Examining the Impact of Gender, Marital Status, Education, Age and Family Size on Employment Status Between 1971-2000 in Indonesia

JASMINE LEE

University of California, Berkeley / College of Letters and Science

Department of Economics

Undergraduate Honors Economics Senior Thesis

Advised by Professor Enrico Moretti

Acknowledgements

I would like to thank Professor Enrico Moretti for offering his generous support and valuable advice on this thesis, and Professor Benjamin Handel and Petra Laura Oreskovic for their guidance in the preliminary exploration of the research focus. I would also like to thank my family for their support.

Jasmine Lee

Berkeley, December 2023

Abstract

Professor Emi Nakamura at the University of Berkeley discussed, in her February 2 guest lecture, that the reason women in the U.S. experienced less layoffs than men during recession is because their jobs are mostly office jobs and have higher demand. Motivated from this investigation, the paper aims to examine this observation with a focus in Asia, particularly through the implementation of micro-data sets from Indonesia. With major economic recessions in the Indonesia's late 1900's, this paper examines temporal and regional trends of gender-specific unemployment rates between 1971 to 1990 in Indonesia. The paper will also investigate the effects of factors such as education, marital status, age, family size and gender on unemployment status taking the effects of recessions into consideration. The primary methodologies implemented in this paper are Linear and Logistic regressions. These regressions find that the listed factors carry a statistically significant impact on the employment status of individuals in Indonesia.

Keywords: Labor economics, logistic regression, linear regression, employment

Contents

1 Introduction

In her economics guest lecture on February 2 of 2023, Professor Emi Nakamura discussed that the reason women in the U.S. experienced less layoffs than men during recession is because their jobs concentrate generally on office jobs and thus, observed higher employment demand. However, it is an Asian custom for married women to be expected to stay home and nurture their children rather than being employed and working. Motivated from Professor Nakamura's studies, this paper aims to investigate the impact of marital status, number of children (family size), level of education, gender and age on an individual's employment status in Indonesia. Specifically, at times of major economic recessions, the paper examines the impact of interested factors, as listed, on employment status as well as their interactions with whether the particular examined year is considered to be in recession. Ultimately, the regressions in this paper aim to identify any observed evidence of major factors impacting an individual's employment status through a focus in Asia, particularly implementing household-collected micro-data set from Indonesia. The primary methodologies implemented in this paper are Linear and Logistic regressions. These regressions find that the listed factors carry a statistically significant impact on the employment status of individuals in Indonesia.

With major economic recessions in the Indonesia's late 1900's, this paper examines temporal and regional trends of gender-specific unemployment rates between 1971 to 1990 in Indonesia. This paper aims to explore and in particular, visualize, the temporal and regional trends in the unemployment rates in Indonesia on a gender-specific level with micro data from 1971 to 1995. The visualizations display comparisons of unemployment rate between genders, regions, and observed years. Indonesia has a total of 38 provinces, nine of which are considered non-official provinces and thus not included in the original data set from Integrated Public Use Microdata Series (IPUMS). Ultimately, the paper aims to answer the following question: What are the temporal trends and regional variations in gender-specific unemployment rates, and how do other factors such as education, age, and number of children interact with gender to shape these patterns over time?

1.1 Brief Overview of Existing Studies and Approaches

Past research, such as "Women, Wealth Effects, and Slow Recoveries" by Masao Fukui, Emi Nakamura, and Jón Steinsson, employs novel methodological approaches to investigate the repercussions of female-biased shocks, particularly those associated with the Gender Revolution, on economic recoveries. The research introduces the concept of "crowding out," which serves as a metric to assess how male employment reacts relative to female employment in the wake of these shocks. The paper highlights empirical challenge of disentangling female-biased shocks from gender-neutral ones, leading to the adoption of cross-sectional gender convergence at the state level as a key source of variation. The cross-sectional analysis observes evidence of gender convergence, particularly in states with initially substantial gender gaps, where more rapid declines are evident Nakamura et al. (2023). This convergence predominantly stems from heightened growth in female employment rates. To address the challenge of accurately distinguishing female-biased shocks, Nakamura et al. (2023) implemented instrumental variables estimates. Proxies for these shocks, such as the gender gap in 1970 and the "Job Opportunity Index," suggest minimal crowding-out effects. Their findings indicate that a one percentage point increase in female employment due to female-biased shocks results in only a modest decrease in male employment. Nakamura et al. used cross-sectional gender convergence and instrumental variables and lends robustness to the empirical findings. Ultimately, this paper lays the groundwork for further exploration of the intricate relationship between gender-related shocks and broader economic outcomes, particularly in the context of economic recoveries. However, these findings rely on microdata from the United States, a country whose culture and economy structure vary from countries in Asia, and as a result, should not be extrapolated to explain unemployment status of individuals in Asia.

Previous study by Wei-hsin Yu explores the role of motherhood in Taiwanese and Japanese households. The study delves into the nuanced employment trajectories of women during their childbearing and early child-rearing years in Japan and Taiwan. It classifies three primary employment patterns: continuous employment, discontinuous employment, and non-participation in the labor force. Similarly, various factors influencing these patterns are examined, such as education, extended-family support, and the number of children. The analysis uncovers notable disparities between Japanese and Taiwanese women, with Taiwanese women exhibiting higher rates of continuous employment compared to their Japanese counterparts. On the contrary, Yu (2005) found that Japanese women are more likely to experience periods of discontinuous employment or complete non-participation in the labor force.

Employing event-history techniques, the study delves deeper into women's labor-force exit following marriage and the birth of their first child. The findings reveal that the presence of a child significantly heightens the likelihood of labor-force exit in both Japan and Taiwan. Furthermore, educational attainment plays a distinct role in each country's employment dynamics. Yu (2005) found that in Japan, higher education is associated with a reduced likelihood of labor-force exit, while in Taiwan, the impact is not statistically significant. Extended-family support emerges as a protective factor, especially in the Taiwanese context, indicating the significance of cultural and familial contexts in shaping women's employment decisions.

The study underscores the complexity of cultural, economic, and familial factors in shaping women's employment trajectories during pivotal life stages in Japan and Taiwan. While Yu's results (2005) shed light onto the existing impact of motherhood, ultimately, the thesis will focus on the impacts of related factors on unemployment on a gender-specific level through the use of logistic and linear regression, and with the lens of economic recessions.

1.2 Overview of Paper

To observe and visualize the temporal and regional trends, the paper leverages household microdata from IPUMS, a public micro-database. The data is then separated by their respective years and region. Within each year, unemployment rate is calculated and compared between genders and displayed by bar graphs in which one bar represents one province in Indonesia. To provide a general view of the data in a time series fashion, line graphs of unemployment rate for each particular province is presented in the paper. While unemployment rate differs visually between genders, it can be seen to ultimately peak in 1995, an identified year of economic recession in Indonesia.

To investigate whether factors such as education levels, marital status, age, gender and family size impacts employment status, the paper uses logistic and linear regression by controlling for interested factors (as listed), create a dummy variable to indicate whether a year is a recession and interactive terms between the dummy variable and factors of interest. The regressions account for year-fixed effects and province-fixed effects that are one-hot encoded, and a linear trend in time is included to account for scale differences between linear and logistic regressions.

2 Literature Review

The research paper, "'Patriarchal reset' in the asia pacific during COVID-19: the impacts on women's security and rights" is written by Australian Social and Political Science researchers Melissa Johnston, Sara E. Davies, Jacqui True and Yolanda Riveros-Morales. In the paper, Johnston et al. explored how COVID-19 negatively impacted the lives of Asian Pacific women disproportionately due to varying factors such as urbanization, governmental social protection resources and family income. Through two survey-based studies of experiences of the "Women, Peace and Security" practitioners, Johnston et al. particularly explored how the rise of domestic violence and the pre-pandemic trend of women making up a higher proportion of healthcare both contributed to the concept of "resetting the patriarchy;" the idea reveals further unequal gender roles and burdens in households during the pandemic. Johnston et al. also found that patriarchy is further reinforced in political economies where countries have sparser social assistance or social protection for their citizens. Utilizing a feminist political economy approach to COVID-19, the research paper worked to demonstrate gender impacts on three spheres in a country: global health organization governance, private patriarchal reinforcement and the presence of patriarchy in diverse intersections of political, social and economic governance. Johnston et. al concluded that in the sphere of global health governance, gender exclusion and resource scarcity due to restrictions on productive health access for women contributed to a reinforcement of patriarchy in healthcare. Furthermore, they discover that unemployment resulting from the pandemic contributed to increasing domestic violence and household insecurities as well as unequal burden on household tasks distribution between genders. Lastly, due to pre-existing male-dominated institutions such as the military and radical capitalism, patriarchy is naturally reinforced in the society.

Through this research paper, Johnston et al. found important evidence of further entrenchment of patriarchy in the public health workforce and private households. The paper also discussed the importance of perceiving the impacts of COVID-19 on women through a feminist outlook given that the majority of current major institutions in Asia Pacific remain male-dominated. However, the research was established on the basis of survey evaluations and analysis, which may be a weaker basis for accurate representation of the true sentiments throughout Asian Pacific countries. Such misrepresentation may be the result of non-response bias and leading questions. The survey method only included 211 respondents, which may be argued that on a per capita representational basis of all women in Asia Pacific, is not fully represented. Furthermore, due to the fact that this paper was written through the lens of a feminist political economy approach to COVID-19, it is plausible that some of the questions on the survey may be leading questions prompting a feminist response from the participants. While Johnston et al. reveals the presence of deepening patriarchy amidst the pandemic, it cannot be extrapolated to sectors beyond the healthcare industry and analysis for layoff rates between genders. This paper justifies the assumption of women carrying unequal weight and moving forward, the thesis intends to find concrete data to support its topic.

2.1 Contributions to Existing Studies

From Johnston et al. research paper, "'Patriarchal reset' in the Asia pacific during COVID-19: the impacts on women's security and rights," the thesis aims to contribute another research methodology that eliminates survey biases and errors through leveraging official datasets. Rather than biasing research analysis on data collected from only 211 survey respondents, the thesis would identify patterns in governmental datasets, such as those from the World Health Organization, World Bank and official government statistical databases. The thesis also identifies related variables such as family income, number of children, marital status, etc, then using these self-identified variables, put them into a correlation matrix to determine how related they are to the layoff rate and labor force participation rate differences between men and women. This approach differs from the methodology utilized in the Johnston et al. paper because it builds analysis on data points rather than surveys that might have biases.

This paper intends to use the correlation coefficient to create logistic and linear regression models. The paper would first determine whether the different layoff rates between male and female workers are indeed statistically significant, and explore whether a historic economic recession caused an increase in layoff rates for women in Asia through building a regression model and utilizing the recession as a variable. This first section of the research aims to establish that gender is indeed a factor when considering how an economic recession affects men and women's work life differently. By showing, through data, that there is a statistical significance between men and women in terms of layoffs rates and labor force participation rates, the paper aims to eliminate the potential error that may come from experimental errors observed in the Johnston et al. paper.

Variable	Label
COUNTRY	Country
YEAR	Year
SAMPLE	IPUMS sample identifier
SERIAL	Household serial number
HHWT	Household weight
GEO1_ID	Indonesia, Province 1971 - 2010 [Level 1; consistent boundaries, GIS]
PERNUM	Person number
PERWT	Person weight
STEPMOM	Probable stepmother
STEPPOP	Probable stepfather
NCHILD	Number of own children in household
AGE	Age
SEX	Sex
MARST (general)	Marital status [general version]
MARSTD (detailed)	Marital status [detailed version]
EDATTAIN (general)	Educational attainment, international recode [general version]
EDATTAIND (detailed)	Educational attainment, international recode [detailed version]
EMPSTAT (general)	Activity status (employment status) [general version]
EMPSTATD (detailed)	Activity status (employment status) [detailed version]
LABFORCE	Labor force participation
CLASSWK (general)	Status in employment (class of worker) [general version]
CLASSWKD (detailed)	Status in employment (class of worker) [detailed version]

Figure 1: 22 Variables extracted from IPUMS Household Dataset

3 Overview of Data

In order to carry out the research, I had initially planned to refer to the yearly data published by the World Bank regarding labor participation rate between genders; however, upon discussing this approach with Petra Oreskovic, I agree with her observations that the general, aggregated trends provided by the World Health Organization or World Bank would not be detailed enough to produce an insightful analysis. As a result, I revised my approach such that I work with microdata points where I can create a regression analysis on detailed data points. One example of this is that rather than building a regression model based on one data point per year from the World Bank's labor participation rate database, I utilize databases where each data point is information of one woman or household. From this thought process, I leveraged the Indonesia Household Microdata from IPUMS International, and specifically extracted variables displayed in Figure 1 for the visualization and regression processes.

3.1 Data Processing for Temporal Visualization

To examine the data on a provincial level, I combined a dataset of geography codes that match to a unique region in Indonesia. The IPUMS dataset categorizes employment states as the following: 0-not in universe, 1-employed, 2-unemployed, 3-inactive, and 9-unknown/missing. For the purpose of this paper, the calculations for unemployment rate only account for those who are employed (with a value of one) or unemployed (with a value of two). For each province in a particular examined year, I obtained the unemployment rate grouped by each province via formula (1).

$$Unemployment\,Rate = \frac{Number\,of\,Unemployed\,People}{Total\,LaborForce} * 100 \tag{1}$$

In context of the IPUMS International Household data and after grouping the data by province, the number of unemployed people are those with value of two for their employment status and the total labor force is the sum of those with employment status of one and two. This process was done for each year with available data from the IPUMS. For visualizations on a yearly and provincial level, the unemployment rates are further separated by male unemployment rate and female unemployment rate. For general visualization where each province is graphed on a time scale from 1971 to 1995, unemployment data is merged among interested years to be displayed as a line graph.

3.2 Data Processing for Linear and Logistic Regressions

To conduct data processing for both linear and logistic regression on a gender-specific level, I filtered and merged my dataset such that the independent dataset, which only contains variables of interest (*MARST, SEX, EDATTAIN,* and *NCHILD*). I then one-hot encoded *EDATTAIN* as the variable is categorized by the following: 0-not in universe, 1-less than primary completed, 2-primary completed, 3-secondary completed, 4-university completed and 9-unknown. If the individual has completed primary school, then they will have a binary number one in the *EDATTAIN2* variable and zeros in all other variables; for simplicity, in this paper, any categories that are zero or nine are omitted. This dataset is then filtered by only individuals of the female gender or male gender, and the results are compared and discussed in section six of this paper. I then split my dataset into training and testing sets before running OLS regression and logistic regression on them.

For linear and logistic regressions with interactions on the years of recession between 1971 and 1995, and controlled for marital status, gender, age, education level and family size, I one-hot encoded *EDATTAIN* and introduce a new controlled variable *AGE*, which captures the working ages of individuals and filter my dataset to only those between 25 and 70. Since the goal is to observe any recession interactions with *MARST, SEX, EDATTAIN* and *NCHILD*, the paper will not analyze those who are most likely to be part of the full-time student or the retired population. As a result, Figure 2 captures the percentage of females, males and general population that are considered "young" (less than 35 years-old) and those on the older side. It can be observed that almost half of the population in the Indonesian dataset are young, with 43.49 percent of females, 46.64 percent of males and 45.58 percent of the general population between 20 and 35 years-old.

4 Temporal and Regional Visualizations

The goal of this section in the paper is to visualize interactions between gender and unemployment rates both on a provincial and temporal level in Indonesia, and ultimately,

	<=20 Age <= 35	35 < Age <= 70
Females	43.39%	56.61%
Males	46.64%	53.36%
Overall	45.58%	54.42%

Figure 2: This figure displays the percentage split of young versus old working individuals on both a general and gender-specified level.

identify patterns or trends of potential factors to changes in unemployment.

4.1 Gender-Specific Visualization Examined in Varying Years

This subsection demonstrates trends and patterns of unemployment rates in Indonesian Provinces as a bar chart for each of the following years: 1971, 1976, 1980, 1985, 1990 and 1995. There are 26 provinces that contain the majority of the Indonesian population examined for each listed years.

4.1.1 1971

In Figure 3 and 4, it can be noticed that generally, in 1971, female unemployment rates are much lower than male unemployment rates. This now begs the question of whether this is due to seeing less female workers in the job market (married Indonesian women choose to stay home) or if female workers truly experience lower unemployment rates. To further explore this differentiation, I will run linear probability models to investigate if factors such as number of children, education and marital status. It is noteworthy to highlight that DKI Jakarta and Gorontalo, Sulawesi Utara both have higher unemployment rate. These two provinces are both cities.

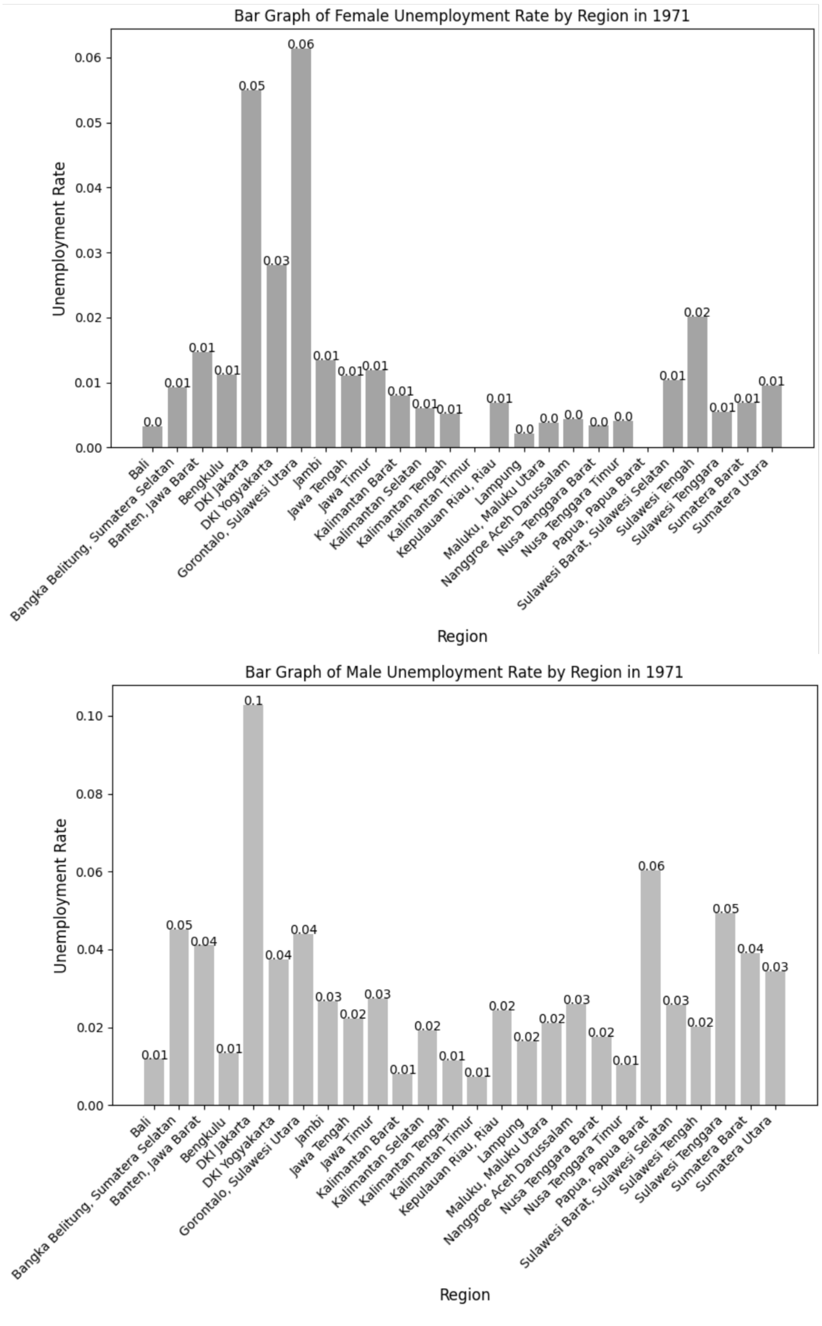

Figure 3: Bar graph comparison between unemployment rates of female and male individuals in Indonesia by provinces in 1971.

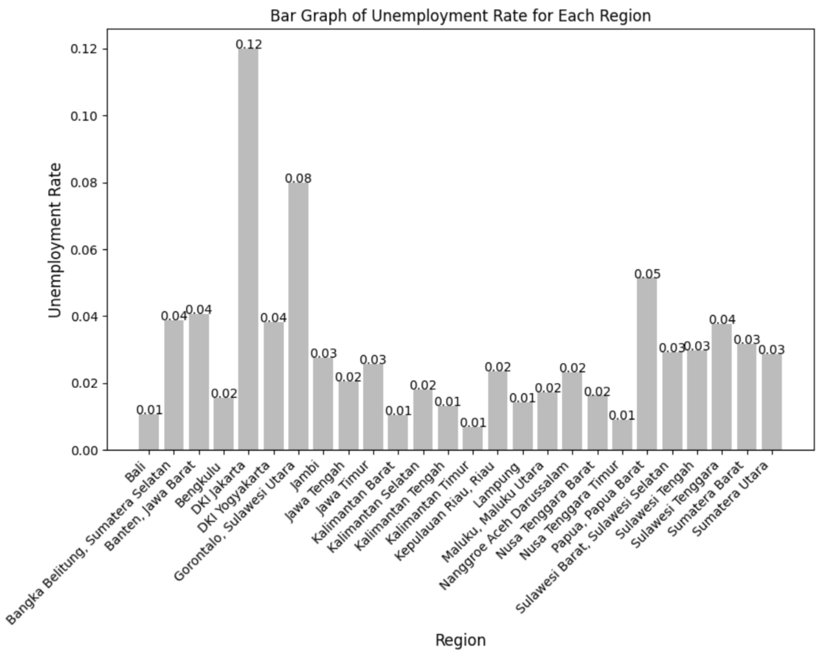

Figure 4: Bar graph of unemployment rate for each province in 1971.

4.1.2 1976

I notice that generally, in 1976, female unemployment rates are still lower than male unemployment rates. This again begs the question of whether this is due to seeing less female workers in the job market (married Indonesian women choose to stay home) or if female workers truly experience lower unemployment rates. To further explore this differentiation, I want to run linear probability models to investigate if factors such as number of children, education and marital status. However, it is noteworthy that Maluku, Maluku Utara has a high unemployment rate for both genders. In 1976, Indonesia experienced religious and political conflicts and has deployed troups in the Maluku Islands. This could be a possible explanation as to the high unemployment rate observed particularly in Maluku.

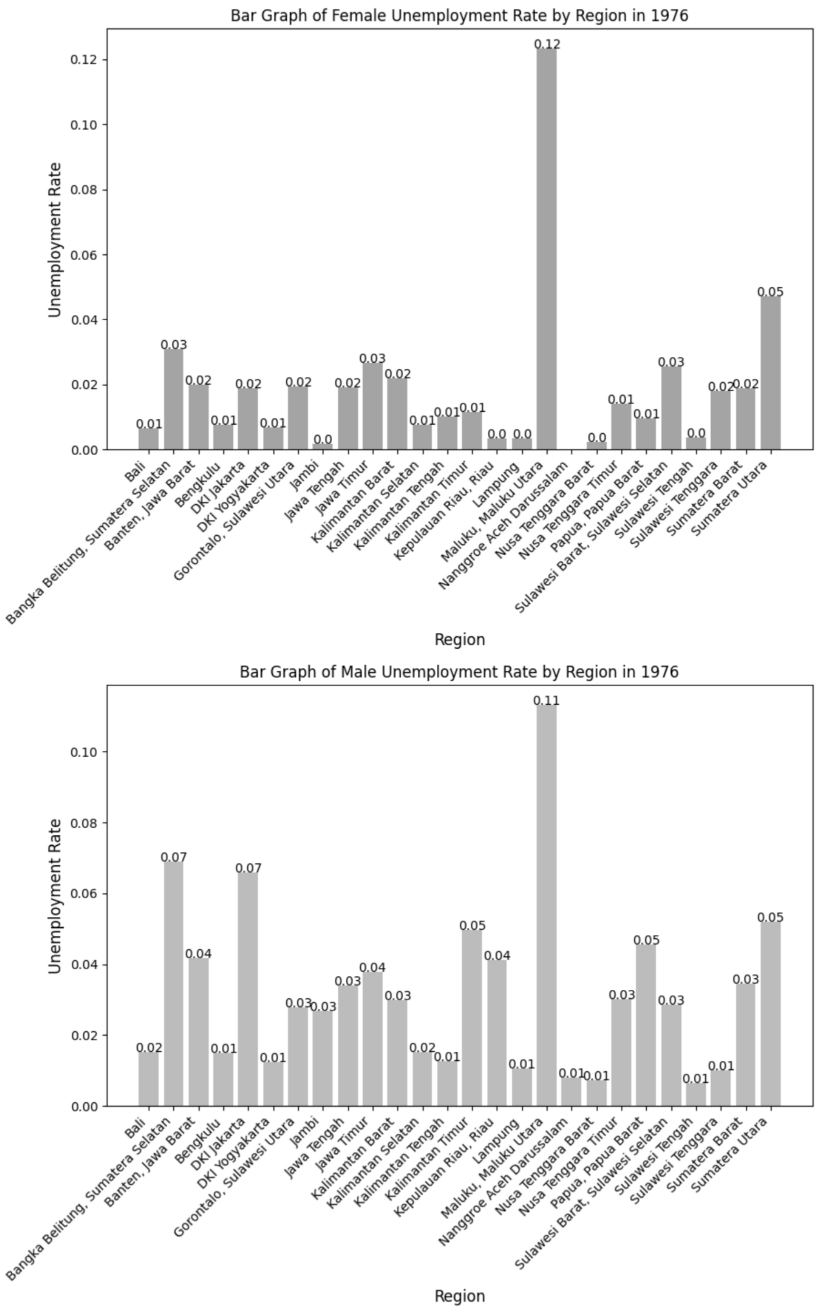

Figure 5: Bar graph comparison between unemployment rates of female and male individuals in Indonesia by provinces in 1976.

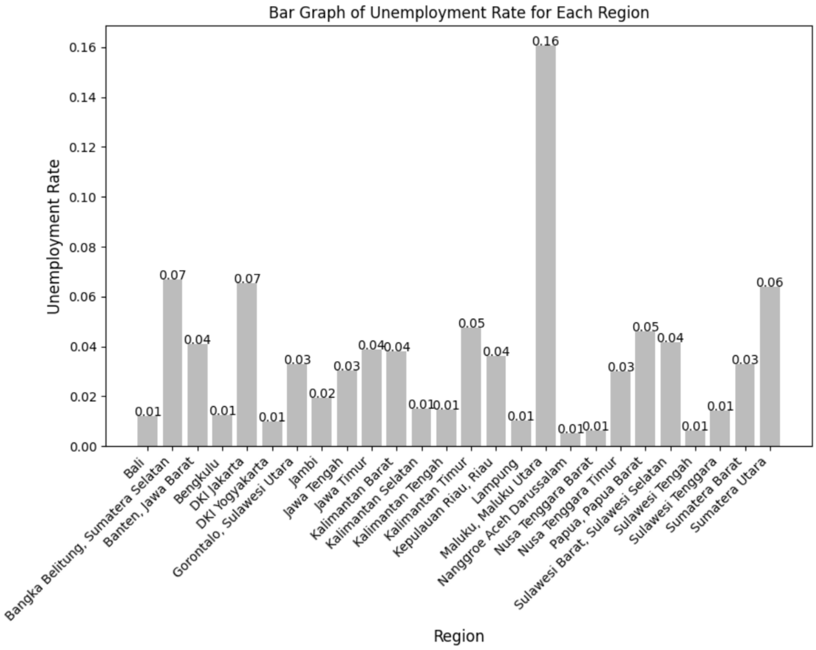

Figure 6: Bar graph of unemployment rate for each province in 1976.

4.1.3 1980

It can be observed in 7 and 8 that generally, in 1980, female unemployment rates are slightly lower than male unemployment rates. I want to investigate if there are any internal factors within the dataset that can point to the closing gap between female unemployment rate and male unemployment rate particularly in 1980. Similarly, I will run a logit model including factors such as education, number of children and marital status.

4.1.4 1985

It can be observed in Figures 9 and 10 that the female unemployment rate for 1985 to be much higher than those observed for the male population. Overall, the unemployment rate hover around 50 percent, so perhaps this is an indication of a recession year in

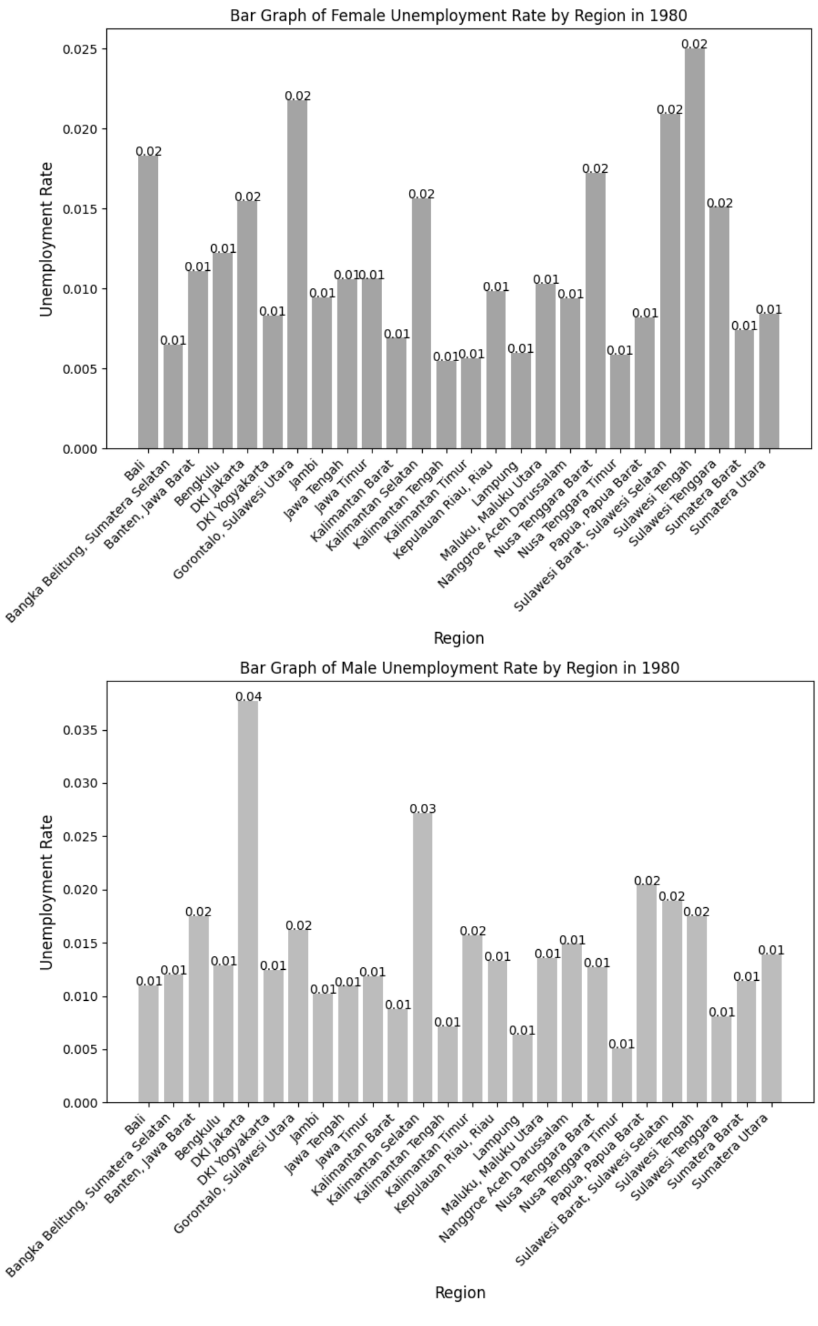

Figure 7: Bar graph comparison between unemployment rates of female and male individuals in Indonesia by provinces in 1980.

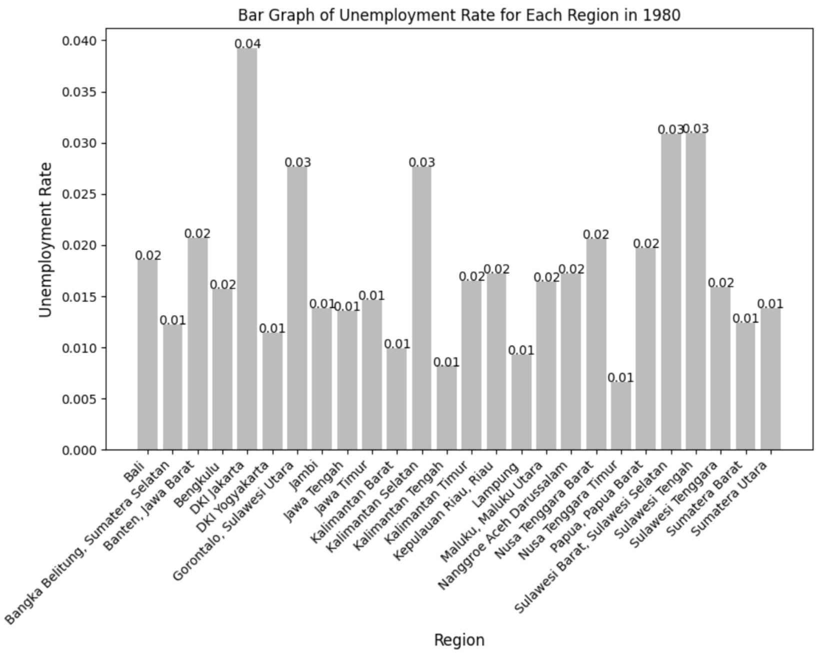

Figure 8: Bar graph of unemployment rate for each province in 1980.

Indonesia rather than flaws in the dataset. In 1985, Indonesia saw violent political struggles and as sited by the journal article "INDONESIA IN 1985: A Year of Trials," author Baladas Ghoshal details the country's violent governmental uproars, which attribute to its economy's recession.

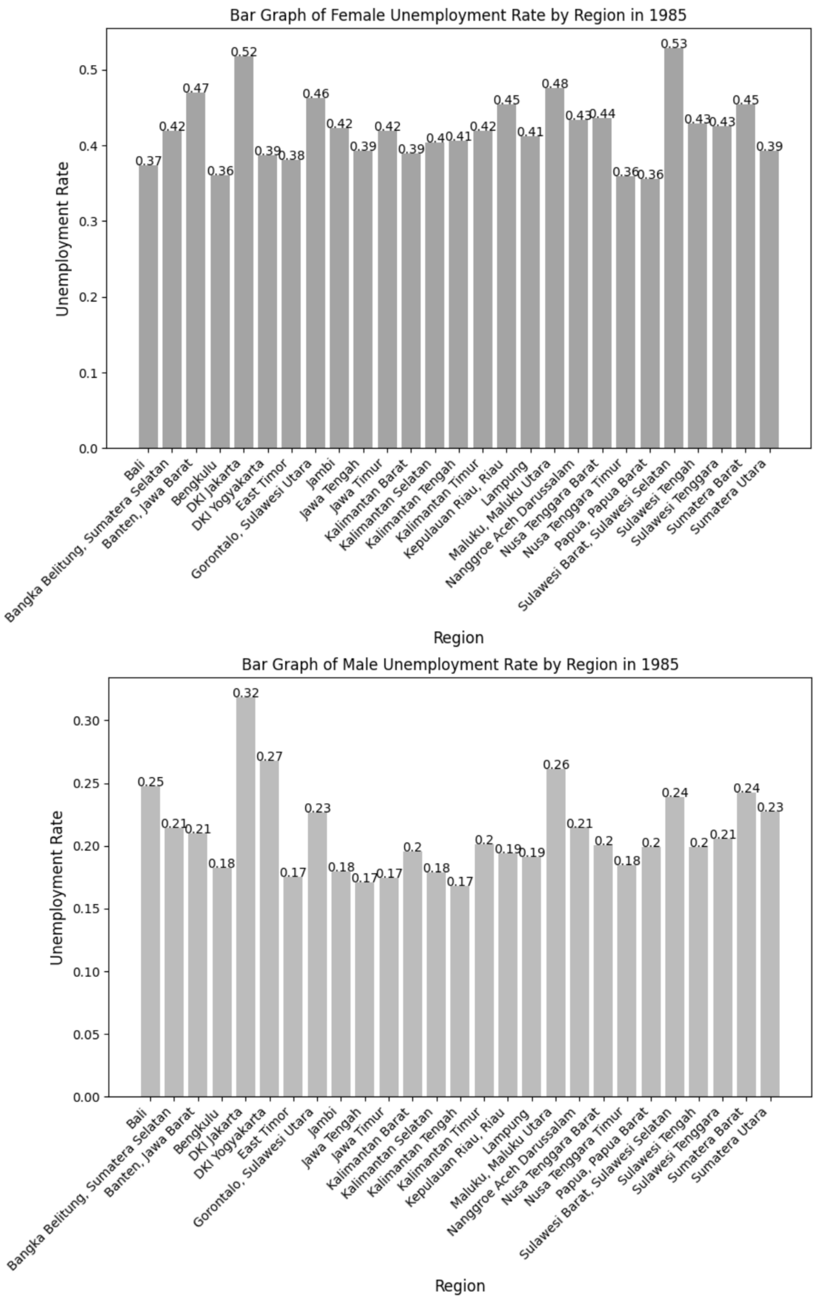

Figure 9: Bar graph comparison between unemployment rates of female and male individuals in Indonesia by provinces in 1985.

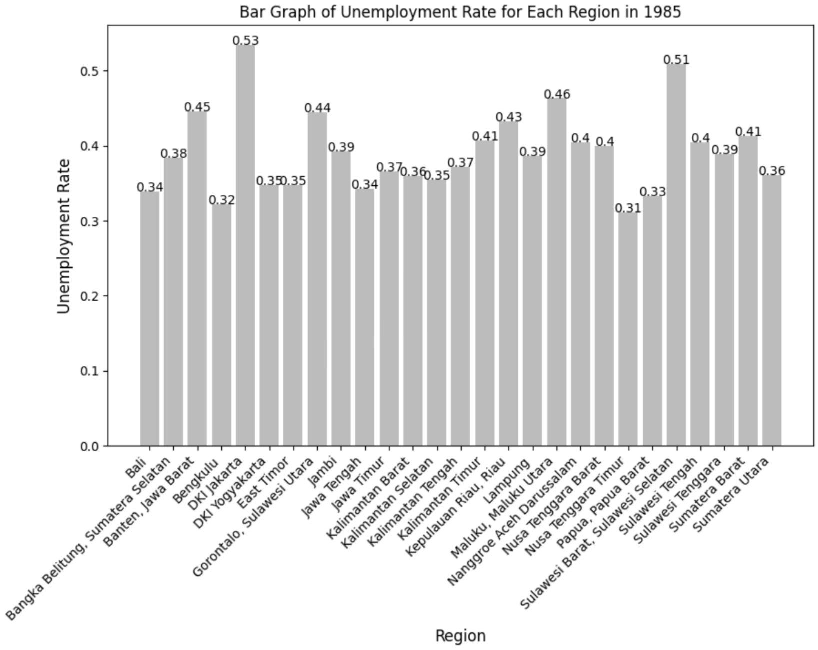

Figure 10: Bar graph of unemployment rate for each province in 1985.

4.1.5 1990

It can be observed that the overall unemployment rates are slightly higher across all regions. This is because Indonesia observed the beginning of an economic recession starting in the 1990s. Future sections Will explore the same features through regression modelling for temporal comparison.

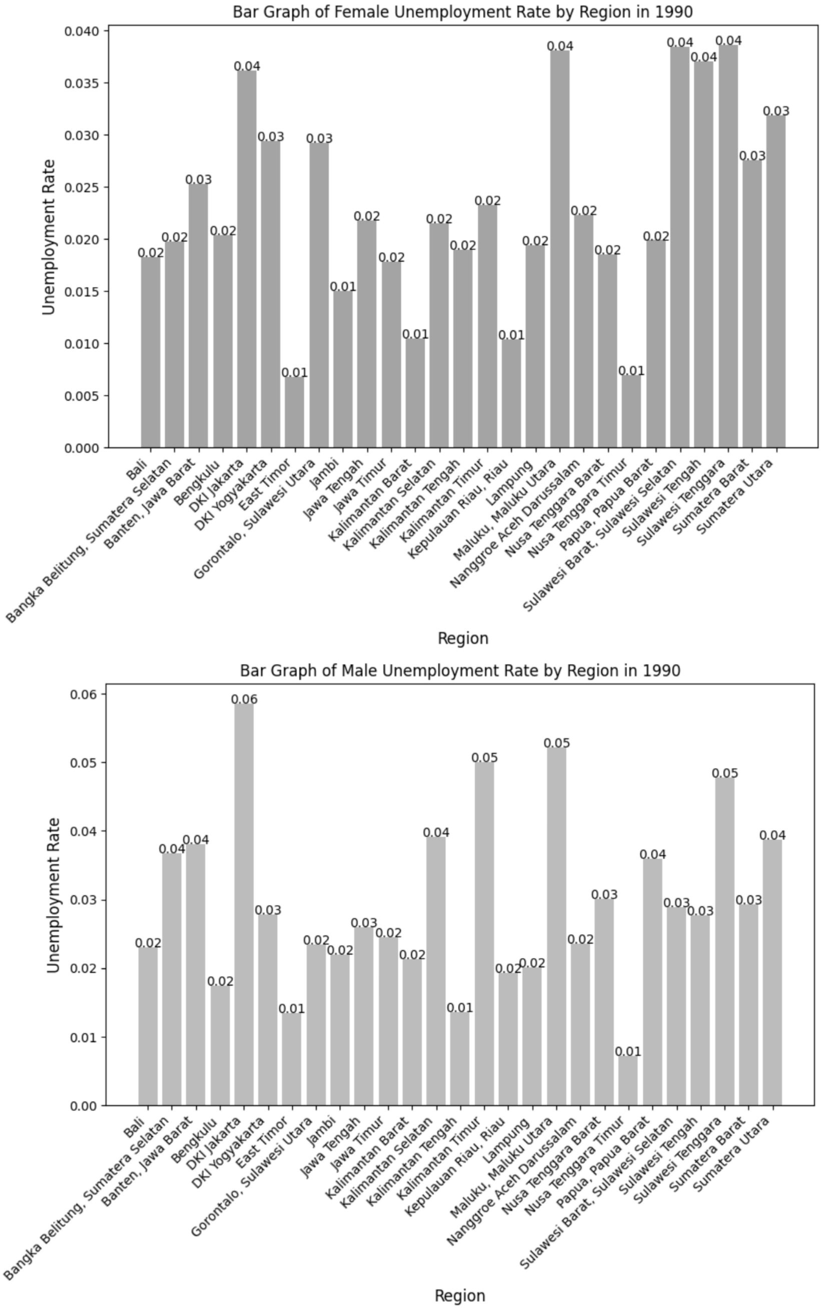

Figure 11: Bar graph comparison between unemployment rates of female and male individuals in Indonesia by provinces in 1990.

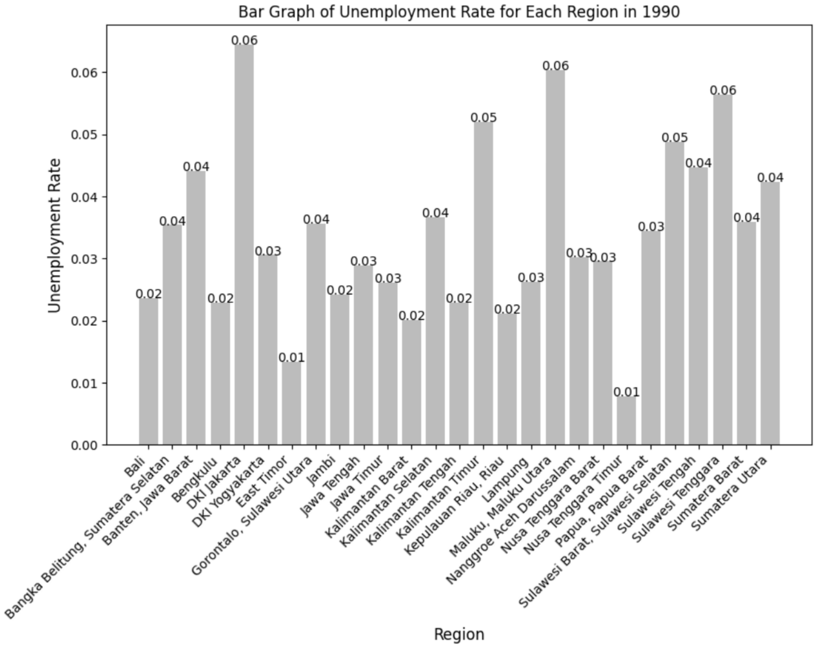

Figure 12: Bar graph of unemployment rate for each province in 1990.

4.1.6 1995

Indonesia observed one of its most serious recession in the late 1990's. The high unemployment rates between genders are an indication of adherence of the expected trend. In fact, unemployment rate for women seem higher than unemployment rate for men across provinces.

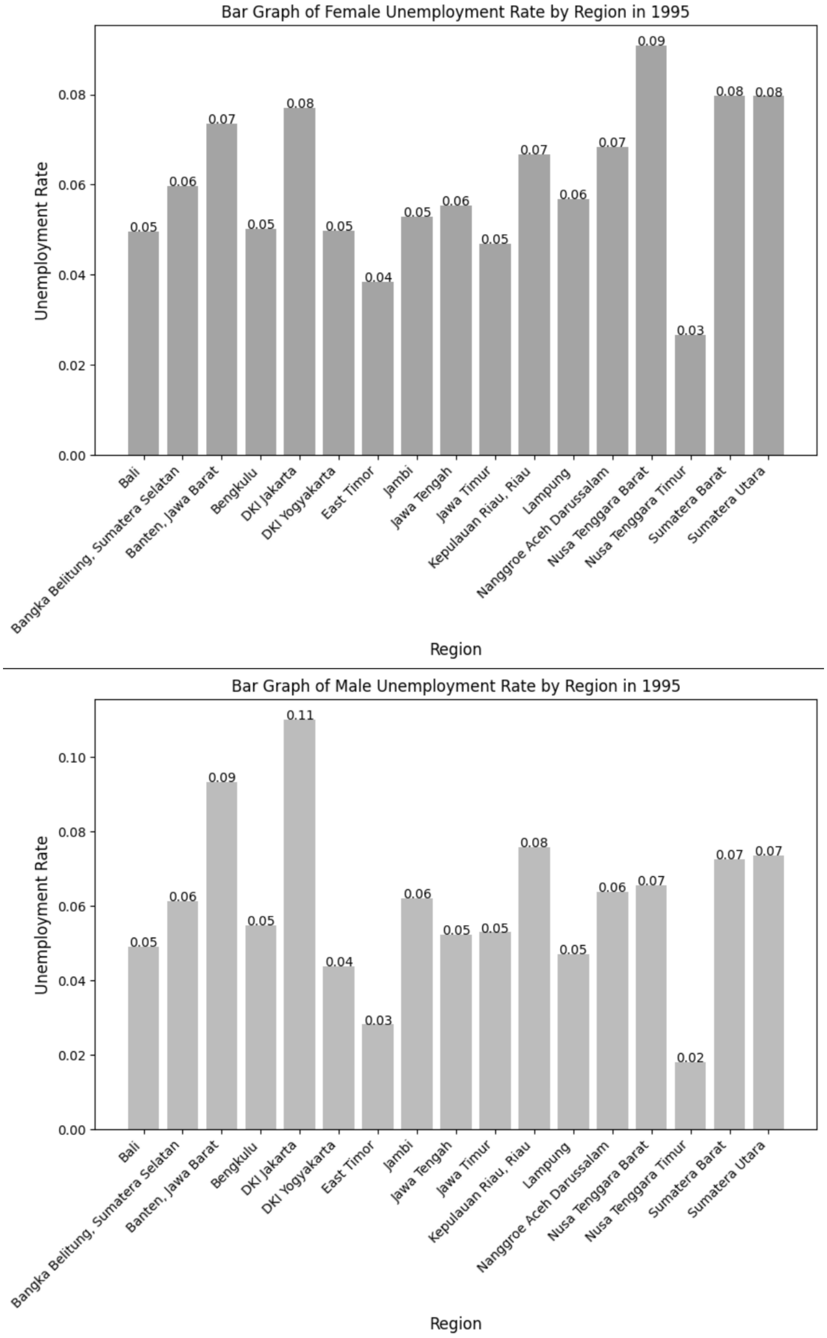

Figure 13: Bar graph comparison between unemployment rates of female and male individuals in Indonesia by provinces in 1995.

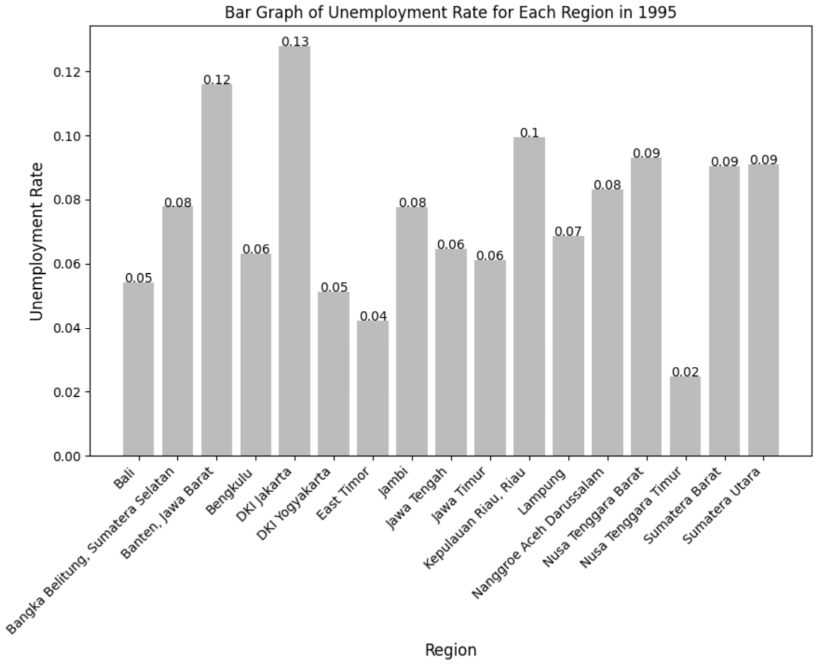

Figure 14: Bar graph of unemployment rate for each province in 1985.

4.2 General Visualizations

Figures 15 and 16 are the line graphs of unemployment rate plotted against years for female individuals, male individuals and the general population. It can be observed that there is a high unemployment rate peak at the year 1985 and a slight upward trend in the year 1995. These are due to economic recessions in Indonesia caused primarily by political and governmental disputes and violent conflicts. The provinces with the highest unemployment rate also tends to be large cities in Indonesia such as DKI Jakarta, for instance, the province with the highest unemployment rate for male individuals. Prior to 1985, female individuals saw similar unemployment rates than male individuals; however, the 1985 recession in Indonesia seems to have a larger impact on females for the majority of the provinces than males. This evidence motivates further exploration of gender and economic recessions.

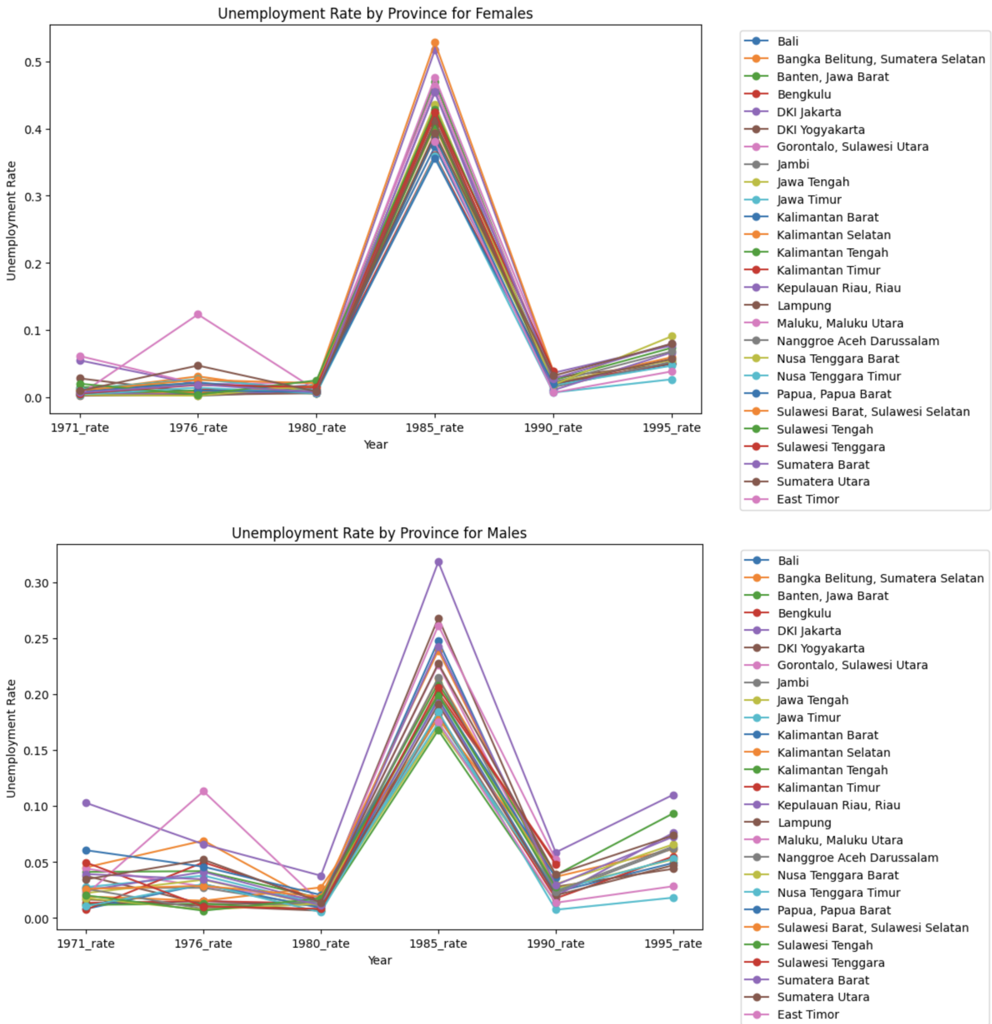

Figure 15: Line graph comparison between unemployment rates of female and male individuals in Indonesia by provinces.

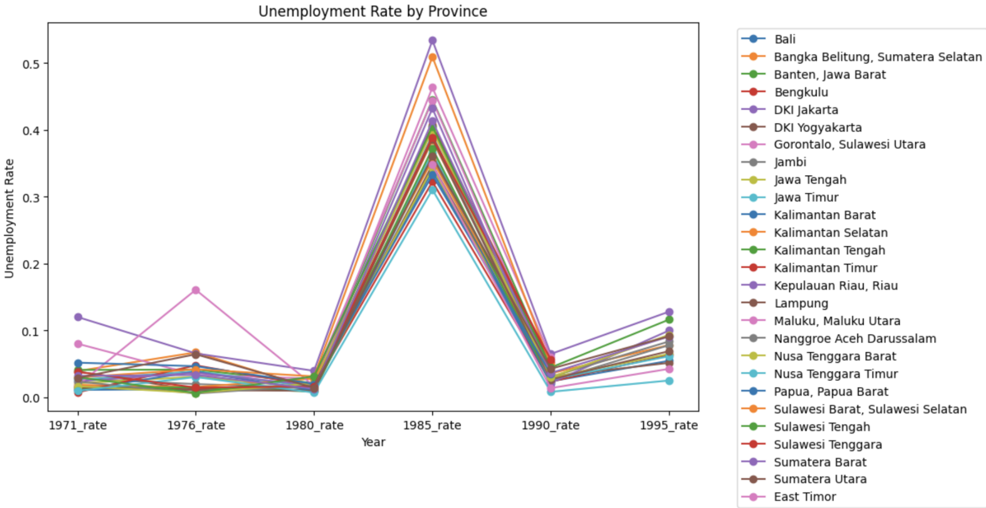

Figure 16: Line graph of unemployment rate for each province.

5 Methodology

5.1 Discussion of Linear Regression

In order to understand the effects of potentially influential factors such as education, age, family size, marital status, and gender on an individual's employment status, I implemented the linear probability model to conduct preliminary exploration of the relationship between selected variables. The linear probability model is a binary regression model that predicts the binary class of the dependent variable based on selected independent variables.

To correctly implement the baseline model, the dependent variable must not only produce binary outcomes and have linear relationship with the explanatory variables, data must also be homoscedastic and independent variables must avoid multicollinearity.

The linear model holds the following general equation:

$$\hat{p} = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + ... + \beta_n \cdot x_n \tag{2}$$

This model assumes that the probability p, as detailed by equation (2), is a linear function of the regressors, which are indicated by the X's. Ultimately, the linear model can predict probabilities that fall beyond the [0,1] range as it assumes that the independent variables have a linear relationship with the probability of the outcome. Due to this, the violation of probability bounds may lead to inconsistent and biased parameter estimates. However, the linear probability baseline model would allow easy interpretation of the results.

5.2 Model 1: Interaction of Recession on Variables of Interest for Linear Regression

Model 1 aims to investigate the effects of an economic recession on *MARST, NCHILD, SEX, EDATTAIN* and *AGE* when regressing against *EMPSTAT*. Model 1 is defined as the following:

$$\begin{aligned} EMPSTAT = {} & \beta_0 + \beta_1 \cdot YEAR + \beta_2 \cdot AGE + \beta_3 \cdot SEX + \beta_4 \cdot NCHILD \\ & + \beta_5 \cdot MARST + \beta_6 \cdot EDATTAIN\_2 + \beta_7 \cdot EDATTAIN\_3 \\ & + \beta_8 \cdot EDATTAIN\_4 + \beta_9 \cdot REC + \beta_{10} \cdot REC \cdot AGE \\ & + \beta_{11} \cdot REC \cdot SEX + \beta_{12} \cdot REC \cdot MARST \\ & + \beta_{13} \cdot REC \cdot EDATTAIN\_2 + \beta_{14} \cdot REC \cdot EDATTAIN\_3 \\ & + \beta_{15} \cdot REC \cdot EDATTAIN\_4 + \beta_{16} \cdot REC \cdot NCHILD \end{aligned} \tag{3}$$

The intercept term, β_0, represents the baseline value of *EMPSTAT* when all other predictor variables are zero. In this context, it might indicate the expected value of *EMPSTAT* when the year (*YEAR*) is at its reference level. The coefficient β_1 associated with *YEAR* suggests the average change in *EMPSTAT* for a one-unit increase in the year, capturing any underlying temporal trends in the data. The coefficients β_2 through β_5 correspond to the variables *AGE*, *SEX*, *NCHILD*, and *MARST*, respectively. These coefficients reveal the impact of each variable on *EMPSTAT*. For instance, a positive coefficient for *AGE* (β_2) implies that as the age of the individual increases, their *EMPSTAT* is expected to increase, holding other variables constant. Conversely, a negative coefficient for *SEX* (β_3) suggests that being male is associated with a decrease in *EMPSTAT*. Similar interpretations can be applied to the other coefficients.

The model includes interaction terms, such as $\beta_{10} \cdot REC \cdot AGE$ and $\beta_{13} \cdot REC \cdot$ *EDATTAIN_2*. These terms capture the joint influence of two variables on *EMPSTAT*. For instance, β_{10} represents the additional impact of the interaction between *REC* and *AGE* on *EMPSTAT*. Similarly, β_{13} captures the joint effect of *REC* and having an educational attainment level of 2 ($EDATTAIN_2$). Interpretation of interaction terms requires careful consideration of the involved variables. The coefficients for education levels ($EDATTAIN_2$, $EDATTAIN_3$, $EDATTAIN_2$) indicate how each level affects *EMPSTAT* relative to the reference level. These interpretations provide a general understanding of how changes in the predictor variables are associated with changes in the predicted variable *EMPSTAT* based on the given coefficients.

5.3 Discussion of Logistic Regression

To further explore the relationship between related explanatory variables and the unemployment status of individuals, I used the logit model to predict employment status. The logit model is a type of regression model commonly used in statistics and econometrics to analyze the relationship between a binary dependent variable and one or more independent variables. It is particularly suited for situations where the outcome of interest is binary, meaning that it can take on only two possible values, such as "success" or "failure," "yes" or "no," or "1" or "0." The logit model allows researchers to estimate the probability of the binary outcome as a function of the independent variables while accounting for the non-linear nature of binary data.

Logit model can be expressed in the following equation:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = B_0 + B_1 x_1 + \ldots + B_k x_k \tag{4}$$

Where *ln* denotes the natural logarithm function, p represents the probability of the binary outcome, usually denoted as the probability of success (coded as 1) in binary data analysis, B_0 is the intercept term, B_1,...,B_k are coefficients of independent variables of x_1,...,x_k and x_1,...,x_k are the independent variables that influence log odds of the binary outcome $EMPSTAT$.

One of the key features of the logit model is its ability to model the log odds or logit transformation of the probability of the binary outcome. The logit transformation is the natural logarithm of the odds, where the odds represent the ratio of the probability of the event occurring to the probability of it not occurring. By taking the logit transformation, the logit model ensures that the relationship between the independent variables and the probability of the binary outcome is linear, making it a valuable tool for understanding the effects of various predictors on the likelihood of an event.

The logit model ensures that the predicted probabilities are bounded between zero and one, making it suitable for modeling probabilities. This characteristic is particularly useful when dealing with binary outcome data, as it guarantees that the estimated probabilities remain within a valid range.

5.4 Model 2: Interaction of Recession on Variables of Interest for Logistic Regression

Model 2 for logit modeling is defined as the following:

$$\begin{aligned}P(EMPSTAT = 1) = e^{-(}&\beta_0 + \beta_1 \cdot YEAR + \beta_2 \cdot AGE_{ip}^t + \beta_3 \cdot SEX_{ip}^t + \beta_4 \cdot NCHILD_{ip}^t \\ &+ \beta_5 \cdot MARST_{ip}^t + \beta_6 \cdot EDATTAIN\_2_{ip}^t + \beta_7 \cdot EDATTAIN\_3_{ip}^t \\ &+ \beta_8 \cdot EDATTAIN\_4_{ip}^t + \beta_9 \cdot REC + \beta_{10} \cdot REC \cdot AGE_{ip}^t \\ &+ \beta_{11} \cdot REC_{ip} \cdot SEX_{ip}^t + \beta_{12} \cdot REC_{ip} \cdot MARST_{ip}^t \\ &+ \beta_{13} \cdot REC_p \cdot EDATTAIN\_2_{ip}^t + \beta_{14} \cdot REC_p \cdot EDATTAIN\_3_{ip}^t \\ &+ \beta_{15} \cdot REC_p \cdot EDATTAIN\_4_{ip}^t + \beta_{16} \cdot REC_p \cdot NCHILD_{ip}^t) + \epsilon_{ip}\end{aligned} \tag{5}$$

In the logistic regression equation P(EMPSTAT = 1) where P represents the probability of the binary outcome *EMPSTAT* being equal to 1, each variable and coefficient plays a crucial role in influencing this probability. I now provide interpretations of the variables and coefficients:

- The intercept term β_0 represents the log-odds of the baseline probability of *EMPSTAT* being 1 when all other predictor variables are zero. It captures the baseline likelihood of the positive outcome.

- The coefficient β_1 is associated with $YEAR$ signifies the average change in the log-odds of $EMPSTAT$ being 1 for a one-unit increase in the variable $YEAR$. It indicates the impact of temporal trends on the likelihood of the positive outcome.

- The coefficient β_2 represents the change in the log-odds of $EMPSTAT$ being 1 for a one-unit increase in AGE. A positive β_2 suggests that as age increases, the log-odds of the positive outcome also increase.

- The coefficient β_4 represents the change in the log-odds of $EMPSTAT$ being 1 for a one-unit increase in the variable $NCHILD$. A positive β_4 implies that a higher number of children is associated with an increase in the log-odds of the positive outcome.

- The coefficient β_5 associated with $MARST$ indicates the change in the log-odds of $EMPSTAT$ being 1 for a change in marital status. Positive or negative β_5 values indicate corresponding increases or decreases in the log-odds.

- The coefficients β_6 to β_8 associated with $EDATTAIN_2$, $EDATTAIN_3$, and $EDATTAIN_4$ represent the impact of different education levels on the log-odds of $EMPSTAT$ being 1 compared to the reference level. The coefficientsβ_9 to β_16 represent the impact of interactions between REC and other variables on the log-odds of $EMPSTAT$ being 1. For example, β_10 captures the additional effect of the interaction between recruitment and age.

- $YEAR$: This variable captures the the effect of time on the probability of employment. The coefficient β_1 represents the change in the log-odds of employment for a one-unit increase in the year. Ultimately this serves as a linear trend in time to adjust for the lack of linearity in the logit regression.

- AGE_{ip}^{t}: This term accounts for the influence of age on employment status. β_2 reflects the change in the log-odds of employment for a one-unit increase in age for an individual i in province p and year t.

- SEX_{ip}^{t}: Represents the impact of gender on employment status. β_3 signifies the change in the log-odds of employment for a one-unit change in the gender variable for individual i in province p and for year t.

- $NCHILD_{ip}^{t}$: Captures the effect of the number of children on employment. $beta_4$ represents the change in the log-odds of employment for a one-unit increase in the number of children for individual i in province p and year t.

- $MARST_{ip}^{t}$:Accounts for the influence of marital status on employment. $beta_5$ signifies the change in the log-odds of employment for a one-unit change in marital status for individual i in province p and year t.

- $EDATTAIN\_2_{ip}^{t}$, $EDATTAIN\_3_{ip}^{t}$, $EDATTAIN\_4_{ip}^{t}$: These variables represent different levels of educational attainment, and the corresponding coefficients ($beta_6$, $beta_7$, $beta_8$) capture the change in the log-odds of employment for a one-unit increase in each level of education for individual i in province p and year t.

- REC: This is the main effect of the recession dummy variable. The coefficient $beta_9$ represents the change in the log-odds of employment when transitioning from a non-recessionary period to a recessionary period, holding all other variables constant.

- The rest of interaction terms between the recession dummy variable and other individual-specific variables can be interpreted with their corresponding coefficients. The corresponding coefficients (β_{10}, β_{11}, etc) indicate how the relationship between employment status and these variables changes during a recession compared to a non-recessionary period. For example, β_{10} explains the effect of age on the log-odds of employment differs during a recession.

Ultimately, the inclusion of the recession dummy variable and its interactions allows the model to capture whether the probability of employment is influenced by economic recessions and whether the effects of other variables vary during recessionary periods. It provides a more nuanced understanding of how economic conditions may impact individual employment probabilities than Model 1.

5.5 Discussion of Panel Data Regression

Panel data, characterized by observations on multiple entities over time, is able to model Ordinary Least Squares (Panel OLS) regression. One compelling motivation to employ Panel OLS is its ability to account for both individual heterogeneity and temporal dynamics simultaneously. By capturing individual-specific effects and time-specific trends, Panel OLS can present intersectional relationships within the data, provide a more nuanced understanding of underlying cross-sectional variations but also temporal trends, and offer a holistic perspective crucial for robust empirical analysis.

Furthermore, Panel OLS regression enhances statistical efficiency by leveraging panel structure to present inherent correlation between observations across entities and time periods. This approach allows higher parameter precision and increased statistical power. This efficiency gain becomes particularly valuable when dealing with datasets that exhibit both cross-sectional and temporal dependencies, as it mitigate the potential biases arising from omitted variable problems and unobserved heterogeneity.

In essence, the adoption of Panel OLS regression is a methodological imperative for this paper to extract comprehensive insights from panel data while ensuring the reliability and validity of their empirical findings.

The following equation presents the general structure of a panel OLS regression:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + \epsilon_{it} \tag{6}$$

- Y_{it}: This denotes the dependent variable, where the subscript i represents the individual entity or cross-sectional unit, and t represents time or the temporal dimension. Thus, Y_{it} refers to the observed value of the dependent variable for entity i at time t.

- β_0: This is the intercept term, representing the constant or baseline value of the dependent variable when all independent variables (X_{it}) are zero. It captures the average or baseline level of Y across all entities and time periods.

- β_1: This is the coefficient associated with the independent variable X_{it}, reflecting the marginal effect of a one-unit change in X_{it} on Y_{it}. It measures the strength and direction of the relationship between the dependent variable and the independent variable.

- X_{it}: This represents the independent variable, and like the dependent variable, it has both a cross-sectional subscript i and a time subscript t, denoting the observed values for entity i at time t.

- α_i: This is the individual or entity-specific fixed effect, capturing unobservable characteristics or heterogeneity that are constant across time. It accounts for the individual-specific variation that is not explained by the observed variables.

- ϵ_i: This term represents the error or disturbance term, accounting for unobserved factors affecting Y_{it} that are not captured by the model. It includes random or stochastic elements and reflects the residual or unexplained part of the dependent variable.

5.6 Model 3: Panel Data

Model 3 of the panel data regression is defined as the following:

$$\begin{aligned} EMPSTAT_{it} = {} & \beta_0 + \beta_1 \cdot YEAR_{it} + \beta_2 \cdot AGE_{it} + \beta_3 \cdot SEX_{it} + \beta_4 \cdot NCHILD_{it} \\ & + \beta_5 \cdot MARST_{it} + \beta_6 \cdot EDATTAIN\_2_{it} + \beta_7 \cdot EDATTAIN\_3_{it} + \beta_8 \cdot EDATTAIN\_4_{it} \\ & + \beta_9 \cdot REC_{it} + \beta_{10} \cdot REC_{it} \cdot AGE_{it} + \beta_{11} \cdot REC_{it} \cdot SEX_{it} \\ & + \beta_{12} \cdot REC_{it} \cdot MARST_{it} + \beta_{13} \cdot REC_{it} \cdot EDATTAIN\_2_{it} \\ & + \beta_{14} \cdot REC_{it} \cdot EDATTAIN\_3_{it} + \beta_{15} \cdot REC_{it} \cdot EDATTAIN\_4_{it} \\ & + \beta_{16} \cdot REC_{it} \cdot NCHILD_{it} + \alpha_i + \varepsilon_{it} \end{aligned} \tag{7}$$

The variables and coefficients can be interpreted as the following:

- $EMPSTAT_{it}$: The dependent variable representing the employment status of individual i at time t.

- β_0: The intercept term represents the average value of $EMPSTAT_{it}$ when all other independent variables are zero. In panel regression, this can be interpreted as the average employment status across all individuals at the reference time point.

- β_1: The coefficient associated with $YEAR_{it}$ indicates the average change in

$EMPSTAT_{it}$ for a one-unit increase in time (year). It captures the overall temporal trend in employment status for all individuals.

- β_2 to β_{16}: These coefficients represent the impact of the corresponding independent variables on the employment status, similar to the interpretation in a cross-sectional regression. For example, β_2 represents the impact of AGE_{it} on $EMPSTAT_{it}$ after accounting for individual and time effects.

- α_i: The individual-specific fixed effects capture unobserved characteristics of each individual that are constant over time. It accounts for individual heterogeneity that does not change over time.

- ϵ_{it}: The error term represents the unobserved factors influencing $EMPSTAT_{it}$ that are not accounted for by the model. In the context of panel regression, it includes individual-specific and time-specific random effects.

6 Results

This section of the paper will present the results of Model 1 and Model 2 in one comparable chart and present the result of Model 3, a panel data regression, on a separate table.

6.1 Interpretation of Linear Regression Results

In Figure 17, the coefficient for $YEAR$ (0.0057) indicates that, on average, each additional year is associated with a 0.0057-unit increase in the expected value of $EMPSTAT$, holding other variables constant. Similarly, the coefficient for AGE

	logit_empstat	EMPSTAT
AGE	0.0126	0.0001
	(0.0013)	(0.0000)
EDATTAIN_2	-88.9872	-2.5756
	(nan)	(0.0238)
EDATTAIN_3	-89.0844	-2.5781
	(nan)	(0.0238)
EDATTAIN_4	-89.0332	-2.5785
	(nan)	(0.0238)
MARST	0.4410	0.0092
	(0.0227)	(0.0005)
NCHILD	0.1166	0.0024
	(0.0062)	(0.0001)
R-squared		0.1877
R-squared Adj.		0.1877
REC	-0.6517	0.0726
	(nan)	(0.0027)
REC*AGE	-0.0070	0.0008
	(0.0016)	(0.0001)
REC*EDATTAIN_2	-0.2365	0.0273
	(nan)	(0.0016)
REC*EDATTAIN_3	-0.1363	0.0289
	(nan)	(0.0018)
REC*EDATTAIN_4	-0.2788	0.0164
	(nan)	(0.0038)
REC*MARST	-0.0625	0.0316
	(0.0277)	(0.0011)
REC*NCHILD	-0.0922	-0.0061
	(0.0079)	(0.0003)
REC*SEX	-1.6629	-0.2654
	(0.0328)	(0.0012)
SEX	-0.7516	-0.0142
	(0.0240)	(0.0006)
YEAR	0.1818	0.0057
	(0.0017)	(0.0000)

Figure 17: Summary result of Linear and Logit regression model (Model 1 and Model 2)

(0.0001) implies a small but positive effect, with each additional year contributing to a 0.0001-unit increase in the expected $EMPSTAT$. Conversely, the coefficient for SEX (-0.0142) highlights that, on average, being male is associated with a decrease of 0.0142 units in the expected value of $EMPSTAT$ compared to being female.

Furthermore, the coefficient for *NCHILD* (0.0024) suggests that having more children is linked to an increase in the expected value of *EMPSTAT* by 0.0024 units. Marital status (*MARST*) also plays a role, as indicated by a positive coefficient of 0.0092, suggesting a change in marital status corresponds to a 0.0092-unit change in the expected value of *EMPSTAT*. The interaction terms, such as *REC***AGE* (0.0008), represent additional effects, signifying the joint impact of recruitment and age on *EMPSTAT*. The overall model's explanatory power, denoted by the R-squared value of 0.1877, implies that approximately 18.77% of the variability in *EMPSTAT* is accounted for by the included variables.

6.2 Interpretation of Logit Regression Results

The logit regression model sheds light on the determinants of the log-odds of *EMPSTAT* being equal to one. The positive coefficient for *YEAR* (0.1818) implies that, on average, each additional unit increase in *YEAR* corresponds to a notable increase in the log-odds of *EMPSTAT* being 1. Similarly, the positive coefficient for *AGE* (0.0126) indicates that, on average, each additional year of age is associated with a modest increase in the log-odds of *EMPSTAT* being one.

Examining the impact of gender, the negative coefficient for *SEX* (-0.7516) suggests that being male ($SEX = 1$) is linked to a significant decrease in the log-odds of *EMPSTAT* being one compared to being female ($SEX = 0$). Additionally, the positive coefficient for *NCHILD* (0.1166) implies that, on average, each additional child (*NCHILD*) corresponds to an increase in the log-odds of *EMPSTAT* being one. The marital status variable (*MARST*) also plays a role, with a positive coefficient of 0.4410, implying that a change in marital status is associated with an impacted increase in the log-odds of *EMPSTAT* being one. Furthermore, the model incorporates interaction terms (*REC*AGE*, *REC*SEX*, *REC*MARST*, etc.) to capture nuanced relationships. For example, *RECAGE* (-0.0070) signifies the additional effect of the interaction between *REC* and *AGE* on *EMPSTAT*.

6.3 Interpretation of Panel Regression Results

The presented panel regression results offer insights into the factors influencing the binary outcome of *EMPSTAT*. Beginning with individual predictors, the coefficient for *SEX* is -0.0085 with a p-value of 0.4521, indicating that gender does not exhibit a statistically significant association with *EMPSTAT*. This implies that, on average, being male does not significantly impact the log-odds of *EMPSTAT* being one compared to being female. In contrast, the variable *REC:SEX* has a more notable effect, with a coefficient of -0.2788 and a statistically significant p-value of 0.0239. This suggests that the interaction between recruitment status (*REC*) and gender has a negative association with the log-odds of *EMPSTAT* being one, implying potential gender-specific influences in the recruitment context.

Moving to the marital status variable (*MARST*), the coefficient is 0.0081 with a p-value of 0.8594, indicating that the main effect of marital status is not statistically significant. However, the interaction term *REC:MARST* has a significant positive effect (coefficient of 0.0292, p-value of 0.0005), suggesting that the joint impact of recruitment and marital status has a positive association with the log-odds of *EMPSTAT* being one. The variable *REC:NCHILD* demonstrates a statistically significant negative association with a coefficient of -0.0072 and a low p-value of 0.0001. This implies that the interaction between recruitment status and the number of children has a negative effect on the log-odds of *EMPSTAT* being one.

6.4 Limitations

This section acknowledges limitations that may influence the generalizability of the findings. The presented models rely on the assumption of linearity in the relationships between predictor variables and the binary outcome. Additionally, the study's findings are contingent on the dataset's representativeness and the comprehensiveness of the included variables, and the observed associations do not imply causation. Future research could address these limitations by incorporating additional relevant variables, employing more sophisticated modeling techniques, and validating findings across diverse datasets, thereby enhancing the robustness and generalizability of the study's conclusions.

7 Conclusion

In conclusion, the linear regression model provides valuable insights into the determinants of the binary outcome variable *EMPSTAT*. Notably, the coefficients associated with *YEAR* and *AGE* suggest that each additional year contributes positively to the expected value of *EMPSTAT*, albeit with a small impact. On the other hand, the coefficient for *SEX* indicates that being male is associated with a decrease in the expected value of *EMPSTAT* compared to being female. The inclusion of interaction terms, such as *REC*AGE*, reveals additional effects, highlighting the joint impact of recruitment and age on *EMPSTAT*. The overall model's explanatory power, denoted by the R-squared value, underscores that approximately 18.77% of the variability in *EMPSTAT* is explained by the included variables. However, interpretations should be made within the study's context, considering statistical significance and precision indicated by standard errors.

Turning to the logit regression model, it elucidates the log-odds of *EMPSTAT* being equal to 1. Key findings include the positive coefficients for *YEAR* and *AGE*, indicating that each additional unit increase in these variables corresponds to a notable increase in the log-odds of *EMPSTAT* being 1. The negative coefficient for *SEX* suggests a significant gender effect, with being male associated with a substantial decrease in the log-odds of *EMPSTAT* being 1 compared to being female. Interaction terms, such as *REC*AGE*, capture nuanced relationships, providing insights into the complex interplay of factors influencing the likelihood of *EMPSTAT* being 1. These results deepen the understanding of the intricate dynamics affecting the binary outcome.

The panel regression results further contribute to the understanding of *EMPSTAT*, emphasizing the importance of considering both main effects and interaction terms. While the main effect of *SEX* is not statistically significant, the interaction term *REC:SEX* reveals a notable gender-specific influence in the recruitment context. Similarly, the interaction between *REC* and marital status (*MARST*) impacts the log-odds of *EMPSTAT* being 1, underscoring the need to account for these joint effects. The negative association between *REC:NCHILD* highlights the relevance of considering the interaction between recruitment status and the number of children.

Bibliography

Brinton, Mary C., et al. “Married Women’s Employment in Rapidly Industrializing Societies: Examples from East Asia.” American Journal of Sociology, vol. 100, no. 5, 1995, pp. 1099–130. JSTOR, http://www.jstor.org/stable/2782272. Accessed 3 Feb 2023.

Caraway, Teri L. “Labor Rights in East Asia: Progress or Regress?” Journal of East Asian Studies, vol. 9, no. 2, 2009, pp. 153–86. JSTOR, http://www.jstor.org/stable/23418972. Accessed 12 April 2023.

Fukui, Masao, et al. “Women, Wealth Effects, and Slow Recoveries.” NBER, 3 Dec. 2018, www.nber.org/papers/w25311.

Ghoshal, Baladas. "Indonesia in 1985: A Year of Trials". Southeast Asian Affairs 1986, edited by Joo-Jock Lim, Singapore: ISEAS Publishing, 1986, pp. 99-121. https://doi.org/10.1355/9789812306760-009

IPUMS CPS, University of Minnesota, www.ipums.org.

Johnston, Melissa, Davies, Sara E., True, Jacqui, and Riveros-Morales, Yolanda (2021). "Patriarchal reset" in the asia pacific during COVID-19: the impacts on women's security and rights. The Pacific Review 36 (3) 1-28. https://doi.org/10.1080/09512748.2021.2005123

Nakamura, Emi, et al. “Are Chinese Growth and Inflation Too Smooth? Evidence from Engel Curves.” American Economic Journal: Macroeconomics, vol. 8, no. 3, 2016, pp. 113–44. JSTOR, http://www.jstor.org/stable/24739426.

Wei-Hsin Yu. “Changes in Women’s Postmarital Employment in Japan and Taiwan.” Demography, vol. 42, no. 4, 2005, pp. 693–717. JSTOR, http://www.jstor.org/stable/4147335. Accessed 11 January 2023.