PPD-558 Class Project: Data and Methods

Yi Ming, Arin Fisher, Pablo Liu

October 22, 2024

# 1. Data

### 1.1 Data Source Description

In this project, we will use this dataset [CLEAR INSIGHT Database](https://www.uscclear.com/insight/datasetcodebook). This dataset is compiled and collected by researchers from the USC Civic Leadership Education and Research (CLEAR) Initiative. The data mainly includes the following part:

* BLS: The Bureau of Labor Statistics’s Occupational Employment and Wage Statistics (OEWS) dataset.
* OPM FedScope Cube: A data cube brings together key dimensions (data elements) of the Federal workforce provided by the Office of Personnel Management (OPM)

### 1.2 Dependent and Explanatory Variables

#### 1.2.1 Dependent Variable:

Our dependent variable is **Vacancy Time Interval (posting\_length)**, which measures recruiting difficulty by counting how many days the position has been vacant.

To construct the **Vacancy Time Interval** (the dependent variable), we calculate the difference between the job's open date and close date, based on these two given variables in the dataset:

* **announcement\_open\_date**: The date when the job vacancy was opened.
* **announcement\_close\_date**: The date when the job vacancy was closed.

We compute the vacancy time interval as the number of days between the open and close dates. This can be done by converting both fields to DateTime format and then subtracting the announcement\_open\_date from the announcement\_close\_date to get the vacancy duration.

#### 1.2.2 Explanatory Variables:

* **total\_applications**: The total number of applications could reflect the level of interest in the position. Fewer applications could signal recruitment difficulty.
* **remote\_work\_available**: A binary variable Indicates whether remote work is an option,. Jobs that offer remote work might be easier to fill, especially in technical fields.
* **grade**: Pay level for each job based on the level of difficulty, responsibility, and qualifications required. The range is from 0 to 15.
* **appointment\_type**: The type of appointment (e.g., permanent, temporary) could affect recruitment, as candidates may be more or less interested depending on the job stability.
* **bls\_occ\_mean\_hmean**: The mean hourly wage for the occupation, which could impact how competitive the position is in the labor market.
* **bls\_occ\_mean\_amean:** The mean annual wage for the occupation
* **bls\_occ\_avg\_jobs1000:** The average number of job postings per 1000 job postings for a given occupation. It shows how prevalent a particular occupation is compared to all other jobs in the dataset.
* **bls\_occ\_avg\_locquot:** The average location quotient (LQ) for a given occupation. LQ is a measure that indicates how concentrated a particular industry or occupation is in a region compared to the national average.

#### 1.2.3 Control Variables:

* **stem**: STEM-related occupation (otherwise = 0). STEM jobs typically face more recruitment challenges due to higher skill requirements.
* **health**: Health-related occupation (otherwise = 0)
* **agency**: Agencies(departments) may have varying levels of recruitment difficulty.
* **job\_series\_title:** The title of the occupation

### 1.3 Stata Summary of the Data

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

job\_postin~d | 1,941,772 1176717 657109.1 48414 2309589

posting\_le~h | 1,941,772 42.94201 85.78723 0 871

department | 0

job\_series~e | 0

job\_series~r | 0

-------------+---------------------------------------------------------

grade\_simp~d | 1,941,772 9.162832 11.09169 0 1008

pay\_plan\_g~y | 1,941,772 .8397732 .3668164 0 1

appointmen~y | 1,941,772 .8559115 .3511795 0 1

total\_appl~s | 1,941,772 199.3298 1109.441 0 42546

stem | 1,941,772 .2005956 .4004461 0 1

-------------+---------------------------------------------------------

health | 1,941,772 .2105706 .407714 0 1

remote\_wor~e | 1,941,772 0 0 0 0

bls\_occ~1000 | 1,941,772 49.36404 45.95839 .049 169.329

bls\_occ\_av~t | 1,941,772 1.024854 .2365573 .25 4.936666

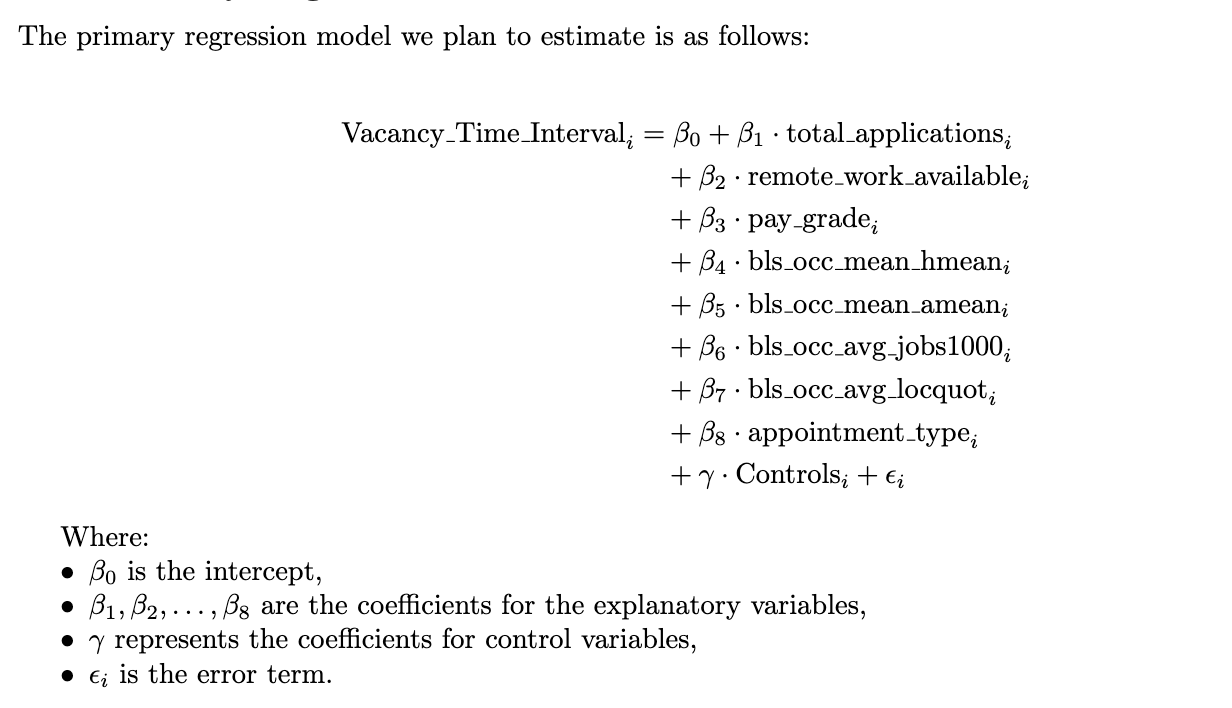
bls\_oc~hmean | 1,941,772 31.65928 10.05249 12.38 69.07

-------------+---------------------------------------------------------

bls\_oc~amean | 1,941,772 66046.43 21083.2 25750 143670

# 2. Empirical Methods

### 2.1 Primary Regression Model



### 2.2 Justification for the Variables

* ​​**total\_applications**: Reflects the level of interest in the position. Fewer applications may indicate recruitment difficulty.
* **remote\_work\_available**: Remote work options could increase the attractiveness of a position, especially for technical jobs.
* **grade**: Higher grade levels may correlate with longer vacancy durations due to increased qualifications and responsibilities.
* **bls\_occ\_mean\_hmean** and **bls\_occ\_mean\_amean**: Wages reflect market competitiveness; higher wages could decrease vacancy time.
* **bls\_occ\_avg\_jobs1000**: Indicates the prevalence of job postings for a given occupation, which could influence vacancy times.
* **bls\_occ\_avg\_locquot**: Measures the regional concentration of occupations, which may impact the pool of available applicants.
* **appointment\_type**: The type of appointment may influence how attractive the job is to potential applicants.

### 2.3 Supplementary Analyses

We plan to conduct diagnostic tests, including checking for multicollinearity using variance inflation factors (VIF) and residual diagnostics for heteroscedasticity using tests such as Breusch-Pagan.