econometrics

MDPI

Article

Relationship Between Coefficients in Parametric Survival Models for Exponentially Distributed Survival Time—Registered Unemployment in Poland

Beata Bieszk-Stolorz

Institute of Economics and Finance, University of Szczecin, 71-101 Szczecin, Poland; beata.bieszk-stolorz@usz.edu.pl

Abstract: Survival analysis is a popular research tool in medicine and demography. It has been used for many years to study the duration of socio-economic phenomena. The aim of this article is to evaluate the relationship between the coefficients of the proportional hazards model (PH) and the accelerated failure time model (AFT), assuming an exponential distribution of survival time. The coefficients of the PH and AFT exponential models have the same magnitude but have opposite signs. It follows that there is a symmetric relation between the coefficients. In the case of exponential PH and AFT models, there is a relation of equality between the parameters describing the quality and fit of the model, as well as between the standard errors of the parameters of both models. In this case also, we can talk about a symmetric relation. The exponential PH model is valid if the exponential AFT model is valid. The study showed that the intensity of starting work was higher in the case of men, people with work experience, people with higher education and young people. The job search time was longer for women, people with no work experience, and people aged 60+, but shorter for people with higher education.

Keywords: parametric survival models; symmetry of coefficients; PH models; AFT models; registered unemployment

check for updates

Academic Editors: Stephen G. Hall and Olivier Parent

Received: 24 September 2024
Revised: 18 December 2024
Accepted: 24 December 2024
Published: 2 January 2025

Citation: Bieszk-Stolorz, B. (2025). 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

1. Introduction

The genesis of survival analysis methods is related to studies of equipment reliability (the duration of operation of machinery and equipment) and demographic studies (the duration of human life). For many years, survival analysis has also been used to study the duration of socio-economic phenomena (Wycinka, 2019; Urbańczyk, 2020; Bieszk-Stolorz & Dmytrów, 2021; Putek-Szeląg & Gdakowicz, 2021; Calabuig et al., 2021), including the study of the duration of unemployment (Nickell, 1979; Moffitt, 1985; Addison & Portugal, 1998; Lalive, 2008; Bieszk-Stolorz, 2017; Doğan, 2019; Bieszk-Stolorz & Markowicz, 2022; Grzenda, 2023). One of the most frequently used models for duration analysis is the proportional hazards (PH) model. However, this model may be inadequate in many situations. An accelerated failure time model (AFT) may provide a better description of the phenomenon. The PH model and its various generalisations are mainly used in medical and biostatistical fields, while the alternative class of regression models, the AFT model, is mainly used in industrial experiments and reliability theory (Orbe et al., 2002). Some authors emphasise that in many situations AFT models better describe and predict the analysed phenomenon (Basha & Gjika, 2022). When proportional hazard assumptions are violated, the AFT model is the appropriate method to apply (Doğan, 2020).

There are many articles in the literature on the comparison of the two models. Often these comparisons involve the Cox proportional hazards model and the AFT model with an arbitrary duration distribution (Orbe et al., 2002; Patel et al., 2006; Faruk, 2018; Khanal et al., 2019; Khamis et al., 2020; Pang et al., 2021; Mishra & Misra, 2022). The comparison of such models makes it impossible to determine the relationship between the coefficients. This is due to the fact that both these models belong to different classes. The Cox proportional hazards (Cox PH) model is a semiparametric one. In this case, no baseline hazard function is specified. This model is particularly attractive when the researcher has only a weak theory supporting a specific parametric model (Landmesser, 2010). In contrast, the AFT model is a parametric one. In such case, an assumption about the duration distribution is made. It is impossible to find a mathematical relationship between the parameters of the semiparametric PH and parametric AFT models. In the literature, there are also issues related to the comparison of parametric PH and parametric AFT models. However, these are applications of these models to the analyses of economic and social phenomena assuming different distributions of duration. There is also no relationship between the PH and AFT models. Many studies confirm that models based on the assumption of an exponential distribution of survival time give a good description. The added value of this study is the determination of the relationship between the parameters of the PH and AFT models assuming an exponential distribution of duration time and the interpretation of this relationship based on the analysis of data on registered unemployment.

The use of both models in the study of socio-economic phenomena confirms that in many cases a good description is provided by models based on the assumption of an exponential distribution of survival time. The two research questions arise:

Q1: What is the relationship between the coefficients of the PH and AFT models in the case of an exponential survival time distribution?

Q2: How does the relationship between the coefficients of the PH and AFT models in the case of an exponential survival time distribution affect the interpretation of these coefficients?

The aim of this article is to evaluate the relationship between the coefficients of the proportional hazards model and the accelerated failure time model, assuming an exponential distribution of survival time.

An example of the application and interpretation of the coefficients of both models is the assessment of duration in registered unemployment. The empirical study was conducted based on individual data from the Poviat Labor Office in Szczecin (Poland) in 2021.

2. Methods of Survival Analysis

Survival (duration) analysis methods are used when the duration of a given phenomenon is studied and is described by a random variable T. These methods allow censored data to be used in the study. It contains observations that do not end with an event before the end of the observation period, or end with an event other than the required one. This approach allows all observed units to be included in the study. The basic function in the survival analysis is the survival (duration) function $S(t)$, described by the formula:

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

where:

t—time to the event,

$F(t)$—cumulative distribution function of random variable T,

$f(t)$—probability density function,
$P(t)$—probability.

The survival function specifies the unconditional probability that the event of interest has not occurred by time t (Aalen et al., 2008).

The second function is the hazard function $h(t)$, describing the intensity of the event at time t:

$$h(t) = \frac{-S'(t)}{S(t)} \tag{2}$$

Figure 1 shows a graph of the survival function $S(t)$ and a geometric interpretation of the hazard function $h(t)$ for time $t > 0$ (Bieszk-Stolorz, 2015). Let x denote the distance between t_1 and t. The non-increasing function $S(t)$ guarantees that $t_1 - t > 0$. We obtain:

$$x = \frac{1}{h(t)} \tag{3}$$

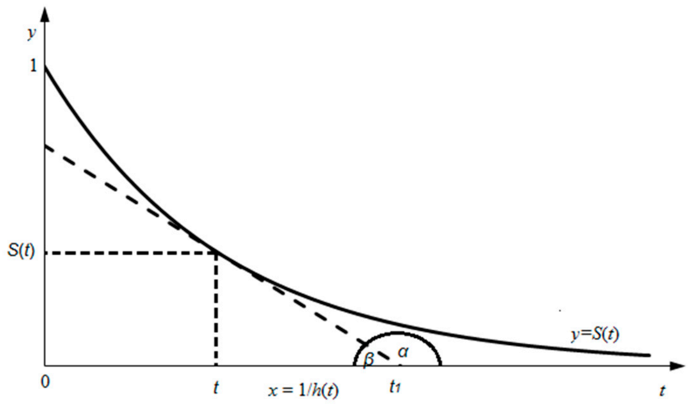

Figure 1. Plot of survival functions $S(t)$ and geometric interpretation of hazards $h(t)$ for $t > 0$.

The hazard function $h(t)$ allows calculation of the value of the hazard function at time t. The cumulative hazard function $H(t)$ is the cumulative amount of hazard up to time t. It is defined by the formula (Klein & Moeschberger, 2003):

$$H(t) = \int_0^t h(u)du \tag{4}$$

While the instantaneous hazard rate $h(t)$ can increase or decrease with time, the cumulative hazard rate $H(t)$ can only increase or remain the same.

The most common methods used in the survival analysis are the proportional hazards (PH) and accelerated failure time (AFT) models.

2.1. Proportional Hazards (PH) Models

The assumption for the proportional hazards (PH) models is that the hazard ratio (HR) is constant at time t. The hazard function at time t in the PH models is given as follows (Qi, 2009):

$$h(t|x) = h_0(t)\exp\left(\sum_{i=1}^{n} \beta_i x_i\right) \tag{5}$$

where:

$h_0(t)$—baseline hazard,
$x = (x_1, x_2, \ldots, x_n)$—vector of covariates,

$\beta = (\beta_1, \beta_2, \ldots, \beta_n)$—vector of coefficients,
n—number of covariates in the model.

The characteristics of hazard function change proportionally to the influence of explanatory variables (Landmesser, 2009). The baseline hazard function is unspecified, so the time-to-event random variable is not assumed to follow any particular distribution. This is one of the attractive features of the PH model.

Hazard ratio is defined as the hazard for one individual divided by the hazard for a different individual (Kleinbaum & Klein, 2005) and is defined by the formula:

$$HR = HR(t, x_i, x_j) = \frac{h(t|x_i)}{h(t|x_j)} \tag{6}$$

where:

$h(t|x_i)$—hazard function,
t—time to the event
$x_i = (x_{i1}^*, x_{i2}^*, \ldots, x_{in}^*)$—vector of covariates for the i-th unit,
$x_j = (x_{j1}, x_{j2}, \ldots, x_{jn})$—vector of covariates for the j-th unit,
n—number of covariates in the model.

In the case of the hazard function of the form (5), HR is defined by the formula:

$$HR(t, x_i, x_j) = \exp\left(\sum_{k=1}^{n} \beta_k (x_i - x_j)\right) \tag{7}$$

Formula (7) is simplified in the case when the i-th and j-th units differ only in the values of the x_k variable. If there is also $x_{ik} - x_{jk} = 1$, then the formula for the hazard ratio is as follows:

$$HR(t, x_i, x_j) = \exp(\beta_k) \tag{8}$$

The term "proportional hazards model" refers to the fact that for two different individuals, the ratio of their hazard functions is constant, assuming that x_i and x_j do not change over time.

2.2. Accelerated Failure Time (AFT) Models

The accelerated failure time (AFT) model is an alternative to the PH model (Pang et al., 2021). Currently, it is not widely used to analyse clinical trial data, although it is quite common in the manufacturing field. Like the PH model, the AFT model describes the relationship between survival probabilities and a set of variables. In the simple case of two groups, it can be written as (Kleinbaum & Klein, 2005):

$$S_1(t) = S_2(\gamma t) \tag{9}$$

where:

γ—the acceleration factor,
S_1, S_2—survival functions for group 1 and group 2, respectively.

The assumptions of the AFT model for the random variable T can be written as follows (Kleinbaum & Klein, 2005):

$$T_2 = \gamma T_1 \tag{10}$$

that is:

$$\gamma = \frac{T_2}{T_1} \tag{11}$$

where:

T_1—survival time of group 1,
T_2—survival time of group 2.

The acceleration factor γ is the key measure of association obtained in the AFT model. After reparameterisation we obtain:

$$\gamma = \exp\left(\sum_{i=1}^{n} \alpha_i x_i\right) \tag{12}$$

where:

$x = (x_1, x_2, \ldots, x_n)$—vector of covariates,
$a = (\alpha_1, \alpha_2, \ldots, \alpha_n)$—vector of coefficients,
n—number of covariates in the model.

Two important conclusions follow from Formulas (9) and (10) (Saikia & Barman, 2017):

1. If $\gamma > 1$, the effect of covariate is decelerated.
2. If $\gamma < 1$, the effect of covariate is accelerated.

Differences between groups are expressed in terms of direct impact on time. For the AFT models, it is assumed that independent variables act additively on the logarithm of time and therefore multiplicatively on time (Patel et al., 2006). The AFT regression model is specified as follows (Basha & Gjika, 2022):

$$\ln T = \alpha_0 + \sum_{i=1}^{n} \alpha_i x_i + \sigma\varepsilon \tag{13}$$

where:

$x = (x_1, x_2, \ldots, x_n)$—vector of covariates,
$a = (\alpha_1, \alpha_2, \ldots, \alpha_n)$—vector of coefficients,
α_0—the intercept,
ε—random errors that does not depend on x_i,
σ—an unknown scale parameter,
n—number of covariates in the model.

This means that the relationship between logarithm of survival time variable and independent variables is linear according to the accelerated failure time model. It can be assumed that the random error follows one of many distributions.

The AFT model assumes that factors interrelate with time directly in the baseline hazard function ($h_0(t)$). In this case, the hazard function at time t can be given as (Qi, 2009):

$$h(t|x) = \frac{t}{\gamma} h_0\left(\frac{t}{\gamma}\right) \tag{14}$$

After applying Formula (12), we obtain:

$$h(t|x) = h_0\left(t \cdot \exp\left(-\sum_{i=1}^{n} \alpha_i x_i\right)\right) \exp\left(-\sum_{i=1}^{n} \alpha_i x_i\right) \tag{15}$$

For the AFT model, the ratio of durations (time ratio) for the i-th and j-th units, characterised by different values of the explanatory variables (x_i and x_j), can be written as follows:

$$\frac{t_i}{t_j} = \gamma = \exp\left(\sum_{i=1}^{n} \alpha_i x_i\right) \tag{16}$$

If the i-th and j-th units differ from each other in the values of only one variable x_k, and $x_{ik} - x_{jk} = 1$, then:

$$\frac{t_i}{t_j} = \gamma = \exp(\alpha_k) \tag{17}$$

2.3. Main Differences Between PH and AFT Models

Not for every duration distribution can we estimate both PH and AFT models. The exponential and Weibull models are the only parametric models that can be written with either a PH or AFT specification. For the Gompertz distribution we can only apply the PH model. For the lognormal, loglogistic, and generalised gamma distributions only the AFT specification exists. The main differences between the PH and AFT models can be divided into two groups.

1. Basic assumptions of the models:

In the PH models, the effect of the influence of variables is multiplicative (proportional) to the hazard. In the AFT models, the effect of variable influence is multiplicative (proportional) to the duration.

2. Different interpretation of coefficients:

The PH models are used to compare hazards (the intensity of an event). AFT models are used to compare survival times. Differences in the interpretation of the coefficients of the PH and AFT models are presented in Figures 2 and 3. Each of them shows survival curves for two groups. On Figure 2 the moment t is fixed and for this moment the hazards $h_1(t)$ and $h_2(t)$ are determined. On the basis of Figures 1 and 2 we obtain:

$$HR = \frac{h_2(t)}{h_1(t)} = \frac{|tt_1|}{|tt_2|} \tag{18}$$

where $|tt_i|$ is the length of the line segment between points t and t_i for i = 1, 2.

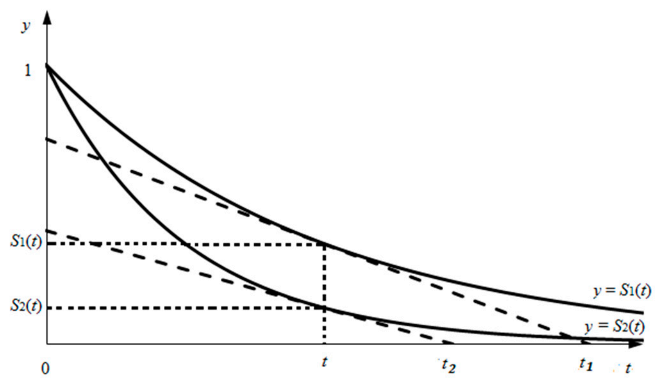

Figure 2. Plots of survival functions for two groups and HR interpretation.

Figure 3 presents the fixed probability of survival $S(t)$, for which the corresponding survival times t_1 and t_2 for groups are determined. On the basis of Figures 1 and 3 we obtain:

$$\gamma = \frac{|0t_2|}{|0t_1|} \tag{19}$$

where $|0t_i|$ is the length of the line segment between points 0 and t_i for i = 1, 2.

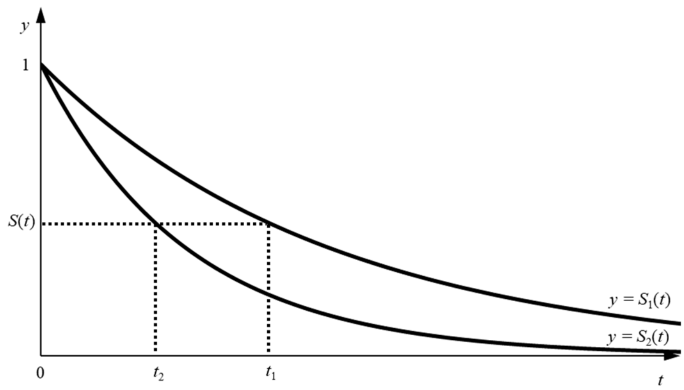

Figure 3. Plots of survival functions for two groups and the interpretation of the γ coefficient.

2.4. Application of Survival Analysis Methods in the Study of Socio-Economic Phenomena

Survival analysis in labour market research began to be used in the 1980s and 1990s. The aim of these analyses was to properly address the activities counteracting long-term unemployment.

Meyer (1990) analysed the impact of the level and duration of unemployment benefits on the duration of unemployment. Using the Kaplan-Meier estimator and other survival analysis models, he showed that the probability of remaining unemployed increased at the end of the period of receiving benefits. Han and Hausman (1990) used the parametric model of proportional hazards and the competing risk model to examine the duration of unemployment until the moment of leaving for a "new job" and the moment of leaving for a "previous job" (recall). They showed that the fact of receiving the benefit had a significant impact on the duration of unemployment, with its impact on hazard in the case of starting a new job was stronger than in the case of returning to the previous job. Research on the duration of unemployment for women and men, with particular emphasis on their marital status, was conducted by Marcassa (2012). Individual data on the unemployed in France in the years 1991–2002 allowed for the construction and analysis of Kaplan–Meier estimators. Married men left unemployment more quickly than unmarried men, and the opposite was true for women. Married women were less likely to quickly start employment than single women. Labour market research using the Kaplan–Meier estimator and the Cox proportional hazards model was conducted in Romania (Babucea & Danacica, 2007; Ciucă & Matei, 2011; Kavkler et al., 2009). Among other factors, the impact of gender, age, education, marital status, the number of children and the place of residence during the time of looking for a job were analysed. The results of the study indicated a large regional variation in the time of leaving unemployment. Women started employment faster than men, but after taking into account marital status, married men were the fastest at finding work, while married women had the least chance. The chance of leaving unemployment increased with education and decreased with age. Not only people with higher education, but also those with vocational education were in a good situation on the labour market in Romania. Tansel and Taşçı (2010) studied the duration of unemployment in Turkey in the years 2000 and 2001. To assess the differences in the time of leaving unemployment, a non-parametric Turnbull estimator was used. Landmesser (2007) applied competing risk models to describe the conditional probability of transition between employment and inactivity and determined the impact of age, education level, place of residence, and type of company ownership on the individual duration of employment.

Other socio-economic phenomena can also be analysed using survival analysis methods: economic activity of the population (Landmesser, 2013), poverty dynamics in urban and rural households (Sączewska-Piotrowska, 2015), credit risk (Matuszyk, 2017; Wycinka, 2019), and probability and intensity of decline and increase in stock prices on the capital market (Bieszk-Stolorz & Dmytrów, 2021). The range of applications of survival analysis methods is very wide and includes political science. These methods have been used in the study of international relations and are used to analyse the duration of post-war peace, civil wars and alliances (Box-Steffensmeier & Zorn, 2001; Box-Steffensmeier et al., 2003; Box-Steffensmeier & Jones, 2004).

3. Symmetry of Coefficients in PH and AFT Models for Exponential Distribution of Survival Time *T*

In the case of parametric survival models, the survival time T follows a specific distribution. For example, there are the following distributions: exponential, Weibull, loglogistic, log-normal, and Gompertz. If we assume that the time variable T is subject to an exponential distribution with the parameter λ, the basic survival functions take the form:

$$\mathrm{S}(t) = \exp(-\lambda t) \tag{20}$$

$$h(t) = \lambda \tag{21}$$

$$h_0(t) = 1 \tag{22}$$

where $\lambda > 0$ and $t \geq 0$.

In order to determine the duration of the phenomenon, Formula (20) should be transformed to determine the time t. After logarithmisation, we obtain:

$$\ln \mathrm{S}(t) = \ln(\exp(-\lambda t)) \tag{23}$$

Using the properties of logarithms, we have:

$$\ln \mathrm{S}(t) = -\lambda t \tag{24}$$

that is:

$$t = -\ln \mathrm{S}(t) \frac{1}{\lambda} \tag{25}$$

In the PH model, the λ coefficient is reparametrised and then defined by the formula:

$$\lambda = \exp\left(\beta_0 + \sum_{i=1}^{n} \beta_i x_i\right) \tag{26}$$

In the AFT model, the coefficient $1/\lambda$ is parameterised. Let it have the form:

$$\frac{1}{\lambda} = \exp\left(\alpha_0 + \sum_{i=1}^{n} \alpha_i x_i\right) \tag{27}$$

The exponential model satisfies both the PH and AFT assumptions. Formula (27) can be written after transformation as:

$$\lambda = \exp\left(-\alpha_0 - \sum_{i=1}^{n} \alpha_i x_i\right) \tag{28}$$

From Formulas (26) and (28), we can read the relationship between the coefficients β_i in the exponential PH model and the coefficients α_i in the exponential AFT model for $i = 1, 2, \ldots, n$. This relationship is as follows:

$$\beta_i = -\alpha_i \tag{29}$$

A thorough justification and derivation of relation (29) was carried out by Machin et al. (2006, pp. 102–103) and Collett (2023, pp. 203, 242–243). This derivation concerns the relationship between the parameters of the PH and AFT models for the Weibull distribution. This distribution for the shape parameter equal to 1 is an exponential distribution.

A binary relation R over a set X is symmetric (Biggs, 2002) if:

$$\forall_{a,b \in X}(aRb \iff bRa) \tag{30}$$

where the notation aRb means that $(a, b) \in R$. This means that R is symmetric, by which we mean that whenever aRb, then also bRa (Enderton, 1977).

The relation between the coefficients defined by Formula (29) satisfies the condition (30), i.e., it is symmetric. The consequence of this relationship is the interpretation of the coefficients. If $\beta_i > 0$, then $\alpha_i < 0$. Conversely, if $\beta_i < 0$, then $\alpha_i > 0$. The coefficients of the PH and AFT exponential models have the same magnitude, but have the opposite signs. The change in sign makes sense because the PH format uses covariates to model the hazard rate whereas the AFT format uses covariates to model the survival times. If the risk or hazard increases then the average survival time decreases (Machin et al., 2006). The relationship given by the Formula (29) also implies a relationship between HR and γ coefficients. Using Formulas (6) and (16) for the exponential distribution, we obtain:

$$HR = \frac{1}{\gamma} \tag{31}$$

Formula (31) implicates the following differences in the interpretation of the coefficients HR and γ:

1. If γ > 1, then the time to the event is higher than in the reference group. In this case, HR < 1, i.e., the intensity of the event, is lower compared to the reference group.
2. If γ < 1, then the time to the event is lower than in the reference group. In this case, HR > 1, i.e., the intensity of the event is higher than in the reference group.
3. In the case, when γ = HR = 1, then the time to the incident and the intensity of the incident are the same as in the reference group.

In the case of exponential PH and AFT models, there is an equality between the parameters describing the quality and fit of the model (Likelihood-ratio test, Akaike's Information Criterion, Log-Likelihood Function, Bayesian Information Criterion) and between the standard errors of the coefficients' estimators of both models. The equality is also a symmetric relation, so one can speak of symmetry between these parameters. The exponential PH model is valid if the exponential AFT model is valid (Olive, 2003).

4. Analysis of Duration in Registered Unemployment—An Empirical Example

The empirical example has been based on the data from the Poviat Labour Office in Szczecin (Poland), in order to illustrate the theoretical considerations presented earlier. The duration of registered unemployment was analysed. The data includes the date of registration, date of deregistration, gender, level of education, age and seniority of the deregistered person, as well as information about the reason for deregistration. These were individual data on 8878 people deregistered from the office in 2021. Deregistration due to starting work was considered to be an event ending full observation. Deregistration due to other reasons (retirement or disability pension, going abroad, starting studies, resigning from the office's mediation, etc.) was considered as the right-censored observation. There were 36.89% of such observations. In this case, the random variable T describes the time from the moment of registration with the office to the moment of starting work (the event ending the observation), i.e., the duration of being in registered unemployment.

Table 1 presents the analysed characteristics of the unemployed with the reference category marked by asterisks. The reference category does not appear in the model, while all other categories are replaced with dichotomous dummy variables. The third column of Table 1 lists the names of individual variables appearing in the PH and AFT models. Table 2 presents the structure of analysed unemployed persons.

Table 1. Characteristics of the unemployed and their categories.

Characteristic	Category	Variable Name
Gender	males *	–
	females	gender
Seniority	without seniority *	–
	with seniority	seniority
Level of education	at most lower secondary *	–
	basic vocational	education2
	general secondary	education3
	vocational secondary	education4
	higher	education5
Age	18–24 *	–
	25–34	age2
	35–44	age3
	45–54	age4
	55–59	age5
	60–64	age6

* Reference category.

Table 2. Structure of the unemployed persons.

Characteristic	Category	Total Size	Starting Work	Right-Censored Observations
Gender	males *	4766	2765	2001
	females	4112	2856	1256
Seniority	without seniority *	3457	1892	1565
	with seniority	5421	3729	1692
Level of education	at most lower secondary *	1929	850	1079
	basic vocational	1648	943	705
	general secondary	1264	823	441
	vocational secondary	1647	1139	508
	higher	2390	1866	524
Age	18–24 *	948	586	362
	25–34	2598	1762	836
	35–44	2442	1631	811
	45–54	1620	1053	567
	55–59	587	360	227
	60–64	683	229	454
Total		8878	5621	3257

* Reference category.

Tables 3 and 4 present the results of coefficient estimation for the exponential PH and AFT models. The calculations were performed in Stata 12.0 software in the Survival analysis module. According to the theoretical considerations presented earlier, the estimators of the coefficients differ from each other only in signs, and the standard deviations are equal. The fit factors of the models are also equal. All parameters are statistically significant, and therefore the analysed characteristics affect the intensity of starting work (PH model) and the time of registration at the labour office (AFT model).

Table 3. Results of the estimation of the PH model coefficients for the exponential distribution of the duration of registered unemployment.

Variable	Coef.	Std. Err.	z	$p >$ ∣z∣	95% Conf. Interval
gender	−0.1884	0.0276	−6.83	0.000	[−0.2424, −0.1344]
seniority	0.0755	0.0305	2.48	0.013	[0.0158, 0.1352]
education2	0.3490	0.0484	7.22	0.000	[0.2543, 0.4438]
education3	0.3506	0.0499	7.03	0.000	[0.2529, 0.4483]
education4	0.5234	0.0460	11.37	0.000	[0.4332, 0.6136]
education5	0.7268	0.0435	16.72	0.000	[0.6416, 0.8121]
age2	−0.2793	0.0493	−5.66	0.000	[−0.3760, −0.1825]
age3	−0.4754	0.0509	−9.34	0.000	[−0.5751, −0.3757]
age4	−0.5887	0.0543	−10.84	0.000	[−0.6951, −0.4822]
age5	−0.6632	0.0695	−9.54	0.000	[−0.7994, −0.5269]
age6	−1.8166	0.0807	−22.51	0.000	[−1.9748, −1.6585]
Const.	−2.3460	0.0515	−45.54	0.000	[−2.4470, −2.2450]
LR chi2(11)	1188.07				
Log likelihood	−12,364.40				
Prob > chi2	0.0000				
AIC	24,752.8				
BIC	24,837.9				

Table 4. Results of the estimation of the AFT model coefficients for the exponential distribution of the duration of registered unemployment.

Variable	Coef.	Std. Err.	z	$p >$ ∣z∣	95% Conf. Interval
gender	0.1884	0.0276	6.83	0.000	[0.1344, 0.2424]
seniority	−0.0755	0.0305	−2.48	0.013	[−0.1352, −0.0158]
education2	−0.3490	0.0484	−7.22	0.000	[−0.4438, −0.2543]
education3	−0.3506	0.0499	−7.03	0.000	[−0.4483, −0.2529]
education4	−0.5234	0.0460	−11.37	0.000	[−0.6136, −0.4332]
education5	−0.7268	0.0435	−16.72	0.000	[−0.8121, −0.6416]
age2	0.2793	0.0493	5.66	0.000	[0.1825, 0.3760]
age3	0.4754	0.0509	9.34	0.000	[0.3757, 0.5751]
age4	0.5887	0.0543	10.84	0.000	[0.4822, 0.6951]
age5	0.6632	0.0695	9.54	0.000	[0.5270, 0.7994]
age6	1.8166	0.0807	22.51	0.000	[1.6585, 1.9748]
Const.	2.3460	0.0515	45.54	0.000	[2.2450, 2.4470]
LR chi2(11)	1188.07				
Log likelihood	−12,364.40				
Prob > chi2	0.0000				
AIC	24,752.8				
BIC	24,837.9				

We can evaluate the fit of the model by using the Cox-Snell residuals (Cox & Snell, 1968). If the model fits the data well, plotting the Cox-Snell residuals (values of the cumulative hazard estimated by the model) against the Nelson-Aalen estimator of cumulative hazard (a non-parametric estimate) would result in a straight line (Collett, 2023). The Nelson-Aalen estimator of the cumulative hazard function is a non-parametric estimate of the cumulative hazard rate and is defined by the formula:

$$H(t) = \sum_{t_i \le t} \frac{d_i}{n_i} \tag{32}$$

where:

d_i—number of events at the time t_i,

n_i—number of units at the time t_i.

Because the line obtained (coloured blue) follows the 45 degree line (coloured red) then we know that the model fits the data well (Figure 4).

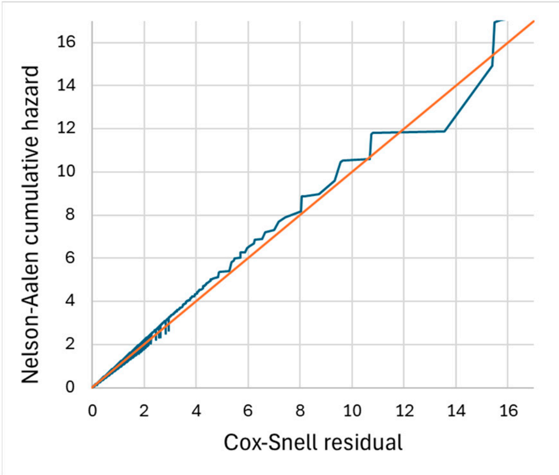

Figure 4. Fitting the model to empirical data using the Cox–Snell residuals.

Table 5 presents the coefficients' estimators of the PH and AFT models. In addition, on the basis of Formulas (7) and (17), HR and γ were determined. The results in Table 4 show the relationship (29) between the coefficients β_i and α_i. The plus sign next to the coefficients indicates that the intensity of exit from unemployment or the duration of unemployment is greater in a given group of people compared to the reference group. Similarly, the minus sign next to the coefficients indicates that the intensity of leaving unemployment or the duration of unemployment are lower in a given group of people compared to the reference group. More detailed information is provided by the HR and γ coefficients. They are inverse to each other, as indicated by the relationship (31). The empirical study shows that the intensity of starting work was 17.7% lower for women than for men. Unemployed people with work experience were 7.8% more likely to find work than those without work experience. The higher the level of education, the greater the intensity of leaving registered unemployment. This intensity was more than twice as high in the case of people with higher education compared to those with, at most, lower secondary education. The intensity of starting work decreased with the age of the unemployed person and for people aged 60+ it was almost 84% lower than for young people aged 18 to 24. Changes in the transition time from unemployment to work were in the opposite direction. For women this time was 20.7% longer than for men. People with work experience had a 7.3% shorter time to exit unemployment compared to people without work experience. The higher the level of education, the shorter the time of being registered at the office. People with higher education searched for a job for about 52% less time than people with, at most, lower secondary education. The time of transition from unemployment to work increased with the age of the unemployed person. The oldest people (i.e., aged 60+) had more than twice as long this time as the youngest people (i.e., aged 18–24). The results indicate that registered young people started work more often and faster than older people. In the light of the study, this was the age group with the best indicators. Individual data show that middle-aged people in 2021 were more likely to give up co-operation with labour offices and less likely to accept jobs offered.

Table 5. Results of estimation of PH, AFT model coefficients and HR, γ parameters for the exponential distribution of the duration of registered unemployment.

Variable	PH Model Coefficients (β_i)	AFT Model Coefficients (α_i)	PH Model HR ($\exp(\beta_i)$)	AFT Model γ ($\exp(\alpha_i)$)
gender	−0.1884	0.1884	0.8283	1.2073
seniority	0.0755	−0.0755	1.0784	0.9273
education2	0.3490	−0.3490	1.4177	0.7054
education3	0.3506	−0.3506	1.4199	0.7043
education4	0.5234	−0.5234	1.6877	0.5925
education5	0.7268	−0.7268	2.0685	0.4834
age2	−0.2793	0.2793	0.7563	1.3221
age3	−0.4754	0.4754	0.6216	1.6087
age4	−0.5887	0.5887	0.5551	1.8016
age5	−0.6632	0.6632	0.5152	1.9409
age6	−1.8166	1.8166	0.1626	6.1512
Const.	−2.3460	2.3460	0.0958	10.444

5. Discussion and Conclusions

This paper investigates the relationship between the coefficients of the PH and AFT models in the context of an exponential duration distribution. Theoretical considerations contributed to answers to both previously asked research questions:

Answer to Q1: In the case of an exponential distribution of survival time, there is a symmetric relation between the coefficients of the PH and AFT models.
Answer to Q2: The symmetric relation between the coefficients of the PH and AFT models (for exponential survival) indicates the opposite direction of the effects of the variables on hazard and the duration of the phenomenon.

The signs of the coefficients in the exponential AFT model are opposite to the signs for the exponential PH model (Patel et al., 2006; Qi, 2009). This relationship is illustrated in the article with an example from the labour market. To interpret the coefficients, their exponential form, i.e., the hazard ratios (HR) and the acceleration factors (γ), were used. In the case of an exponential distribution of duration, HR and γ are inverse to each other. A negative effect on unemployment duration implies a positive effect on the hazard rate from unemployment and vice versa. The obtained results of the empirical study confirm the research conducted by Dendir (2006). A greater survival time indicates a longer time before an event occurs, a greater hazard rate indicates a shorter time before an event would occur (Gelfand et al., 2016).

The AFT model is rarely used to analyse survivorship data, but offers a potentially useful statistical approach. This approach is based on the survival curve and not on the hazard function. The AFT model's deceleration rates provide a more intuitive measure of treatment effect than the hazard ratio and are robust to departures from modelling assumptions (Swindell, 2009). The results of the AFT models may be easier to interpret as the covariate effects are directly expressed in terms of time ratio (Khanal et al., 2014). There is also a dilemma as to which of the PH models is better: the Cox regression model (semiparametric) or parametric model? Research indicates that the parametric model is more informative than the Cox model (Bradburn et al., 2003). Research conducted by Pourhoseingholi et al. (2007) indicated that the Cox PH model and exponential PH model were similarly good models in multivariate analysis and yielded the same conclusions in univariate analysis. Exponential distribution is simpler for use and interpretation because it is independent of the length of time (Rashidi & Mohammadian, 2011). However, it seems that there is no single model that is much better than the others.

The considerations presented in the article are an introduction to broader research on the relationships between the coefficients of the PH and AFT models in the case of durations other than exponential. A certain limitation here is that not both models exist for every distribution. The condition of having opposite signs for the coefficients of these models is certainly satisfied. However, the absolute values of the coefficients are different, which implies a relation other than a symmetric relation.

Additionally, the analysis of the duration of unemployment has certain social implications. The study of the duration of unemployment has made it possible to identify groups of people who are most at risk of unemployment. These are people who have had a long time searching for a job and leave the employment office register less intensively. The study presented in the article might indicate that these are women, people with no professional experience, older people and those with a lower level of education. State social policy should be targeted at such groups of people. It is important that studies of this type cover the smallest possible territorial units. This is the only way to identify areas at risk of high unemployment. Empirical analysis was carried out for data from year 2021. Formally, it was still a period of ongoing pandemic. However, the indicators pointed to an improvement in the situation and the return of positive trends in the labour market in Poland. According to the Statistics Poland, from February 2021 to December 2021, the registered unemployment rate steadily decreased and reached 5.8% (Statistics Poland, n.d.). Thus, a level similar to the pre-pandemic level had been reached. Recognising the problems of young people, public employment services in Poland are taking action to activate them and support them in entering the labour market. Young people are among the main addressees of activation programmes carried out by labour offices. Data for the entire EU show that in 2021 the youth unemployment rate in the 15–29 age group in Poland was lower than the European Union average (7.2% vs. 13%). This is certainly the result of the activation measures implemented. They should be continued on a regular basis. In Poland, the Youth Guarantee programme has been implemented since 2014. Activities activating young people carried out by labour offices are one of those undertaken in this programme. The research confirms the need to continue support programs for young job seekers.

One of the main research limitations is access to individual data. Statistical offices only have aggregated data, which prevents the use of most methods from the area of survival analysis. In Poland, in order to obtain individual data, research cooperation with Poviat labour offices is necessary. Only these offices have collected individual data. Another limitation of this study is the fact that it concerns registered unemployment.

The theoretical goal of future research will be to try to find the relationship between the coefficients of the PH and AFT models for other duration distributions. The empirical objective of future research will be to assess the intensity of (PH models) and time to exit unemployment (AFT models) using semiparametric and parametric models.

Funding: This research received no external funding.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study or due. Requests to access the datasets should be directed to the Poviat Labour Office in Szczecin (Poland).

Conflicts of Interest: The author declares no conflicts of interest.

References

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

Addison, J. T., & Portugal, P. (1998). Some specification issues in unemployment duration analysis. *Labour Economics*, *5*(1), 53–66. [CrossRef]

Babucea, A. G., & Danacica, D. E. (2007). *Using kaplan—Meier curves for preliminary evaluation the duration of unemployment spells, MPRA paper 7854*. University Library of Munich. Available online: https://core.ac.uk/download/pdf/7296501.pdf (accessed on 14 November 2024).

Basha, L., & Gjika, E. D. (2022). Accelerated failure time models in analyzing duration of employment. *Journal of Physics: Conference Series*, *2287*, 012014. [CrossRef]

Bieszk-Stolorz, B. (2015). Geometric method of determining hazard for the continuous survival function. *Folia Oeconomica Stetinensia*, *15*, 22–33. [CrossRef]

Bieszk-Stolorz, B. (2017). Cumulative incidence function in studies on the duration of the unemployment exit process. *Folia Oeconomica Stetinensia*, *17*, 138–150. [CrossRef]

Bieszk-Stolorz, B., & Dmytrów, K. (2021). Evaluation of changes on world stock exchanges in connection with the SARS-CoV-2 pandemic. survival analysis methods. *Risks*, *9*(7), 121. [CrossRef]

Bieszk-Stolorz, B., & Markowicz, I. (2022). The impact of the COVID-19 pandemic on the situation of the unemployed in Poland. A study using survival analysis methods. *Sustainability*, *14*(19), 12677. [CrossRef]

Biggs, N. L. (2002). *Discrete mathematics* (2nd ed.). Oxford University Press.

Box-Steffensmeier, J. M., & Jones, B. S. (2004). *Event history modeling a guide for social scientists*. Cambridge University Press.

Box-Steffensmeier, J. M., & Zorn, C. J. W. (2001). Duration models and proportional hazards in political science. *American Journal of Political Science*, *45*, 972–988. [CrossRef]

Box-Steffensmeier, J. M., Reiter, D., & Zorn, C. J. W. (2003). Nonproportional hazards and event history analysis in international relations. *The Journal of Conflict Resolution*, *47*, 33–53. [CrossRef]

Bradburn, M., Clark, T. G., Love, S., & Altman, D. (2003). Survival analysis part II: Multivariate data analysis—An introduction to concepts and methods. *British Journal of Cancer*, *89*, 431–436. [CrossRef]

Calabuig, J. M., Jiménez-Fernández, E., Sánchez-Pérez, E. A., & Manzanares, S. (2021). Modeling hospital resource management during the COVID-19 pandemic: An experimental validation. *Econometrics*, *9*(4), 38. [CrossRef]

Ciucă, V., & Matei, M. (2011). Survival rates in unemployment. *International Journal of Mathematical Models and Methods in Applied Sciences*, *5*, 362–370.

Collett, D. (2023). *Modelling survival data in medical research* (4th ed.). CRC Press.

Cox, D. R., & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, *30*, 248–265. [CrossRef]

Dendir, S. (2006). Unemployment duration in poor developing economies: Evidence from urban Ethiopia. *The Journal of Developing Areas*, *40*, 181–201. [CrossRef]

Doğan, F. İ. (2019). Unemployment benefits and unemployment duration in France and Poland. *Marmara Journal of European Studies*, *27*, 191–216. [CrossRef]

Doğan, F. İ. (2020). Social transfers and unemployment duration: An empirical evidence of the EU and Turkey. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *42*, 247–264. [CrossRef]

Enderton, H. B. (1977). *Elements of set theory*. Academic Press.

Faruk, A. (2018). The comparison of proportional hazards and accelerated failure time models in analyzing the first birth interval survival data. *Journal of Physics: Conference Series*, *974*, 012008. [CrossRef]

Gelfand, L. A., MacKinnon, D. P., DeRubeis, R. J., & Baraldi, A. N. (2016). Mediation analysis with survival outcomes: Accelerated failure time vs. proportional hazards models. *Frontiers in Psychology*, *7*, 423. [CrossRef]

Grzenda, W. (2023). Estimating the probability of leaving unemployment for older people in Poland using survival models with censored data. Statistics in Transition. *New Series*, *24*, 241–256. [CrossRef]

Han, A., & Hausman, J. A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, *5*, 1–28. [CrossRef]

Kavkler, A., Danacica, D. E., Babucea, A. G., Bicanic, I., Bohm, B., Tevdovski, D., Tosevska, K., & Borsic, D. (2009). Cox regression models for unemployment duration in Romania, Austria, Slovenia, Croatia, and Macedonia. *Romanian Journal of Economic Forecasting*, *10*, 81–104.

Khamis, A., Hamat, C. A. C., & Abdullah, M. A. A. (2020). Modeling students' performance using Cox and parametric accelerated failure time models. *Scientific Research Journal*, *8*, 44–49. [CrossRef]

Khanal, S. P., Sreenivas, V., & Acharya, S. K. (2014). Accelerated failure time models: An application in the survival of acute liver failure patients in India. *International Journal of Science and Research*, *6*, 161–166.

Khanal, S. P., Sreenivas, V., & Acharya, S. K. (2019). Comparison of Cox proportional hazards model and lognormal accelerated failure time model: Application in time to event analysis of acute liver failure patients in India. *Nepalese Journal of Statistics*, *3*, 21–40. [CrossRef]

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis techniques for censored and truncated data* (2nd ed.). Springer.

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

Lalive, R. (2008). How do extended benefits affect unemployment duration? A regression discontinuity approach. *Journal of Econometrics*, *142*, 785–806. [CrossRef]

Landmesser, J. M. (2007). Wykorzystanie modeli ryzyka konkurencyjnego w celu szacowania czasu trwania aktywności zawodowej. *Prace Naukowe Akademii Ekonomicznej im. Oskara Langego we Wrocławiu. Taksonomia*, *14*, 542–550.

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

Landmesser, J. M. (2010). A dynamic approach to the study of unemployment duration. *Metody Ilościowe w Badaniach Ekonomicznych*, *11*, 212–222.

Landmesser, J. M. (2013). *Wykorzystanie metod analizy czasu trwania do badania aktywności ekonomicznej ludności w Polsce*. Wydawnictwo SGGW.

Machin, D., Cheung, Y. B., & Parmar, M. (2006). *Survival analysis: A practical approach* (2nd ed.). John Wiley & Sons.

Marcassa, S. (2012). *Unemployment duration of spouses: Evidence from France. THEMA working papers 2012-31*. Université de Cergy-Pontoise.

Matuszyk, A. (2017). *Credit scoring*. CeDeWu.

Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica*, *58*, 757–782. [CrossRef]

Mishra, N., & Misra, S. (2022). Comparison of Cox proportional hazard model and accelerated failure time (AFT) models: An application to under-five mortality in Uttar Pradesh. *Indian Journal of Applied Research*, *12*, 83–87. [CrossRef]

Moffitt, R. (1985). Unemployment insurance and the distribution of unemployment spells. *Journal of Econometrics*, *28*, 85–101. [CrossRef]

Nickell, S. (1979). Estimating the probability of leaving unemployment. *Econometrica*, *47*, 1249–1266. [CrossRef]

Olive, D. J. (2003). Available online: http://parker.ad.siu.edu/Olive/survbook.pdf (accessed on 12 July 2024).

Orbe, J., Ferreira, E., & Núñez-Antón, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in Medicine*, *21*, 3493–3510. [CrossRef]

Pang, M., Platt, R. W., Schuster, T., & Abrahamowicz, M. (2021). Spline-based accelerated failure time model. *Statistics in Medicine*, *40*, 481–497. [CrossRef] [PubMed]

Patel, K., Kay, R., & Rowell, L. (2006). Comparing proportional hazards and accelerated failure time models: An application in influenza. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry*, *5*, 213–224. [CrossRef] [PubMed]

Pourhoseingholi, M. A., Hajizadeh, E., Dehkordi, B. D. M., Safaee, A., Abadi, A., & Zali, M. R. (2007). Comparing Cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pacific Journal of Cancer Prevention*, *8*, 412–416. [PubMed]

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. [CrossRef]

Qi, J. (2009). *Comparison of proportional hazards and accelerated failure time models* [Ph.D. Thesis, University of Saskatchewan].

Rashidi, T. H., & Mohammadian, A. (Kouros). (2011). Parametric hazard functions: Overview. *Transportation Research Record*, *2230*, 48–57. [CrossRef]

Sączewska-Piotrowska, A. (2015). Poverty duration of households of the self-employed. *Econometrics. Ekonometria*, *1*, 44–55. [CrossRef]

Saikia, R., & Barman, M. P. (2017). A review on accelerated failure time models. *International Journal of Statistics and Systems*, *12*, 311–322.

Statistics Poland. (n.d.). Available online: https://stat.gov.pl/en/poland-macroeconomic-indicators/ (accessed on 16 November 2024).

Swindell, W. R. (2009). Accelerated failure time models provide a useful statistical framework for aging research. *Experimental gerontology*, *44*, 190–200. [CrossRef]

Tansel, A., & Taşçı, H. M. (2010). Hazard analysis of unemployment duration by gender in developing county: The case of Turkey. *Labour*, *24*, 501–530. [CrossRef]

Urbańczyk, D. M. (2020). Competing risks models for an enterprises duration on the market. *Folia Oeconomica Stetinensia*, *20*, 456–473. [CrossRef]

Wycinka, E. (2019). Competing risk models of default in the presence of early repayments. *Econometrics. Ekonometria. Advances in Applied Data Analytics*, *23*, 99–120. [CrossRef]