemployment duration of French graduates in 1996. The authors find that the average spell of tertiary education graduates is roughly 4 months. In addition, the mean duration of graduates from Business schools is approximately 2 months, and those graduated in Law, Economics and Social Sciences are more likely to experience longer spells. In Albania, Alikaj and Shehaj (2015) use Living Standards Measurement Survey (LSMS, 2012) to investigate the determinants of school to work transition. The authors estimate a Weibull model and conclude that the unemployment duration of those employed in high profile jobs (e.g., professionals and technicians) have higher probabilities of exiting from unemployment.

This paper builds on theoretical framework of the job-search theory[1]. Furthermore, borrowing from non-parametric methods of duration analysis the odds of exiting from unemployment given the duration time are estimated. Regarding estimation methods, both parametric and non-parametric approaches are used to determine the effect of various demographic factors including age, gender, education attainment, residence and occupation on unemployment duration.

This paper is organized as follows. Section 2 provides an overview of the labour market situation in Albania. Section 3 presents the building blocks of duration analysis and the proportional hazard models, and section 4 presents the data and estimation results. Section 5 concludes.

2. An Overview of the Albanian Labour Market

This section presents an overview of the labour market in Albania with focus on the labour market position of youth (those aged from 15-29). In the last two decades, the labour market in Albania has mainly assumed the following two features: high unemployment, especially among youth and low female participation rate (see Figures 1 & 3). Çela and Kamberi (2015) argue that the first feature is a result of the continuous mismatch of the supply and demand for labour market skills. That is partially caused by the low involvement of employers in the education system of Albania (Hoxhaj, 2017). On the causes of the low

[1] for details please see Gorenca (2018), paper submitted to the International Institute for Private, Commercial and Competition Law in May, 2018.

employability of females, Shehaj et al. (2015) claim that the economic downturn, combined with social factors such as childcare, patriarchy and discouragement have affected negatively the participation rate of females in Albania.

Figure 1. Labour Force Participation rate by gender and age groups

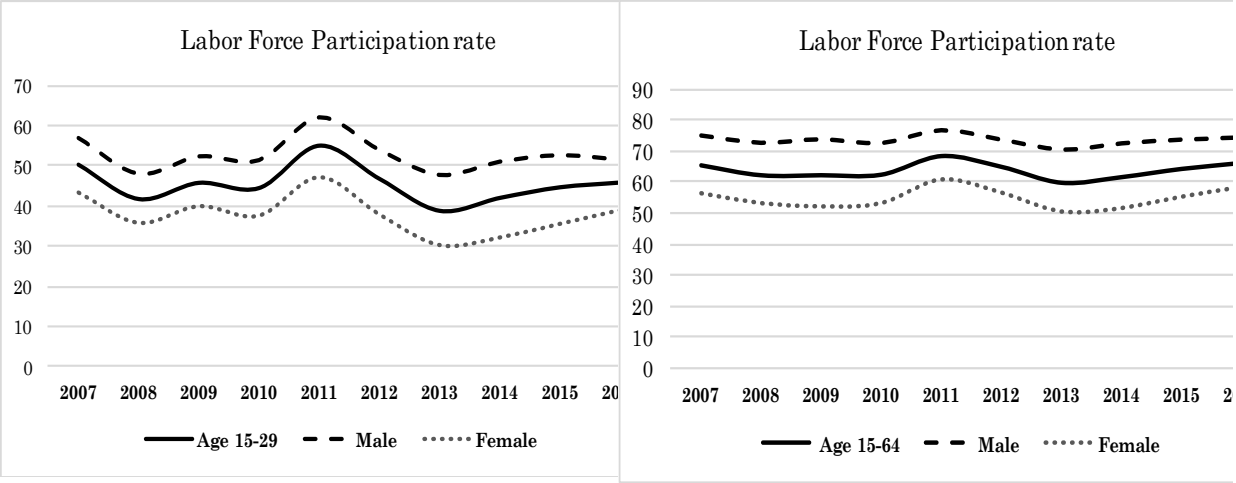


Figure 2. Employment rate by gender and age groups

Figure 3. Unemployment rate by gender and age groups

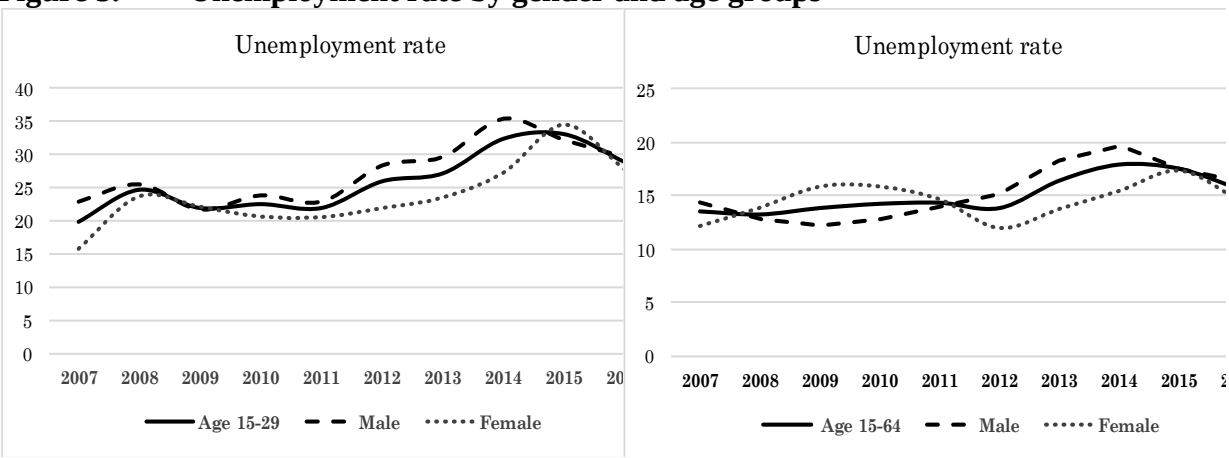
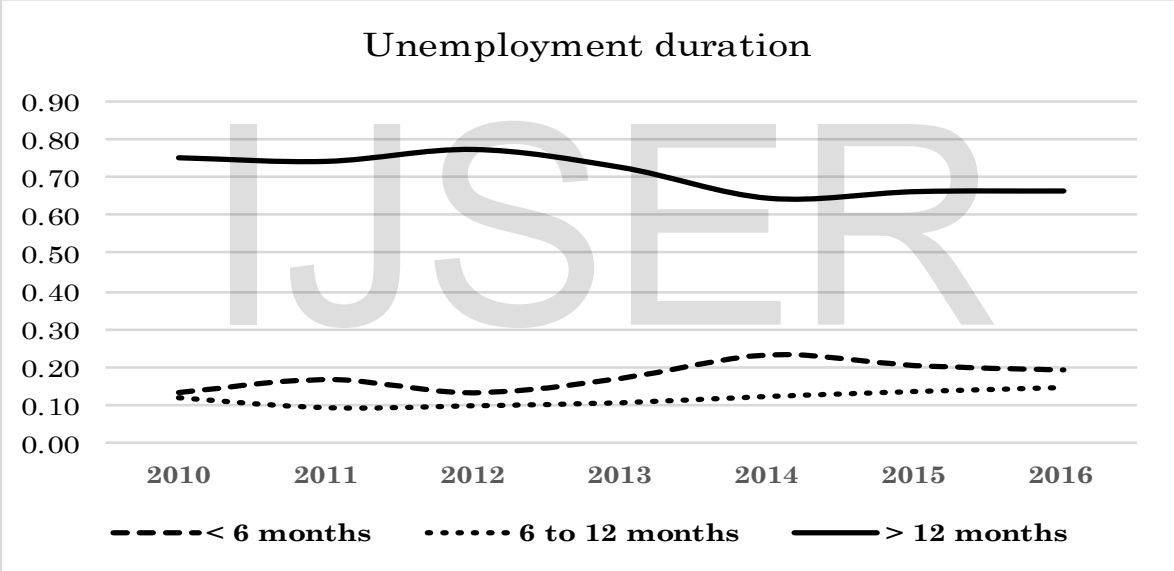


Figure 1 indicates the labour force participation rate (LFP) from 2007 to 2016 by gender and age groups. The LFP rate of those eligible to work (15-64) is more stable than the youth LFP rate. From 2013 onwards, the LFP rates of both age groups follow a gradually increasing

trend. In both left and right panels, the female participation rate is lower compared to male participation rate. In addition, the changes in the female LFP rate are likely to determine the changes in the aggregate LFP. That implies that the LFP of males is more stable. Figure 2 presents the employment rate by gender and age groups. The employment of females is lower than the employment of males for both age groups. However, since 2013, the increase in the employment rate of females is higher compared to incline in the employment rate of males. In addition, young individuals (aged from 15-29) have lower employment rates compared to all those eligible to work. Figure 3 indicates the unemployment rate by gender and age groups. Since the aftermath of the global financial crisis, which gradually assumed the features of an all-encompassing economic crisis, the unemployment rate among youth has gradually inclined till 2015. Afterward, the series follow a downward trend. In addition, from 2011 onwards, females have lower unemployment rate than males.

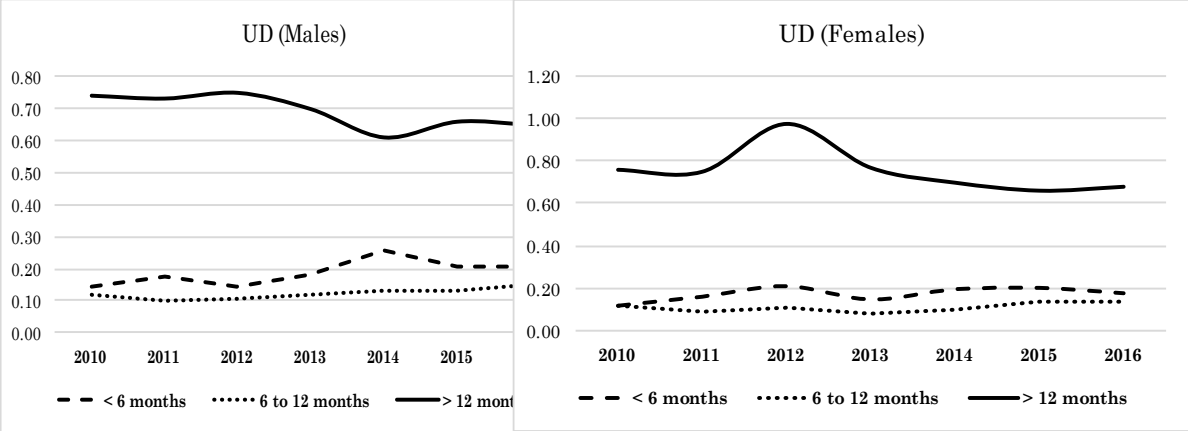
Figures 4, 5 and 6 present the unemployment duration measured in spells' intervals for all unemployed jobseekers, males and females, respectively. Throughout the examined period, there are more jobseekers with long-term spells (more than 1 year) than unemployed people with short-term spells. Furthermore, there are more unemployed people with spells of 6 to 12 months than jobseekers with spells of less than 6 months.

Figure 4. Unemployment Duration (UD) (2010-2016)



With respect to Figure 5, there are more males with long-term unemployment duration than females. From 2014 onwards, the share of females with long-term unemployment spells has remained at a constant level (around 60 percent). We observe a similar trend for males also. Another feature of the labour market in Albania, is the stability in the share of those with short-term unemployment duration (less than 6 months) for both genders. The latter implies that on average, long-term unemployment prevails in the labour market of Albania.

Figure 5. Unemployment Duration (UD) by gender



3. The Econometrics of Duration Models

In essence, duration models in unemployment explain the time an individual spends in unemployment. The empirical application in this paper is mostly based on Cameron and Trivedi (2005). The base function in any duration analysis is the hazard function which defines the concept of duration dependence. Let T be a continuous random variable denoting unemployment duration. Its cumulative distribution function [2] (*cdf*), F(t), equals the probability that the unemployment spell ends before *t* time units, F(t)=Pr [T<t]. Cahuc and Zylberberg (2004) define the hazard function as the instantaneous probability of exiting from unemployment:

$$\varphi(t)dt = \Pr[t \leq T \leq t + dt | T \geq t] \quad \textbf{(1)}$$

Using the definition of conditional probability Eq. **(1)** can be written as:

$$\varphi(t)dt = \frac{\Pr[t \leq T \leq t+dt]}{\Pr[T \geq t]} = \frac{dt\, f(t)}{1-F(t)} \quad \textbf{(2)}$$

Intuitively, Eq. (2) reads as the conditional probability that the unemployment duration lies in the [t, t+dt] interval. The denominator can be additionally written as Pr [T<t], which is the *cdf* of the random variable T. The numerator simplifies to the probability distribution function (pdf) of T, f(t)[3]. In addition, the denominator represents the *survival function*, S(t) = 1-F(t). Hence, $\varphi(t)$ = f(t)/S(t).

Another useful primitive of duration models is the integrated hazard, Φ(t). Writing the hazard function as the negative change in the logarithmic form of S(t), and then integrating over the [0, t] area the integrated hazard can be written as:

$$\Phi(t) = -\ln S(t) \quad \textbf{(3)}$$

Additionally, Cameron and Trivedi (2005) define the expression in (3) as the cumulative hazard function. Previously, we defined the hazard function as the change in ln[S(t)]. Hence, ln[S(t)] is the integral over [0, t] of the hazard function. Applying the exponential

[2] Usually, the *cdf* is unknown. However, the economic theory provides knowledge about the specification of the hazard function.

[3] From the fundamental theorem of calculus: $F(y) = \int_{-\infty}^{x} f(y)dy$ and in an interval [t, t+dt], $Pr[t \leq X < t + dt] = \int_{t}^{t+dt} f(x)dx \approx dt\, f(x)$

transformation to both sides, S(t) takes the form:

$$S(t) = \exp[-\int_0^t \varphi(\xi)\, d\xi\,] \quad (4)$$

The estimators of the the cumulative hazard function and the survival function we consider in this paper are Nelson-Aalen and Kaplan-Meier estimators, respectively. It is important to note that both estimators are nonparametric and work in discrete time settings. In the absence of censoring, the survival function estimator would be 1 - $\widehat{F(t)}$.

From the class of parametric estimations in transition analysis, we consider the proportional hazard (PH) models. The general form of PH models can be written as:

$$\varphi(t|\boldsymbol{x},\boldsymbol{\beta}) = \varphi_0(t)\, \theta(\boldsymbol{x},\boldsymbol{\beta}) \quad (5)$$

where $\varphi_0(t)$ denotes the baseline hazard, a function of time only. $\theta(\boldsymbol{x},\boldsymbol{\beta})$ is a function of the covariates in $\boldsymbol{x}$. Cameron and Trivedi (2005) argue that the most common specification of $\theta(\boldsymbol{x},\boldsymbol{\beta})$ is the exponential specification. The baseline hazard can follow several distributions including gamma, Gompertz, Weibull and exponential. The hazard under weibull and exponential distributions of $\varphi_0(t)$ takes the forms $\alpha t^{\alpha-1} exp(x'\beta)$ and $exp(x'\beta)$, respectively. To parameterize the models, logarithmic transformations are applied to both specifications. Hence, the parametric weibull and exponential PH models can be written as in Eqs. (6) and (7) respectively:

$$\log \varphi(t|\boldsymbol{x},\boldsymbol{\beta}) = \log a + (\alpha - 1)\log t + \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (6)$$

$$\log \varphi(t|\boldsymbol{x},\boldsymbol{\beta}) = \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (7)$$

Note that the intercept in the weibull model is $\log a + (\alpha - 1)\log t + \beta_0$ and in the exponential model is β_0. The following section presents the data and estimation results.

4. Data and Empirical results

The data sample is drawn from the Living Standards Measurement Survey (LSMS) of Albania for the year 2012. LSMS is a multi-purpose households survey. It is considered as one of the main data sources to measure living standards and poverty indicators based on household consumption. Moreover, the survey provides sufficient and relevant tools to help policy makers in monitoring and developing social programs.

We restrict the sample to only those aged from 19 to 35 years old. In addition, we exclude all individuals with student status in 2012. The survey does not provide a measure of unemployment duration. However, we construct the unemployment spell measure given the information on first employment year and the respondent's age at last year of her studies. We adjust the unemployment spell by subtracting half year to all observations given that on average, the academic year in Albania ends in June. Further, we drop all cases with unemployment spells longer than 19 years given that the minimum working age is 16. The determinants of unemployment duration we consider in this analysis are a set of demographics including age, gender and residence, human capital attributes including education attainment, and occupation dummies. Table I and II present the sample characteristics.

Roughly 68 percent of the sample is male and 32 percent female. The majority (or approximately 38 percent) has completed the primary education or less. While roughly 31 percent have earned a general high school degree, only 4 percent have completed vocational education. Those with high education degrees comprise 28 percent of the sample. Regarding

residence, only 49 percent live in urban areas and the majority or approximately 51 percent come from other residential areas such as rural areas among others.
With respect to occupations, the majority or 29 percent work in skilled agriculture jobs. Those who work as professionals and clericals are roughly 17 and 3 percent respectively. Managers and those who work in armed forces comprise each approximately 1 percent of the sample. Roughly 6 percent are operators and 5 percent are technicians. The rest works in elementary occupations (7 %) and craft and trades related jobs (14.27 %).

TABLE I. SAMPLE CHARACTERISTICS (CATEGORICAL VARIABLES)

Variable	Frequency	Percentage
Gender		
Males	994	32.47
Females	478	67.53
Education Attainment		
Primary or less	559	37.98
General secondary	455	30.91
Vocational	53	3.60
Higher education	405	27.51
Residence		
Other	747	50.75
Urban	725	49.25
Occupation		
Professionals	244	16.58
Managers & Admin.	17	1.15
Service workers	224	15.22
Clerical	39	2.65
Operative workers	94	6.39
Technicians	71	4.82
Skilled agriculture	430	29.21
Elementary occupations	103	7.00
Craft and related trades	210	14.27
Armed forces	19	1.29

Table II presents the descriptive statistics of age and unemployment spell. The mean age is 28 with a standard deviation of 4.32 years. The average spell is 7 years with a standard deviation of 4.53 years. There is not much difference between the 50th percentile (6.17) and the mean. That is, we do not observe outliers in our sample. The maximum unemployment duration is 18.83 years.

TABLE II. SAMPLE CHARACTERISTICS (CONTINOUS VARIABLES)

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	1,472	28.19	4.32	19	35
Spell	1,472	6.99	4.53	0.08	18.83

Figure 6 presents the estimated hazard function (first panel) and the cumulative hazard function (second panel). This non-parametric estimation of the hazard ratios and survival function is done under data censoring. That is, we construct a dummy indicator which takes value 1 if the failure time is known and 0 otherwise.
The hazard rate shows that the probability of finding a job increases from 1 percent to roughly 4 percent over 15 time periods. It increases faster after the 15th period reaching 5.5 percent at the maximum observed spell. It is a non-linear and non-monotonic function. In general, it is an increasing function for the first transition to employment. The second panel presents the Nelson-Aalen cumulative hazard curve. It is estimated by summing the hazard rates over time.

Figure 6. Hazard ratios and cumulative hazard function

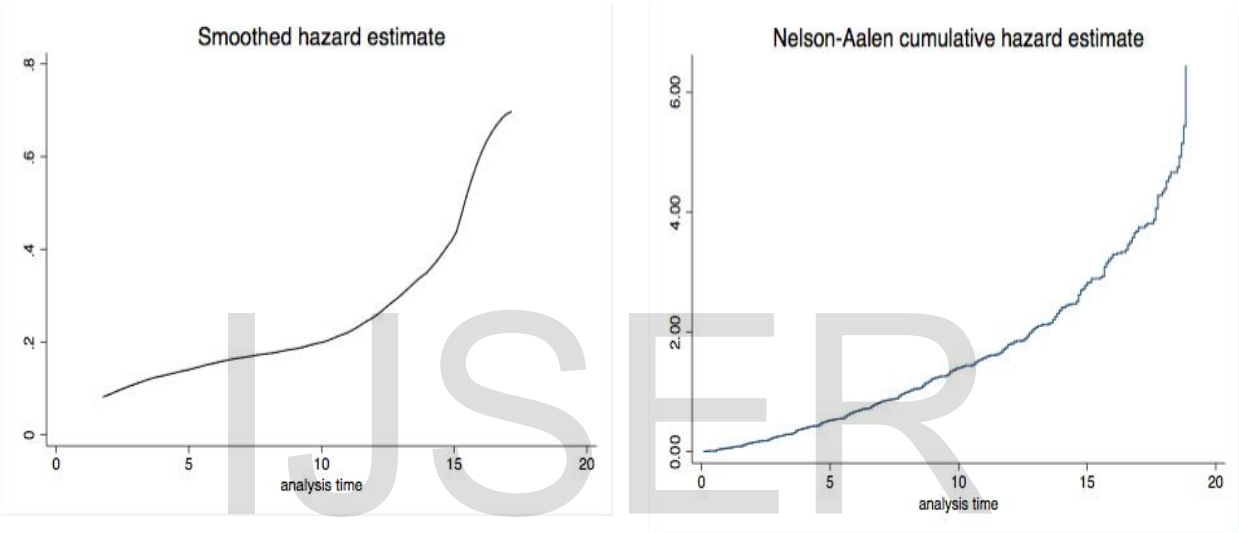
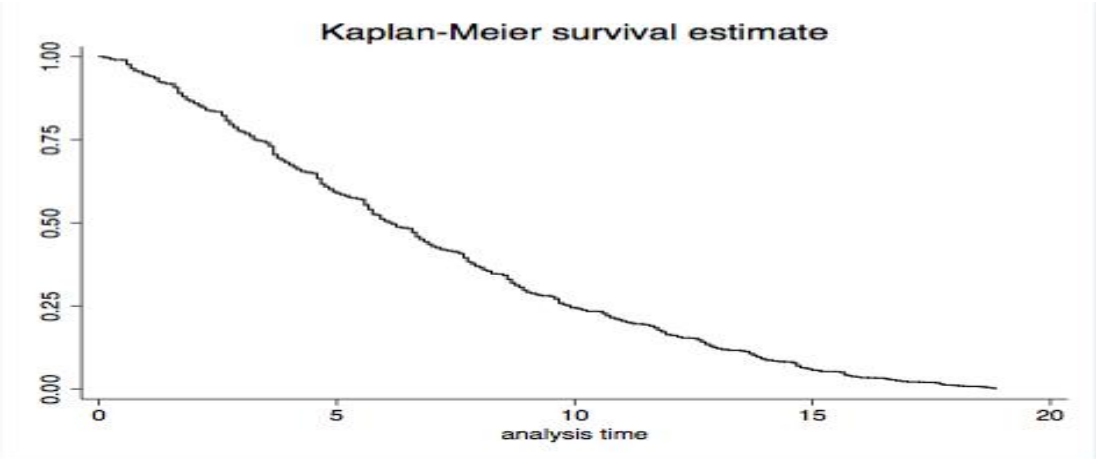


Figure (7) shows the estimated survival curve using the Kaplan-Meier estimator. The curve shows the survival probability. It is a decreasing function of time. In the first period the probability is 1. At the 5th year it declines to roughly 60 percent, and after 10 years to 25 percent. At the mean spell, the survival probability is 43 percent. The survival probability gradually vanishes at the last observed spell.

Figure 7. Estimated survival function



The following analysis presents the parametric estimations of hazard rates, the exponential and the weibull distributions, i.e., the baseline hazard follows an exponential and weibull distribution. It is worth noting that both regression models are proportional hazard models with different specification of the baseline hazard. Table III presents the estimated models. The failure time column shows the effect of covariates on the time when record ends. A positive coefficient means lower unemployment duration and vice versa. The hazard rates are interpreted as follows: a hazard rate higher than 1 means a "high failure rate", which in the context of the school to work transition means a shorter duration of first unemployment.
With respect to the weibull model, the unemployment duration of males is shorter than that of females. Moreover, males have 11.3 percent higher hazard rates, i.e., males are more likely to find a job compared to females. In contrast, the results indicate an insignificant effect of gender on unemployment spells in the exponential model. Age has a negative effect on unemployment duration in the weibull model. That is, the odds of exiting unemployment decrease with age; if age increases by one year failure time decreases by 17.3 percent. In the exponential specification, the failure time does also decelerate by a marginal increase in age. However, the effect is smaller, i.e. an increase of 1 year in age leads to 9.1 percent lower hazard rates. In both specifications, the indicator of urban residence has no significant effect on the failure time.

TABLE III. PARAMETRIC ESTIMATIONS (WEIBULL & EXPONENTIAL DISTRIBUTIONS)

	Weibull			Exponential		
Eplanatory variables (*X*)	Hazard rate	failure time	S.E	Hazard rate	failure time	S.E
Gender (female)						
Male	1.113	0.120**	0.061	1.058	0.056	0.064
Age	0.827	-0.190***	0.007	0.909	-0.094***	0.006
Residence *(other)*						
Urban	1.023	0.023	0.068	1.017	0.016	0.066
Schooling (general secondary or less)						
Vocational	1.493	0.401***	0.143	1.223	0.201	0.142
Higher education	4.120	1.416***	0.108	2.037	0.711***	0.102
Occupation (elementary occupations)						
Professionals	1.760	0.565***	0.148	1.352	0.301**	0.146
Armed forces	2.103	0.744***	0.251	1.456	0.376	0.250
Managerial	1.233	0.210	0.267	1.080	0.077	0.266
Craft & related trades	1.192	0.176	0.114	1.105	0.099	0.114

Technicians	1.486	0.396***	0.161	1.179	0.164	0.160
Services	1.340	0.292***	0.115	1.158	0.147	0.115
Clerical	1.368	0.313*	0.191	1.181	0.166	0.190
Skilled agriculture	1.701	0.564***	0.106	1.402	0.338***	0.107
Operatives	1.701	0.158	0.138	1.107	0.102	0.137
Constant	1.462	0.483*	0.219	1.452	0.373*	0.215
Observations		1,472			1,472	
AIC		2588			3484	

Reference category in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

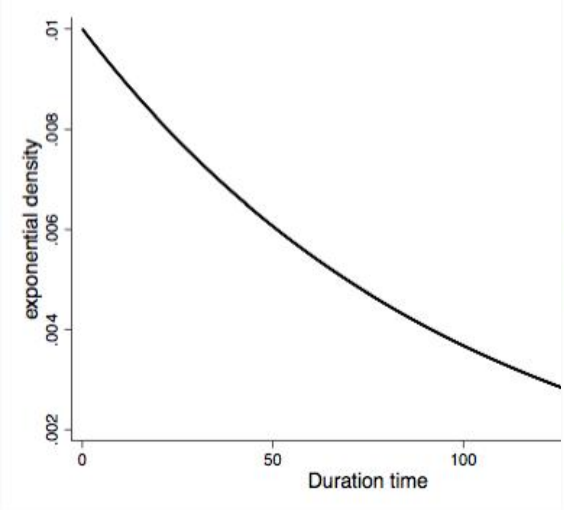
In the weibull model, those who have earned a vocational education degree have shorter spells than those who have completed gymnasium or those with lower education attainment. Vocational education graduates have approximately 49 percent higher hazard rates. In the exponential specification, this variable is insignificant. In both models, the unemployment duration of university graduates is shorter than the duration of general high-school graduates or those with lower education attainment. Regarding occupation dummies, in both specifications professionals have shorter spells than those employed in elementary occupations. While professionals have 76 percent higher hazard rates than the reference category in the weibull model, the estimate is 48 percent higher hazard rates in the exponential model. Those who work as armed forces have 103 percent higher hazard rates than those employed in elementary occupations. In contrast, the estimate reveals insignificance of the armed forces indicator in the exponential specification.

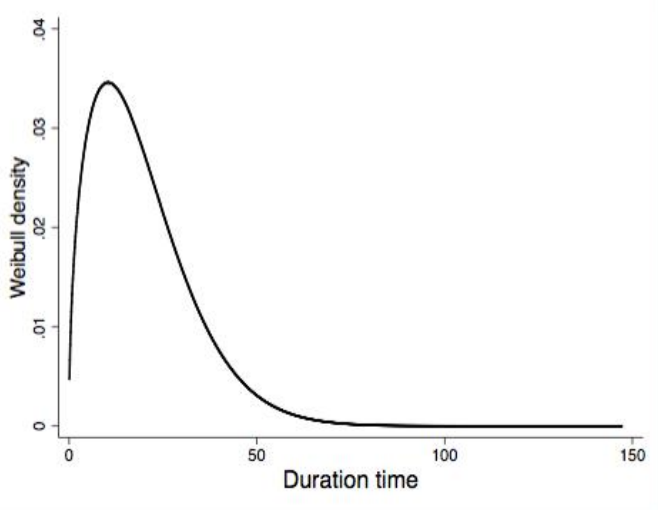
The indicators of managerial, operative and craft positions are insignificant in both models. In the weibull model, technicians and service workers have 48 and 34 percent higher hazard rates compared to elementary occupations employees, respectively. In the weibull model, the hazard rates of those employed in clerical jobs are 36.8 percent higher compared to those of elementary occupations. Lastly, the indicator of skilled agriculture occupations is highly significant in both models. The hazard rates of those employed in skilled agriculture occupations are 70 (weibull) and 40 percent (exponential) higher than the hazard rates of elementary occupation workers'.

Intuitively, the results of the weibull model match better with expectations. Technically, the model choice is in general based on the Akaike's Information Criteria (AIC). The best model would be the one with the lower AIC value. However, the Akaike's Information Criteria does not suggest the weibull model to better fit the data. To this extent, further diagnosis emerges. Hence, the selection of the preferred model can be based on implications of each model regarding the duration dependence (shape of the hazard function) of a spell of unemployment. Figure 8 presents the hazard probability distribution function using the exponential and weibull distributions.

In the case of the exponential distribution (first panel) the hazard pdf is decreasing. However, the exact spikes and dips in duration dependence are not observed. In contrast, the weibull distribution provides a better fit to the data. This is owing to the baseline hazard specification of the weibull model which involves the duration time. Conversely, the exponential specification does not involve time in the baseline hazard.

Figure 8. Hazard pdf with exponential and weibull distributions





5. Conclusion

This paper aims to investigate the school to work transition, and to determine the effect of socio-economic factors on unemployment duration. Using LSMS 2012 for Albania, and parametric and non-parametric methods, the results indicate that the probability of finding a job increases from 1 percent to roughly 4 percent over 15 time periods and reaches 5.5 percent by the end of the analysis time. The weibull model provides a better fit to the data compared to other parametric models, and its results suggest an important role of several factors such as age, education, gender and occupation on unemployment duration. Moreover, males are more likely to find a job than females. Regarding age, the older an individual becomes, the lower the chances to exit from unemployment are. Rural and urban locations do not result in different probabilities of finding a job and thus terminating unemployment duration. Professionals, technicians, armed forces, those in the service sector and skilled agriculture workers have higher odds of exiting from unemployment compared to those employed in elementary occupations.

It is worth noting that the socio-economic factors examined in this study do not provide sufficient source of motivation for policymaking. Instead, our estimates offer a comparative overview of the job-finding probability among individuals with different socio-economic characteristics. Further examination concerns future research.

References

Alikaj, E., & Shehaj, E. (2015). Determinants of transition from school to work in Albania. In Proceedings of 4th UBT Annual International Conference on Business, Technology and Innovation (p. 65).

Cahuc, P., & Zylberberg, A. (2004). *Labor economics*. MIT press.

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

Canals,V. and Diebolt, C. 2002. *France*. In H. Ehlert and H. Cordier 2002, 13–54

Çela, A., & Kamberi, G. (2015). *Albanian Youth 2015*. Tirana: Friedrich Ebert Stiftung.

Gorenca, A. (2018). An application of Weibull and Exponential distributions in school to work transition models. Forthcoming in Sixth Intenational Conference on Education and Social Sciences – Global Challenge 2018, International Insitute for private – Commercial and Competition Law (IIPCCL), University of Byalistok, Poland.

Gorenca, A. (2018). An Empirical Investigation of School to Work Transition Using the Gompertz Model. In Proceedings of the 4th international Scientific Conference: Knowledge based sustaibable economic development – Eraz 2018. Sofia, Bulgaria.

Hoxhaj, M. (2017). Youth unemployment in Albania, causes and consequences. *EuroEconomica*, *36*(2).

INSTAT. (2014). *Vjetari Statistikor 2010-2014*. Tirana: MEDIA - PRINT SH.P.K.

INSTAT. (2016). *Tregi i Punës 2016*. Tirana: MEDIA - PRINT SH.P.K.

INSTAT. (2017). *Vjetari Statistikor 2017*. Tirana: MEDIA - PRINT SH.P.K.

Jaunky, V. C., & Khadaroo, A. J. (2007, June). The School-to-Work Transition for University Graduates in Mauritius: A Duration Model Approach. In *Second Baltic-Nordic Conference on Survey Sampling, University of Finland, Kuusamo, Finland* (pp. 2-7).

Lassibille, G., Gómez, L. N., Ramos, I. A., & de la O Sánchez, C. (2001). Youth transition from school to work in Spain. *Economics of Education Review*, *20*(2), 139-149.

McCall, J. J. (1970). Economics of Information and Job Search. *The Quarterly Journal of Economics*, 84(1), 113-126.

Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *The review of economic studies*, *61*(3), 397-415.

Shehaj, E., Bimbashi, E., Rizvanolli, A., and others. (2015). *Women's Inclusion in Socio-Economic Life: Determinants of Female Labour Force Participation in Western Balkans.* ISBN: 978-9928-114-43-3. Tirana: Faculty of Economics, University of Tirana.