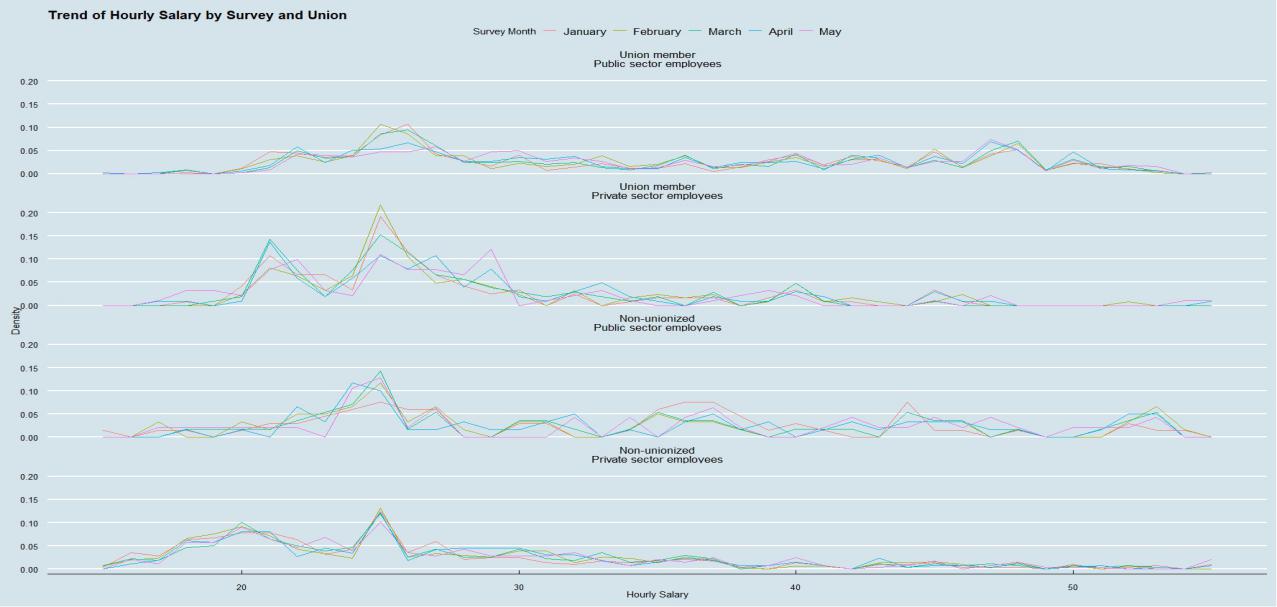
Union Benefit Analysis with Labour Force Survey

Union utilize the collective bargaining process to negotiate with the employer and thereby effectively present union members as a group of a worker who is looking for jobs, compared to the individual worker. In countries like Canada and US, the coverage of the Union is significantly less than those of its Europe counterparts and the union is largely distributed within the public sector. By contrast, in a country such as Scandinavia, with overwhelming coverage of unions, the union essentially becomes the wage maker. In Canada, the power of a union likely lies in if individuals will have a strong ability to obtain a fair wage on their own. (Bhuller et al) With a significant number of immigrant inflow and the relative scarcity of white-collar middle-class jobs in Canada, my assumption is the union will significantly help its member who is in the low to medium-income range to earn more compared to their un-unionized counterpart.

Now with the assumption let’s dive into the data. The data are from the Canada Labour Force Survey (LFS) from 2023 from January to May. The focus is within the province of British Columbia. Firstly, look at the wage distribution by the union and the survey month as below:

Graph. 1- Density distribution of hourly salary by survey and union

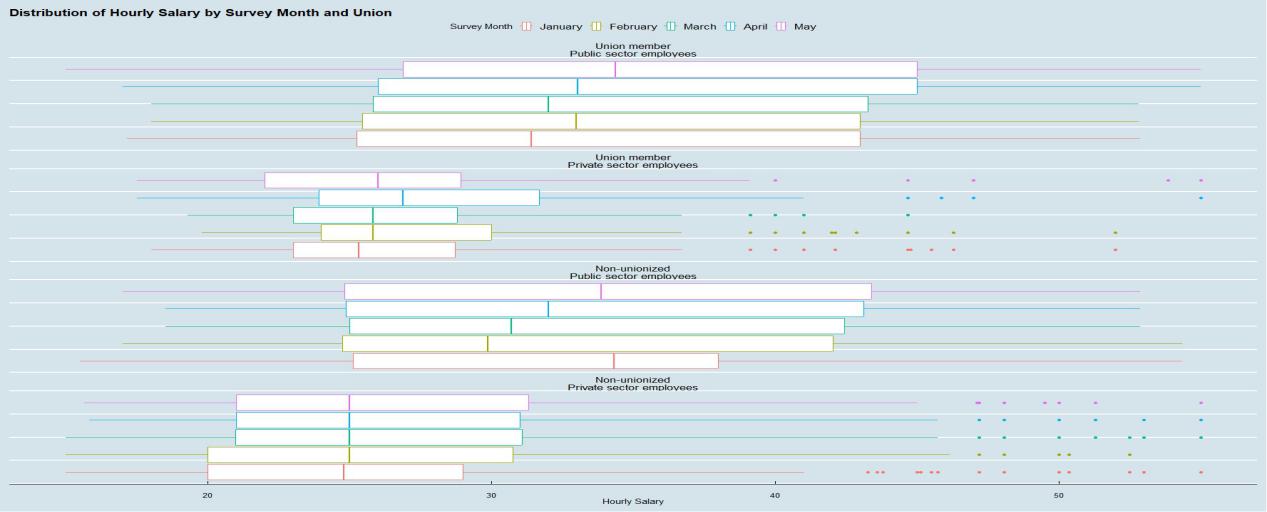


\* Hourly salary (wage) distribution density is displayed on the left and the hourly salary is on the bottom side of the photo. For an easy presentation of the data, this graph only shows the wage range from 15-50. Graph based on ggplot2 package in R and please see corresponding text file for coding details.

It can be seen in the chart, the density reaches the peak for unionized public servants at around $20-$28, and for unionized private session (the second at the top), it does show a higher density near $30 with a highest peak around $25, while the tail in near $20 is also higher. The non-unionized public sector has several peaks with the highest one being near $25, this is similar to employees who are covered under the collective agreement but not union members (not in this graph). Those two groups of individuals can play various roles in different levels of pay such as pay and benefits consultant or deputy ministers. For the non-unionized private sector, we see a plateau from $15 to $25 with a peak around $25 and the density quickly falls to a low level after the $25 per hour. Given in any organization, it is reasonable to assume, there will be a number of low-skill jobs that usually fall within the minimum wage level ($15-$20) (Retail Council of Canada), it can be seen from the graph, the union likely to push the minimum pay to a higher range. Compare to public sector unions who are more tend to cause wages to center in the $20-$25 range, the private sector union tend to push wages higher in each category of job and level of pay without making wage to center in a particular range. However, another explanation for the unionized public sector having fewer minimum wage employees may be due to outsourcing such jobs to unionized employees or contractors. Besides, if we look at unionized private employees, we can see a significant deviation in wages in the May survey, especially in the $20-$25 range and the pink line (May Survey) drifted to a higher wage near $30 from near $25 in January and February and this may due to new collective agreement take effect in April or May 2023 (Lamb et al).

In short, the union in the public sector tends to increase wages in the $15-$25 level and cause a concentration in this range and the union in the private sector may push wages in a wider range without too much concentration. Compare to density, when we further review the quotients in a box plot.

Graph. 2- Box plot distribution of hourly salary by survey and union



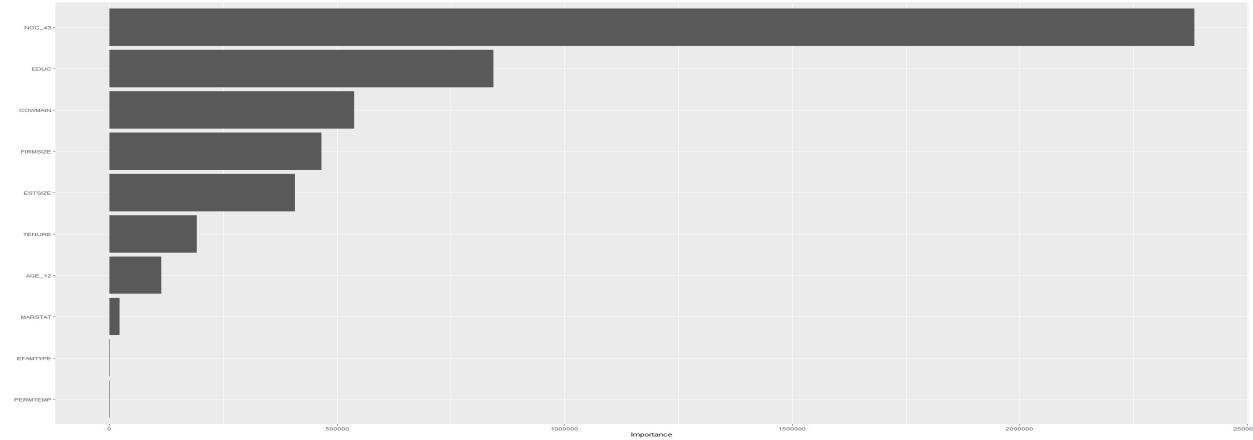
In this box plot, we can see public sector has a higher average pay and the unionization only make a small positive change. Union increase the wage in the private sector for a moderate amount as well. Additionally, there is a good number of out-of-the-box points for the private sector. This indicates higher pay inequality in the private sector.

In summary, the union tends to increase the wage in both the private and public sectors. Unlike the more centred public sector pay structure, the pay structure of the private sector tends to be more sparse. Thereby, the Union is more likely to center the wage in the $20-$25 range than to generate an increase in the wage compared to the private sector.

Based on the discussion in session one above, the union and sector (public, private etc) will affect the wage determination. The first step I took is to review the LFS data set and remove columns that are related to job searching (not related to the topic), most of the family information that should not determine wage and columns that may introduce multicollinearity such as the Age\_6, which has strong correlation with the Age\_12. Similarly, the Noc\_10 is removed due to Noc\_43.

Then, I run a recursive partitioning and regression trees (Breiman, L) under the R package ‘rpart’ to identify factors that correlate with wage the most as below graph:

Graph.3- Recursive Partitioning and Regression Trees



\* From top to bottom: NOC\_43,EDUC,COWMAIN,FIRMSIZE,ESTSIZE,TENURE,AGE\_12,MARSTAT,EFAMTYPE, and PERMTEMP (only the top 10 are presented).

The NOC\_43 as the occupation plays a major factor and then education, sector (public, private etc), the union status and age.

Based on this finding and the data available on the LFS data, I create the formula below:

**HRLYEARN ~AGE\_12+SEX+EDUC+MARSTAT+COWMAIN+IMMIG+PERMTEMP+UNION+FIRMSIZE+TENURE+UHRSMAIN +SURVMNTH+NOC\_43**

Further, it is preferable to say how those factors affect hourly salary (HRLYEARN) by percentage rather than the linear and absolute same amount to each employee but rather by percentage. A good number of papers also support it. (Card et al, Lamb et al). Thereby, a second log-based formula was born:

**log(HRLYEARN) ~ AGE\_12+SEX+EDUC+MARSTAT+COWMAIN+IMMIG+PERMTEMP+UNION+FIRMSIZE+TENURE+UHRSMAIN +SURVMNTH+NOC\_43**

The union and other factor act as dummy variables and the numerical variables are the employment length (Tenure) and the usual work hour (UHRSMAIN). It’s reasonable to assume that with employees accumulating more experience within the company, such individual will cultivate their skill set and can be promoted to a managerial role.

Since the majority of the variable is a factor, it is good to start with a simple linear model, I would like to call it glm with the original hourly wage as the regressand. Comparably, a glmlog model with a log10 transformed hourly wage is created for comparison. The AIC for the glm model arrives at 25018.7 and it is 1118.1 for the glmlog model and BIC is at 25465.2 and 1350.5 respectively. Based on the theory, the smaller the AIC/BIC is the better the model is, the log-transformed model works better in terms of model fit and expected out-of-sample error. This is also supported by the relevant research (Card et al, Lamb et al).

After determining, the log transformation is more suitable for this evaluation, it is time to revisit our models and variables. With 13 regressors, some regressors may not have a significant effect on wage, especially after the occupation (NOC\_43) play a big part in the estimation. Additionally, Since this data covers a number of NOC\_43 and survey months, the data is essentially a panel data set and a panel linear regression is run based on the log-transformed hourly wage. The AIC is at -375.4.

For this step, I pick up the elastic net regressions family and especially the Least Absolute Shrinkage and Selection Operator (LASSO). This regression method looks for variables that contribute little to the overall model accuracy and shrink their coefficient to be exactly zero. This is a great method for model selection and avoid over-fitting the data. surprisingly, the hours of the main job (UHRSMAIN) are removed from the model. It supports the assumption that the employer pays for the skill set by the hour not by per employee. The data quality of the UHRSMAIN can also be the cause or it is masked by the Age factor, where employees in their 50-54 years of old are expected to make 0.04% more than a individual in their 35-39 years of old. In contrast, each hour of paid overtime hour (PAIDOT) seems to decrease wages by approximately 0.001% percent and one hour of unpaid seems to increase salary by 0.006%. This is a reverse of causality, and unpaid overtime maybe already counted in the wage determination. We should be careful of a regressand that potentially affects the regressor, such as with higher wages encouraging employees to stay longer. This will cause an endogeneity issue. (George 2010). Being in the public sector, increase the wage by 0.05% compared to the private sector. A union will, according to the model, reduce the wage by 0.35%. A non-union number but covered by a union contract or collective agreement will have a 0.64% increase in pay compared to non-union members. This may be due to that the executive positions are usually excluded from the collective agreement but they will be affected by a collective agreement. Finally, The employment length turn out to increase wages by 0.001% per year. The LASSO model has an AIC value of -108.4.

Table 1. Model Comparison

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Regressand | HRLYEARN | log(HRLYEARN) | | | |
| Model | Linear | Log Linear | LASSO | Panel | GAM (see below session) |
| Private Sector | -2.427 | -0.062 | ~0 | -0.062 | -0.060 |
| Non Union | 3.028 | 0.053 | N/a | 0.053 | 0.051 |
| AIC | 25,018.710 | 1118.1 | -108.4 | -375.4. | -363.0 |

\*~ 0 means approximately to be 0. The unionized group is considered as the base group excerpt for the LASSO where the opposite happens and it is reported as -0.035.

Finally, to study if a more sophisticated model will capture more patterns in the data set especially the ineffective variable UHRSMAIN. I decided to use the Generalized Additive Model (GAM) The new formula become:

**log(HRLYEARN)~AGE\_12+SEX+EDUC+MARSTAT+COWMAIN+IMMIG+PERMTEMP+UNION+FIRMSIZE+s(TENURE)+s(UHRSMAIN)+SURVMNTH+NOC\_43**

The S() here standards for cubic spline. The model slightly improves the AIC to be -363.0 compared to the regular linear model and the LASSO, but it is lower than the panel model at -375.4. All in all, consider the LASSO is more efficient in variable selection, avoiding over-fitting and has one of the lowest (best) AIC levels. The LASSO model, thereby, is considered the best overall. For out-of-sample Root Mean Squared Error(RMSE), the LASSO achieves a 0.23 RMSE, the linear model achieves a 0.28 RMSE.

For the testing, the variance inflation factor of the data set is all within the range of 1-2, this indicates no significant multi-colinearity. The studentized Breusch-Pagan test shows consistent heteroskedasticity in the model residues. The Shapiro-Wilk normality test on models consistently rejects the residue normality. Durbin Watson test shows no significant auto-correlation. Due to the heteroskedasticity, the heteroskedasticity consistent error is used to evaluate coefficient significance.

Besides the Lasso model explained in the former session, all the coefficient is listed in the appendix. For more detail such as the CART trees and coefficient test results please see the corresponding R code attached to this assignment.

**Further action to improve this study:**

Due to the time and scope limitation of this written assignment, there are a few things I would like to continue to work on.

1. More data exploration such as cluster analysis and random forest analysis. This will help us better understand how variables affect each other and how we select factors.
2. Cross-effect between variables. For example, We know the potential decrease in the wage level from the research and a higher level of wage in the public sector, but how exactly does the union affect a public session? A cross-effect between the public sector and the union can be of interest.
3. The handling of the missing NA data and factor handling could improve. In this assignment, the NA is treated by imputation when treating NA as a new level of factor is considered better. The overwhelming number of the factor used in this study such as NOC 43 caused some calculations to fail, a potential over-fit, and small coefficient values and thereby the study didn’t produce corresponding RMSE results for the Panel model and the GAM model.
4. The use of Log transformation, although is supported by literature and the data, is known to glorify the test results, especially the out-of-sample error RMSE.
5. Witting and formatting in this file can be further improved.

Reference:

Bhuller, Manudeep, Karl Ove Moene, Magne Mogstad, and Ola Lotherington Vestad. “Facts and Fantasies about Wage Setting and Collective Bargaining.” *Journal of Economic Perspectives* 36, no. 4 (November 1, 2022): 29–52. https://doi.org/10.1257/jep.36.4.29.

Abacus Data Network. “Labour Force Survey, 2023,” June 9, 2023. <https://abacus.library.ubc.ca/dataset.xhtml?persistentId=hdl:11272.1/AB2/IJU1QK.>

Support. “Minimum Wage by Province | Retail Council of Canada.” Retail Council of Canada, July 25, 2023. https://www.retailcouncil.org/resources/quick-facts/minimum-wage-by-province/.

Lamb, Danielle, Rafael Gomez, and Milad Moghaddas. “Unions and Hazard Pay for COVID‐19: Evidence from the Canadian Labour Force Survey.” *British Journal of Industrial Relations* 60, no. 3 (November 1, 2021): 606–34. <https://doi.org/10.1111/bjir.12649.>

Breiman, L. (1984). Classification and Regression Trees (1st ed.). Routledge. <https://doi.org/10.1201/9781315139470>

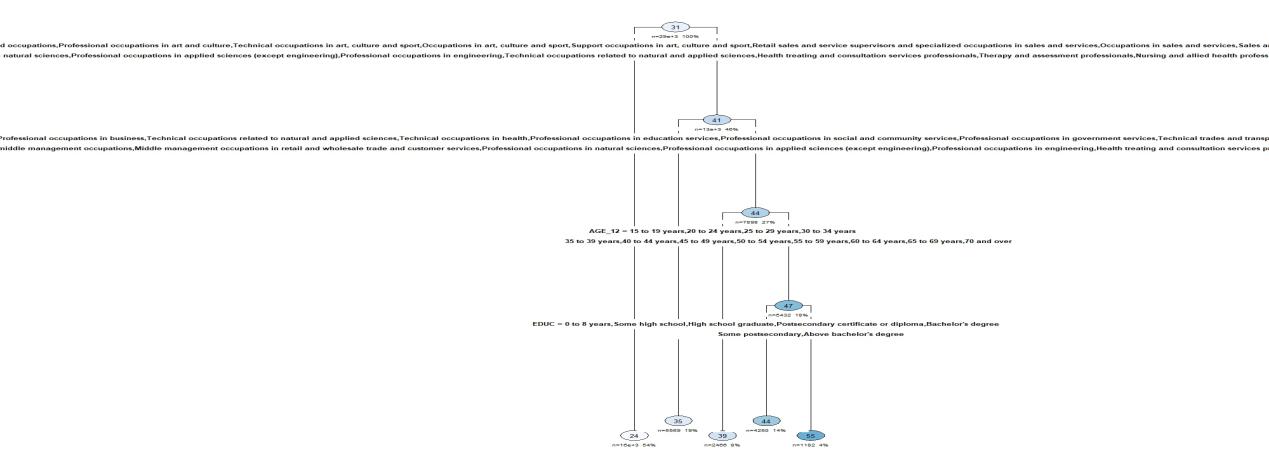
Card, David, Thomas Lemieux, and W. Craig Riddell. “Unions and Wage Inequality: The Roles of Gender, Skill and Public Sector Employment.” *Canadian Journal of Economics* 53, no. 1 (February 1, 2020): 140–73. https://doi.org/10.1111/caje.12432.

George, Linda. *Handbook of Aging and the Social Sciences*. Academic Press, 2010.

Appendix: Table of regression results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | HRLYEARN | log(HRLYEARN) | | | |
|  | Linear Model | Linear Model | | Panel Model | GAM Model |
| AGE\_1220 to 24 years | 1.090 | 0.038 | 0.038 | | 0.034 |
|  | (2.218) | (0.051) | (0.051) | | (0.051) |
| AGE\_1225 to 29 years | 3.626\* | 0.146\*\*\* | 0.146\*\*\* | | 0.134\*\*\* |
|  | (2.184) | (0.050) | (0.050) | | (0.050) |
| AGE\_1230 to 34 years | 3.779\* | 0.162\*\*\* | 0.162\*\*\* | | 0.137\*\*\* |
|  | (2.205) | (0.050) | (0.050) | | (0.051) |
| AGE\_1235 to 39 years | 5.569\*\* | 0.199\*\*\* | 0.199\*\*\* | | 0.174\*\*\* |
|  | (2.219) | (0.051) | (0.051) | | (0.051) |
| AGE\_1240 to 44 years | 7.203\*\*\* | 0.228\*\*\* | 0.228\*\*\* | | 0.207\*\*\* |
|  | (2.230) | (0.051) | (0.051) | | (0.051) |
| AGE\_1245 to 49 years | 7.279\*\*\* | 0.238\*\*\* | 0.238\*\*\* | | 0.219\*\*\* |
|  | (2.242) | (0.051) | (0.051) | | (0.052) |
| AGE\_1250 to 54 years | 8.520\*\*\* | 0.246\*\*\* | 0.246\*\*\* | | 0.228\*\*\* |
|  | (2.231) | (0.051) | (0.051) | | (0.052) |
| AGE\_1255 to 59 years | 6.725\*\*\* | 0.208\*\*\* | 0.208\*\*\* | | 0.190\*\*\* |
|  | (2.234) | (0.051) | (0.051) | | (0.052) |
| AGE\_1260 to 64 years | 4.938\*\* | 0.170\*\*\* | 0.170\*\*\* | | 0.151\*\*\* |
|  | (2.241) | (0.051) | (0.051) | | (0.052) |
| AGE\_1265 to 69 years | 4.312\* | 0.130\*\* | 0.130\*\* | | 0.116\*\* |
|  | (2.422) | (0.055) | (0.055) | | (0.056) |
| AGE\_1270 and over | 3.638 | 0.123\*\* | 0.123\*\* | | 0.101\* |
|  | (2.552) | (0.058) | (0.058) | | (0.058) |
| SEXFemale | -1.534\*\*\* | -0.029\*\* | -0.029\*\* | | -0.029\*\* |
|  | (0.509) | (0.012) | (0.012) | | (0.012) |
| EDUCSome high school | -1.631 | -0.060 | -0.060 | | -0.060 |
|  | (3.018) | (0.069) | (0.069) | | (0.068) |
| EDUCHigh school graduate | -1.574 | -0.067 | -0.067 | | -0.062 |
|  | (2.736) | (0.062) | (0.062) | | (0.062) |
| EDUCSome postsecondary | -0.659 | -0.020 | -0.020 | | -0.014 |
|  | (2.833) | (0.065) | (0.065) | | (0.064) |
| EDUCPostsecondary certificate or diploma | -0.397 | -0.013 | -0.013 | | -0.011 |
|  | (2.694) | (0.061) | (0.061) | | (0.061) |
| EDUCBachelor's degree | 1.918 | 0.046 | 0.046 | | 0.049 |
|  | (2.711) | (0.062) | (0.062) | | (0.061) |
| EDUCAbove bachelor's degree | 3.894 | 0.065 | 0.065 | | 0.070 |
|  | (2.757) | (0.063) | (0.063) | | (0.062) |
| MARSTATLiving in common-law | -0.506 | -0.002 | -0.002 | | -0.004 |
|  | (0.591) | (0.013) | (0.013) | | (0.013) |
| MARSTATWidowed | -2.953\* | -0.050 | -0.050 | | -0.042 |
|  | (1.564) | (0.036) | (0.036) | | (0.035) |
| MARSTATSeparated | -1.093 | -0.021 | -0.021 | | -0.014 |
|  | (1.054) | (0.024) | (0.024) | | (0.024) |
| MARSTATDivorced | -2.269\*\*\* | -0.051\*\*\* | -0.051\*\*\* | | -0.054\*\*\* |
|  | (0.758) | (0.017) | (0.017) | | (0.017) |
| MARSTATSingle, never married | -0.252 | -0.006 | -0.006 | | -0.003 |
|  | (0.500) | (0.011) | (0.011) | | (0.011) |
| COWMAINPrivate sector employees | -2.427\*\*\* | -0.062\*\*\* | -0.062\*\*\* | | -0.060\*\*\* |
|  | (0.551) | (0.013) | (0.013) | | (0.012) |
| IMMIGImmigrant, landed more than 10 years earlier | -0.948 | -0.013 | -0.013 | | -0.018 |
|  | (0.759) | (0.017) | (0.017) | | (0.017) |
| IMMIGNon-immigrant | 1.786\*\* | 0.063\*\*\* | 0.063\*\*\* | | 0.059\*\*\* |
|  | (0.714) | (0.016) | (0.016) | | (0.016) |
| PERMTEMPTemporary, seasonal job | -3.652 | -0.175\*\* | -0.175\*\* | | -0.192\*\* |
|  | (3.393) | (0.077) | (0.077) | | (0.077) |
| PERMTEMPTemporary, term or contract job | 1.726\*\* | 0.005 | 0.005 | | 0.015 |
|  | (0.867) | (0.020) | (0.020) | | (0.020) |
| PERMTEMPTemporary, casual or other temorary jobs | -1.450\* | -0.039\*\* | -0.039\*\* | | -0.024 |
|  | (0.802) | (0.018) | (0.018) | | (0.018) |
| UNIONNot a member but covered by a union contract or collective agreement | 4.030\*\*\* | 0.086\*\*\* | 0.086\*\*\* | | 0.084\*\*\* |
|  | (1.235) | (0.028) | (0.028) | | (0.028) |
| UNIONNon-unionized | 3.028\*\*\* | 0.053\*\*\* | 0.053\*\*\* | | 0.051\*\*\* |
|  | (0.509) | (0.012) | (0.012) | | (0.012) |
| FIRMSIZE20 to 99 employees | -0.875 | -0.003 | -0.003 | | -0.002 |
|  | (0.622) | (0.014) | (0.014) | | (0.014) |
| FIRMSIZE100 to 500 employees | -2.431\*\*\* | -0.022 | -0.022 | | -0.028\* |
|  | (0.708) | (0.016) | (0.016) | | (0.016) |
| FIRMSIZEMore than 500 employees | -0.732 | 0.012 | 0.012 | | 0.011 |
|  | (0.699) | (0.016) | (0.016) | | (0.016) |
| TENURE | 0.017\*\*\* | 0.001\*\*\* | 0.001\*\*\* | |  |
|  | (0.003) | (0.0001) | (0.0001) | |  |
| UHRSMAIN | -0.010 | 0.0003 | 0.0003 | |  |
|  | (0.020) | (0.0005) | (0.0005) | |  |
| SURVMNTHFebruary | -0.276 | -0.005 |  | | -0.005 |
|  | (0.535) | (0.012) |  | | (0.012) |
| SURVMNTHMarch | 0.089 | 0.004 |  | | 0.004 |
|  | (0.551) | (0.013) |  | | (0.012) |
| SURVMNTHApril | 0.612 | 0.022\* |  | | 0.026\*\* |
|  | (0.549) | (0.013) |  | | (0.012) |
| SURVMNTHMay | 0.591 | 0.024\* |  | | 0.025\*\* |
|  | (0.556) | (0.013) |  | | (0.013) |
| NOC\_43Specialized middle management occupations | -23.202\*\*\* | -0.490\*\*\* |  | | -0.471\*\*\* |
|  | (5.940) | (0.135) |  | | (0.134) |
| NOC\_43Middle management occupations in retail and wholesale trade and customer services | -5.088 | -0.084 |  | | -0.066 |
|  | (9.212) | (0.210) |  | | (0.208) |
| NOC\_43Middle management occupations in trades, transportation, production and utilities | -25.542\*\*\* | -0.477\*\*\* |  | | -0.463\*\*\* |
|  | (7.163) | (0.163) |  | | (0.162) |
| NOC\_43Professional occupations in finance | -33.580\*\*\* | -0.709\*\*\* |  | | -0.672\*\*\* |
|  | (6.351) | (0.145) |  | | (0.144) |
| NOC\_43Professional occupations in business | -29.950\*\*\* | -0.639\*\*\* |  | | -0.612\*\*\* |
|  | (6.067) | (0.138) |  | | (0.137) |
| NOC\_43Administrative and financial supervisors and administrative occupations | -36.355\*\*\* | -0.826\*\*\* |  | | -0.799\*\*\* |
|  | (5.933) | (0.135) |  | | (0.134) |
| NOC\_43Administrative occupations and transportation logistics occupations | -40.224\*\*\* | -0.941\*\*\* |  | | -0.911\*\*\* |
|  | (5.902) | (0.134) |  | | (0.133) |
| NOC\_43Administrative and financial support and supply chain logistics occupations | -40.131\*\*\* | -0.949\*\*\* |  | | -0.923\*\*\* |
|  | (5.874) | (0.134) |  | | (0.133) |
| NOC\_43Professional occupations in natural sciences | -32.130\*\*\* | -0.700\*\*\* |  | | -0.662\*\*\* |
|  | (7.178) | (0.164) |  | | (0.162) |
| NOC\_43Professional occupations in applied sciences (except engineering) | -29.220\*\*\* | -0.583\*\*\* |  | | -0.552\*\*\* |
|  | (6.249) | (0.142) |  | | (0.141) |
| NOC\_43Technical occupations related to natural and applied sciences | -28.901\*\*\* | -0.587\*\*\* |  | | -0.562\*\*\* |
|  | (7.176) | (0.164) |  | | (0.162) |
| NOC\_43Health treating and consultation services professionals | -5.163 | -0.235\* |  | | -0.201 |
|  | (5.996) | (0.137) |  | | (0.136) |
| NOC\_43Therapy and assessment professionals | -24.236\*\*\* | -0.499\*\*\* |  | | -0.463\*\*\* |
|  | (5.942) | (0.135) |  | | (0.134) |
| NOC\_43Nursing and allied health professionals | -24.069\*\*\* | -0.486\*\*\* |  | | -0.451\*\*\* |
|  | (5.877) | (0.134) |  | | (0.133) |
| NOC\_43Technical occupations in health | -30.440\*\*\* | -0.648\*\*\* |  | | -0.614\*\*\* |
|  | (5.882) | (0.134) |  | | (0.133) |
| NOC\_43Assisting occupations in support of health services | -39.139\*\*\* | -0.911\*\*\* |  | | -0.880\*\*\* |
|  | (5.865) | (0.134) |  | | (0.133) |
| NOC\_43Professional occupations in education services | -26.197\*\*\* | -0.472\*\* |  | | -0.480\*\* |
|  | (8.354) | (0.190) |  | | (0.189) |
| NOC\_43Professional occupations in social and community services | -32.120\*\*\* | -0.688\*\*\* |  | | -0.671\*\*\* |
|  | (5.910) | (0.135) |  | | (0.134) |
| NOC\_43Professional occupations in government services | -30.154\*\*\* | -0.629\*\*\* |  | | -0.602\*\*\* |
|  | (5.952) | (0.136) |  | | (0.135) |
| NOC\_43Paraprofessional occupations in legal, social, community and education services | -40.089\*\*\* | -0.940\*\*\* |  | | -0.908\*\*\* |
|  | (5.860) | (0.134) |  | | (0.132) |
| NOC\_43Care providers and public protection support occupations and student monitors, crossing guards and related occupations | -40.224\*\*\* | -0.985\*\*\* |  | | -0.955\*\*\* |
|  | (5.955) | (0.136) |  | | (0.135) |
| NOC\_43Occupations in art, culture and sport | -50.047\*\*\* | -1.357\*\*\* |  | | -1.327\*\*\* |
|  | (11.660) | (0.266) |  | | (0.264) |
| NOC\_43Support occupations in art, culture and sport | -41.068\*\*\* | -0.997\*\*\* |  | | -0.946\*\*\* |
|  | (6.382) | (0.145) |  | | (0.144) |
| NOC\_43Retail sales and service supervisors and specialized occupations in sales and services | -38.888\*\*\* | -0.911\*\*\* |  | | -0.896\*\*\* |
|  | (6.624) | (0.151) |  | | (0.149) |
| NOC\_43Occupations in sales and services | -42.486\*\*\* | -1.054\*\*\* |  | | -1.027\*\*\* |
|  | (6.243) | (0.142) |  | | (0.141) |
| NOC\_43Sales and service representatives and other customer and personal services occupations | -31.564\*\*\* | -0.687\*\*\* |  | | -0.676\*\*\* |
|  | (6.419) | (0.146) |  | | (0.145) |
| NOC\_43Sales and service support occupations | -41.235\*\*\* | -1.020\*\*\* |  | | -0.990\*\*\* |
|  | (5.925) | (0.135) |  | | (0.134) |
| NOC\_43Technical trades and transportation officers and controllers | -33.792\*\*\* | -0.708\*\*\* |  | | -0.697\*\*\* |
|  | (7.156) | (0.163) |  | | (0.162) |
| NOC\_43General trades | -38.468\*\*\* | -0.852\*\*\* |  | | -0.796\*\*\* |
|  | (6.391) | (0.146) |  | | (0.145) |
| NOC\_43Helpers and labourers and other transport drivers, operators and labourers | -47.815\*\*\* | -1.261\*\*\* |  | | -1.239\*\*\* |
|  | (7.461) | (0.170) |  | | (0.169) |
| NOC\_43Supervisors, central control and process operators in processing, manufacturing and utilities and aircraft assemblers an | -31.094\*\*\* | -0.595\*\* |  | | -0.542\*\* |
|  | (11.691) | (0.266) |  | | (0.264) |
| Constant | 60.306\*\*\* | 3.927\*\*\* |  | | 3.970\*\*\* |
|  | (6.898) | (0.157) |  | | (0.156) |
| Observations | 3,348 | 3,348 | 3,348 | | 3,348 |
| R2 |  |  | 0.146 | |  |
| Adjusted R2 |  |  | 0.127 | | 0.592 |
| Log Likelihood | -12,437.350 | 223.749 |  | | 181.506 |
| Akaike Inf. Crit. | 25,018.710 | 1118.1 | -375.4. | | -363.0 |
| UBRE |  |  |  | | 0.053 |
| F Statistic |  |  | 15.534\*\*\* (df = 36; 3276) | |  |
| Note: 1. the T value in this table is not adjusted for heteroskedasticity and please refer back to attached the R code for accurate T value and significance. 2. AIC in this table is updated to reflect the correct AIC value. 3. The NOC\_43 and SURVMNTH in the panel Model are fixed effect variables and it is not presented here. | \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 | | | | |

CART Tree:



\* the cut-off line in the graph is NOC\_43 and please run the R code file if interested.