Social and Economic Factors Influencing Occupational Choices in Armenia: Multiclass Logistic Regression Analyses

Gishyan K.
University of Bath (United Kingdom)
kg634@bath.ac.uk
Martirosyan D.
École Polytechnique Fédérale de Lausanne(Switzerland)
davit.martirosyan@epfl.ch

Զբաղվածության վերաբերյալ նախընտրությունները ձևավորող սոցիալ-տնտեսական գործոնները Հայաստանում. բազմաչափ լոգիստիկ ռեգրեսիոն վերլուծություն

Գիշյան Կ.
Բաթի համալսարան (Մեծ Բրիտանիա)
kg634@bath.ac.uk
Մարտիրոսյան Դ.
Լոզանի դաշնային պոլիտեխնիկական դպրոց (Շվեյցարիա)
davit.martirosyan@epfl.ch

Ամփոփում: Հոդվածում դիտարկվում են Հայաստանի Հանրապետությունում բարձրագույն կրթություն (բակալավրի կամ մագիստրոսի աստիճան) ունեցող անհատների՝ զբաղվածության վերաբերյալ նախընտրությունները ձևավորող գործոնները: Հետազոտության նպատակն է բացահայտել, թե ինչպես են սոցիալ-տնտեսական գործոններն ազդում անհատների մասնագիտական ընտրության վրա: Չնայած հետազոտության հիմնական առարկան զբաղվածության հիմնական/ընդհանուր մոդելն է, այնուամե-նայնիվ, զբաղվածության վերաբերյալ սեռային/գենդերային նախընտրությունների տարբերությունները բացահայտելու նպատակով բերված են նաև առանձնացված արդյունքներ կանանց և տղամարդկանց համար: Համապատասխան գնահատականներ ստանալու նպատակով Հայաստանի Հանրապետության աշխատուժի 2017թ. հետազոտության տվյալների հիման վրա կիրառվել են բազմաչափ լոգիստիկ ռեգրե-սիայի մոդելներ, իսկ մոդելի համապատասխանության որակը գնահատելու համար՝ դասակարգման չափորոշիչներ: Առանձնացվել են զբաղվածության նախընտրության սեռային տարբերիչ առանձնա-հատկությունները տղամարդկանց և կանանց համար, ինչպես նաև հստակեցվել են հետազոտության հետագա ուղղությունները:

Վճռորոշ բառեր՝ բազմաչափ լոգիստիկ ռեգրեսիա, դասակարգում, մասնագիտական ընտրություն

Социально-экономические факторы, формирующие предпочтения о занятости в Армении: многомерный логистический регрессионный анализ

Гишян К.
Батский университет (Великобритания)
kg634@bath.ac.uk
Мартиросян Д.
Федеральная политехническая школа Лозанны (Швейцария)
davit.martirosyan@epfl.ch

Аннотация: В данной статье исследуются основные факторы, влияющие на выбор профессии гражданами Армении, получившими степень бакалавра или магистра. Предметом исследования является определение влияния на данное решение социально-экономических факторов. Помимо основной модели, также были рассмотрены отдельные модели для мужчин и женщин с целью выявления возможных различий между гендерами. В качестве модели была применена многоклассовая логистическая регрессия, обученная на данных опроса рабочей силы Армении в 2017 году, а оценка качества результатов осуществлялась с помощью классификационных метрик. В завершении, предоставлены полученные наборы признаков для мужчин и женщин, показывающие разницу, которая существует между гендерами при выборе профессии, и идентифицированы направления дальнейшего исследования.

Ключевые слова: многоклассовая логистическая регрессия, классификация, выбор профессии

1 Introduction and Literature Review

Career choice has become a complex science with the advent of informational technology and job competition (Quinter & Edwards, 2011). The selection of a particular occupation determines the current earnings as well as the future earnings of individuals which are closely linked with household consumption, health, and general status in the society (Harper & Haq, 1997; Freeman, 1971). In addition, Kerka (1998) states that the choice of the career is influenced by factors such as personality, interests, globalization, and information resources. In an educational setting, certain variables are believed to have particular importance. According to Blau & Kahn (1996) and Bayard et al. (2003), the suitability for the job for high-school students is greatly influenced by the ethnic background, level of achievement, job characteristics, and the choice of science subjects.

The results of the multinomial logistic regression analysis provided by Nasir (2005) demonstrate that education is a significant variable for all of the categories except craft and production workers, and the likelihood of having a higher paying job such as managerial, professional, teaching and medical rather than agricultural increases with the years of education. Nasir (2005) also states that marriage does not influence the occupational choice, however having younger children in the family increases the odds of being in teaching, agriculture, and craft rather than in agriculture. Furthermore, the results show that the odds of women being involved in medical, teaching and craft related occupations are high, and the odds of men being in managerial and administrative, sales and production-related occupations are high relative to agricultural-related jobs. There is also evidence (Zveglich & van der Meulen Rodgers, 2004) that in comparison to females, males are more likely to be employed in higher-paying occupations.

This paper provides a set of key social and economic variables that impact the occupational choices of those Armenians who have obtained either a bachelor's or a master's level of education. Further estimation of male and female models indicate that occupational differences exist in choosing a profession after graduation. By observing the results and the goodness of fit, the model for the female category provides the highest result in terms of F1 score (0.54), followed by the overall model (0.52) and the model for the male category (0.3). The model for the female category also provides a relatively high recall measure. Given the true values, 61.6% of the time the model correctly classifies the desired categories. The model for the male category provides the lowest classification measure of the three models.

2 Data and Methodology

We conduct the analysis using a logit model which is directly derived from multinomial logistic regression. Multinomial logistic regression, also known as Softmax Regression, has the following form:

$$\Pr[(Y)]_i = c) = \frac{\exp\left(\beta_c{}^T X_i\right)}{\sum_{j=1}^{K} \exp\left(\beta_j{}^T X_i\right)}$$

The equivalent logit model can be written in the following way:

$$\frac{\ln(\Pr(Y_i = 1))}{\Pr(Y_i = K)} = \beta_1{}^T X_i$$

$$\dots$$

$$\frac{\ln(\Pr(Y_i = K-1))}{\Pr(Y_i = K)} = \beta_{k-1}{}^T X_i$$

where K is the number of possible target variable outcomes and also the pivot class, and $K-1$ is the number of independent binary logistic regression models built. In the binary case where $K = 2$, softmax regression reduces to logistic regression, showing that it is a generalization of the logistic regression ('Softmax Regression').

Our study is based on the Armenian Labour Force Survey Questionnaire results for the year 2017 provided by the Armenian National Statistical Committee. The survey covers the entire territory of the Republic of Armenia, where the surveyed units are randomly selected private households (*Instructions for Filling in the Labour Force Survey Questionnaire*, 2016). It overall provides comprehensive information for 28,464 individuals with respective social and economic characteristics and their working behaviors.

All the multiclass categorical variables in the dataset have been converted into dummy variables. The final set of significant variables in the models are presented below:

- *Gender* – "Male" is the base category
- *Part-Time* – "Full-Time" is the base category
- *Married* – "Never Married" is the base category
- *Income-2* – Wage category of "55,000" AMD (minimum wage), "up to 55,000" is the base category
- *Income-3* – Wage category of "55,001-100,000" AMD, "up to 55,000" is the base category
- *Education Relation-2* – Category of "No, my education/qualification is higher", "Yes, it Corresponds" is the base category

- *Job Kind-2* – Category of "Temporary /Seasonal" jobs, "Permanent/Unlimited" is the base category
- *Job Kind-3* – Category of "Casual/One-time" jobs, "Permanent/Unlimited" is the base category
- *Marz-2* – Category of "Aragatsotn" region, "Yerevan" is the base category
- *Marz-9* – Category of "Syuniq" region, "Yerevan" is the base category
- *Job Engagement-5* – Category of "1-2 years" job engagement, "Up to 3 months" is the base category

In the survey, there is a question about how much income the participant received last month ("Exact Wage"). If one abstains from mentioning an exact numeric amount, an approximate wage is specified for belonging to the relevant income group as part of another variable ("Approximate wage"). Due to this fact, we encounter data sparsity in the "Exact Wage" and "Approximate wage" variables. To overcome this problem, we divide the "Exact Wage" variable into categories corresponding to the categories of "Approximate wage" variable and combine these variables into one column, which includes 10032 observations. In addition, the variables corresponding to the educational level and the profession obtained from formal education are highly correlated, thus the latter is removed.
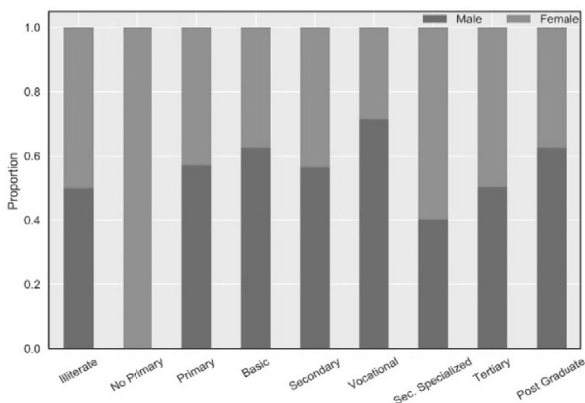

Figure 1: Education Categories and Gender Proportion

As the focus of the paper is to find out the occupational determinants of the individuals who have completed a secondary specialized level of education (Bachelor's or Master's level), we take another subset of the data to get only observations corresponding to this level of education which leaves us with 2028 observations, 1214 of which are female (60%) and the rest are male. From Figure 1, we can see that given the highest educational

qualification of the individual is "No Primary", all of these individuals are female. In addition, in the "Vocational" education section, females have the lowest proportion, which indicates that males are more prone to getting a short-term institutionalized education, completing a diploma or a certificate program. From Figure 2, one may also observe that the biggest proportion of individuals (34%) are in the "Agricultural" section, which is an interesting finding given the level of education that these individuals hold.

Figure 2: Occupation Categories and Gender
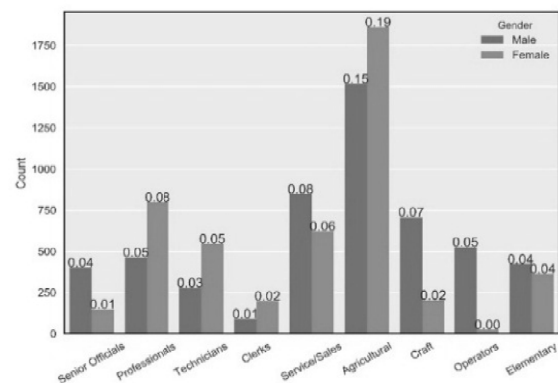
3   Results
3.1 Overall Significant Model

In the logit model, "Legislator, Senior Officials, Managers" category is taken as the base category for the dependent variable, and the empirical results provide some insightful findings (see Table 1). The "Married" variable, is significant for all except for the "Professionals" and "Skilled Agriculture" categories. It also shows that if a person is married, compared to an unmarried person the odds are less for working in the remaining occupations, which might be an indication that non-married people are generally more favored by the employers.

Working in a part-time position increases the odds for a person to work in a technician-professional, skilled-agricultural or an elementary-occupational position, but decreases the odds for working in an operator-assembling position, which indicates that this specialty requires more full-time employees. The "Income-2" variable represents the AMD 55,000 minimum wage category. As the minimum wage regulations differ depending on the employment status of the person, for example, the minimum wage law may not apply to self-employed individuals, wages below the minimum wage threshold are also present in the data. It shows that earning a minimum wage rather than a lower than the minimum wage decreases the chance of having

Table 1: Overall Significant Logit Model

Variables	Professionals	Technicians Professionals	Clerks	Service and Sales	Skilled Agriculture	Craft	Operators Assemblers	Elementary Occupations
				Main Occupation Categories				
const	-7.617***	-2.321***	-2.805**	-0.024	-4.617***	-0.602	5.596***	-2.799***
	(0.000)	(0.098)	(0.061)	(0.977)	(0.010)	(0.548)	297.736)	(0.061)
Gender	2.456***	2.038***	2.013***	0.656**	0.948***	-0.137	-2.461***	0.911**
	(11.658)	(7.627)	7.482	(1.924)	(2.58)	(0.872)	(0.085)	(2.486)
Part-Time	1.199	1.129**	0.357	0.626	3.29***	0.829	-2.616***	1.508***
	(3.315)	(3.092)	(1.428)	(1.877)	(26.931)	(2.290)	(0.073)	(4.518)
Married	0.490	-0.763*	-0.845*	-1.178***	-0.042	-0.836*	-1.001*	-1.182**
	(1.632)	(0.467)	(0.429)	(0.308)	(0.959)	(0.434)	(0.365)	(0.307)
Income-2	-0.867	1.686**	1.908**	1.492**	-1.424	0.00	0.43	-1.18**
	(0.420)	(5.394)	(6.736)	(4.444)	(0.241)	(1.000)	(1.537)	(6.843)
Income-3	1.323*	1.04***	0.614	0.74**	-1.20**	1.00***	0.718*	0.65*
	(3.756)	(2.831)	(1.848)	(2.097)	(0.301)	(2.719)	(2.051)	(1.916)
Education Relation-2	-3.900***	-1.416***	-0.340	0.978***	1.712***	1.136***	1.391***	2.133***
	(0.020)	(0.243)	(0.712)	(2.660)	(5.538)	(3.114)	(4.018)	(8.442)
Job Type-2	-0.000	0.956	-1.11***	2.026***	4.889***	3.087***	0.658	2.651***
	(1.00)	(2.602)	(0.330)	(7.584)	(131.87)	(21.911)	(1.930)	(14.172)
Job Type-3	0.000	2.240*	-0.00	1.776	-2.370***	4.535***	-0.000	-0.404
	(1.000)	(9.388)	(1.000)	(5.906)	(0.094)	(93.203)	(1.000)	(0.668)
Marz-2	-2.343***	-0.827	-1.790**	-1.488***	-3.160***	-2.354***	-2.130**	-3.623***
	(0.096)	(0.438)	(0.167)	(0.227)	(0.042)	(0.095)	(0.119)	(0.027)
Marz-9	0.000	0.069	-0.655	1.257	0.066	1.603*	0.379	-2.567***
	(2.471)	(1.156)	(0.226)	(0.420)	(0.304)	(1.114)	(0.307)	(0.187)
Job-Engagement-5	0.000	0.000	-0.431	2.595***	1.121**	1.457**	1.718***	0.707
	(0.796)	(4.912)	(2.287)	(10.215)	(1.937)	(4.557)	(4.467)	(3.44)
Pseudo R Square	0.278							
Recall Score	0.585							
Precision Score	0.527							
F1 Score	0.523							

an elementary-occupational working job, which indicates that people earning at least as much as the minimum wage are more likely to employed in other occupations, such as technicians-professional, clerks service and sales-related jobs. "Income-3" variable is significant for all occupational categories except for the clerks, and it shows that earning a salary from this range decreases the chance of being in a skilled agricultural position, which can be regarded as a sign that skilled agricultural workers are more employed in higher paying jobs. "Education Relation-2" variable represents those individuals who think their education/qualification is higher than their current job's qualification, and it shows that people who think accordingly are less likely to be employed in a professional or a technician-professional position. The variable has positive coefficients for all other categories except for the clerks. Having a temporary or a seasonal job ("Job Kind-2") increases the odds of being employed in service and sales, skilled-agricultural, craft, and elementary-occupational positions, while decreases the odds of working as a clerk. Casual-one-time job ("Job Kind-3") shows that such a job increases the odds of being in a technician-professional or belonging to a craft-occupational category, but decreases the odds of working in a skilled-agricultural position, which is represented by seasonal job characteristics.

The "Marz-2" variable represents the Aragatsotn region, and it negatively impacts all the categories except for the clerk. The coefficients show that being from this region, compared to those

Table 2: Logit Model for the Occupational Choices for Men

Variables	Professionals	Technicians Professionals	Clerks	Service and Sales	Skilled Agriculture	Craft	Operators Assemblers	Elementary Occupations
				Main Occupation Categories				
const	0.00 (1.00)	0.029 (1.029)	0.584 (1.792)	0.966 (2.628)	3.944*** (0.019)	-0.182 (0.834)	0.715 (2.044)	0.00 (1.00)
Part-Time	-2.675*** (0.069)	0.917 (2.502)	-1.369 (0.254)	0.078 (1.081)	3.492 *** (32.849)	0.885 (1.089)	-0.786 (0.456)	-0.001 (0.990)
Married	0.000 (1.000)	-0.624 (0.536)	0.000 (1.000)	-0.820 (0.441)	-0.018 (0.982)	-0.327 (0.721)	-0.115 (0.891)	-1.217* (0.296)
Income-2	0.000 (1.00)	2.137** (8.471)	0 (1.00)	2.382*** (10.825)	1.817 (6.153)	-0.197 (0.821)	-0.579 (0.561)	3873*** (48.07)
Income-3	1.790*** (5.992)	0.083 (1.086)	0.308 (1.361)	0.482 (1.619)	-1.406*** (0.245)	0.446 (1.561)	0.265 (1.304)	0.718 (2.049)
Education Relation-2	-1.028** (0.358)	-0.200 (0.818)	-3.845*** (0.021)	0.575 (1.777)	1.609*** (4.998)	0.869* (2.383)	1.297*** (3.658)	1.013* (2.753)
Job Type-2	-0.000 (1.000)	0.693 (2.000)	0.000 (1.000)	2.466*** (11.774)	5.224*** (185.725)	3.679*** (39.598)	-1.263*** (0.283)	3.250*** (25.778)
Job Type-3	0 (1.000)	3.542*** (34.548)	0.000 (1.000)	0 (1.000)	-0.833 (0.435)	0.000 (1.000)	0.000 (1.000)	0 (1.000)
Marz-2	-0.124 (0.883)	-0.975 (0.377)	-2.825** (0.059)	0.000 (1.000)	0 (0.435)	-4.520*** (0.011)	-3.890*** (0.020)	-3.903*** (0.020)
Marz-9	0 (1.000)	0.069 (1.072)	-0.655 (0.520)	1.257 (3.513)	0.066 (1.068)	1.603* (4.968)	0.379 (1.460)	-2.567 (0.077)
Job-Engagement-5	0 (1.000)	0 (1.000)	-0.431 (0.65)	2.596*** (13.397)	1.121** (3.069)	1.457** (4.294)	1.718*** (5.573)	0.706 (2.027)
Pseudo R Square	0.183							
Recall Score	0.351							
Precision Score	0.251							
F1 Score	0.3							

people who are from Yerevan, decreases the odds of working in all remaining occupational positions. Being from the Syunik region ("Marz-9") increases the odds of working in a craft position and decreases the odds of having an elementary occupation, compared to the individuals from Yerevan. Working in the current job for 1-2 years ("Job Engagement-5") increases the odds of being in sales and services, skilled-agricultural, craft, and operator-assembler positions, and is insignificant for the remaining variables. As "Up to 3 months" is the base variable for this category, it shows that shorter job engagement is not favored in such jobs. "Gender" variable is significant for all occupations except for the craft, and the statistical significance for the remaining professions indicates that male and

female Bachelor and Master graduates behave differently when choosing their occupation. To further assess the importance of the variables that may differ based on gender we build two separate multinomial logit models, and the results are discussed in the next section.

3.2 Models for Males and Females
For the separate models, the significance of the variables differs. The "Part-Time" variable is highly significant and increases the odds for both genders being in the skilled-agricultural field. The "Married" variable for the female gender positively affects the odds of having a "Professional" occupation rather than being a senior official or a legislator. Earning a minimum wage as a male

Table 3: Logit Model for the Occupational Choices for Women

Variables	Professionals	Technicians Professionals	Clerks	Service and Sales	Skilled Agriculture	Craft	Operators Assemblers	Elementary Occupations
const	0.000 (1.000)	1.827 (6.215)	0.000 (1.000)	0 (1.000)	-2.673** (0.069)	-2.242 (0.106)	0.000 (1.000)	-1.537 (0.215)
Part-Time	0.944 (2.570)	0.721 (2.057)	0.479 (1.514)	0.755 (2.127)	3.080 *** (21.736)	1.197 (3.311)	-1.612 (0.200)	1.325 (3.760)
Married	0.988** (2.685)	-0.097 (0.907)	0.035 (1.036)	-0.595 (0.551)	-0.506 (1.659)	-0.521 (0.594)	0.000 (1.000)	-0.197 (0.821)
Income-2	0.000 (1.000)	0.585 (1.794)	1.146 (3.144)	0.430 (1.537)	-4.721*** (0.009)	0.000 (1.000)	-1.457 (0.233)	0.000 (1.000)
Income-3	1.007* (2.737)	1.114* (3.048)	1.132* (3.101)	0.835 (2.306)	-1.531** (0.216)	1.488** (4.430)	1.023 (2.780)	0.513 (1.670)
Education Relation-2	-0.841* (0.431)	-2.130*** (0.119)	0.000 (1.000)	1.374** (3.951)	1.940*** (6.960)	2.117*** (8.302)	0.00 (1.000)	2.351*** (10.491)
Job Type-2	0.000 (1.000)	0.000 (1.000)	-0.421 (0.657)	2.262** (9.597)	5.073*** (159.571)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)
Job Type-3	0.000 (1.000)	4.808*** (122.456)	0.000 (1.000)	2.560 (12.932)	-2.050** (0.129)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)
Marz-2	-1.770* (0.170)	0.780 (2.182)	-0.746 (0.474)	-1.302 (0.272)	-3.089** (0.046)	0.000 (1.000)	-1.950 (0.142)	-5.138*** (0.006)
Marz-9	-2.583*** (0.076)	0.063 (1.065)	-0.665 (0.514)	-2.154 (0.116)	0.066 (0.156)	0.000 (1.000)	-2.942*** (0.053)	-2.397** (0.091)
Job-Engagement-5	-1.286 (0.276)	-0.108 (0.898)	0.015 (1.015)	0.719 (2.052)	-1.360 (0.257)	1.033 (2.810)	1.153 (3.169)	0.000 (1.000)
Pseudo R Square	0.245							
Recall Score	0.616							
Precision Score	0.509							
F1 Score	0.547							

increases the odds of being in a technician-professional, service and sales, and an elementary-occupational position. For women, however, the variable is insignificant for all except for the skilled-agricultural occupation, where earning a minimum wage decreases the odds of working in such a field. "Education Relation-2" is completely insignificant for the females for the "Clerk" category, however, very significant for the males, showing that those males who think that their educational level is higher than their job current job position have a very little probability of having a clerk occupation. Another interesting difference comes from the "Job Engagement-5" variable, showing that males who have been working in their current positions for 1-2 years are likely to be in the service and sales, skilled-agricultural, craft,

operator, and assembler-positions, however, the respective variable is insignificant for all female categories, showing that for Bachelor's or Master's female graduates, this variable has no statistical importance. Further results can be found in Table 2 and Table 3.

4 Conclusion and Discussion

In this paper, we contribute by identifying a set of features that are important in determining a person's occupation given he or she has completed a Bachelor's or a Master's level of study. We show that the "Gender" variable is significant in the main model, thus we build logistic regression models for males and females separately. The results show that differences exist based on gender on having a certain occupation, and the Pseudo R Square and F1

classification score indicate the overall importance and better goodness of fit of the female model compared to the male one (see Table 3). The paper establishes the basis for further research for deeper understanding the gender preferences for choosing a specific occupation. Further analysis may be conducted by combining observations from multiple years for a more complete model and outcomes. Variance Inflation Factor analysis, and baseline classification models such as KNN and Decision Tree Classifier, can be constructed for evaluating the performance of the Logistic regression as a multilabel classification model.

References
1. **Bayard, K., Hellerstein, J., Neumark, D., & Troske, K.** (2003). New evidence on sex segregation and sex differences in wages from matched employee-employer data. *Journal of labor Economics*, *21*(4), 887–922.
2. **Blau, F. D., & Kahn, L. M.** (1996). Wage structure and gender earnings differentials: an international comparison. *Economica*, S29–S62.
3. **Freeman, R.** (1971). The market for college trained manpower.
4. **Harper, B., & Haq, M.** (1997). Occupational attainment of men in britain. *Oxford Economic Papers*, *49*(4), 638–650. Retrieved from http://www.jstor.org/stable/2663697
5. *Instructions for filling in labour force survey questionnaire.* (2016). State Council of Statistics of RA. Retrieved from https://www.armstat.am/file/doc/99506713.pdf
6. **Kerka, S.** (1998). *Career development and gender, race, and class. eric digest no. 199.* ERIC.
7. **Nasir, Z. M.** (2005). An analysis of occupational choice in pakistan: A multinomial approach. *The Pakistan Development Review*, *44*(1), 57–79. Retrieved from http://www.jstor.org/stable/41260703
8. **Quinter, M., & Edwards, K.** (2011). Factors influencing students career choices among secondary school students in kisumu municipality, kenya. *Journal of emerging trends in educational research and policy studies*, *2*(2), 81–87
9. Softmax Regression. (n.d.). UFDL. Retrieved from http://ufldl.stanford.edu/tutorial/supervised/Softmax Regression/
10. **Zveglich, J. E., & van der Meulen Rodgers, Y.** (2004). Occupational segregation and the gender wage gap in a dynamic east asian economy. *Southern Economic Journal*, *70*(4), 850–875. Retrieved from http://www.jstor.org/ stable/4135276