ECONOMETRICS
EKONOMETRIA
Advances in Applied Data Analysis

Econometrics. Ekonometria.
Advances in Applied Data Analysis
Year 2025, Vol. 29, No. 3
ISSN 2449-9994
journals.ue.wroc.pl/eada

The Cumulative Residual Entropy in Assessing the Duration of Unemployment

Beata Bieszk-Stolorz

University of Szczecin, Poland

e-mail: beata.bieszk-stolorz@usz.edu.pl

ORCID: 0000-0001-8086-9037

Quote as: Bieszk-Stolorz, B. (2025). The Cumulative Residual Entropy in Assessing the Duration of Unemployment. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, *29*(3), 15-29.

DOI: 10.15611/eada.2025.3.02

JEL: C38, C41, J64

Abstract

Aim: The aim of the article was to use cumulative residual entropy (*CRE*) to assess the informational value of data on deregistration from the Poviat Labour Office in Szczecin (Poland) from 2007 to 2024. The events of declining cooperation with the labour office and starting work were analysed.

Methodology: The study used survival analysis methods with the assumption of an exponential distribution of duration. The *CRE* was calculated for the specified distribution parameters. Hierarchical clustering was used to identify clusters of years with similar *CRE* values. Using the dynamic time warping method, the *CRE* time series and the unemployment rate were compared.

Results: The study revealed a similarity between the registered unemployment rate and the development of entropy of unemployment duration. High unemployment rates corresponded to high entropy values, and vice versa. During periods of shock (caused by crises) in the labour market, the *CRE* for both reasons for deregistration took on extreme values.

Implications and recommendations: Analyses related to the duration in unemployment are more informative when the unemployment rate is lower. The labour market, as a system, is then characterised by less uncertainty.

Originality/value: There is a lack of research on the application of entropy by using survival analysis in studies regarding the labour market. The study took into account the relationship between the hazard value and the *CRE* for the exponential duration distribution.

Keywords: entropy, survival analysis, exponential distribution, unemployment

1. Introduction

Since Shannon's research (Shannon, 1948a, 1948b) on the mathematical theory of communication, entropy has been used as a key tool in information theory and engineering. The idea is used in research on complexity of a system and to examine the information content of probability distributions. Entropy is a general measure, which is why many definitions and applications have been proposed in various fields of science, such as demography, probability and statistics, finance, actuarial sciences, and economics (Olbryś, 2022). One area of application of entropy is the analysis of human life expectancy (Zografos, & Nadarajah, 2005; Meyer, & Ponthière, 2020a, 2020b). Initially, methods of survival analysis were used in demography and theory of reliability, whereas currently these methods are also used in relation to a random variable describing the duration of any socio-economic occurrence. Research on the entropy of the duration of socio-economic phenomena is scarce. The research presented in this article fills this gap. The study explored the applicability of entropy in labour market research in conjunction with survival analysis methods. In labour market analysis, the event of employment most often ends the observation, however data from labour offices in Poland indicate a huge number of people declining the labour office's mediation services – hence the interest in unemployed persons who, for unknown reasons, leave the labour office register.

The discussed topic is innovative for three reasons:

1. There is a lack of research on the application of entropy in labour market analysis using survival and hazard functions.
2. The study analysed unemployed people who refuse to cooperate with the labour office. The results of the study were compared with an analysis of the time until starting work.
3. The study took into account the relationship between the hazard value and the cumulative residual entropy for the exponential distribution of duration.

The aim of this article was to use the cumulative residual entropy to assess the informational value of data on deregistration from the labour office from the point of view of the duration of unemployment. The study was conducted on the basis of data from the Poviat Labour Office in Szczecin (Poland). Duration models were applied with the assumption of the exponential distribution of unemployment duration separately for each of the analysed years, namely 2007-2024. The final event considered was the termination of cooperation with the labour office. The results obtained were compared with a similar study on the event of registered unemployed persons starting work. The research period was long (18 years) and included the entire population of persons registered with the Poviat Labour Office in Szczecin (a total of 299,669). The article refers to the results of the study by Bieszk-Stolorz (2025a) concerning the analysis of starting work, thus it is a continuation of earlier research. In this case the event of declining cooperation with the office in terms of job search mediation was analysed.

2. Literature Review

The subject matter of this study focused on two groups of methods: survival analysis and entropy analysis. Survival analysis, initially associated with reliability theory and demography, has been used for many years to study the duration of socio-economic phenomena. These methods can be used to analyse the economic activity of the population (Landmesser, 2009a), credit risk (Wycinka, 2019), duration of unemployment (Meyer, 1990), duration of companies (Markowicz, 2015), poverty dynamics (Sączewska-Piotrowska, 2015), and the real estate market (Putek-Szeląg, & Gdakowicz, 2021). Survival analysis methods are also used in labour market research. In such studies, the event ending the observation is most often considered to be starting work. A preliminary analysis of individual data obtained from the Labour Office in Szczecin allowed to conclude that the beneficiaries' willingness to find a job was not the main reason for their decision to register as confirmed by earlier studies by Bieszk-Stolorz (2017a, 2017b). The cumulative incidence function revealed that declining cooperation with the office was the most likely reason for deregistration, whilst taking up employment was in second place. Survival analysis methods (Bieszk-Stolorz, 2019) allowed for the identification of groups of people who were least likely to start work such as men, those with short work experience, registered for the first time, with the lowest level of education, aged 60-64 before 12 months and aged 18-24 after 12 months. At the same time, these were the people who most strongly declined cooperation with the office.

The literature presents numerous studies on lifetime entropy which measures the amount of information obtained at the moment of an event (e.g. death) (Pierce, 1980; Meyer, & Ponthière, 2020a). In the case of human life, Shannon's entropy index allows people to better understand the risks associated with duration of their life. The value of lifetime entropy changes with age. For young people, lifetime entropy is high because there are many possible scenarios for the continuation of life. In old age, the number of possible scenarios for the continuation of life becomes smaller, and human lifetime entropy decreases (Meyer, & Ponthiere, 2020b).

An important and interesting proposal for the application of entropy in survival analysis is entropy of the population (Keyfitz & Caswell, 2005). This is often referred to as the entropy of the life table, also known as Keyfitz entropy (Keyfitz, 1977). The entropy of the population determines how relative changes in mortality rates affect the relative change in the expected life span of the population (Fernandez, & Beltrán-Sánchez, 2015). The literature mentions the following indicators of uncertainty associated with survival studies: Wiener's entropy index (Wiener, 1965), Hill's entropy index (Hill, 1993), Shannon's lifetime entropy index (Meyer's entropy) (Meyer, & Ponthière, 2020a).

An analysis of the studies referring to the entropy of human life expectancy demonstrates their universality, which means that these models can be used to analyse the duration of any socio-economic phenomenon. In the case of human life, the event that ends the observation is the death of the observed individual. The death of a human being is a negative event, and the longer the time leading up to this event, the better. In the case of duration analysis, there are two possible situations. If the event ending the observation is starting work (a positive event), the shorter the time, the better for the unemployed person, yet in the case of declining cooperation with the labour office, it is not possible to clearly assess this event in terms of positive/negative in relation to the duration of unemployment, whereas in the event of 'resignation', no specific reason is given.

3. Research Methodology and Data

3.1. Basic Functions in Survival Analysis

The duration of the phenomenon can be described as random variable *T*. With this approach, survival analysis methods can be used to study the duration. The basic function in survival analysis is the non-decreasing positive survival (duration) function *S* which specifies the unconditional probability that the event of interest has not happened by time *t* (Aalen et al., 2008). The survival function is described by the formula (Kleinbaum, & Klein, 2005):

$$S(t) = 1 - F(t) = P(T > t) = \int_t^{\infty} f(u)du, \quad (1)$$

where
$F(t)$ – cumulative distribution function of random variable *T*, $f(t)$ – probability density function.

The second important function is hazard function *h*, describing the intensity of the event at time *t* (Kleinbaum, & Klein, 2005):

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (2)$$

One can think of some important relationships between functions *h*, *S* and *f*, and two of them deserve more attention. These are:

$$h(t) = \frac{f(t)}{S(t)}, \tag{3}$$

$$h(t) = -\frac{S\prime(t)}{S(t)}. \tag{4}$$

The duration analysis uses individual data on each examined unit, hence it was possible to use incomplete (censored) data in the study, which have an impact on the shape of the survival curve *S*(*t*).

3.2. Entropy in the Survival Analysis

The idea of entropy was introduced into information theory by Shannon in 1948. If *X* is a finite discrete random variable for which the probability mass function is defined by formula $p_i = P(X = x_i)$, for $i = 1, \ldots, n$, where $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$. The Shannon entropy *H* of random variable *X* is defined as (Shannon, 1948a):

$$H = -\sum_{i=1}^{n} p_i \ln p_i. \tag{5}$$

Subsequently, the concept of entropy was extended to the case of continuous variables. If *X* is a random variable with a continuous distribution with probability density function *f*, then Shannon entropy *H* is defined by the following formula (Shannon, 1948b):

$$H = -\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx. \tag{6}$$

This measure is commonly referred to as Shannon's information measure. The assumption of continuous density function *f* means that entropy is also referred to as differential entropy or continuous entropy (Di Crescenzo, & Paolillo, 2021). If probability density function *f* of a continuous non-negative random variable X satisfies the condition $f(x) = 0$ for $x \leq 0$, then (6) takes the form:

$$H = -\int_{0}^{\infty} f(x) \ln f(x)\, dx \tag{7}$$

and represents a measure of uncertainty for X.

Entropy *H* gives the expected uncertainty contained in *f* regarding the predictability of the value of random variable *X*, i.e. it measures the concentration of probabilities (Di Crescenzo, & Toomaj, 2015; Ebrahimi, & Pellerey, 1995). Since distributions with low entropy have higher concentration, they carry more information than distributions with high entropy.

In discrete cases, Shannon entropy has a very important property. It is always non-negative, and for a certain event the entropy equals zero. One of the most significant drawbacks of this measure appears in the case of continuous distribution entropy where it can take negative values (Shannon, 1948b; Ebrahimi, 1996). This means that entropy as a measure of uncertainty is useless (Shrahili et al., 2022).

To solve this problem, Wang et al. (2003) and Rao et al. (2004) proposed a new measure of uncertainty, named cumulative residual entropy (*CRE*) (Abbasnejad et al., 2010; Rezaei, & Yari, 2021). In many practical applications, *CRE* is more interesting than the Shannon entropy. For example, if the random variable is the lifetime of a machine, it is more interesting to know whether the lifetime exceeds *t* than whether the lifetime equals *t*. Cumulative residual entropy is based on survival function *S* and defined by the formula:

$$CRE = -\int_{0}^{t} S(u) \ln S(u)\, du, \tag{8}$$

where
$S(t)$ – survival function.

This measure always takes non-negative values.

3.3. Exponential Distribution of Duration

Various approaches related to survival time distribution are used in research, such as non-parametric models (Nelson-Aalen estimator, Kaplan-Meier estimator), semi-parametric models (Cox hazard model) and parametric models based on various assumptions regarding the form of the survival function. In the case of known distribution of the duration of the phenomenon, parametric methods are preferred as they are better than other methods (non-parametric or semi-parametric) (Zhang, 2016). Survival time studies use, among others, the gamma, Weibull, exponential, and log-normal distributions. The Gompertz distribution is often used in demography as it provides a convenient way to describe human life span. In epidemiology and veterinary medicine, the most important parametric forms are the Weibull and exponential distributions (Stevenson, 2007; Montaseri et al, 2016). Due to certain advantages, the exponential distribution is especially useful in survival analysis. First, it assumes a constant risk rate over time. Second, this distribution is characterised by simplicity and lack of memory. This means that future life expectancy is the same regardless of current age. A constant risk rate is rare not only in human survival studies, but also in those involving animals. The applicability of the exponential distribution is limited in these cases (Gross, 1995). The exponential model is, however, taken into account in studies on the duration in unemployment and job search time (Landmesser, 2009b; Güell, & Lafuente, 2022; Basha, & Gjika, 2022). In the case of persons registered with labour offices, many of them start working within the first few months of registration. A preliminary analysis of the data indicated that the time from registration at the office to declining cooperation is also strongly right-skewed. This is often an exponential distribution (Bieszk-Stolorz, 2025a, 2025b).

The basic functions describing the distribution of duration regarding exponential distribution are included in Table 1.

Table 1. Basic functions describing the distribution of duration regarding exponential distribution for $\lambda > 0$

Function	Formula	Formula number
Probability density function	$f(t) = \lambda e^{-\lambda t}$	(9)
Cumulative distribution function	$F(t) = 1 - e^{-\lambda t}$	(10)
Survival function	$S(t) = e^{-\lambda t}$	(11)
Hazard function	$h(t) = \lambda$	(12)

Source: author's own study.

The distribution is supported on the interval [0, ∞), and λ is called the rate parameter. Using formulas (7) and (9), Shannon entropy can be determined for an exponential distribution, thus one obtains:

$$H = 1 - \ln \lambda. \quad (13)$$

It results from formula (13) that for $\lambda > \mathrm{e}$ (e – Euler's number) entropy H takes negative values, therefore for exponential distribution, a fundamental flaw of this measure may become apparent (Bieszk-Stolorz, 2025a). In this case, it was worth using cumulative residual entropy, which for the assumed distribution (formulas (8) and (11)) was defined by the formula in (Hooda, & Sharma, 2010):

$$CRE = \frac{1}{\lambda}. \quad (14)$$

The cumulative residual entropy, given by formula (14) is always positive (because $\lambda > 0$), and is a better measure of entropy for exponential distribution.

In the case of exponential distribution, based on formulas (12) and (14), an important relationship between hazard and cumulative residual entropy can be formulated, namely:

$$CRE = \frac{1}{h(t)}. \quad (15)$$

Based on formula (15), the following conclusions can be drawn:

1. For an exponential distribution of duration, the cumulative residual entropy is the inverse of the hazard.
2. For an exponential distribution of duration, the cumulative residual entropy values decrease as the hazard increases.
3. For an exponential distribution of duration, the information value of the system increases with the intensity of the event.

3.4. Data Used in the Research

The research used individual data on unemployed persons who deregistered from the Poviat Labour Office in Szczecin (Poland) between 2007 and 2024. The need to use such data resulted from the specific nature of the survival analysis methods, and the main problem was access to such data concerning every person registered with the labour office and obtained from the SYRIUSZ IT system, whereas before 2007 the PULS IT system was used in labour offices in Poland. Naturally, data from the old system was transferred to the new one, but this resulted in some inaccuracies, especially regarding the reason for deregistration, hence this study was conducted from 2007 onwards.

The data obtained from the poviat (county) labour office included the date of registration, the date of deregistration and information about the reasons for it. The time of registration of an unemployed person at the office was determined, and was random variable *T*. The office register lists many reasons for deregistration, with some of them related to taking up work, retirement, moving abroad, continuing education, or death. However, a significant proportion of deregistrations indicate that the person has ceased to co-operate with the office, namely:

- refusal without reasonable cause to accept a suitable job offer or other gainful employment, to perform intervention works or public works, or to participate in training, an internship or vocational preparation at the workplace;
- unwillingness to start work for a period of at least 10 days (failure to appear at the office on the appointed date);
- the person did not appear at the Labour Office on the appointed date and did not provide a valid reason for his/her absence;
- unemployed person's request to be removed from the register.

These reasons were considered as declining cooperation with the office and were accepted as an event ending the observation. In this case, deregistration for other reasons is a right-censored observation. Table 2 contains information on the number and structure of unemployed persons.

Table 2. Total number of people deregistered from the Poviat Labour Office in Szczecin and the percentage of the de-registered due to starting work and declining cooperation in 2007-2024

Year	Total number of deregistered persons	Deregistered persons (in %) due to	
		starting work	declining cooperation
2007	23,745	34%	59%
2008	17,232	32%	62%
2009	19,398	36%	55%
2010	17,613	41%	51%
2011	15,194	39%	49%
2012	15,570	39%	48%
2013	23,762	46%	44%
2014	24,443	45%	46%
2015	25,568	43%	48%
2016	23,447	42%	48%
2017	19,697	40%	50%
2018	14,873	42%	49%

Year	Total number of deregistered persons	Deregistered persons (in %) due to	
		starting work	declining cooperation
2019	12,680	44%	47%
2020	7772	69%	20%
2021	8878	63%	27%
2022	9537	51%	40%
2023	9259	58%	33%
2024	11,001	53%	40%

Source: author's own study.

There were four stages of the study. In the first stage, for each year analysed (2007–2024), random variable *T* was determined and describes the time from registration to deregistration from the labour office. Declining cooperation with the office is a complete observation coded as 1, whilst deregistration for other reasons (censored observation) was coded as 0. In the second stage of the study cumulative residual entropy (*CRE*) was determined. The third stage involved the use of hierarchical clustering and Ward's method to identify clusters with similar *CRE* values. In the fourth stage, the *DTW* measure was used to compare time series of *CRE* and unemployment rates (current and previous year's) In each stage the results for the event of ceasing cooperation were compared with the event of starting work (Bieszk-Stolorz, 2025a).

4. Empirical Results

The first stage of the study involved processing data obtained from the Poviat Labour Office in Szczecin. Based on the available data (date of registration and date of deregistration) the period (time) of registration at the office (random variable *T*) was determined. In addition, the total average registration time and the average time until declining cooperation were determined (Table 3). These averages (dashed lines) were compared with the percentage of the deregistered in Figure 1. These values were collated with the unemployment rate in Szczecin (current and in the previous year), and were compared with analogous data concerning the event of starting work (Table 3, Figure 1).

Table 3. Average time of registering an unemployed person at the Poviat Labour Office in Szczecin in 2007-2024

Year	Average registration time (months)			Registered unemployment rate in Szczecin (%)	
	total	starting work	declining cooperation	current	previous year's
2007	12.5	13.8	18.9	6.5	11.8
2008	11.7	9.8	12.4	4.3	6.5
2009	4.9	4.2	4.6	8.5	4.3
2010	6.1	6.3	5.5	9.7	8.5
2011	7.4	7.6	6.6	9.9	9.7
2012	7.9	7.7	7.2	11.0	9.9
2013	8.5	7.5	9.0	10.6	11.0
2014	8.9	8.0	8.8	9.3	10.6
2015	8.4	7.4	8.5	6.8	9.3
2016	8.4	6.6	8.8	4.7	6.8
2017	7.0	5.6	6.3	3.1	4.7
2018	5.7	4.8	5.3	2.6	3.1
2019	5.3	4.1	4.8	2.4	2.6
2020	5.4	4.4	4.3	3.9	2.4
2021	7.7	6.3	8.4	3.3	3.9
2022	7.7	6.1	7.8	3.1	3.3
2023	7.1	5.2	7.8	3.6	3.1
2024	7.3	5.1	8.3	3.4	3.6

Source: author's own study.

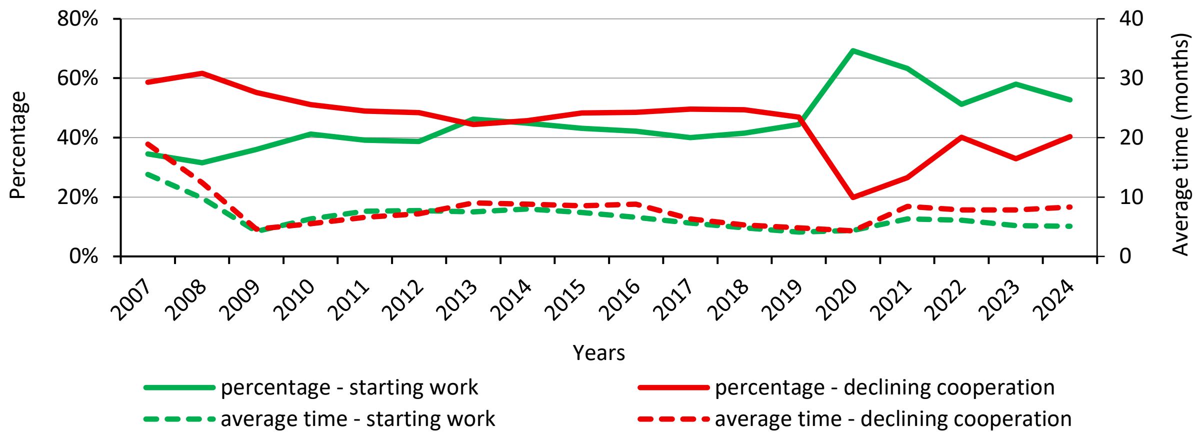

Fig. 1. Percentage of people starting work and declining the mediation services of the Poviat Labour Office in Szczecin and the average time to deregistration in 2007-2024

Source: author's own study.

The figures in Table 2 indicate a high percentage of persons deregistered for reasons other than starting work. Analysis of the raw data indicated that this was influenced by the percentage of persons who declined cooperation with the labour office, which was lower than for those starting work only in 2013 and 2020-2024. Figure 1 shows the percentage of people starting work and declining cooperation with the labour office (solid lines), with the period 2020-2021 being particularly noteworthy, when the number of people starting work through the office increased significantly. At the same time, the number of people declining cooperation with the office decreased. Szczecin is a large voivodeship city in the north-western part of Poland (386,706 inhabitants in 2024), located about 16 km from the German border. This has an impact on the specific nature of the Szczecin labour market, where many people find employment in Germany. The short distance, higher wages and open borders influence the decision to commute to work every day. Frequently, such people register with the labour office, but start working abroad and often fail to fulfil their obligation to deregister, hence the high percentage of people who did not report to the labour office on the appointed date and did not notify the office of a valid reason for their absence. In the years 2020-2021 during the pandemic, the borders of EU countries were closed, therefore people commuting to work in Germany were forced to look for employment in Poland. The following years 2022-2024 represent the post-pandemic period and also of the Russian-Ukrainian war. The number of persons deregistered from the labour office increased compared to the preceding pandemic period (especially in 2024). The percentage of persons deregistered for work decreased, but still remained at a high level compared to the period before the pandemic. The percentage of those declining the office's mediation increased compared to 2020-2021 (the pandemic), but was lower than in the period 2007-2019.

In the second stage of the research, estimators of parameter λ for the exponential distribution of the duration of unemployment were determined (Table 4). Two events were taken into account, i.e. declining mediation and starting work. Using formula (14), the cumulative residual entropy was determined for both events.

The analysis of formula (14) revealed that the declining values of the parameter λ corresponded with the increasing values of the cumulative residual entropy and vice versa (see Table 4). This results from the fact that for an exponential distribution of survival time, larger values of parameter λ correspond with smaller values of the survival time – the larger parameter λ, the faster the survival curve decreases and the faster the deregistration occurs, both due to declining mediation and starting work. Analogous regularity in the case of parameter λ and unemployment rates does not exist.

Interesting conclusions can be drawn from the analysis of formula (15). The greater the intensity of declining mediation and starting work, the lower the value of the cumulative residual entropy. This means that as the intensity of deregistration (for work or for declining cooperation) increases, the information value of the system increases.

Table 4. Results of the estimation of parameter λ for the exponential distribution of duration and the *CRE* value for deregistration due to starting work and to declining cooperation with the labour office

Year	Parameter λ		*CRE*	
	starting work	ceasing cooperation	starting work	ceasing cooperation
2007	0.0205	0.0349	48.7489	28.6707
2008	0.0271	0.0529	36.9525	18.9023
2009	0.0729	0.1115	13.7169	8.9651
2010	0.0672	0.0834	14.8704	11.9925
2011	0.0528	0.0660	18.9253	15.1494
2012	0.0492	0.0616	20.3431	16.2400
2013	0.0541	0.0520	18.4971	19.2456
2014	0.0505	0.0515	19.7953	19.4091
2015	0.0512	0.0574	19.5200	17.4177
2016	0.0504	0.0578	19.8554	17.2922
2017	0.0570	0.0706	17.5523	14.1661
2018	0.0726	0.0863	13.7695	11.5887
2019	0.0848	0.0895	11.7878	11.1695
2020	0.1283	0.0368	7.7926	27.2035
2021	0.0827	0.0347	12.0895	28.8557
2022	0.0667	0.0522	14.9880	19.1377
2023	0.0819	0.0464	12.2134	21.5580
2024	0.0724	0.0554	13.8199	18.0592

Source: author’s own study.

The third stage of the research involved clustering years according to *CRE* values for the events of employment and declined cooperation. For this purpose, hierarchical clustering applying Ward’s method (Ward, 1963) was used. The distance matrix was determined by means of the Euclidean metric. The number of clusters for both events was the same and equalled 3 (Figures 2 and 3). This was determined in the R environment (R Core Team, 2025) using the `NbClust` package (Charrad et al., 2014).

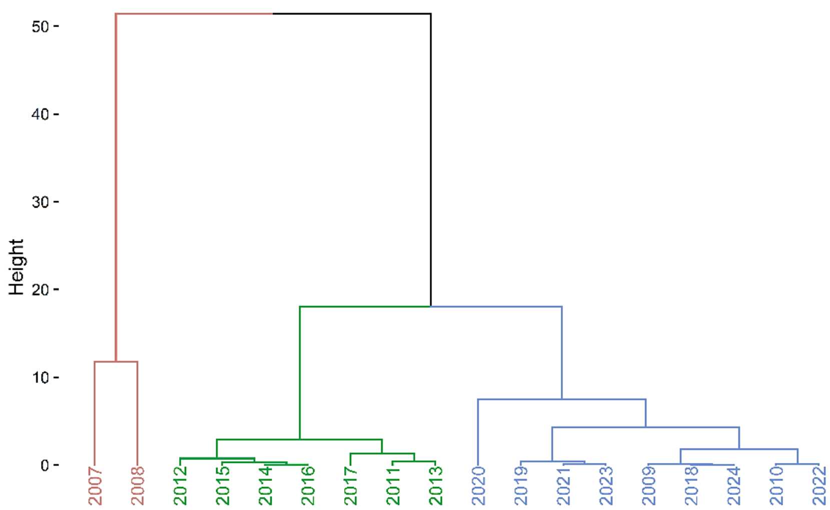

Fig. 2. Results of hierarchical clustering using Ward’s method for the event of starting work

Source: author’s own study.

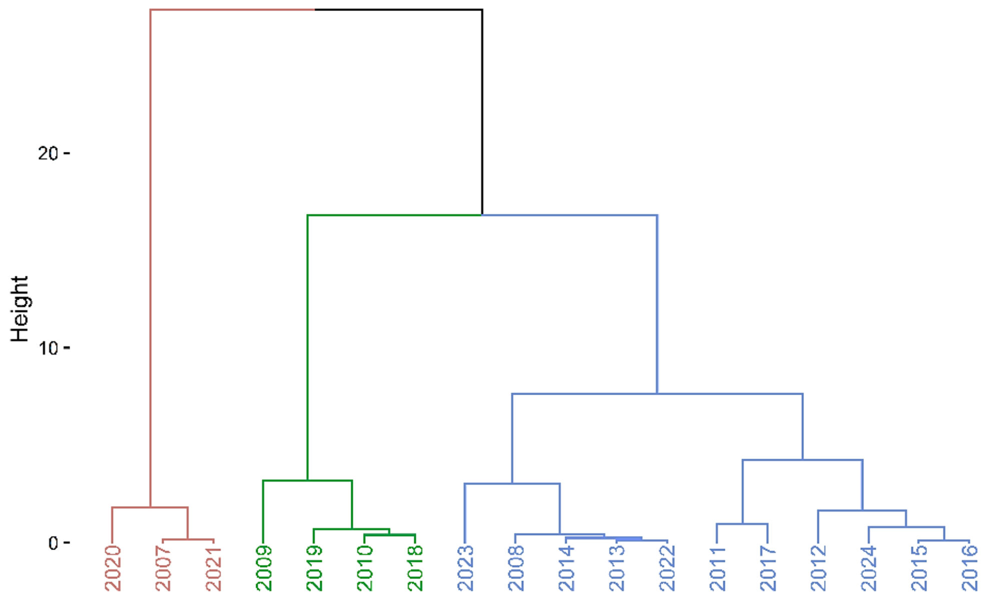

Fig. 3. Results of hierarchical clustering using Ward's method for the event of declining cooperation with the labour office

Source: author's own study.

The clusters in the case of starting work are presented in Figure 2:

- Cluster 1 – years 2007-2008 – high values of CRE.
- Cluster 2 – years 2011-2017, 2022-2024 – medium values of CRE.
- Cluster 3 – years 2009-2010, 2018-2021 – low values of *CRE*.

In the case of declining cooperation, the optimal number of clusters was also three (Figure 3):

- Cluster 1 – years 2007, 2020-2021 – high values of CRE.
- Cluster 2 – years 2008, 2011-2017, 2022-2024 – medium values of CRE.
- Cluster 3 – years 2009-2010, 2018-2019 – low values of *CRE*.

Table 5. Summary of clusters for events: starting work and declining cooperation

Year	Reason of deregistration	
	starting work	declining cooperation
2007		
2008		
2009		
2010		
2011		
2012		
2013		
2014		
2015		
2016		
2017		
2018		
2019		
2020		
2021		
2022		
2023		
2024		

Source: author's own study.

Table 5 summarises the clusters determined for both events. Clusters with high *CRE* values are marked red, those with medium *CRE* values are marked orange, and those with low *CRE* values are marked green. In the periods 2009-2019 and 2022-2024, the *CRE* levels of values for both events were the same, however in 2009-2010 and 2018-2019, they were low, and in 2011-2017 and 2022-2024 – medium. This was a period of relative economic stability, when the percentages of people starting work and declining cooperation with the labour office were similar. In 2007-2008, the *CRE* values were also similar, yet characterised by high or medium *CRE* values (declining cooperation in 2008), at the beginning of the global financial crisis in 2007-2009. The greatest variation was during the pandemic years in 2020-2021, when the *CRE* values for taking up employment were low, whereas the *CRE* values for declining cooperation were high during the crisis caused by the outbreak of the COVID-19 pandemic. It can therefore be concluded that during shock periods (caused by crises) in the labour market, the cumulative residual entropy for both reasons for deregistration took extreme values. The Russian-Ukrainian war (2022-2024) did not affect the differentiation of deregistrations between events due to employment and declined cooperation, during this time being at an average level.

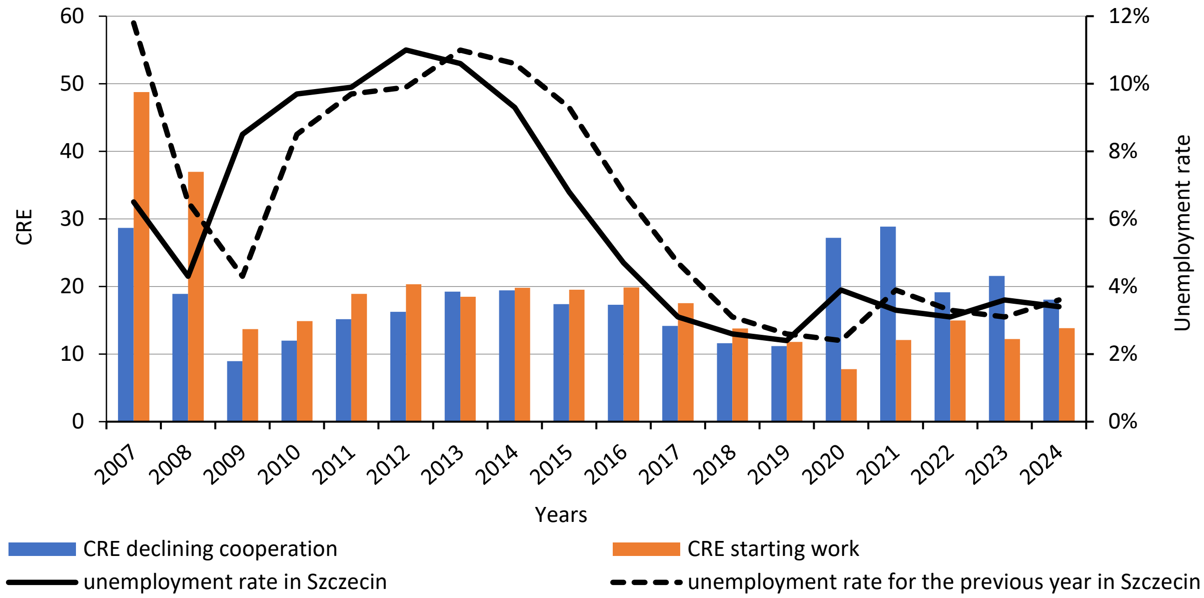

Fig. 4. The cumulative residual entropy determined for persons deregistered due to starting work and due to declining cooperation in 2007-2024

Source: author's own study.

Thus no clear pattern was observed for the cumulative residual entropy values in relation to the occurrence of a crisis situation, both for starting work and declining cooperation. In the fourth stage of the research, the cumulative residual entropy (Figure 4) was compared with the registered unemployment rate in Szczecin (solid line). A preliminary analysis of the plots confirmed the assumption of a relationship between the unemployment rate and *CRE*, which seems particularly evident in the period 2007-2021. In order to accurately assess this relationship, Figure 5 presents unemployment rates moved back one year (dashed line – unemployment rate from the previous year). The changes in CRE values (in the cases of employment and declining cooperation) occur in a similar way to the changes in the unemployment rate. For the previous year's unemployment rate, this similarity was more visible. The *DTW* distance was used to confirm these observations. Table 6 presents *DTW* distances, median and mean differences (in years) between the *CRE* (for working and declining cooperation) and unemployment rates. It can be seen that the analysed pattern was disrupted during 2022-2024, therefore DTW distances were determined for two time series: the periods 2007-2021 and 2007-2024. *DTW* distances were smaller in the case of starting work, which indicates that changes in *CRE* for starting work were more consistent with changes in the unemployment rate than changes in *CRE* for declining cooperation. On the contrary, a one-year shift in the unemployment rate increased the similarity between the analysed time series for both reasons of de-registration. This fact was also confirmed by the medians and means of the shifts. In the period 2007-2021 a strong similarity was observed between the cumulative residual entropy (for starting work and declining cooperation) and the registered unemployment rate. However, if one analyses the second time series, i.e. the years 2007-2024, there was similarity for starting work, yet there was no analogous regularity for declining cooperation. This suggests that the period of the Russian-Ukrainian war caused some disruptions in the labour market in Szczecin. In both analysed periods the high unemployment rates corresponded with high entropy values and vice versa, hence it can be concluded that during periods of high unemployment, the distributions of the duration of unemployment carry less information than during periods of low unemployment (measured by the unemployment rate). Note the differences in *CRE* values for both events analysed in crisis situations – at the beginning of the financial crisis (2007-2008), distributions related to declining cooperation were more informative, whilst at the beginning of the COVID-19 pandemic, distributions related to employment were more informative.

Table 6. *DTW* distances, median and mean shifts (in years) between *CRE* (work, declining cooperation) and unemployment rate

CRE	2007-2021		2007-2024	
	Registered unemployment rate	Previous year's registered unemployment rate	Registered unemployment rate	Previous year's registered unemployment rate
DTW distances				
declining cooperation	16.498	13.868	19.630	19.061
work	14.034	10.328	15.872	12.128
Median shifts				
declining cooperation	1	0	7	6
work	1	1	1	1
Mean shifts				
declining cooperation	0.760	0.607	5.667	5.323
work	0.792	0.440	0.759	0.355

Source: author's on study.

5. Discussion and Conclusions

The study demonstrates the application of entropy in the analysis of socio-economic phenomena using survival analysis methods. In studies on human life expectancy, entropy is sometimes identified as an alternative measure to further survival time, whilst this concept is presented here in its classical sense, i.e. as a measure of the information content of a system. Survival analysis methods are often used in the study of socio-economic phenomena. The use of survival entropy in relation to such phenomena appears extremely rarely in the literature, and this study concerned the duration in unemployment. Models of duration were described and analysed assuming exponential survival time. The research focused on two events: starting work and declining cooperation with the labour office, whose primary task is assistance in finding a job, yet persons who declined the services of the Poviat Labour Office in Szczecin in their search for work have for many years constituted a very high percentage of the deregistered unemployed. This figure is similar to or higher than the percentage of those starting work, and therefore constitutes a major social problem.

In crisis situations on the labour market, the informational value of the system changes in terms of reaching extreme entropy values – these can be minimal or maximal values. The cumulative residual entropy values allowed the author to divide the analysed period into three clusters covering crisis periods in different ways. In the case of the financial crisis (2007-2008), entropy values decreased for both unemployed people starting work and those declining cooperation, hence the informational value of the system increased, however this was higher for the latter reason than for starting work. For the

COVID-19 pandemic (2020-2021) the situation was different: the entropy value for both events increased, meaning that the informational value of the system decreased. However, this time the informational value was lower for declining cooperation than for starting work, whereas for the crisis situation there was no single direction of change in entropy values.

The study revealed a similarity between the registered unemployment rate and the development of entropy of unemployment duration. In the analysed period, high unemployment rates corresponded to high entropy values and vice versa, therefore during periods of low unemployment, the distributions of unemployment duration were more informative than during periods of high unemployment (measured by the unemployment rate). It is also interesting that extending the analysis to the years 2022-2024 (during the Russian-Ukrainian war) indicated the occurrence of disturbances on the labour market in Szczecin. The regularity regarding entropy for starting work and the unemployment rate was similar, yet this regularity was disturbed regarding the cases of declining the labour office's mediation. This observation is a good starting point for further research.

This leads to an important conclusion. Analyses related to the duration of unemployment are more informative when the unemployment rate is lower. High unemployment rates in the labour market may reduce the informative value of these methods. In a situation of low unemployment, the labour market is more orderly (in terms of entropy), and events such as starting work and declining mediation by the unemployed are less chaotic. The labour market, as a system, is then characterised by lesser uncertainty.

The study was not without limitations such as access to individual data. The data collected by Statistics Poland were aggregated, whilst only Poviat Labour Offices in Poland have access to individual data which can be made available as anonymous information for scientific purposes. The second limitation regards the fact that the study only examined registered unemployment. Information about unemployed persons from Poviat Labour Offices does not fully reflect the situation on the labour market in Poland.

References

Aalen, O. O., Borgan Ø., & Gjessing H. K.(2008). *Survival and event history analysis: A process point of vi*ew. Springer.

Abbasnejad, M., Arghami, N. R., Morgenthaler, S., & Borzadaran, G. M. (2010). On the dynamic survival entropy. *Statistics & Probability Letters*, *80*(23-24), 1962-1971. https://doi.org/10.1016/j.spl.2010.08.026

Basha, L., & Gjika, E. (2022). Accelerated failure time models in analyzing duration of employment. *Journal of Physics: Conference Series*, *2287*, 012014. https://doi.org/10.1088/1742-6596/2287/1/012014

Bieszk-Stolorz, B. (2017a). Funkcja skumulowanej częstości i modele hazardu w ocenie konkurujących form wyjścia z bezrobocia. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Research Papers of Wrocław University of Economics, 469*, 21-31. https://doi.org/10.15611/pn.2017.469.02

Bieszk-Stolorz, B. (2017b). Cumulative Incidence Function in Studies on the Duration of the Unemployment Exit Process. *Folia Oeconomica Stetinensia*, *17*(1), 138-150. https://doi.org/10.1515/foli-2017-0011

Bieszk-Stolorz, B. (2019). The Modified Lunn-McNeil Model in the Assessment of Intensity of Exiting from Unemployment. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, *23*(1), 77-89. https://doi.org/10.15611/eada.2019.1.06

Bieszk-Stolorz, B. (2025a). Entropy in the Assessment of the Labour Market Situation in the Context of the Survival Analysis Methods. *Entropy*, *27*(7), 665. https://doi.org/10.3390/e27070665

Bieszk-Stolorz, B. (2025b). Relationship Between Coefficients in Parametric Survival Models for Exponentially Distributed Survival Time–Registered Unemployment in Poland. *Econometrics*, *13*(1), 1. https://doi.org/10.3390/econometrics13010001

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, *61*(6), 1-36. https://doi.org/10.18637/jss.v061.i06

Di Crescenzo, A., & Paolillo, L. (2021). Analysis and applications of the residual varentropy of random lifetimes. *Probability in the Engineering and Informational Sciences*, *35*(3), 680-698. https://doi.org/10.1017/S0269964820000133

Di Crescenzo, A., & Toomaj, A. (2015). Extension of the past lifetime and its connection to the cumulative entropy. *Journal of Applied Probability*, *52*(4), 1156-1174. https://doi.org/10.1239/jap/1450802759

Ebrahimi, N. (1996). How to Measure Uncertainty in the Residual Life Time Distribution. *The Indian Journal of Statistics, Series A (1961-2002), 58*(1), 48-56.

Ebrahimi, N., & Pellerey, F. (1995). New partial ordering of survival functions based on the notion of uncertainty. *Journal of Applied Probability*, *32*(1), 202-211. https://doi.org/10.2307/3214930

Fernandez, O. E., & Beltrán-Sánchez, H. (2015). The entropy of the life table: A reappraisal. *Theoretical Population Biology*, *104*, 26-45. https://doi.org/10.1016/j.tpb.2015.07.001

Gross, A. J. (1995). Applications in Survival Analysis. In: K. Balakrishnan, & A. P. Basu (Eds.), *Exponential distribution: Theory, methods and applications* (pp. 498-508). Gordon and Breach.

Güell, M., & Lafuente, C. (2022). Revisiting the determinants of unemployment duration: Variance decomposition à la ABS in Spain. *Labour Economics*, *78*, 102233. https://doi.org/10.1016/j.labeco.2022.102233

Hill, G. (1993). The entropy of the survival curve: An alternative measure. *Canadian Studies in Population*, *20*(1), 43-57. https://doi.org/10.25336/P6830H

Hooda, D. S., & Sharma, D. K. (2010). Exponential survival entropies and their properties. *Advances in Mathematical Sciences and Applications*, *20*, 265-279.

Keyfitz, N. (1977). What difference would it make if cancer were eradicated? An examination of the Taeuber paradox. *Demography*, *14*(4), 411-418.

Keyfitz, N., & Caswell, H. (2005). *Applied Mathematical Demography*. Springer.

Kleinbaum, D. G., & Klein, M. (2005). *Survival analysis a self-learning text*, 2nd ed. Springer.

Landmesser, J. (2009a). The survey of economic activity of people in rural areas-the analysis using the econometric hazard models. *Acta Universitatis Lodziensis. Folia Oeconomica*, *228*, 385-392.

Landmesser, J. M. (2009b). Econometric analysis of unemployment duration using hazard models. *Studia Ekonomiczne*, *1-2*(60-61), 79-92.

Markowicz, I. (2015). Duration analysis of firms–cohort tables and hazard function. *International Journal of Business and Social Research*, *5*(11), 36-47.

Meyer, B. D. (1990). Unemployment Insurance and Unemployment Spells. *Econometrica*, *58*(4), 757-782. https://doi.org/10.2307/2938349

Meyer, P., & Ponthière, G. (2020a). Human Lifetime Entropy in a Historical Perspective (1750-2014). *Cliometrica*, *14*, 129-167. https://doi.org/10.1007/s11698-019-00185-y

Meyer, P., & Ponthiere, G. (2020b). Threshold ages for the relation between lifetime entropy and mortality risk. *Mathematical Social Sciences*, *108*, 1-7. https://doi.org/10.1016/j.mathsocsci.2020.07.004

Montaseri, M., Charati, J. Y., & Espahbodi, F. (2016). Application of Parametric Models to a Survival Analysis of Hemodialysis Patients. *Nephro-Urology Monthly*, *8*(6), e28738. https://doi.org/10.5812/numonthly.28738

Olbryś, J. (2022). Entropy-Based Applications in Economics, Finance, and Management. *Entropy*, *24*(1), 1468. https://doi.org/10.3390/e24101468

Pierce, J. R. (1980). *An introduction to information theory: Symbols, signals & noise*. Dover Publications.

Putek-Szeląg, E., & Gdakowicz A. (2021). Application of Duration Analysis Methods in the Study of the Exit of a Real Estate Sale Offer from the Offer Database System. In: K. Jajuga, K. Najman, & M. Walesiak (Eds.), *Data Analysis and Classification. Methods and Applications* (pp. 153-169). Springer. https://doi.org/10.1007/978-3-030-75190-6_10

R Core Team (2025). R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing; Vienna, Austria. Retrieved from https://www.R-project.org

Rao, M., Chen, Y., Vemuri, B. C., & Wang, F. (2004). Cumulative residual entropy: a new measure of information. *IEEE Transactions on Information Theory*, *50*(6), 1220-1228. https://doi.org/10.1109/TIT.2004.828057

Rezaei, R., & Yari, G. (2021). Keyfitz entropy: Investigating some mathematical properties and its application for estimating survival function in a life table. *Mathematical Sciences*, *15*(3), 229-240. https://doi.org/10.1007/s40096-020-00354-5

Sączewska-Piotrowska, A. (2015). Poverty duration of households of the self¬employed. *Econometrics. Ekonometria*, *1*(47), 44-55. https://doi.org/10.15611/ekt.2015.1.03

Shannon, C. E. (1948a). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379-432. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E. (1948b). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(4), 623-656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Shrahili, M., El-Saeed, A. R., Hassan, A. S., Elbatal, I., & Elgarhy, M. (2022). Estimation of Entropy for Log-Logistic Distribution under Progressive Type II Censoring. *Journal of Nanomaterials*, *2022*, 2739606. https://doi.org/10.1155/2022/2739606

Stevenson, M. (2007). *An introduction to survival analysis*. EpiCentre, IVABS.

Wang, F., Vemuri, B. C., Rao, M., & Chen, Y. (2003). A New & Robust Information Theoretic Measure and Its Application to Image Alignment. In: C. Taylor, & J. A. Noble (Eds.), *Information Processing in Medical Imaging. IPMI 2003. Lecture Notes in Computer Science* (Vol. 2732, pp. 388-400). Springer. https://doi.org/10.1007/978-3-540-45087-0_33

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236-244. https://doi.org/10.1080/01621459.1963.10500845

Wiener, N. (1965). *Cybernetics or Control and Communication in the Animal and the Machine*. MIT Press.

Wycinka, E. (2019). Competing risk models of default in the presence of early repayments. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, *23*(2), 99-120. https://doi.org/10.15611/eada.2019.2.07

Zhang, Z. (2016). Parametric regression model for survival data: Weibull regression model as an example. *Annals of Translational Medicine*, *4*(24), 484. https://doi.org/10.21037/atm.2016.08.45

Zografos, K., & Nadarajah, S. (2005). Survival exponential entropies. *IEEE transactions on information theory*, *51*(3), 1239-1246. https://doi.org/10.1109/TIT.2004.842772

Skumulowana entropia rezydualna w ocenie czasu trwania w bezrobociu

Streszczenie

Cel: Celem artykułu jest wykorzystanie miary entropii *CRE* do oceny wartości informacyjnej danych dotyczących wyrejestrowania z Powiatowego Urzędu Pracy w Szczecinie (Polska) w latach 2007-2024. Analizie poddano zdarzenia polegające na rezygnacji ze współpracy z urzędem oraz na podjęciu pracy.

Metodyka: W badaniu wykorzystano metody analizy przeżycia przy założeniu wykładniczego rozkładu czasu trwania. Dla wyznaczonych parametrów rozkładów obliczono *CRE*. Do wyodrębnienia grup lat o podobnych wartościach *CRE* zastosowano grupowanie hierarchiczne Warda. Za pomocą miary *DTW* porównano szeregi czasowe *CRE* i stopy bezrobocia.

Wyniki: Badanie wykazało podobieństwo pomiędzy kształtowaniem się entropii czasu trwania w bezrobociu a stopą bezrobocia rejestrowanego. Wysokim wartościom stopy bezrobocia odpowiadały wysokie wartości entropii i odwrotnie. W okresach szokowych (spowodowanych kryzysami) na rynku pracy *CRE* dla obu powodów wyrejestrowania przyjmowała wartości ekstremalne.

Implikacje i rekomendacje: Analizy związane z czasem trwania bezrobocia mają większą wartość informacyjną, gdy na rynku jest niższa stopa bezrobocia. Rynek pracy jako system charakteryzuje się wówczas mniejszą niepewnością.

Oryginalność/wartość: W literaturze brakuje badań nad zastosowaniem entropii w analizie rynku pracy z wykorzystaniem analizy trwania. W badaniu uwzględniono zależności między wartością hazardu a *CRE* dla wykładniczego rozkładu czasu trwania.

Słowa kluczowe: entropia, analiza przeżycia, rozkład wykładniczy, bezrobocie