SURVIVAL ANALYSIS FOR INCOME AFFLUENCE DURATION IN POLAND

Anna Sączewska-Piotrowska

Abstract

Poverty is a phenomenon analyzed all over the world. Preventing and combating poverty is a very serious problem of modern social policy. At the other end of income distribution is the affluence. The range of affluence, the determinants of this phenomenon, affluence duration, and territorial division of the affluence are analyzed relatively rare. It should be noted that knowledge of the affluence is very important – the affluent have a huge influence on economy, politics, culture, and they are the reference group for poorer groups in the society. The aim of the paper was to assess the affluence duration of households in Poland. The attention was paid on the income dimension in 2000-2015. Additionally, there were indicated the factors associated with exits from the affluence. In the analysis were included age, sex, and education of the household head. The analogous analysis was conducted for non-affluence spells – assessment of duration out of affluence and indication of the factors associated with entries to the affluence. Kaplan-Meier estimators and discrete-time event history models were used in the analysis.

Key words: income affluence, Kaplan-Meier estimator, discrete-time event history model, household

JEL Code: C22, D31

Introduction

The affluence is a phenomenon not analysed as often as poverty. One-dimensional or multi-dimensional, temporary or persistent, objective or subjective – these are only selected problems related to poverty phenomenon analyzed all over the world. A particurarly important problem of social policy is a duration of poverty. The duration of affluence should also be considered because this is an important issue from the point of view of economy, politics, etc. The affluent people are a very influential group, they are also the reference group for poorer groups in the society. The aim of the paper was to analyze the affluence duration of households in Poland between 2000 and 2015. Additionally, there were indicated the factors associated with exits from the affluence. In the analysis were included age, sex, and education of the household head. The analogous analysis was conducted for non-affluence spells – assessment of duration out of affluence and indication of the factors associated with entries to the affluence. Methods and models of event history analysis were used to achieve the goals of the paper. These methods and models are used in the poverty duration analysis. One of the precursors using survival analysis in poverty study were Bane and Ellwood (1986). Poverty duration in the United States was analyzed by Stevens (1999) – she conducted an analysis of multiple spells incorporating spell duration, individual and household characteristics, and unobserved heterogeneity. Finnie and Sweetman (2003) estimated entry and exit models exploring in Canada the relationship between poverty transitions and sex, family status and other personal and situational attributes. Clement (2006) analyzed the dynamics of poverty in Russia estimating poverty exit, entry and re-entry rates, and estimating logistic discrete-time models to identify the factors associated with poverty transitions. Based on German Socio-Economic Panel, Corak et al. (2008) estimated poverty rates, rates of entry to and exit from poverty, and the duration of time of time spent in and out of poverty among children. Based on longitudinal data (1991-2006) the poverty persistence in Britain was estimated by Devicienti (2011). In subsequent years poverty duration was also analyzed in other countries, for example, Canavire-Bacarreza and Robles (2017) focused on poverty in Peru. To our knowledge, there are no analyses focusing on affluence duration. This paper is one of the first attempts of estimating affluence exit and entry rates, and factors associated with affluence exit and affluence entry. In this paper, the analysis of affluence duration focuses on income aspect. Therefore, this is one-dimensional analysis and there are not taken into account another aspects of life (e.g. luxury consumption).

1 Material and methods

In the analysis, the database from Social Diagnosis project was used (Council for Social Monitoring, 2016). The database contains information about households in Poland from 2000 to 2015. The Social Diagnosis project is a panel study and each subsequent wave involves all available households from the previous wave and households from a new representative sample. Eight waves have been conducted from 2000 to 2015 (starting from 2003 subsequent waves of the panel took place every two years). In each wave of the panel were included several thousand households (from 3005 households in 2000 to 12359 households in 2011).

The data about household income were used in the analysis. The household was considered affluent when its income was higher than 200% of the median income. This solution was adopted by many authors, for example by Brzeziński (2010). There was calculated equivalised income in order to take account of the differences in a household's size and its composition. There was used the modified OECD (Organisation for Economic Co-operation and Development) equivalence scale. This scale assigns 1 to the first adult of the household, 0.5 to each subsequent adult aged 14 or more and 0.3 to children (a person under 14).

One of the core concepts in survival analysis is survival function $S(t)$ and hazard function $h(t)$. Survival function express the probability that the survival time T is equal to or greater than some time t. The survival function is given by:

$$S(t) = P(T \geq t). \tag{1}$$

The hazard function is the instantaneous risk that the event occurs in time interval $[t, t + \Delta t]$, given survival at or beyond time t (Mills, 2011).

A classic technique allowing to estimate the survival function at time t is non-parametric Kaplan-Meier estimator given by (Kaplan and Meier, 1958):

$$\hat{S}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right), \tag{2}$$

for $t_k \leq t < t_{(k+1)}$, $k = 1,2, \ldots, r$, where r are the ordered event times $t_1 < t_2 < \cdots < t_r$, and where d_j is the number of events at time j, n_j is the number of individuals at risk at time j including censored survival times. Kaplan-Meier estimator adjusts the estimated survival time to account for the presence of right-censored observations, which occurs when the event is not experienced by the last observation.

If the hazard function is assumed to be constant between successive event times, the hazard per unit time can be found by further dividing by the time interval. If there are d_j events at the jth event time, $t_{(j)}$, $j = 1,2, \ldots, r$, and n_j at risk at time $t_{(j)}$, the hazard function in the interval from $t_{(j)}$ to $t_{(j+1)}$ can be estimated by (Collett, 2004):

$$\hat{h}(t) = \frac{d_j}{n_j \left(t_{(j+1)} - t_{(j)} \right)} \tag{3}$$

Instead of measure the time to event (in our case – exit from or enter to affluence), each time interval (period) between waves can be modeled, by creating a binary variable, which indicates whether the event occurred in the interval j. We introduce subscript i (in our case – household i), then h_{ij} defines the conditional probability that household i experiences the event in time period j, given that this household did not experience the event prior to j. Taking into account predictor X_1 hypothesized to be associated with event occurrence, discrete-time hazard is given by (Singer and Willett, 1993):

$$h_{ij} = P(T_i = j | T_i \geq j, X_{1ij} = x_{1ij}), \tag{4}$$

One of the most common models used for discrete time is a logit model, which may be written according to the formula (Singer and Willett, 1993):

$$logit(h_{ij}) = [\alpha_0 + \alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \cdots + \alpha_l D_{(l-1)ij}] + \beta_1 X_{1ij}, \tag{5}$$

where l is the number of time periods when an event can occur, $D_{1ij}, D_{2ij}, \dots, D_{(l-1)ij}$ are dummy variables indicating time intervals (each time indicator is set to 1 in the time period it represents and 0 otherwise) and one of the time indicators must be dropped from the model to avoid complete linear dependence among the time indicators. Variable X_1 can be constant over time or time-varying. The function $\alpha_0 + \alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \cdots + \alpha_l D_{(l-1)ij}$ represents the baseline logit hazard function, i.e. the value of logit hazard when all predictors in the model are 0 (in our model is only one predictor X_1).

For estimated coefficient β_1 the odds ratio $\exp(\beta_1)$ is calculated. In the case of dichotomous variables, the odds ratio allows to compare the odds of event occurrence for two groups in every time period.

To conduct analysis (Kaplan-Meier estimator) R program (R Core Team, 2018) with survival package (Therneau, 2015) was used. Discrete-time event history models were also estimated using R program, but with glm() function.

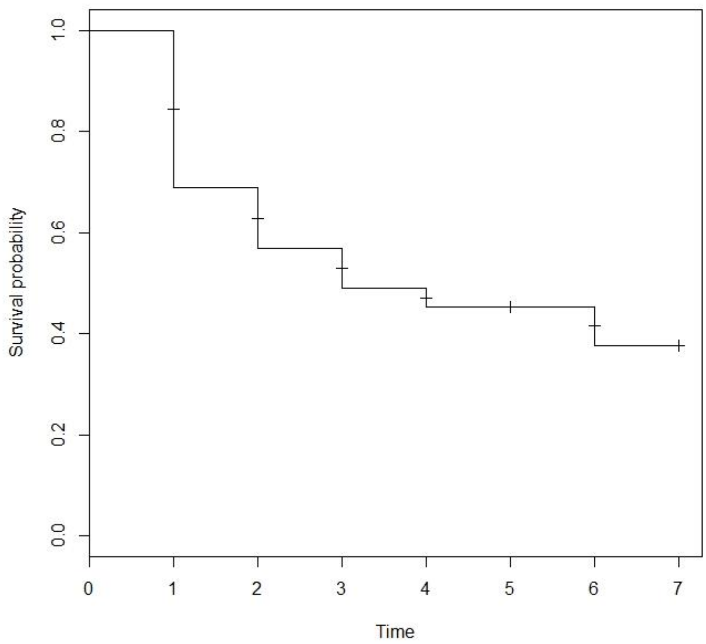
2 Results

The dynamics of income affluence of households was analyzed using the division into affluence and non-affluence spell durations, i.e. the analysis focused on time to affluence exit and time to affluence entry. The attention was paid on the time to the first event, it means on the time to the first exit and on the time to the first entry. In each case, Kaplan-Meier survival functions and discrete-time event history logit models were estimated. Logit models were estimated in two versions: base model (Model 1) and model with variables connected with characteristics of the household head (Model 2).

2.1 Affluence exits

Kaplan-Meier survival function of staying in affluence is shown in Fig. 1. Exact values of survival and hazard functions are presented in Tab. 1.

Fig. 1: Kaplan-Meier survival function of staying in affluence



Source: own work based on Council for Social Monitoring (2016).

Tab. 1: Estimated survival and hazard functions of staying in affluence

Time	1	2	3	4	5	6	7
Survival function	0.688	0.568	0.489	0.452	0.452	0.377	0.377
Hazard function	0.312	0.174	0.139	0.076	0.000	0.166	0.000

Source: own work based on Council for Social Monitoring (2016).

The estimated value for the first time interval $\hat{S}(t_1) = 0.688$ means that 68.8% of households survived in affluence two years or more. Probability of survival in affluence for a long period of time is relatively high – 37.7% of households within 12 years did not exit from affluence. Mean survival time in affluence (area under survival curve) is 4.026, i.e. approximately 8 years. It can be stated that, starting from the fourth interval of time, the affluence was abandonded by a small number of households, the majority of households leave this state earlier. Therefore, in further analysis the attention was paid on the first three intervals of income affluence duration. The base model (Model 1) and model with individual characteristics of the household head (Model 2) were estimated to identify the odds of affluence exit and to determine the factors associated with affluence exit (Tab. 2).

Tab. 2: The odds of affluence exit and factors associated with affluence exit in discrete-time event history analysis

Predictors	**Model 1**			**Model 2**		
	Odds Ratios	*CI*	*p*	*Odds Ratios*	*CI*	*p*
(Intercept)	0.45	0.40 – 0.51	**<0.001**	0.40	0.24 – 0.66	**<0.001**
Affluence exit after						
one period in affluence	ref.			ref.		
two periods in affluence	0.46	0.34 – 0.63	**<0.001**	0.50	0.37 – 0.69	**<0.001**
three periods in affluence	0.36	0.21 – 0.61	**<0.001**	0.39	0.23 – 0.67	**0.001**
Education of household head						
lower secondary or below				ref.		
basic vocational				0.75	0.51 – 1.11	0.153
secondary				1.00	0.69 – 1.44	0.984
tertiary				0.78	0.53 – 1.14	0.198
Age of household head						
34 and less				ref.		
35-44				1.30	0.84 – 2.01	0.241
45-59				1.11	0.76 – 1.64	0.582
60 and above				1.10	0.73 – 1.66	0.639
Sex of household head						
Male				1.17	0.92 – 1.49	0.209
Female				ref.		
Observations	1727			1727		
Cox & Snell's R^2 / Nagelkerke's R^2	0.022 / 0.032			0.023 / 0.034		
AIC	2091.8			1987.4		

Abbreviations: ref. is a reference category, CI is a confidence interval

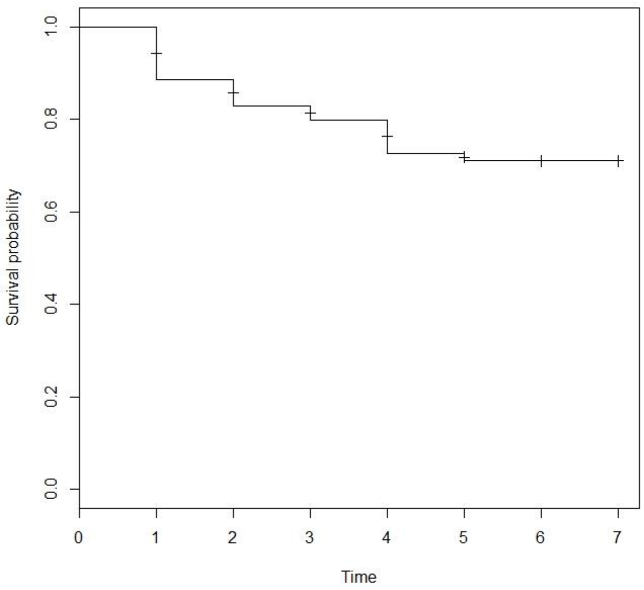
Source: own work based on Council for Social Monitoring (2016).

In Model 1 the odds of affluence exit after two periods (i.e. four years) in affluence are approximately 46% of the odds of exit after one period, and the odds of exit after three periods are about 36% of the odds of exit after one period. Based on Model 2 it can be stated that none of the included variables (sex, age, and education of the household head) is significantly associated with affluence exit. Despite this, Model 2 is better fitted than Model 1 (lower value of AIC in Model 2 than in Model 1).

2.2 Affluence entries

Kaplan-Meier survival function of staying out of affluence is presented in Fig. 2. Exact values of survival and hazard functions are shown in Tab. 3.

Fig. 2: Kaplan-Meier survival function of staying out of affluence



Source: own work based on Council for Social Monitoring (2016).

Tab. 3: Estimated survival and hazard functions of staying out of affluence

Time	1	2	3	4	5	6	7
Survival function	0.885	0.830	0.798	0.727	0.710	0.710	0.710
Hazard function	0.115	0.063	0.038	0.089	0.023	0.000	0.000

Source: own work based on Council for Social Monitoring (2016).

Based on conducted analysis it can be stated that 88.5% of households survived out of affluence state two years or more, and 71% of households survived out of affluence at least 12 years. Mean survival time out of affluence is 5.660, i.e. approximately 11 years. It means that survival time out of affluence is definitely higher than survival time in affluence. It should be clearly emphasized that most events (entries to affluence) took place in three initial time intervals and therefore in estimated discrete-time event history models only time indicators relating to the first three time intervals were included (Tab. 4).

Tab. 4: The odds of affluence entry and factors associated with affluence entry in discrete-time event history analysis

Predictors	Model 1			Model 2		
	Odds Ratios	*CI*	*p*	*Odds Ratios*	*CI*	*p*
(Intercept)	0.12	0.10 – 0.14	**<0.001**	0.04	0.02 – 0.08	**<0.001**
Affluence entry after						
one period out of affluence	ref.			ref.		
two periods out of affluence	0.54	0.38 – 0.78	**0.001**	0.57	0.39 – 0.83	**0.003**
three periods out of affluence	0.32	0.17 – 0.61	**0.001**	0.32	0.16 – 0.64	**0.001**
Education of household head						
lower secondary or below				ref.		
basic vocational				1.01	0.55 – 1.86	0.982
secondary				1.61	0.91 – 2.85	0.101
tertiary				2.90	1.65 – 5.09	**<0.001**
Age of household head						
34 and less				ref.		
35-44				1.16	0.65 – 2.08	0.611
45-59				1.69	0.99 – 2.89	0.055
60 and above				1.32	0.76 – 2.30	0.319
Sex of household head						
Male				1.55	1.10 – 2.16	**0.011**
Female				ref.		
Observations	2444			2444		
Cox & Snell's R^2 / Nagelkerke's R^2	0.010 / 0.021			0.027 / 0.060		
AIC	1474.2			1418.0		

Abbreviations: ref. is a reference category, CI is a confidence interval

Source: own work based on Council for Social Monitoring (2016).

In Model 1 the odds of affluence entry after four years out of affluence are approximately 54% of the odds of entry after one period, and the odds of entry after three periods are about 32% of the odds of entry after one period. The fit of Model 2 was better than Model 1 (lower AIC value in Model 2 than in Model 1). Two variables related to characteristics of the household head was statistically significant. Households with high-educated heads had 2.90 times higher odds to entry to affluence than households with low-educated head. Sex of household head also was statistically significant – male-headed households had 1.55 times higher odds to entry to affluence than female-headed households.

Conclusion

Based on the conducted analysis it can be stated that households survive out of affluence for a longer period of time than in affluence. The odds of affluence exit and the odds of affluence entry decrease with time spend in affluence and out of affluence, respectively. The odds of affluence entry are associated with education and sex of household head. It should be emphasized the role of tertiary education of the household head (compared to the lower secondary or below) significantly increasing the odds of affluence entry. There is no association between sex, age, and education of household head and affluence exit.

In further research, the attention should be paid on the indication of different factors which are associated with affluence exits and affluence entries. These factors may relate to some events in the household (e.g. death of a household member, the birth of the new household member, loss of a job by one of the household members) or some characteristics of household (e.g. place of resident).

References

Bane, M.J., Ellwood, D. (1986). Slipping into and out of poverty: The dynamics of spells. *Journal of Human Resources*, 21(1), 1-23.

Brzeziński, M. (2010). Income affluence in Poland. *Social Indicators Research*, 99(2), 285-299.

Canavire-Bacarreza, G., Robles, M. (2017). Non-parametric analysis of poverty duration using repeated cross section: an application for Peru. *Applied Economics*, 49(22), 2141-2152.

Clement, M. (2006). The dynamics of poverty in Russia: An analysis in terms of exit and entry. *Revue d'Etudes Comparatives Est-Ouest*, 37(2), 135-168.

Collett, D. (2004). *Modelling Survival Data in Medical Research*. Boca Raton: Chapman and Hall/CRC.

Corak, M., Fertig, M., Tamm, M. (2008). A portrait of child poverty in Germany. *Review of Income and Wealth*, 54(4), 547-571.

Council for Social Monitoring (2016). *Integrated database*. http://www.diagnoza.com (11.09.2016).

Devicienti, F. (2011). Estimating poverty in Britain. *Empirical Economics*, 40(3), 657-686.

Finnie, R., Sweetman, A. (2003). Poverty dynamics: empirical evidence for Canada. *Canadian Journal of Economics-Revue Canadienne d'Economique*, 36(2), 291-325.

Kaplan, E.L., Meier, P. (1958). Non-parametric estimation from incomplete observations. *Journal of American Statistical Association*, 53(282), 457-481.

Mills, M. (2011). *Introducing survival and event history analysis*. Los Angeles-London-New Dehli-Singapore-Washington, DC: Sage Publications.

R Core Team (2018). *R: A language and environmet for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Singer, J.D., Willett, J.B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18(2), 155-195.

Stevens, A.H. (1999). Climbing out of poverty, falling back in – Measuring the persistence of poverty over multiple spells. *Journal of Human Resources*, 34(3), 557-588.

Therneau, T. (2015). *A package for survival analysis in S*, version 2.38. https://CRAN.R-project.org/package=survival.

Contact

Anna Sączewska-Piotrowska
University of Economics in Katowice
ul. 1 Maja 50, 40-287 Katowice, Poland
Mail: anna.saczewska-piotrowska@ue.katowice.pl