Modeling Uncertainty in Unemployment Duration - Project Abstract

Sneha Sarkar
A0304787U

Liu Xin Ying
A0188304A

Prerana Chakraborty
A0305018R

Lee Jia Yun Amanda
A0296790L

Akshaya Vajpeyarr
A0307243M

1 Introduction

We aim to model the uncertainty surrounding the duration an individual remains unemployed before securing a job. Specifically, we want to analyze how various socio-economic factors—such as age, education level, and retrenchment rates—affect the likelihood of re-entering the workforce and try to model the uncertainty of it.

1.1 Why?

This model can provide valuable information for policy makers, job seekers, and organizations. Individuals/next generation students can make informed career decisions based on employment trends. In addition, companies and hiring managers can use these data to identify potential labor market gaps and optimize hiring strategies.

This project contributes to economic growth and enhances our understanding of labor market dynamics as well as current and near-future job markets, especially as industries evolve in a rapidly changing employment landscape, helping the working population upskilled and aligned with market demands.

Beyond just predicting the length of unemployment, this project also seeks to understand the underlying causes of unemployment and identify potential policy interventions to reduce it. This model allows predictive intervention, helping governments not only react to labor market trends, but shape them by creating proactive policies that reduce the duration of unemployment and improve the stability of the workforce. This model could also be used as a personalized career assistance tool, as it could help a job seeker and could suggest actions to improve re-employment chances, such as skill development, networking strategies, or job market trends to focus on.

2 Related Work

2.1 Uncertainty in Employment in Singapore

Various work has been done in the study of factors impacting unemployment and the job market in Singapore. In particular, Tan et al. [8], in their study of Singaporean student perspectives, highlight the need to consider the impact of the COVID-19 pandemic on such uncertainty. Further, Appold [9] identifies a weakening position of university graduates in Singapore's labor market, a trend that introduces increased uncertainty in unemployment duration. Finally, Wong [11] underscores the transformative shifts in Singapore's employment patterns, emphasizing the rise of non-standard employment, the gig economy, and the increasing demand for continuous skills upgrading, all of which contribute to heightened uncertainty in unemployment duration. Taken together, these highlight the complex relationships between different factors that we will need to take into account for our model.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

2.2 Bayesian Analysis of Unemployment Duration Data in the Presence of Right and Interval Censoring

Ganjali and Baghfalaki [10] address the critical issue of censoring in unemployment duration data, employing a Bayesian framework to handle both right and interval censoring. This methodological approach is highly relevant for modeling uncertainty in unemployment duration, as real-world datasets often suffer from incomplete observations. The paper demonstrates how a Bayesian analysis can effectively incorporate prior knowledge and handle the uncertainty introduced by censoring, leading to more robust and accurate estimates of unemployment duration. Their work highlights the importance of accounting for censoring mechanisms when modeling unemployment duration, particularly in dynamic labor markets where data collection may be incomplete or subject to varying observation intervals. By adopting a Bayesian approach, researchers can better capture the inherent uncertainty and provide more reliable predictions of unemployment duration.

2.3 Survival analysis for unemployment modelling

A related work [12] aimed to model the probability of job seeker finding a job as a function of time. As the entries in the Slovenian unemployment records are right censored, survival analysis is a suitable approach for parameter estimation. Survival analysis models time-to-event data, making it useful for estimating unemployment duration. Traditional models, such as the Kaplan-Meier estimator and Cox proportional hazards model, manage censored data but struggle with non-linear dependencies. Ganjali et al.(2014) [13] conducted a Bayesian analysis of unobserved heterogeneity in unemployment duration, addressing challenges posed by right and interval censoring. Their study utilized Accelerated Failure-Time (AFT) models with log-logistic, log-normal, and Weibull distributions, incorporating latent variable models from the exponential family to account for unobserved individual differences. Using Markov Chain Monte Carlo (MCMC) sampling in WinBUGS, they estimated posterior distributions and assessed model fit via Deviance Information Criterion (DIC3). Their findings highlighted significant variation in unemployment duration across demographic groups, emphasizing the importance of modeling heterogeneity in employment dynamics. However, the reliance on MCMC methods makes the approach computationally intensive, limiting scalability to large, high-dimensional datasets, necessitating more efficient inference techniques. To address this, Variational Bayes inference enhances survival models by approximating posterior distributions while ensuring computational efficiency through the Evidence Lower Bound (ELBO). The survival function $S(t)$ and risk function $h(t)$ quantify event probabilities, while Accelerated Failure-Time (AFT) models express log-survival time as:

$$\log T = y = \beta^T \mathbf{x} + \sigma W. \tag{1}$$

which can also be read as:

$$y = \ h_z(\mathbf{x}) + \sigma W. \tag{2}$$

where z is the set of parameters for the non linear function. To model complex relationships, artificial neural networks (ANNs) replace traditional risk functions, allowing non linear learning of socio-economic interactions affecting unemployment. The ANN-based function:

$$h_z(\mathbf{x}) = g_\theta(\mathbf{x}) \tag{3}$$

maps high-dimensional features to survival time estimates while integrating VB inference for uncertainty quantification. This combination of VB and ANNs enhances survival analysis by enabling robust uncertainty estimation in employment duration modeling.

3 Approach/Methodology

Using available data, prior knowledge, and posterior inference, we seek to understand how each feature affects the duration of unemployment. Our objectives are to estimate an individual's expected unemployment duration based on socio-economic characteristics, and offer actionable insights to help reduce this duration. Below, we describe the models we plan to experiment and the criteria we will use to evaluate them. We will also consider the possibility of combining the methods to create a hybrid model. The study will utilize the following variables.

Observed Variables Age, sex, nationality, citizenship, years of experience in the field, highest education attainment, grades obtained, job-related courses, extra certifications, parental education level, number of connections, job search frequency, house location, social media availability (e.g., LinkedIn), expected salary, success rate of intake for this industry/job role, co-curricular activities etc.

Hidden/Latent Variables General unemployment trends, job market availability across different job scopes and industries, popularity of specific industries or job scopes, genetics, IQ, EQ etc.

3.1 Approaches

3.1.1 Directed Acyclic Graph (DAG) Bayesian Network Approach

Our first approach leverages a Directed Acyclic Graph (DAG) Bayesian Network to model the uncertainty in unemployment duration. This method enables us to capture the causal relationships between socio-economic variables and their influence on an individual's time to re-employment.

The model includes observed variables related to the individual, such as age, education, job experience, as well as observed economic variables like unemployment trends and job market availability. Additionally, we introduce hidden variables such as genetics, IQ, and EQ, which may significantly affect employability but are unobserved. In order to learn the parameters for such unobserved factors (which hence do not have data associated with them), we will seek to find proxy variables such as education or family background to make estimations.

To perform inference and uncertainty estimation, we plan to integrate the Sum-Product Algorithm (SPA), which efficiently computes marginal probabilities over the Bayesian Network. SPA will allow us to estimate the probable duration of unemployment of an individual, taking into account the dependencies between variables. By analyzing the posterior distributions, we can generate personalized recommendations on actions an individual can take to reduce the duration of unemployment, such as acquiring additional skills, expanding their network, or adjusting salary expectations. For the Bayesian Network implementation, we will most likely use the **pgmpy** library.

3.1.2 Hidden Markov Model (HMM) Approach

We plan to use a Hidden Markov Model (HMM) to represent the changes in a person's employment status over time. HMMs are particularly useful for modeling hidden states, variables such as market demand, that influence the duration of unemployment, but cannot be directly observed. In this approach, we will use observable features such as education level, job search frequency, and work experience to estimate the probability of transitioning between different states, such as "actively searching," "not searching," or "employed." These transitions will help us model the dynamic nature of unemployment and identify which features increase or decrease the likelihood of reemployment, given the hidden state of an individual.

A better understanding of how the impact of observable factors can differ based on an individual's current state could also allow for more targeted intervention, where those helping job seekers can more accurately improve a person's chances of reentering the job market by focusing on certain aspects or transitioning job seekers to a different state. To implement the HMM, we will likely use the **hmmlearn** library.

3.1.3 Monte Carlo Markov Chain (MCMC) Approach

We can also consider Monte Carlo Markov Chain (MCMC) methods to simulate different potential unemployment durations. MCMC allows sampling from complex probability distributions, making it useful when we lack a straightforward method to compute exact solutions. Using MCMC, we aim to generate multiple possible career paths based on features such as skills, connections, and job search effort to estimate a range of potential outcomes. This could model how choices such as upskilling or networking could impact the likelihood of finding employment. For the MCMC implementation, we will likely use **PyMC3**.

3.2 Datasets

We will make use of available datasets on data.gov, an open data portal with data compiled from various government agencies in Singapore. These include various data sets on the unemployment rate, duration, and job vacancies according to the age, education, field of study, and industry of the individual. Using these datasets to learn the parameters of our model will allow the models to be applicable in a Singaporean context. A list of potential datasets have been included in our reference list. Furthermore, we will validate assumptions on conditional probabilities using data that is available on Kaggle. Such data is not Singapore-centric but can be insightful for our understanding of how universal factors such as age and education can impact unemployment duration, especially if the local data is not available on data.gov.

3.3 Evaluation Criteria

The approaches will then be evaluated through the below criteria

Predictive Accuracy: How well does each model estimate the actual duration of unemployment? We will likely use Root Mean Squared Error (RMSE) to evaluate. This will be done on a subset of data set aside initially as the test set, and will not be used for learning parameters, but used to test the learned models.

Uncertainty Qualification: How well does the model capture variability and confidence in predictions? We will look at Confidence Intervals to evaluate the range of uncertainty.

Interpretability/Explainability: Can policymakers, job seekers, or economists understand the model's decisions? We will do feature importance analysis (ie. which variables most influence unemployment duration) and evaluate transparency in generating recommendations. Simpler models with fewer parameters will be easier to explain.

Actionable Recommendations: Does the model provide clear, actionable suggestions for reducing unemployment duration? This will be evaluated by the number of actionable insights provided and whether those actions are feasible for individuals to implement.

Computational Efficiency: How much time and computational power does the model require? We will measure training time (e.g., CPU/GPU requirements) and memory usage during the model fitting and prediction phases.

4 Social and Ethical Impact Statement

Reducing Long-Term Unemployment & Economic Hardship : Understanding the factors affecting unemployment duration could help job seekers re-enter the workforce faster, reducing financial stress and economic inequality.

Enhancing Fairness in Employment Opportunities : The findings could help to ensure equal access to job opportunities and retraining programs, and particularly improve access to such resources for those who need it most.

Improving Workforce Resilience : The model could help to improve planning of reskilling programs and economic policies.

Ethical Use of Predictive Insights (misuse of prediction) : There's a risk that employers or policymakers misuse predictions, e.g., businesses might avoid hiring individuals from groups associated with longer unemployment durations. In order to prevent this, findings from results should only be applied in policy interventions, and not made known to potential employers

5 AI Tool Use

The team used ChatGPT to refine the English language in this paper.

6 References

[1] Ministry of Manpower (2024), *Unemployed Residents Aged 15 Years and Over by Highest Qualification Attained and Duration of Unemployment*, data.gov,https://data.gov.sg/datasets/d_a0ca632fd1d6ff841f0e47298a9ab589/view

[2] Ministry of Manpower (2024), *Unemployed Residents Aged 15 Years and Over by Age and Duration of Unemployment*, data.gov, https://data.gov.sg/datasets/d_db95e15ceffaa368a043310479dc7d57/view

[3] Ministry of Manpower (2024), *Median Duration of Unemployment*, data.gov, https://data.gov.sg/datasets/d_c01a3210fb10f1a52676f97498d4ec2c/view

[4] Ministry of Manpower (2024), *Rate of Re-entry Into Employment for Retrenched Residents (Six Months Post-Retrenchment), Annual*, data.gov, https://data.gov.sg/datasets?query=unemployment&page=1&resultId=2119

[5] Ministry of Education (2024), *Graduate Employment Survey - NTU, NUS, SIT, SMU, SUSS & SUTD*, data.gov, https://data.gov.sg/datasets/d_3c55210de27fcccda2ed0c63fdd2b352/view

[6] Ministry of Education (2025), *Statistics on Local Graduates Working in Jobs Related to Field of Studies*, moe.gov.sg, https://www.moe.gov.sg/news/parliamentary-replies/20250108-statistics-on-local-graduates-working-in-jobs-related-to-field-of-studies

[7] Ministry of Manpower (2024), *Job Vacancy by Industry and Occupational Group, Annual*, data.gov, https://data.gov.sg/collections/688/view

[8] Tan, E., Mwagwabi, F., Lim, T., & Lim, A. (2020). *Graduate employability concerns amidst a crisis: Student perspectives from Singapore on COVID-19*. Journal of Education for Work, 37(5-6), 352-366.

[9] Appold, S. J. (2018). *The Weakening Position of University Graduates in Singapore's Labor Market: Causes and Consequences*. Social Sciences, 7(12), 257.

[10] Ganjali, M., & Baghfalaki, T. (2012). *Bayesian Analysis of Unemployment Duration Data in the Presence of Right and Interval Censoring*. Communications in Statistics - Theory and Methods, 41(15), 2738-2751.

[11] Wong, S. C. (2018). *Transformation of Employment Patterns and Need for Career Services in Modern Singapore*. Singapore Economic Review, 63(S1), 229-246.

[12] Boškoski, P., Perne, M., Rameša, M., & Boshkoska, B. M. (2021). *Variational Bayes survival analysis for unemployment modelling* Knowledge-Based Systems, 229, 107335.

[13] Ganjali, M., Baghfalaki, T., Berridge, D., Shahid Beheshti University, Shahid Beheshti University, & Lancaster University. (2009). *A Bayesian analysis of unobserved heterogeneity for unemployment duration data in the presence of interval censoring*. In International Econometric Review (IER) (p. 25).