

Prediction of the probability of employment termination by people over the age of 50 using parametric survival models

Wioletta Grzenda¹

Abstract

The paper proposes the use of survival analysis methods to predict the probability of transition from employed to unemployed or economically inactive of people aged over 50, after different periods of time. The study considers the periods of three, five and ten years. In addition, the use of parametric survival models for the prediction has allowed to indicate these characteristics of respondents that have the greatest impact on the probability of transition under consideration. The analysis is based on the data from the Labour Force Survey from 2018. Two models were considered due to the differences in the retirement age for women and men. For both women and men, gamma model turned out to be the model best fitted to empirical data. Apart from the age of the respondents, the probability of employment termination was most influenced by such features as education, form of employment and place of residence. In addition, it was received that if an individual reaches retirement age and does not terminate employment then, in the next three years they still have a high probability of remaining economically active.

Keywords: economic activity, parametric survival models, prediction, employment termination

JEL Classification: J64, C41

1. Introduction

Survival analysis methods are increasingly used to study the economic activity of a population (Volker *et al.*, 2015; Auer and Fossati, 2019; Grzenda, 2019). They enable the modelling of the transitions of individuals between various states such as employed, unemployed, and economically inactive while taking into account the duration of the period an individual remains in a given state. Using parametric survival analysis models (Kalbfleisch and Prentice, 2011), it is possible to study the influence of various factors on the transition between distinguished states. This paper proposes to extend this approach by predicting the probability of employment termination for individuals.

In view of the aging of the Polish society (Leszko *et al.*, 2015), which results in the decrease in labour supply, it is important to search for answers not only to the question of what factors influence the transition of older people to the economic inactivity, but also what are the chances

¹ Corresponding author: SGH Warsaw School of Economics/Collegium of Economic Analysis, Institute of Statistics and Demography, al. Niepodległości 162, 02-554 Warszawa, wgrzend@sgh.waw.pl.

that a given person does not end their economic activity as soon as they reach retirement age. The results obtained in this study contribute to the discussion on the statutory retirement age of Poland, which is one of the lowest in Europe. In 2017, new act decreasing the retirement age to 60 years for women and 65 years for men came into force in Poland, amending the earlier act from 2012 on the equalization of retirement age for women and men and the gradual increase of the retirement age to 67 years starting from January 1, 2013. Such activities are not conducive to improving the rates of economically activity older people in Poland. In 2018, employment rate of older workers in Poland for age group 55-64 was 48.9%, compared to 71.4% in Germany, 72% in Norway and 65.1% in the Czech Republic (Eurostat, 2020).

Demographic changes taking place in Poland require greater economic activation of the elderly. The legal arrangements should encourage the people, who have already reached retirement age, but are still ready to work, to continue their economic activity.

2. Research method

In this study, parametric models that include a very broad class of survival models were considered (Swindell, 2009; Kalbfleisch and Prentice, 2011). A model that best represents the process of transition of people aged over 50 from employed to unemployed or economically inactive was sought. In this case, the most frequently considered probability distributions are: exponential distribution, Weibull distribution, gamma distribution, log-logistic distribution, and log-normal distribution. Each of them has been implemented in the empirical part of our study. The model with the use of which one can most accurately describe the process of transition from employed to unemployed or economically inactive has been identified to be the gamma model. Next, based on this model, using the survival function, a prediction of the probability of employment termination after the selected periods of time was made.

The gamma model is a generalization of the Weibull model and the log-normal model. The density function with parameters α , $\alpha > 0$ and ξ , $\xi \in \mathbf{R}$ for this model is given by the formula:

$$f(t_i|\alpha, \xi) = \frac{1}{\Gamma(\alpha)} t_i^{\alpha-1} \exp(\alpha\xi - t_i \exp(\xi)). \quad (1)$$

Let $\mathbf{t} = (t_1, t_2, \dots, t_n)^T$ be a vector of survival times, independent and with the same distribution, and $\mathbf{v} = (v_1, v_2, \dots, v_n)^T$ a vector of censoring variables. Then, the likelihood function for the parameters vector (α, ξ) in the gamma model under consideration is given by the formula:

$$\begin{aligned}
 L(\alpha, \xi; \mathbf{t}, \mathbf{v}) &= \prod_{i=1}^n [f(t_i|\alpha, \xi)]^{v_i} [S(t_i|\alpha, \xi)]^{1-v_i} \\
 &= \frac{1}{(\Gamma(\alpha))^d} \exp\left(d\alpha\xi + \sum_{i=1}^n v_i(\alpha \ln(t_i) - t_i \exp(\xi))\right) \\
 &\quad \cdot \prod_{i=1}^n t_i^{-v_i} (1 - IG(\alpha, t_i \exp(\xi)))^{1-v_i},
 \end{aligned} \tag{2}$$

where $d = \sum_{i=1}^n v_i$ and $IG(\alpha, t_i \exp(\xi)) = \frac{1}{\Gamma(\alpha)} \int_0^{t_i \exp(\xi)} u^{\alpha-1} \exp(-u) du$.

The vector of explanatory variables is introduced to this model using the ξ parameter, i.e. $\xi_i = \mathbf{x}_i \boldsymbol{\beta}$, where $\mathbf{x}_i^T = [x_{i1}, \dots, x_{ik}]$ denotes the vector of independent variables for the i -th individual, $i = 1, \dots, n$ and $\boldsymbol{\beta} = [\beta_1, \dots, \beta_k]$ denotes the parameters vector. The parameters of this model are estimated using a Newton-Raphson algorithm.

3. Empirical data and models estimation

In this study, a data set derived from the Labour Force Survey (LFS) was used. The LFS is a quarterly panel survey with a rotational sample selection scheme. In our study, research sample comprised individuals who were surveyed for two consecutive quarters in 2018. These are people from sample numbers 75-77 and 79-81. In the first stage of the analysis, in line with the research objective, people who have ever worked after 2011 were selected. The choice of year 2011 resulted from the fact that the research goal was the analysis of the economic activity of the elderly in recent years. In addition, a broader survey was used in LFS starting from this year. Furthermore, for the study women aged 51 to 70 and men aged 56 to 73 were chosen.

Different age groups for men and women were considered due to the differences in the statutory retirement age for men and women in Poland. On the other hand, the choice of the lower age limit of the respondents resulted from the fact that the average age of economic deactivation is generally lower than the statutory retirement age. As a consequence of the aforementioned decisions, 9,540 women and 7,855 men were considered in the study. At the time of the LFS survey, 60.22% of women and 59.10% of men were still working. For people, who were still working at the time of the LFS survey, time was defined as the number of months since January 2011 or since the time of starting work, if it was taken after 2011. For people who were not working at the time of the LFS survey, time was defined as the number of months starting from 2011 until the employment termination. Table 1 presents the characteristics of the respondents.

Table 1. Sample characteristics in percent

Variable	Description	Categories	Men	Women
age_group_m	Age group of men at the time of the survey	1 = from 56 to 58 years old	21.06	-
		2 = from 59 to 61 years old	22.72	-
		3 = from 62 to 64 years old	21.74	-
		4 = from 65 to 67 years old	18.57	-
		5 = from 68 to 70 years old	11.38	-
		6 = 71 and above (ref.)	4.52	-
age_group_w	Age group of women at the time of the survey	1 = from 51 to 53 years old	-	16.93
		2 = from 54 to 56 years old	-	17.65
		3 = from 57 to 59 years old	-	18.86
		4 = from 60 to 62 years old	-	20.37
		5 = from 63 to 65 years old	-	12.35
		6 = from 66 to 67 years old	-	9.86
		7 = 68 and above (ref.)	-	3.98
education	Level of education	1 = higher	17.43	23.82
		2 = post-secondary and secondary professional	22.84	28.40
		3 = secondary general	2.11	7.85
		4 = basic vocational	41.49	24.17
		5 = primary school (ref.)	16.13	15.77
marital_status	Marital status	0 = unmarried, a widower, a widow or divorced (ref.)	14.91	27.41
		1 = married	85.09	72.59
employment	Type of employment	1 = salaried employee	72.62	79.88
		2 = self-employed or helping family member (ref.)	27.38	20.12
place_residence	Class of place of residence during the survey	1 = city of 100 thousand residents and more	34.41	37.51
		2 = city of 20-100 thousand residents	16.94	17.55
		3 = city under 20 thousand residents	12.78	12.87
		4 = rural areas (ref.)	35.86	32.08
region	Region of Poland	1 = Central	13.21	13.79
		2 = Southwest	10.85	10.36
		3 = South	10.48	11.17
		4 = Northwest	18.55	17.03
		5 = North	20.75	20.88
		6 = East (ref.)	26.16	26.76
child	The presence of a child under 15 years old in a household	0 = no (ref.)	95.40	95.28
		1 = yes	4.60	4.72
elderly_person	The presence of elderly person over 75 years old in a household	0 = no (ref.)	92.82	91.43
		1 = yes	7.18	8.57

The following parametric survival analysis models were estimated in the study: exponential model, Weibull model, gamma model, log-logistic model and log-normal model. Based on the statistics (AIC), (AICC) and (BIC), the gamma model turned out to be the model best fitted to the data. At the stage of model construction, it turned out that one set of variables should be included in a model for women and another one in a model for men. Table 2 presents the results of the model estimation for women and Table 3 for men, respectively.

Based on the results obtained for women (Table 2), it can be concluded that as expected women from the first four age groups, i.e. aged 51 to 62, delayed the transition from working to unemployed or economically inactive compared to women from the oldest age group i.e. women from 68 to 70 years old. The transition to unemployed or economically inactive state was influenced by at least secondary education, with the best-educated women delaying the moment of employment termination compared to women with primary education. Women with at least tertiary education more than twice more delayed the termination of employment than women with secondary vocational or post-secondary education compared to women with primary education. In the case of the variable describing the place of residence, living in cities with more than 100,000 inhabitants had a significant impact on employment termination. Inhabitants of such cities delayed the transition to unemployed or inactive status compared to women living in the countryside. The region of residence feature had a limited impact on the probability of transition studied, due to the large values of the p -value. Nevertheless, for the variable describing the region of Poland, it was obtained that the inhabitants of the South and North regions accelerated the moment of quitting their work compared to the inhabitants of the East region, and the residents of the other regions compared to the inhabitants of the East region delayed the transition to the unemployed or inactive state.

Table 2. Results of the estimation of a Weibull model for women

Variable	Parameter estimate	Standard error	t statistic	p-value	
intercept	4.3151	0.1022	1783.93	<.0001	
age_group_w	1	1.4172	0.0687	425.30	<.0001
age_group_w	2	1.1586	0.0628	340.87	<.0001
age_group_w	3	0.7883	0.0576	187.21	<.0001
age_group_w	4	0.2551	0.0538	22.50	<.0001
age_group_w	5	-0.0858	0.0459	3.49	0.0617
age_group_w	6	-0.1397	0.0411	11.55	0.0007
education	1	0.5541	0.0393	198.69	<.0001
education	2	0.2160	0.0278	60.25	<.0001
education	3	0.0881	0.0333	6.99	0.0082
education	4	0.0018	0.0235	0.01	0.9401
place_residence	1	0.0726	0.0240	9.17	0.0025

Variable		Parameter estimate	Standard error	t statistic	p-value
place_residence	2	0.0101	0.0240	0.18	0.6728
place_residence	3	0.0268	0.0286	0.88	0.3480
region	1	0.0304	0.0283	1.15	0.2830
region	2	0.0285	0.0337	0.72	0.3968
region	3	-0.0526	0.0287	3.36	0.0670
region	4	0.0024	0.0302	0.01	0.9354
region	5	-0.0295	0.0303	0.95	0.3300
child	1	0.1010	0.0349	8.38	0.0038
elderly_person	1	0.0681	0.0395	2.97	0.0846
employment	1	-0.2765	0.0250	122.57	<.0001
scale		0.4322	0.0476	-	-
shape		1.6037	0.2294	-	-

In addition, in the case of women, the presence of a child under 15 years old in a household and the presence of an elderly person over 75 years old in a household influenced the transition from working state to unemployed or inactive state. In both cases, it turned out that both the presence of a child under 15 years of age and of a person aged over 75 years delayed employment termination compared to women living in households in which there were no such persons. The form of employment also had an impact on the transition from working to unemployed or inactive state. Employed workers have accelerated their transition from working to unemployed or economically inactive compared to self-employed or assisting a family member women.

Table 3. Results of the estimation of a Weibull model for men

Variable		Parameter estimate	Standard error	t statistic	p-value
intercept		4.4614	0.0662	4535.22	<.0001
age_group_m	1	1.3271	0.0655	411.03	<.0001
age_group_m	2	1.0168	0.0607	281.04	<.0001
age_group_m	3	0.6213	0.0552	126.56	<.0001
age_group_m	4	0.2522	0.0466	29.25	<.0001
age_group_m	5	-0.0073	0.0447	0.03	0.8697
education	1	0.4630	0.0401	133.07	<.0001
education	2	0.2197	0.0317	47.91	<.0001
education	3	0.1252	0.0640	3.82	0.0506
education	4	0.0507	0.0250	4.11	0.0427
place_residence	1	0.0940	0.0252	13.90	0.0002
place_residence	2	0.0295	0.0269	1.20	0.2726
place_residence	3	0.0044	0.0292	0.02	0.8797
marital_status	1	-0.0970	0.0252	14.82	0.0001
employment	1	-0.4269	0.0271	248.39	<.0001
scale		0.4337	0.0354	-	-
shape		1.6336	0.1650	-	-

For men, less variables in a model resulted in a better fit of the model to empirical data (Table 3). Similarly to the case of women, it was obtained that men from the first four age groups, i.e. from 56 to 67 years old, delayed the transition from working to unemployed or inactive status compared to men from the oldest age group i.e. from 71 to 73 years old. For the education feature, it was obtained that having any higher level of education than a primary education had a positive effect on economic activity. At the same time, men with higher education delayed the termination of employment to the greatest extent compared to men with primary education. In the case of a variable describing the place of residence, as in the case of women, only living in cities with more than 100,000 inhabitants had a significant impact on the employment termination. Men living in such cities delayed the transition to unemployed or economically inactive compared to men living in the countryside. In addition, it was received that married men accelerated the moment of employment termination compared to unmarried men. In the case of the form of employment feature, it was received that employees have accelerated the transition from working to unemployed or inactive compared to men who were self-employed or assisting family members. In addition, for men the value of parameter estimation at this variable was almost twice as large as compared to women.

Next, based on the estimated models, the probability of employment termination was predicted. Sample results for selected variables are presented in Table 4.

Table 4. The results of the prediction of the probability of quitting work for women and men

Variable values								Prediction of the probability					
age_group		education		place_resid.		employ.		3 years		5 years		10 years	
W	M	W	M	W	M	W	M	W	M	W	M	W	M
1	1	1	4	4	4	1	1	0.018	0.059	0.037	0.121	0.101	0.319
1	1	4	2	1	4	1	2	0.035	0.025	0.074	0.052	0.200	0.139
1	1	1	1	1	1	1	1	0.016	0.029	0.033	0.059	0.091	0.158
2	2	1	4	1	3	2	1	0.018	0.091	0.038	0.186	0.102	0.482
2	2	1	2	1	3	1	2	0.023	0.039	0.088	0.081	0.131	0.213
2	2	4	4	2	2	1	1	0.056	0.100	0.117	0.206	0.316	0.528
3	2	2	4	3	4	1	2	0.080	0.057	0.166	0.118	0.440	0.311
3	3	4	4	3	4	1	2	0.094	0.087	0.195	0.179	0.511	0.465
4	3	4	2	4	1	1	1	0.209	0.126	0.428	0.258	0.907	0.643
4	3	4	4	4	4	2	1	0.141	0.159	0.292	0.325	0.718	0.768

The probability values presented in Table 4 were determined on the basis of all information about the individuals considered, which was included in the models, while these tables present only selected characteristics for these individuals. Both tables reveal how much the probability of finishing employment during the period of three, five and ten years varies depending on

individual characteristic. As an example, the probability of terminating employment within ten years by women aged 60-62 having basic vocational education and living in rural areas is equal 0.907 for salaried employees, but by far lower and equal 0.718 for self-employed women. For the second age group we may even observe that the probability of finishing employment within 5 years by some women from this group is higher than the probability of finishing employment within 10 years by other women from the same group age. In the case of women, it can be seen that the probability of remaining economically active is strongly influenced by the education and the place of residence of a woman. In addition, it was found that, regardless of age, women who are self-employed or assisting a family member are unlikely to terminate employment compared to women who are employed. The predictions received for men aged 56 to 58 years are the most diverse due to education feature. The form of employment has a big impact on the probability of employment termination for men. This can be clearly seen in the case of the oldest age group presented in the table (62-64 years). The self-employed men were almost half as likely to terminate their employment. In addition, there was no such large variation of probability due to various places of residence among men as for women. It is worth noting here that for both men and women it is only after reaching a certain age that the probability of employment termination increases significantly, the values before this age are at a similar level and are largely dependent on the individual characteristics of individuals.

4. Conclusions

One of the most important conclusions received in this study was the indication of the age after which Poles most often ended their economic activity. It was received that this is not the current statutory retirement age. Neither women nor men most often ended their economic activity at the current statutory age, but at 63 years for women and 68 for men. Of course, when interpreting this result, it is necessary to take into account both the 4-year transition period of extending the retirement age as well as the rights of some professional groups to early retirement.

The second important conclusion is that women living in households in which there were children under 15 years of age and persons over 75 years of age did not give up their economic activity. This shows that changes are taking place in taking care of those who require it. Previous studies showed that informal care negatively affected employment, in particular for women (Kotsadam, 2011; Wilińska *et al.*, 2019). In addition, it was found that well-educated

older people who live in large cities are more economically active, and when they are running their own business they tend to delay the time they quit their jobs.

Performing prediction of the probability of transition from working to unemployed or economically inactive with the use of parametric survival models has made it possible to include individual employee characteristics in the model. This approach allowed both to indicate the profile of individuals who withdraw fastest from the labour market and those who remain economically active despite reaching retirement age, as well as enabled determining the probability of termination of employment for each employee separately.

Due to the low economic activity of older Poles compared to the inhabitants of other European countries and the progressing aging of the Polish society, the results obtained in this study can be helpful in developing effective programs for the economic activation of people aged 50+, by precisely defining the target group of these programs. Moreover, it may be helpful in forecasting labour supply.

References

- Auer, D., Fossati, F. (2019). The absent rewards of assimilation: how ethnic penalties persist in the Swiss labour market. *The Journal of Economic Inequality*, 17(2), 285-299.
- Grzenda, W. (2019). Socioeconomic aspects of long-term unemployment in the context of the ageing population of Europe: the case of Poland. *Economic research – Ekonomska istraživanja*, 32(1), 1561-1582.
- Eurostat (2020). Employment rate of older workers, age group 55-64. <https://ec.europa.eu/eurostat/databrowser/view/tesem050/default/table?lang=en>.
- Kalbfleisch, J.D., Prentice, R.L. (2011). *The statistical analysis of failure time data* (Vol. 360). John Wiley & Sons.
- Kotsadam, A. (2011). Does informal eldercare impede women's employment? The case of European welfare states. *Feminist Economics*, 17(2), 121-144.
- Leszko, M., Zajac-Lamparska, L., Trepala, J. (2015). Aging in Poland. *The Gerontologist*, 55(5), 707-715.
- Swindell, W.R. (2009). Accelerated failure time models provide a useful statistical framework for aging research. *Experimental gerontology*, 44(3), 190-200.
- Volker, D., Zijlstra-Vlasveld, M.C., Brouwers, E.P.M., Van Lomwel, A.G.C., Van Der Feltz-Cornelis, C.M. (2015). Return-to-work self-efficacy and actual return to work among long-term sick-listed employees. *Journal of occupational rehabilitation*, 25(2), 423-431.
- Wilińska, M., Grzenda, W., Perek-Biała, J. (2019). Grandmothers and Non-grandmothers in the Polish Labor Market: The Role of Family Issues. *Journal of Family Issues*, 40(12), 1677-1704.