Modeling unemployment duration, determinants and insurance premium pricing of Malaysia: insights from an upper middle-income developing country

Wooi Chen Khoo1 · Kim Leng Yeah2 · Shun Yi Hong3

Received: 28 January 2021 / Accepted: 8 July 2022 / Published online: 2 August 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract
This paper investigates the behavior and determinants of unemployment in Malaysia, an upper middle-income country with relatively low unemployment but rising labor market mismatches, and derives a model-based insurance premium pricing schedule that broadens the country’s unemployment insurance coverage. The unemployment profile is best fitted with Burr distribution, and using survival analysis, the average unemployment duration is found to be the highest for the youths and it decreases gradually after age 20. The results of the Cox proportional hazards model show that age, gender, and education are significant explanatory variables of unemployment. While all age groups are statistically significant in explaining unemployment, the youngest cohort has the least likelihood of being hired, while the older groups are found to have a lower hazard ratio that could be attributed to work experience. The significant gender effect is also characterized by the female workforce taking a longer time to exit unemployment as shown by its lower hazard rate than male workers. Similar to findings for other countries, those with higher education have the highest chance of gaining employment. Fitting the estimated parameters to a standard unemployment insurance premium pricing framework based on the Capital Asset Pricing Model (CAPM), the analysis shows that the coverage of the existing national unemployment insurance scheme can be expanded to all income categories at lower premium prices. The widening of the coverage will contribute to the strengthening of the country’s social safety net in addition to reducing inequalities in the post-COVID-19 period.

Keywords Unemployment duration · Employability insurance scheme · Survival analysis

JEL classification C14 · C26 · J60 · J64

* Wooi Chen Khoo
wooichenk@sunway.edu.my

Extended author information available on the last page of the article
Introduction

Unemployment remains a primary and recurring concern in most countries regardless of development stage and unemployment level. The multifaceted issues range from massive job losses and anemic recovery following the 2007–2009 global financial crisis to the high youth unemployment in the middle-eastern countries that has been cited as a key factor for the Arab Spring uprising. More contemporary issues include unequal access to job opportunities in low unemployment economies and the potential large job displacements caused by automation and digital technologies and skills mismatches in middle and high-income countries.

An upper-middle-income country since the early 1990s, Malaysia’s unemployment rate has been relatively low and stable, averaging 3.3% over the last 2 decades. By contrast, the average rate for upper-middle countries is 5.6%, while that for high-income nations is higher at 6.8%. Notwithstanding the country’s low unemployment rate, the Malaysian government introduced national unemployment insurance in 2017. There is however limited statistical modeling of unemployment behavior both in Malaysia and the developing world.

This study fills the gap by estimating Malaysia’s unemployment distribution and duration, two crucial inputs in the design of unemployment insurance schemes. The study also examines key factors influencing unemployment. Despite the country’s low unemployment rate, a comparison of key determinants with those of countries with high unemployment rates will have significant policy implications in addressing unemployment issues that are common as well as unique to each country.

The objectives of the unemployment analysis are twofold. First, this paper provides a deeper insight into the patterns of the unemployment rate. The data analysis shows that the Burr distribution form is appropriate for the model fitting as it outperformed other positively skewed distribution functions. Second, the survival analysis via hazard function corresponding to Burr distribution is used to estimate unemployment duration which is a key input in pricing unemployment insurance premium. To motivate the enhancement of the country’s recently launched national unemployment insurance scheme, the study applies the derived statistical estimates to estimate a model-based unemployment insurance premium rates by wage classes.

In survival analysis, the hazard function defines instantaneous death. We apply this definition to explain unemployment duration. Unemployment duration is defined by the conditional probability of exiting unemployment, which means that an unemployed person will find employment at time t, given that at time t-1, the person was unemployed. This is illustrated by the hazard function or commonly known as hazard rate. This paper explores the relationship between two covariates. The hazard ratio is used to measure the relativity. If the ratio is equal to 1, there is an equal possibility of being unemployed at time t-1 and employed at time t. If it is more than 1, it means that the employment at time t has a higher chance to happen, while for a ratio lower than 1, the situation of being unemployed will remain longer than being employed. This statistical approach is employed to analyze the underlying unemployment situation in Malaysia and provide an empirical estimation.
of unemployment duration, determinants, and insurance premium pricing to support evidence-based policy decisions.

The rest of the paper is organized as follows. The relevant literature is reviewed in the section “Literature review”. The section “Data sources and description” describes the labor survey data sourced from the Department of Statistics Malaysia (DOSM). The section “Modeling and estimation methods” details the model and estimation methods for fitting unemployment rates and deriving unemployment duration using the Burr distribution function. This section also explores covariates relationships using Cox proportional hazard regression models and enumerates the construction of the unemployment insurance scheme based on the Equivalence Principle. The section “Results and discussion” discusses the estimation results and the section “Conclusion” concludes.

Literature review

Unemployment is intertwined with a country’s macroeconomic management where it is commonly aimed at securing sustained growth, price stability, and full employment. Unemployment below the natural rate reflects underutilization of a country’s human resources and, consequently, lower gross domestic product (GDP) growth potential. Besides a loss of tax revenue to the government, unemployment is also associated with social and political problems such as crime, vandalism, and popular uprisings. Among the earliest to relate unemployment to macroeconomic impact is Malinvaud (1982) who suggested a national unemployment insurance system to minimize unemployment risk. Beenstock (1985) enhanced the unemployment insurance scheme to further consider an individual’s characteristics, specifically, working behavior. He proposed that the premium paid should fairly reflect different risk groups. He then constructed a competitive pricing model for unemployment insurance. The model was subsequently enhanced with stochastic properties by Blake and Beenstock (1988), taking inflation into the consideration as a factor in determining premium pricing. Other studies such as Pollak (2012) established the relationship between human capital depreciation and unemployment insurance policy.

Some existing study is found applying survival analysis to analyze the unemployment spells and the measurement of the effects of the unemployment condition. Lakuma et al. (2016) applied survival analysis for a regional unemployment study in Uganda. They reported that the cohort with higher education surprisingly gains employment slower than those with lower education across the region. Women have a lower possibility to exit unemployment and the chances of being employed increase according to age. While Malaysia is advisable to focus on skill development for the age cohort of 19–21, in Uganda, it is suggested that the investment in training skills should be given to the cohort younger than 25 years old. Bajram (2013) found that re-employment happens in the first 59 months with the frailty model in survival analysis. Bowers (1980) applied transition probabilities to estimate the unemployment duration. In the transition probabilistic model, he defined three states—employed, unemployed, and not in the labor force. Salant (1977) discussed the Pareto distribution, while Chuang and Yu (2010) proposed Weibull
distribution for estimating the unemployment duration in Taiwan’s labor force. In an enhancement to the estimation technique, Simwa et al. (2016) applied a mixture of the Weibull distribution to accommodate the heterogeneity in the US unemployment data. The non-parametric Kaplan–Meier estimator has been used to estimate the unemployment duration.

Given that the existing literature is limited and focused primarily on developed economies, this paper seeks to address the lack of statistical modeling in unemployment studies for developing countries. This research provides an opportunity to extend the coverage using Malaysia as a case study. The country’s status as an upper-middle-income country with a labor market characterized by low unemployment but a high number of unskilled foreign workers could shed light on the structural characteristics influencing unemployment behavior. As unemployment is an uncertain event, one can consider applying probabilistic models to measure it numerically. Unemployment duration, defined as the length of time in which a person seeking a job remains unemployed, therefore can be modeled by conditional probabilities. We applied a parametric approach in this study where the Burr distribution is fitted to the data set given its superiority over other forms in terms of flexibility.

While there are limited probabilistic model-based studies on unemployment, several researchers have applied other estimation approaches. Aitkin and Healey (1985) applied logit models to unemployment rates. Pereira et al. (2018) focused on the application of Bayesian hierarchical models to the Portuguese labor force survey data and made comparisons with the multinomial model and Beta model. Time-series analysis and forecasting are popular approaches to modeling unemployment. Holmes et al. (2013) applied a high-dimensional vector autoregressive analysis to study the behavior of unemployment rates in the US over time and across space. Christos (2005) forecasted the UK unemployment rate with 30 years data set. In time-series modeling, the researchers found that the autoregressive conditional heteroscedasticity models yield the best forecasts. Claveria (2019) forecasts the unemployment rate in eight European countries with autoregressive integrated moving average models. Kreiner and Duca (2020) show that the machine learning approaches outperform time-series models to provide better forecasts at shorter horizons. When applied to actual data, his results indicate that the machine learning model can identify the turning points of the unemployment rate. In comparison, the survival analysis remains the advantage to analyze the unemployment issue as it can measure the effect of the covariates.

Given Malaysia’s low unemployment rate, a comparison of unemployment duration and the key determinants of unemployment with those of countries with high unemployment rates will have significant policy implications in addressing both unemployment issues that are common as well as unique to each country. Most researchers focus on the analysis of the determinants such as age, gender, education attainment level, and region. The east European countries, such as Romania, Bulgaria, Hungary, and Slovenia, are of interest, because they have a high unemployment rate among the countries in the region. Romania has an average unemployment rate of 6.64% from 1999 to 2018, among the lowest in the region. Bulgaria on the other hand recorded the highest average unemployment rate at 10.94% for the past 20 years. Hungary and Slovenia have recorded 7.36% and 6.95%, respectively.
Danacica and Paliu-Popa (2017) investigated the determinants of unemployment in the East European countries where unemployment reached its highest level in 2011 following a recession. Focusing on explanatory variables such as gender and education, the authors found that in Romania, women experienced a 14% lower exit rate than men. In other words, it is more challenging for women to leave their unemployment state. The study also found that 22.8% of the unemployed are educated at the high school level, 9.1% with higher education, and 10.8% with unknown education. Those with a college education have the highest chance to find employment, followed in declining order by those with university education, high school, and unknown education. Similar results have been found by Ciuca and Matei (2011) where those with higher education have the lowest probability of unemployment. The study also revealed the 36–45 age cohort faced the most difficulty in finding employment, while it is easier for those aged 15–25 years. The study revealed that among the key variables tested, education has the highest power in explaining unemployment. In contrast, our finding for Malaysia shows education and gender as having an equal impact on unemployment with both groups facing the same possibility to exit unemployment.

Kavler et al. (2009) reported the unemployment exit rate is 32% higher for men than for women in Croatia followed by Austria (29.9%), Slovenia (20.8%), and Romania (16.3%). Malaysia shares similar features whereby women take longer to exit unemployment and face more hurdles in finding jobs. The study on East European countries found older workers to be disadvantaged in finding employment in contrast to Malaysia where the unemployment duration is less than 30 days for the age group greater than 50 years. Another difference found in our study is that in Malaysia the unemployed with higher academic qualifications find it easier to leave unemployment as evidenced by the shorter unemployment duration for those with master’s degrees compared to those with bachelor’s degrees.

There are various other studies on the influence of factors such as age, gender, race, living region, and education attainment. In some other European countries. D’Agostino (2000) found that gender has an impact on unemployment in Belgium, Greece, France, Spain, Denmark, and Portugal, but not Italy. In Italy, gender does not exert much influence on unemployment. The effect of location or place of living area is prominent in Italy where those living in the south have greater difficulties in finding jobs. The study also shows the age variable has the least impact in Italy, the United Kingdom (UK), and Spain, but in Portugal, France, and Denmark the older cohorts find more difficulties in exiting unemployment. Higher education levels shorten the unemployment duration in the UK, Belgium, and Ireland, while in Greece and Spain, the educational level does not have a strong effect on the expected duration. On the other hand, Baussola and Mussida (2017) and Mussida and Fabrizi (2014) reported contradictory findings. They found gender has an impact on unemployment outflow.

In the UK, the analysis of the British Household Panel Survey by Boheim and Taylor (2000) indicates that female exit rates are longer than men. Women need 9 months to leave unemployment, whereas the men take only 6 months. Another key finding is that men aged over 25 are entering unemployment with an increased education level. Highly educated women are significantly more likely to enter part-time
employment from unemployment than men. Most studies report that women face greater challenges in looking for a job, the exception being the study by Hoffman (1991) who found white females in the US are more likely to leave unemployment, particularly the higher educated white females who have better opportunities to secure jobs. Mirroring the finding for European countries, the US study found that older persons are less likely to exit unemployment.

There is a paucity of similar studies for developing countries where unemployment is expected to vary widely in scope and intensity depending on an individual country’s economic performance, population growth, and labor market characteristics. Slow-growing economies face higher unemployment, while the more dynamic ones are likely to face different unemployment challenges.

**Data sources and description**

The data are sourced from the Labor Force Surveys conducted by the Department of Statistics once every 3 years for the period 1997–2015. There are 15 variables enumerated according to the International Standard Classification of Education (ISCED) classification that is analyzed, covering a total of 427,714 sample points. Table 1 provides the labor force count and participation rate by gender. The key variables are highlighted below:

*Age:* A key variable in the employability modeling and analysis, the working age is based on the universally adopted range of 15–64 years, while youths are defined as those in the 15–24 age group.

*Employment status:* There are three categories in the labor force summary (see Supplementary Table 10 for the definition). The three labor force categories are Employed (code 1), Unemployed (code 2), and Outside Labor Force (code 3).

*State:* There are 13 states and 3 separately coded federal territories resulting in a total of 16 localities (see Supplementary Table 11 for the state label).

*Educational attainment:* There are four categories in the educational attainment variable as defined in Supplementary Table 12. The education system is divided into preschool education, primary, secondary, pre-university, and tertiary education. To enable a more focused analysis this study employs four categories: no formal education (code 1), primary (code 2), secondary (code 3), and tertiary (code 4) education.

*Ethnicity:* The main ethnic groups are Malay (code 1), other Bumiputra (code 2), Chinese (code 3), Indians (code 4), and Others (code 5).

*Gender:* An equal proportion of males (code 1) and females (code 2) are drawn from the survey. Table 1 shows the labor force participation by gender.

| Code | Classification      | Sample size | Male    | Female   |
|------|---------------------|-------------|---------|----------|
| 1    | Employed            | 260,173     | 38.4%   | 22.4%    |
| 2    | Unemployed          | 8174        | 1.2%    | 0.7%     |
| 3    | Outside labor force | 159,367     | 10.4%   | 26.9%    |
Highest certificate obtained: There are nine categories under this explanatory variable, comprising, primary school achievement test (UPSR) or equivalent (code 1), lower secondary evaluation (PT3/PMR/SRP) or equivalent (code 2), the Malaysian certificate examination (SPM) or equivalent (code 3), Malaysian higher school certificate (STPM) or equivalent (code 4), certificate less than 6 months (code 5), diploma (code 6), degree (code 7), no certificate (code 8), and not applicable (code 9).

Field of study: This variable consists of nine fields of study. They are general programs (code 1), education (code 2), humanities and arts (code 3), social sciences, business and law (code 4), science, mathematics and computing (code 5), engineering, manufacturing and construction (code 6), agriculture and veterinary (code 7), health and welfare (code 8) and services (code 9). Figure 1 shows the unemployment rate for the cohort with tertiary education.

The data have been filtered and cleaned before the analysis. The descriptive statistics on the mean, median, first quartile, third quartile, range, standard deviation, and skewness of the unemployment rate for each variable are shown in Table 2. The standard deviations are small. It means that the data are not likely to disperse away from the mean and so it is reliable. The findings show that the explanatory variable has an impact on unemployment. The effects of these explanatory variables are tested with the Cox proportional hazards model.

Modeling and estimation methods

This section describes the selection of the distribution specification for the unemployment rate followed by the application of hazard function and hazard ratio in survival analysis to estimate unemployment duration. To investigate the key factors influencing unemployment, the Cox proportional hazards modeling technique
Table 2  Descriptive statistics of the unemployment rate for each variable

| Variable                     | Code | UR (%) | Statistic (%) |
|------------------------------|------|--------|---------------|
|                              |      |        | Max           | Min     | Q1    | Median | Q3    | Mean     | Range   | SD²     |
| Gender                       | 1    | 2.38   | 5.08         | 0.00    | 0.99  | 2.61   | 3.68  | 2.44     | 5.08    | 1.36    |
|                              | 2    | 1.44   | 3.60         | 0.00    | 0.55  | 1.99   | 2.64  | 1.71     | 3.60    | 1.11    |
| State (location)             | 01   | 1.73   | 5.01         | 0.00    | 0.78  | 2.85   | 3.90  | 2.50     | 5.01    | 1.62    |
|                              | 02   | 1.89   | 1.79         | 0.00    | 0.30  | 0.67   | 1.04  | 0.69     | 1.79    | 0.44    |
|                              | 03   | 1.53   | 1.65         | 0.00    | 0.29  | 0.47   | 0.67  | 0.51     | 1.65    | 0.30    |
|                              | 04   | 0.81   | 0.77         | 0.00    | 0.14  | 0.31   | 0.47  | 0.32     | 0.77    | 0.20    |
|                              | 05   | 2.01   | 0.87         | 0.00    | 0.05  | 0.09   | 0.19  | 0.13     | 0.87    | 0.15    |
|                              | 06   | 1.63   | 0.72         | 0.00    | 0.00  | 0.05   | 0.09  | 0.06     | 0.72    | 0.07    |
|                              | 07   | 1.32   | 1.20         | 0.00    | 0.13  | 0.23   | 0.32  | 0.25     | 1.20    | 0.17    |
|                              | 08   | 1.83   | 7.06         | 0.00    | 0.97  | 2.44   | 4.17  | 2.72     | 7.06    | 1.89    |
|                              | 09   | 1.76   | 3.85         | 0.00    | 0.13  | 0.76   | 2.02  | 1.12     | 3.85    | 1.08    |
|                              | 10   | 1.88   | 1.03         | 0.00    | 0.12  | 0.31   | 0.52  | 0.33     | 1.03    | 0.25    |
|                              | 11   | 1.92   | 1.21         | 0.00    | 0.09  | 0.27   | 0.47  | 0.30     | 1.21    | 0.26    |
|                              | 12   | 3.26   | 1.16         | 0.00    | 0.10  | 0.28   | 0.42  | 0.30     | 1.16    | 0.24    |
|                              | 13   | 2.47   | 0.81         | 0.00    | 0.00  | 0.14   | 0.29  | 0.17     | 0.81    | 0.16    |
|                              | 14   | 1.80   | 0.81         | 0.00    | 0.08  | 0.17   | 0.33  | 0.23     | 0.81    | 0.20    |
|                              | 15   | 3.02   | 0.75         | 0.00    | 0.08  | 0.25   | 0.43  | 0.27     | 0.75    | 0.22    |
|                              | 16   | 1.37   | 1.08         | 0.00    | 0.09  | 0.20   | 0.36  | 0.27     | 1.08    | 0.24    |
| Ethnic group                 | 1    | 1.95   | 1.47         | 0.00    | 0.12  | 0.31   | 0.46  | 0.32     | 1.47    | 0.25    |
|                              | 2    | 2.65   | 0.89         | 0.00    | 0.00  | 0.11   | 0.23  | 0.16     | 0.89    | 0.17    |
|                              | 3    | 1.35   | 1.19         | 0.00    | 0.14  | 0.39   | 0.66  | 0.43     | 1.19    | 0.33    |
|                              | 4    | 2.31   | 0.93         | 0.00    | 0.06  | 0.19   | 0.39  | 0.23     | 0.93    | 0.20    |
|                              | 5    | 1.33   | 1.51         | 0.00    | 0.18  | 0.52   | 0.84  | 0.52     | 1.51    | 0.38    |
| Variable                               | Code | UR (%) | Statistic (%)         |
|---------------------------------------|------|--------|-----------------------|
|                                       |      |        | Max  | Min  | Q1   | Median | Q3   | Mean | Range | SD   |
| Education attainment                  | 1    | 0.95   | 1.25 | 0.00 | 0.13 | 0.35   | 0.53 | 0.36 | 1.25  | 0.28 |
|                                       | 2    | 1.04   | 1.57 | 0.00 | 0.06 | 0.14   | 0.26 | 0.21 | 1.57  | 0.27 |
|                                       | 3    | 2.14   | 0.56 | 0.00 | 0.00 | 0.05   | 0.12 | 0.08 | 0.56  | 0.11 |
|                                       | 4    | 2.36   | 0.22 | 0.00 | 0.00 | 0.05   | 0.09 | 0.05 | 0.22  | 0.06 |
| Highest certificate obtained          | 1    | 1.49   | 1.87 | 0.00 | 1.87 | 0.00   | 12.93| 27.24| 0.20  | 0.26 |
|                                       | 2    | 1.96   | 5.24 | 0.00 | 5.24 | 0.00   | 15.03| 35.59| 0.40  | 0.84 |
|                                       | 3    | 2.38   | 4.84 | 0.00 | 4.84 | 0.00   | 0.00 | 24.60| 0.32  | 0.77 |
|                                       | 4    | 2.60   | 1.91 | 0.00 | 1.91 | 0.00   | 0.00 | 7.30 | 0.09  | 0.23 |
|                                       | 5    | 3.47   | 1.47 | 0.00 | 1.47 | 0.00   | 0.00 | 7.27 | 0.08  | 0.19 |
|                                       | 6    | 2.00   | 1.81 | 0.00 | 1.81 | 0.00   | 0.00 | 10.60| 0.11  | 0.24 |
|                                       | 7    | 1.61   | 2.15 | 0.00 | 2.15 | 0.00   | 23.14| 48.87| 0.32  | 0.37 |
|                                       | 8    | 0.97   | 0.38 | 0.00 | 0.38 | 0.00   | 0.00 | 5.64 | 0.04  | 0.07 |
| Field of study                        | 9    | 0.96   | 0.32 | 0.00 | 0.32 | 0.00   | 0.00 | 0.00 | 0.01  | 0.04 |
|                                       | 1    | 1.49   | 100.00| 0.00 | 100.00| 0.00   | 0.00 | 0.00 | 0.64  | 7.09 |
|                                       | 2    | 1.41   | 2.27 | 0.00 | 2.27 | 0.00   | 0.00 | 0.00 | 0.08  | 0.26 |
|                                       | 3    | 3.11   | 5.17 | 0.00 | 5.17 | 0.00   | 0.00 | 19.43| 0.32  | 0.79 |
|                                       | 4    | 3.48   | 4.40 | 0.00 | 4.40 | 0.00   | 0.00 | 36.13| 0.46  | 0.96 |
|                                       | 5    | 4.00   | 7.10 | 0.00 | 7.10 | 0.00   | 0.00 | 58.16| 0.51  | 1.14 |
|                                       | 6    | 3.18   | 16.67| 0.00 | 16.67| 0.00   | 0.00 | 0.00 | 0.47  | 1.80 |
|                                       | 7    | 2.24   | 2.17 | 0.00 | 2.17 | 0.00   | 0.00 | 0.00 | 0.07  | 0.25 |
|                                       | 8    | 3.45   | 2.77 | 0.00 | 2.77 | 0.00   | 0.00 | 0.00 | 0.10  | 0.32 |
|                                       | 9    | 5.94   | 2.70 | 0.00 | 2.70 | 0.00   | 0.00 | 0.00 | 0.08  | 0.33 |

1 UR unemployment rate
2 SD standard deviation
is used. The Equivalence Principle method is used to determine the premium in the insurance field based on the principle that the expected value of the future loss function is equal to zero.

**Distribution specification for unemployment rate**

It is observed that the distribution of Malaysia’s unemployment rate is positively skewed, as shown in Fig. 2. Given that the Burr distribution is characterized by unimodality and strong positive skewness, it is tested against other unemployment distribution functions for the working-age population (15 – 65 years). The probability density function (pdf) is given by

\[ f(x) = \frac{k c}{\alpha} \left( \frac{x}{\alpha} \right)^{c-1} \left[ 1 + \left( \frac{x}{\alpha} \right)^{c-k} \right]^{-(k+1)}, \]

where \( k > 0, \alpha > 0, c > 0, k, \alpha, c \) are the parameters of Burr distribution. The parameters are derived using maximum-likelihood estimation. The random variable \( x \) is age considered throughout the year (time) from 1997 to 2015.

**Survival analysis**

**Hazard function for unemployment duration**

Unemployment duration is estimated using the hazard function given by the mathematical expression

\[ \lambda(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt)}{dt \cdot S(t)}. \]  

(1)

In general terms, Eq. (1) gives the failure rate of an item given that the item has survived for a period of time \( t \). In this study, the term ‘survive’ refers to being unemployed. Therefore, Eq. (1) is interpreted as the chance of an unemployed individual at \( x \) the age to be unemployed for another interval \( dt \) given that the person has been unemployed for a period of time \( t \). This is the underlying idea of unemployment duration estimation.

To estimate the unemployment duration for a person aged \( x \), we consider the alternative expression of the hazard function. For simplification, Eq. (1) is expressed as

\[ \lambda(t) = \frac{f(t)}{S(t)}, \]

where \( f(t) \) is the probability density function for unemployment rate and \( S(t) \) is the survival function which represents no change in status, in other words, remaining unemployed.
The Cox proportional hazards model was introduced by a British statistician, David Cox in 1972 when he published the paper entitled “Regression analysis and life tables”. Since then, the use of the model has been extremely significant in biomedical research. Cox’s proportional hazards model is mainly concerned with investigating the effects of the explanatory variables. The hazard function to measure the unemployment scenario is represented by

$$\lambda_i(t) = \lambda_0(t) \cdot \exp[\beta_i x_i]; \ i = 1, 2, ..., 7,$$

where $x_i$ are covariates ($x_1 = \text{age}$, $x_2 = \text{gender}$, $x_3 = \text{education}$, $x_4 = \text{ethnic}$, $x_5 = \text{state}$, $x_6 = \text{highest certificate obtained}$, and $x_7 = \text{field of study}$), $\lambda_i(t)$ is the hazard function of the individual explanatory variables $i$, and $\lambda_0(t)$ is the baseline function which corresponds to an observation (unemployment rate) with no covariates.

The quantiles $\exp(\beta_i)$ represents the hazard ratio. The hypotheses to test the significance of the effects of the respective coefficients are as follows:

$$H_0 : \text{coefficient is not significant, } \beta = 0$$

$$H_1 : \text{coefficient is significant, } \beta \neq 0.$$

A critical $p$ value of less than $\alpha = 0.05$ is adopted to reject the null hypothesis which implies that the coefficient is an influencing factor for unemployment.

The Hazard Rate (HR) is also used to measure the relationship between two covariates. An HR larger than one indicates that a covariate is positively associated with unemployment, an HR less than one indicates a covariate is negatively
associated with unemployment, and an HR equal to one indicates no effect between the covariates and unemployment. In short, the HR can be interpreted by

\[ HR = 1 : \text{no effect} \]
\[ HR < 1: \text{negative associated} \]
\[ HR > 1: \text{positive associated.} \]

**Unemployment insurance premium pricing framework**

The Equivalence Principle is a method to determine the premium in the insurance field. The idea behind the principle is that the expected value of the future loss function is equal to zero. It is applied here to determine the premium rate for unemployment insurance. An insurance policy is a financial agreement between an insurance company and its policyholders where the insurer provides security to the policyholder by paying some benefits during the event of an occurrence, as well as covering expenses associated with maintaining the policy contract. On the other hand, the insured consents to pay a premium to the insurance company to secure these benefits. Under the Equivalence Principle, the expected loss of an insurance policy to the insurer is assumed to be zero. Therefore, the expected present value of future premium (PVFP) income paid by the insured is equal to the expected present value of future benefit (PVFB) and the expected present value of future expenses (PVFE) paid by the insurer, as shown as follows:

\[
E(\text{PVFP}) = E(\text{PVFB}) + E(\text{PVFE}).
\]

The expression of the expected PVFB, expected PVFP and the expected PVFE is given by

\[
E(\text{PVFB}) = \left( \sum W_i \times S \right) \times U \times N
\]

\[
E(\text{PVFP}) = P \times \sum_{k=0}^{n-1} \left( \frac{1}{1 + r_b} \right)^k
\]

\[
E(\text{PVFE}) = 0.20 \times P \times \sum_{k=0}^{n-1} \left( \frac{1}{1 + r_b} \right)^k,
\]

where \( k \) represents the numbers of annual premiums payable at the beginning of each year. We assume the annual expenses charges amount to 20% of the annual premiums collected. It is noted that \( r_b \) represents the risk-adjusted rate of return derived from the Capital Asset Pricing Model (CAPM). It is calculated using the following expression:

\[
r_b = r_f + \beta(r_m - r_f), \tag{2}
\]
where $\beta$ denotes the correlation between unemployment rate and market rate of return, which is expressed by $\beta = \frac{\text{cov}(r_u, r_m)}{\text{var}(r_m)}$. The definition for the parameters in the Equivalence Principle framework is given in detail in Table 3. The computation of the parameters is discussed in the next section.

**Results and discussion**

The fitting of unemployment distribution is discussed first followed by unemployment duration, the key variables influencing unemployment, and unemployment insurance premium pricing.

**Distribution fitting of unemployment rate**

The fitting of the probability density function is a popular approach to explaining data patterns. The histogram in Fig. 2 shows that the unemployment data fit well with positively skewed distributions. The mathematical formula of Burr distribution and several other continuous distribution functions such as gamma, lognormal, logistic, log-logistic, and Weibull distributions are shown in Table 4. The log-likelihood function is differentiated and solved numerically to obtain the parameter estimation. Log-likelihood, Akaike Information Criteria (AIC), and Bayesian Information Criteria (BIC) are used in the model selection. The results in Table 5 show that the Burr distribution has the best fit and its hazard function is then used to estimate unemployment duration.

Figure 2 shows the distribution fittings for the unemployment rate and the histogram indicates the unemployment count is left-skewed. Youth unemployment (17–24 years) accounts for 60.3% of total national unemployment. It affirms the need for a greater policy focus on youth employment despite the country’s overall low unemployment rate.

**Estimation of unemployment duration**

The unemployment duration is measured using the hazard function. The hazard function, defined as instantaneous death rate in this study, describes the condition of an unemployed person is employed at time $t$ provided that the person is unemployed at time $t-1$. Figure 3 shows that the hazard rate increases at early young age of employment and then decreases gradually after age 20. Figure 4 shows the computed results of unemployment duration.

The estimated unemployment duration shows a declining pattern with increasing age. The decline in unemployment duration with rising age is consistent with the observation that work experience is key to returning quickly to the workplace. Another plausible reason is that a person who is retiring or close to retiring will consider taking up any offer, either on part-time or contract employment (Ministry of Human Resources, 2019). The relatively young retirement age which was raised from 55 to 60 years in 2000 compared to an average of 65 years in developed
countries, in tandem with rising life expectancy, are identified as the major drivers of the labor force working past the mandatory retirement age.

Further motivating this trend is the relatively young demographic profile of the country and its steady economic expansion, two major factors underlying the strong demand for skilled and experienced staff. The shorter unemployment duration for the older labor force is attributable to a combination of supply and demand factors. The employability of experienced staff also explains in part the puzzle of country’s high graduate unemployment despite its low national unemployment rate.

The age 20 cohort shows the highest hazard rate estimated at 0.1457. Annualizing it (365 days), the unemployment duration is estimated at 53 days. This means that for the age 20 cohort, the maximum duration of unemployment is about 2 months. The result highlighted the contribution of this study. The youth unemployment issue merits policymakers’ attention as crime and unemployment are positively associated (Fougere, et al. 2009). The youth unemployment is not unique in Malaysia. Similar patterns are observed in many countries in the world. Ciuca and Matei (2011) found that unemployment duration of 5.13 months for the less than 25 year age group. In Korea, Lim and Lee (2019) reported that 70% of the youths take 8 months or less to find a job after their graduation. In contrast, Danacica (2015) analyzed the (re) employment of young generation in Romania. For the three age groups of 15–19,

Table 3  Definition for the parameters in Equivalence Principle and CAPM

| Parameter | Definition |
|-----------|------------|
| U         | Average of unemployment rate (%) |
| P         | Annual premium (RM) |
| S         | Monthly salary in Malaysia (RM) |
| $r_b$     | Risk-adjusted rate of return from CAPM (%) |
| $r_f$     | Risk-free rate of interest (%) |
| $r_m$     | Market rate of return (%) |
| $r_u$     | Unemployment rate (%) |
| $W_t$     | Allowance payment given during unemployment spell (RM) |
| N         | Number of unemployment spell allowed for each insured |

Table 4  Probability density function of Weibull, lognormal, log-logistic, logistic, and gamma

| Distribution     | Probability density function |
|------------------|-----------------------------|
| Weibull          | $f(x|a,b) = \frac{b}{x^a} \left(\frac{1}{x^a}\right)^{b-1} e^{-\left(\frac{1}{x^a}\right)^b}, \ x \geq 0$ |
| Lognormal        | $f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(\log x - \mu)^2}{2\sigma^2} \right\}, \ x > 0$ |
| Logistic         | $f(x|\mu,\sigma) = \frac{\exp\left\{ \frac{-x}{\sigma}\right\}}{\sigma\left(1+\exp\left\{ \frac{-x}{\sigma}\right\}\right)}, \ x \in (-\infty, \infty)$ |
| Gamma            | $f(x|a,b) = \frac{1}{\Gamma(a)} x^{a-1} e^{-\frac{x}{b}}, \ x \geq 0$ |
| Log-logistic     | $f(x|\mu,\sigma) = \frac{1}{\sigma z} \left(1+e^{-z}\right)^{\frac{-z}{\sigma}}; \ x \geq 0, z = \frac{\log(x)-\mu}{\sigma}$ |
20–24, and 25–29 years examined, the author found that the chances to exit unemployment increase for the young generation below 30 years old.

The unemployment duration estimates are a novel contribution to policy discussion in Malaysia. The government currently assumes a maximum of 6-month unemployment duration for allowance payment under the employment insurance scheme (EIS) that was launched in January 2018. The estimated unemployment duration indicates that the EIS may not be reflective of the current unemployment situation.

Factors influencing unemployment

The factors influencing unemployment are explored using Cox proportional hazards model. The regression results are summarized in Table 6. The variances of the coefficients (covariance) are small for all estimates, indicating low variability from the average. The coefficient estimate $\beta$ is used to calculate the hazard rate represented by $e^{\beta}$. The $p$ value indicates the significance of the effects of the estimated coefficients.

From Table 6, it is observed that all age groups are found to be statistically significant in affecting national unemployment. We consider the hazard ratio to interpret the condition for the change from being unemployed to employed. For instance, the 31–36 year age group has a 64% probability to be employed. The hazard ratio of the 25–30 years age group exceeds almost 184%. This means that an
unemployed individual in this age group can quickly get a job. For the age group less than 18 years, the likelihood of being hired is only 6%, and the unemployment is contributed by the category which has less than 6-month certificate in educational attainment, or those who took general programs in their field of study. The statistical results suggest that the youths possessing full certificate or specific study program have lower unemployment compared to those equipped with general programs in their field of study. The low probability can be attributed to a majority of this age group still in school. The 19–24 years cohort has approximately 37% probability to be employed, while the 31–36 year cohort has a lower hazard ratio that could be attributed to longer working experience.

The results affirm the government’s concern over the relatively high youth unemployment. The policy response includes the provision of wage incentives to employees who have been out of work for a year and hiring incentives to employers to encourage youth employment. The government has also intensified skill enhancement programs to help unemployed workers return to the workplace. The current unemployment insurance scheme that pays 6 months of the allowance paid to the unemployed according to the respective percentage is considerably longer than that found in this study.

Malaysia is not alone in facing the unemployment issue. Similar unemployment condition has been observable over the past 4 decades. Katz (1974) reported that for the age cohort of 25–34, the length of unemployment (in weeks) increases in tandem with the decline in the years of schooling. It can be noticed that for the years of schooling of less than 8, the length of unemployment is 18.6 weeks. By contrast, it is about 14.2 weeks of unemployment for those with more than 12 years of schooling. For the age cohort of more than 55 onwards, the unemployment length does not vary with the respective schooling years. Those with more than 12 years of schooling averaged 21.6 weeks of unemployed which is about the same for those with less than 8 years of schooling at 22.6 weeks. Hoffman (1991) found that the age group of 19–24 has the greatest impact on unemployment. The results of the present study are consistent with the cited studies in demonstrating the importance of the age variable in influencing unemployment patterns.

On the gender effect, the p value for male is statistically significant, indicating that the male is impacted more by unemployment. This is attributable to the lower labor force participation rate of females as a significant number opt to remain as

![Fig. 4 Estimated unemployment duration by age](image-url)
full-time housewives or are not seeking employment due to their husbands’ ability to provide for the family. Some female employees work part-time as they have to care for their children. The part-time employees are classified as outside the labor force or as unemployed. As shown in Table 1, the share of females outside the labor force is 26.9%, which is more than double compared to male at only 10.4%. Further supporting the greater impact of unemployment on males is the higher hazard rate at 0.5322 compared to females at 0.0387, respectively. The figures confirm that the male is easier to get a job compared to female by almost 50%. In Russia, married women have significantly longer unemployment duration than married men (Foley, 1997). In Turkey, the report by Tansel and Tasci (2010) shows that the duration dependence of the exit rate from unemployment is different for men and women. The study found a U-shaped duration dependence for men, while for women, no duration dependence was found. For women, the percentage of leaving unemployment is lower than for men. A similar finding is obtained for Malaysia. In Finland, the women who are keen to look for employment are related to the number of children they have (Gonzalo and Saarela 2000). An increase in the number of children corresponds to a decrease in the women’s exit rate into employment. A study on Russia and Ukraine (Kupets 2006) found that children do not affect the women’s unemployment duration. By contrast, a different pattern is observed in the US where female has a shorter length of unemployment and a higher possibility to exit unemployment compared to male (Hoffman 1991).

On the influence of location, the results show that regional effects on unemployment are statistically significant but less pronounced compared to the level of education. An interesting outcome is that Sabah state with the highest unemployment rate is found to have an 11% higher chance to end unemployment compared to other states in Malaysia. It can be explained by the development programs mounted by the state government to enhance employability leading to a wider involvement of undergraduates and graduates from institutions of higher learning. Another plausible factor is the state’s smaller population relative to its land mass and the size of economic activity.

The influence of region on unemployment has been reported in several other studies. In Romania, Danacica and Paliu-Popa (2017) found that the West Region has the shortest median survival time, while South-Oltenia Region has twice the median survival time. The study also found that the median survival duration is 754 days for rural areas. In urban, the median survival duration takes up 443 days. In Romania, Cuica and Matei (2011) show that the average duration of unemployment by counties is 6 months.

Due to its multiethnic composition, ethnicity in Malaysia is an important variable in explaining the variation in the unemployment rate and duration. The percentage to exit unemployment for the majority Malay ethnic group is estimated at 8.9% which is substantially higher than the ethnic Chinese (3.4%) and ethnic Indians (0.01%). There are limited studies on the race effect. Hoffman (1991) investigated the unemployment exit rate of black and white Americans. He found that white females have the highest chance of leaving unemployment.

Besides age, gender, region, and ethnicity, existing research has identified education attainment as a key explanatory factor for unemployment, although the results
Table 6  Cox proportional hazards model for the factor investigation to the unemployment

| Variable     | Category         | $\hat{\beta}$ | p value | Result          |
|--------------|------------------|----------------|---------|-----------------|
| Age          | All              | 1.1484         | 0.0000  | Reject $H_0$    |
|              | <18              | 0.0684         | 0.0000  | Reject $H_0$    |
|              | 19–24            | 1.3726         | 0.0000  | Reject $H_0$    |
|              | 25–30            | 2.8400         | 0.0000  | Reject $H_0$    |
|              | 31–36            | 1.6446         | 0.0000  | Reject $H_0$    |
|              | > 37             | 1.0061         | 0.1682  | Do not reject $H_0$ |
| Gender       | All              | 0.8563         | 0.0000  | Reject $H_0$    |
|              | Male             | 0.5322         | 0.4808  | Do not reject $H_0$ |
|              | Female           | 0.0387         | 0.0000  | Reject $H_0$    |
| Education    | All              | 0.8124         | 0.0000  | Reject $H_0$    |
|              | None             | 0.01394        | 0.0000  | Reject $H_0$    |
|              | Primary          | 0.06460        | 0.0000  | Reject $H_0$    |
|              | Secondary        | 0.3251         | 0.0000  | Reject $H_0$    |
|              | Tertiary         | 0.7189         | 0.0000  | Reject $H_0$    |
| Ethnic       | All              | 1.1037         | 0.0000  | Reject $H_0$    |
|              | Malays           | 0.08906        | 0.0000  | Reject $H_0$    |
|              | Other bumiputra  | 0.0042         | 0.0000  | Reject $H_0$    |
|              | Chinese          | 0.0340         | 0.0000  | Reject $H_0$    |
|              | Indians          | 0.0010         | 0.0000  | Reject $H_0$    |
|              | Others           | 0.0788         | 0.0000  | Reject $H_0$    |
| State        | Johor            | 0.0045         | 0.0000  | Reject $H_0$    |
|              | Kedah            | 0.0466         | 0.0000  | Reject $H_0$    |
|              | Kelantan         | 0.0045         | 0.0000  | Reject $H_0$    |
|              | Melaka           | 0.0056         | 0.0000  | Reject $H_0$    |
|              | Negeri Sembilan  | 0.0329         | 0.0000  | Reject $H_0$    |
|              | Pahang           | 0.0240         | 0.0000  | Reject $H_0$    |
|              | Pulau Pinang     | 0.0143         | 0.0000  | Reject $H_0$    |
|              | Perak            | 0.0160         | 0.0000  | Reject $H_0$    |
|              | Perlis           | 0.0982         | 0.0000  | Reject $H_0$    |
|              | Selangor         | 0.0618         | 0.0000  | Reject $H_0$    |
|              | Terengganu       | 0.0524         | 0.0000  | Reject $H_0$    |
|              | Sabah            | 0.1089         | 0.0000  | Reject $H_0$    |
|              | Sarawak          | 0.0269         | 0.0000  | Reject $H_0$    |
|              | KL (Federal Territory) | 0.0490 | 0.0000  | Reject $H_0$    |
|              | Labuan (Federal Territory) | 0.0896 | 0.0000  | Reject $H_0$    |
vary depending on the prevailing economic conditions. In developed countries such as the United Kingdom, Belgium, and Ireland, higher education levels shorten the unemployment duration. The chance to leave unemployment for those with tertiary education is estimated at approximately 72%. However, in East European countries as found in Romania, the higher the education level, the more difficult it is to find a job. In Greece and Spain, similar studies have reported that educational level has fewer effects on the expected unemployment duration, attesting to the influence of other variables such as economic structure and conditions in influencing employability.

In this study, education attainment across all levels is found to have a statistically significant effect on unemployment. Those with no formal education (no certificate) have a greater impact compared to those who received formal education. Among those with formal education, the lower secondary certificate category is more unlikely to exit unemployment. Interestingly, a diploma certificate stands a better chance to leave unemployment compared to a degree level, reflecting the strong industry need for technical and vocational trained rather than university graduates. On the influence of the field of study, it is observed that the labor force with agriculture and veterinary training has a high 80% probability to find employment, followed by those in education and services, respectively. The fields of social sciences, business, and law have about half the possibility to leave unemployment, rendering them with the lowest hazard ratios in the study. In Europe, Núñez and Livanos

| Table 6 (continued) |
|---------------------|
| Variable | Category | $\hat{\beta}$ | $p$ value | Result |
| Highest certificate obtained | UPSR or equivalent | 0.6107 | 0.0000 | Reject $H_0$ |
| | PT3/PMR/SRP or equivalent | 0.4526 | 0.0000 | Reject $H_0$ |
| | SPM or equivalent | 0.6365 | 0.0000 | Reject $H_0$ |
| | STPM or equivalent | 0.6416 | 0.0000 | Reject $H_0$ |
| | Certificate | 0.7234 | 0.0000 | Reject $H_0$ |
| | Diploma | 0.7933 | 0.0000 | Reject $H_0$ |
| | Degree | 0.7045 | 0.0000 | Reject $H_0$ |
| | No certificate | 1.0048 | 0.8657 | Do not reject $H_0$ |
| Field of study | Not applicable | 0.9687 | 0.2720 | Do not reject $H_0$ |
| General programs | 0.8679 | 0.0015 | Reject $H_0$ |
| Education | 0.8936 | 0.0078 | Reject $H_0$ |
| Humanity and arts | 0.7632 | 0.0000 | Reject $H_0$ |
| Social sciences, business and law | 0.5254 | 0.0000 | Reject $H_0$ |
| Science, mathematics and computing | 0.6807 | 0.0000 | Reject $H_0$ |
| Engineering, manufacturing and construction | 0.7713 | 0.0000 | Reject $H_0$ |
| Agriculture and veterinary | 0.9116 | 0.0242 | Reject $H_0$ |
| Health and welfare | 0.8107 | 0.0000 | Reject $H_0$ |
| Services | 0.8795 | 0.0044 | Reject $H_0$ |
(2010) found that the disciplines that are most effective in exiting unemployment are health and welfare, education, and engineering.

**Model-based unemployment insurance premium pricing**

This analysis is aimed at motivating enhancements to the country’s recently launched national unemployment insurance scheme. The statistical estimates derived earlier are used to formulate a model-based unemployment insurance premium. The Equivalence Principle has been used to compute the premium pricing. See Table 7 for the estimated parameters in the Equivalence Principle calculation. The current EIS has been implemented since the year 2018. Only employees earning not more than RM4000 a month are eligible for the mandatory coverage. Premium payments amounting to 0.2% of the monthly basic salary are collected from employees and employers. As shown below, this coverage can be extended based on the statistical parameters derived from the unemployment analysis.

The proposed premium schedule is calculated based on the premium pricing framework described earlier. We calculate the risk-free rate of interest and market rate of return based on the median. For the unemployment rate, the first moment (expected mean) of Burr distribution is used. Table 8 shows the monthly allowance payment during the unemployment duration. We proposed an unemployment duration of a maximum of 3 months based on our findings. Some assumptions underlying the proposal include:

- 20% of the premium is charged for the expenses
- The chargeable monthly salary is constant
- 3 unemployment spells are allowed for the insured
- CAPM is considered for the expected risk-adjusted return on premiums
- The mortality rate is not considered during the period of insurance coverage.

Figure 5 shows the comparison between the proposed premium and the current EIS. Based on the current EIS, it is seen that a worker with a monthly salary of RM1,000, the worker pays an annual premium of RM24.00. The computed schedule shows an annual premium of RM7.95. The wide pricing gap highlights the need to examine the underlying cost structure including profit margin, reserves accumulation, and claims-paying ability. The salary proportion is now reduced to 0.066% compared to 0.2% and there is no cap on eligibility based on wage level. The median household income has been taken into consideration in the proposed scheme. The premium paid by the respective household group is also lower compared to the existing scheme. The bottom 20% income group (B20) pays RM23.85 for the annual premium for an average monthly household income of RM3000. The annual premiums for middle-income (M40) and top income (T20) groups are computed at RM49.30 and above RM80.00 per annum, respectively. Figure 6 displays the household group income in Malaysia as of the year 2019. The detailed premium schedule is given in Supplementary Table 15.
Study implications

The findings highlighted that age has the most influence on unemployment in Malaysia among the covariates investigated in this study. The youth group (19–30 years) is found to be the most vulnerable given its higher unemployment rate and longer unemployment duration. To address this key unemployment issue, the government has launched various initiatives such as job and skills matching and entrepreneurship, including seeking employment and entrepreneurial opportunities in business start-ups and small- and medium-sized enterprises. In the 2022 budget announcement, the Ministry of Finance Malaysia allocated RM4.8 billion to fund the job guarantee program (JaminKerja) to create 600,000 job opportunities. The budget will also target 220,000 trainees for various training and upskilling programs with an additional allocation of RM1.1 billion. On reskilling and upskilling, a dedicated statutory body, the Human Resource Development Corporation (HRD Corp) has launched a digital platform called e-LATiH that will provide them access to more than 300 online courses to meet the demands of various industries. In addition, e-LATiH also provides a job matching platform to job seekers and employers. Given the crucial role of the young generation in nation-building, these initiatives are important to address the social ills and underutilization of human capital due to high youth unemployment as affirmed in this study.

Table 7  Estimated parameters in Equivalence Principle

| Parameter | Values |
|-----------|--------|
| $U$       | 3.00%  |
| $r_b$     | 3.92%  |
| $r_f$     | 3.93%  |
| $r_m$     | 0.51%  |
| $r_u$     | 3.20%  |
| $W_t$     | Refer to Table 8 |
| $N$       | 3      |

Table 8 Allowance payment for the respective unemployment duration

| Unemployment duration, $t$ (month) | Allowance payment, $W_t$ (%) |
|-----------------------------------|------------------------------|
| 1                                 | 80                           |
| 2                                 | 50                           |
| 3                                 | 30                           |

The current unemployment insurance scheme implemented in Malaysia considers charges of 0.2% from employers and 0.2% from employees respectively, with the coverage limited to employees whose monthly salary is not more than RM4000 a month. This is a major drawback of the current EIS. Based on the statistical inference, the unemployment scheme derived in this study requires a lower insurance premium and no imposition of salary limits, thereby allowing for universal coverage.
of the country’s entire workforce. The proposed insurance scheme is, therefore, more reasonable that can be extended to every household in the country.

**Conclusion**

Unemployment is an overarching development issue regardless of a country’s income level and a crucial policy challenge irrespective of the prevailing level and duration of joblessness. This study is the first extensive analysis using statistical tools to investigate the determinants of unemployment in Malaysia, an upper-middle-income country. It models the distribution of unemployment and provides statistical support for extending the coverage of the existing unemployment insurance scheme. The Burr distribution is used to fit the unemployment data and the Cox proportional hazards model is applied to determine the key influencing variables. The study sheds new light on the challenges facing Malaysia despite its low unemployment situation. Based on the statistical inference, a feasible unemployment insurance scheme is developed to provide wider coverage at a lower cost to employees.

The study found that the 19–24 year age group has the highest unemployment rate, affirming the need for the youth unemployment issue to take center stage in policy formulation as well as in enhancing the design of the existing national unemployment insurance scheme. The statistical evidence shows that it is easier for the male cohort to exit unemployment than for women, suggesting that greater policy focus on female unemployment will have a greater overall impact on development. The study also found significant differences across states, whereby Sabah has the highest rate of unemployment, but further analysis shows that the regional effect is less pronounced than education level. Consistent with the findings in other countries, those with higher education have a better chance to exit unemployment.

![The comparison between the proposed premium and current EIS](image)

**Fig. 5** The comparison between the proposed premium and current EIS
Based on fundamental design principles, the unemployment insurance premium framework applied in the study suggests that the existing scheme can be expanded with lower premiums charged and without any cap on wages for employees to be eligible for coverage. The proposed insurance scheme framework developed in this paper can be used to motivate enhancements to the existing scheme after considering other factors such as age, education level, and field of study in a dynamic risk-based insurance scheme. To design a more flexible risk-based unemployment insurance scheme, the significant covariates identified in the Cox regression hazard model can be used as a guide to incorporate other risk factors.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43546-022-00287-2.

Acknowledgements The authors would like to thank the valuable comments from the reviewers to improve the manuscript. The authors are grateful to the Sunway University (IRGS: INT-2019-SMS-02) for the financial support towards this research.

Author contributions KWC: conceived the research, compiled the analysis, and wrote the manuscript. YKL: contributed to the manuscript writing. HSY: performed data cleaning and filtering.

Funding The study secured financial support from the Sunway University Internal Research Grant Scheme through a competitive process.

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability Software application: MATLAB R2018a, Excel Macro.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical approval The data used in this study is secondary data provided by the Department of Statistics Malaysia (DOSM).

Consent to participate The ethics approval and consent to participate is not applicable.
Consent for publication  The data are provided by the national statistical agency, Department of Statistics Malaysia (DOSM). Citation has been done to acknowledge the contribution.

References

Aitkin M, Healey R (1985) Statistical modelling of unemployment rates from the EEC lablur force surve -

Byram N (2013) Distribution of unemployment spells and survival analysis: the case of Bosnia and Her-

Baussola M, Mussida C (2017) Regional and gender differentials in the persistence of unemployment in

Beenstock M (1985) Competitive unemployment insurance pricing. Geneva Pap Risk Insur Issues Pract

Blake D, Beenstock M (1988) The stochastic analysis of competitive unemployment insurance premiums.

Böheim R, Taylor MP (2000) Unemployment duration and exit states in Britain. ISER Working Paper

Bowers N (1980) Probing the issues of unemployment duration. Mon Labor Rev 103(7):23–32

Christos F (2005) Forecasting the UK unemployment rate: model comparison. Int J Appl Econ Quant

Chuang HL, Yu MT (2010) Pricing unemployment insurance—an unemployment duration adjusted

Claveria O (2019) Forecasting the unemployment rate using the degree of agreement in consumer unem-

Cox D (1972) Regression models and life-tables. J R Stat Soc. Ser B (methodol) 34(2):187–220

D'Agostino A, Mealli F (2000) Modeling short unemployment in Europe. ISER Working Paper Series,

Dănăcică DE (2015) Youth (re) employment probability in Romania. Procedia Econ Financ 22:148–153

Dănăcică DE, Paliu-Popa L (2017) Determinants of unemployment spells and exit destinations in Roma-

Foley MC (1997) Determinants of Unemployment Duration in Russia. Working Papers 779, Economic

Fougere D, Pouget J, Kramarz F (2009) Youth unemployment and crime in France. J Eur Econ Assoc

Gonzalo MT, Saarela J (2000) Gender differences in exit rates from unemployment: evidence from a

Hoffman EP (1991) Estimation of length of job search by survival analysis. East Econ J 17(4):393–401

Holmes MJ, Otero J, Panagiotidis T (2013) Modelling the behavour of unemployment rates in the US

Katz A (1974) Schooling, age, and length of unemployment. Ind Labor Relat Rev 27(4):597–605

Kavler A, Danacica D, Babucea AG, Bicanic I, Bohm B, Tvedovski D, Tosevskka K, Borsic D (2009) Cox

Kreiner A, Duca JV (2020) Can machine learning on economic data better forecast the unemployment

Kupets O (2006) Determinants of unemployment duration in Ukraine. J Comp Econ 34:228–247

Lakuma CP, Marty R, Kuteesa A (2016) Survival analysis of regional unemployment in Uganda: evi-

Lim JY, Lee YM (2019) Exit duration and unemployment determinants for Korean graduates. J Labour

SN Business & Economics
Malinvaud E (1982) Wages and unemployment. Econ J 92(365):1–12
Ministry of Human Resources (2019) National strategic development plan on ageing population: inclusion and employment of Malaysia’s ageing population. Research report. Institute of Labour Market Information and Analysis (ILMIA) Publication
Mussida C, Fabrizi E (2014) Unemployment outflows: the relevance of gender and marital status in Italy and Spain. Int J Manpow 35(5):594–612
Núñez I, Livanos I (2010) Higher education and unemployment in Europe: an analysis of the academic subject and national effects. High Educ 59(4):475–487
Pereira S, Turkman F, Correia L (2018) Spatio-temporal analysis of regional unemployment rates: a comparison of model-based approaches. Revstat Stat J 16(4):515–536
Pollak A (2012) Unemployment, human capital depreciation, and unemployment insurance policy. J Appl Economet 28(5):840–863
Salant SW (1977) Search theory and duration data: a theory of sorts. Q J Econ 91(1):39–57
Simwa et al (2016) Application of Burr XII mixture distributions to model unemployment duration in pricing unemployment insurance assuming USA data. Int J Stat Distrib Appl 2(3):27–34
Tansel A, Tasci HM (2010) Hazard analysis of unemployment duration by gender in a developing country: the case of Turkey. Labour 24(4):501–530

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Wooi Chen Khoo1 ⋆ · Kim Leng Yeah2 · Shun Yi Hong3

Kim Leng Yeah
klyeah@sunway.edu.my

Shun Yi Hong
15003817@imail.sunway.edu.my

1 Department of Applied Statistics, School of Mathematical Sciences, Sunway University, No. 5, Jalan Universiti, Bandar Sunway, 47500 Petaling Jaya, Selangor Darul Ehsan, Malaysia

2 Jeffrey Sachs Center On Sustainable Development, Sunway University, No. 5, Jalan Universiti, Bandar Sunway, 47500 Petaling Jaya, Selangor Darul Ehsan, Malaysia

3 Department of Actuarial Science and Risk, School of Mathematical Sciences, Sunway University, No. 5, Jalan Universiti, Bandar Sunway, 47500 Petaling Jaya, Selangor Darul Ehsan, Malaysia