RIETI Discussion Paper Series 24-E-034

# Unraveling the Determinants of Overemployment and Underemployment among Older Workers in Japan: A machine learning approach

**ZHANG, Meilian**
Chinese University of Hong Kong

**YIN, Ting**
RIETI

**USUI, Emiko**
Hitotsubashi University

**OSHIO, Takashi**
Hitotsubashi University

**ZHANG, Yi**
Central University of Finance and Economics

Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 24-E-034
**March 2024**

# Unraveling the Determinants of Overemployment and Underemployment Among Older Workers in Japan: A machine learning approach

Meilian Zhang[a], Ting Yin[b,c], Emiko Usui[d], Takashi Oshio[e], Yi Zhang[f]

**Abstract**

Overemployment and underemployment being widely existing phenomena, much less is known about their determinants for older workers. We innovatively employ machine learning methods to determine the important factors driving overemployment and underemployment among older workers in Japan. Those with better economic conditions, worse health, less family support, and unfavorable job characteristics are more likely to report overemployment, whereas increasing age, less disposable income, shorter current work hours, holding a job with a temporary nature, and low job and pay satisfaction are predictive to underemployment. Cluster analysis further shows that reasons for having work hour mismatches can be highly heterogeneous within both overemployed and underemployed groups. Subgroup analyses suggest room for pro-work policies among 65+ workers facing financial stress and lacking family support, female workers with unstable jobs and low spousal income, and salaried workers working insufficient hours.

**Keywords:** Overemployment, underemployment, determinants, machine learning, random forest, K-means clustering

**JEL classification:** J22, J28, C52, C55



* This study is conducted as a part of the project "Economic Analysis on the Problem of an aging population and a declining birthrate in China and Japan in the COVID-19 Pandemic" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The authors thank the JSTAR (Japanese Study of Aging and Retirement) team for providing data. JSTAR was conducted by the Research Institute of Economy, Trade and Industry (RIETI), Hitotsubashi University, and the University of Tokyo. Yi Zhang is grateful for the support from the National Natural Science Foundation of China (grant number: 72303269), the Joint Usage and Research Center, Institute of Economic Research, Hitotsubashi University (grant IDs: IERPK2213, IERPK2312), and the Program for Innovation Research at Central University of Finance and Economics. We wish to thank Steven Lehrer, Jian Wang, Keisuke Kawata, and the participants at the RIETI Seminar (November 2023), for their helpful comments and suggestions. We also thank the RIETI staff for their kind help and cooperation. Any errors are our own. There are no conflicts of interest.

[a] Department of Economics, The Chinese University of Hong Kong, Meilian.Zhang@link.cuhk.edu.hk
[b] Center for Intergenerational Studies, Institute of Economic Research, Hitotsubashi University. Email: yin-ting@ier.hit-u.ac.jp
[c] Research Institute of Economy, Trade and Industry, IAA. Email: yin-ting@rieti.go.jp
[d] Center for Intergenerational Studies, Institute of Economic Research, Hitotsubashi University. Email: usui@ier.hit-u.ac.jp
[e] Center for Intergenerational Studies, Institute of Economic Research, Hitotsubashi University. Email: oshio@ier.hit-u.ac.jp
[f] Corresponding author at China Center for Human Capital and Labor Market Research, Central University of Finance and Economics, P.R China. Email: yi.zhang@cufe.edu.cn

## 1. Introduction

A neoclassical intertemporal labor supply model where individuals optimally choose their hours of work and pensionable ages generally predicts no "overemployment" (working more hours than desired) or "underemployment" (working less hours than desired) among younger and older workers (Reynolds and Aletraris 2006, 2010; Kaufman 2016; Usui et al. 2016; Paz 2022). Existing evidence, however, shows that overemployment and underemployment widely exist in both developed (e.g., Reynolds and Aletraris 2007; Bell and Rutherford 2013; Campbell and van Wanrooy 2013; Sewdas et al. 2017; Bell and Blanchflower 2021; Chambers et al. 2021; Reich 2023) and developing countries (e.g., Fernandez and Shiang 2017; Li et al. 2022; Dvouletý 2023). Japan, where older male workers are known for their high labor-force attachment, is no exception in this respect. Hara and Sato (2008) find that 45% of Japanese workers report to be overemployed and 6% are underemployed. Usui et al. (2016) find that self-employed male workers at the pensionable age tend to report overemployment while their salaried counterparts tend to report underemployment.

Both underemployment and overemployment induce welfare loss, particularly for older workers. Underemployment, also referred to as "hidden unemployment" or "involuntary part-time employment", is linked with under-utilization of human resource (Dvouletý 2023), insufficient income and unrealized pension expectations (Bell and Rutherford 2013), and low satisfaction about work (Wilkins 2007; George et al. 2012; Raykov 2014; Roh et al. 2014). Overemployment, on the other hand, is found to disrupt work-life balance (Yamaguchi 2008), increase absenteeism (Reich 2023), lower subjective wellbeing (Wooden et al. 2009; Angrave and Charlwood 2015; Pagan 2017; Bell and Blanchflower 2019), worsen mental health (Otterbach et al. 2021; De Moortel 2022), and even cause *karoshi*, i.e., death from overwork (Iwasaki et al. 2006; Asgari et al. 2016). Therefore, it is of broad interest of researchers and policymakers to know the determinants of underemployment and overemployment.

Previous literature across countries uses various methods to explore the determinants of underemployment and overemployment. A majority of the literature focuses on younger adults or working age population, typically using logistic or probit regression to examine the association between a pre-assumed set of variables and overemployment or underemployment incidence. The factors identified are rather context-specific. For example, Ruiz-Quintanilla and Claes (1996), using probit regressions, find that for young office technology workers and machine operators in European countries, underemployment is associated with individual level characteristics (education, occupation group) and labor market and societal variables (initial labor market experience, perception of labor market conditions). Golden and Gebreselassie (2007) use multinomial logistic regressions and find that overemployment among American workers is associated with gender, job type, income level, workweek length, and stage of workers' life cycle. Otterbach (2010) uses data from 21 countries and point out that work hour mismatch is interrelated with macroeconomic conditions (e.g., unemployment rates, GDP per capita, average weekly work hours, and income inequality). Yamaguchi (2008) and Fernandez and Shiang (2017), analyzing samples of Japanese workers aged between 20 and 50 and workers in Penang of Malaysia, both find that overemployment not only correlates with individual characteristics (e.g., age, ethnicity, education, occupation, gender), but also with household characteristics (e.g., number of children, married with young children, parents cohabiting), and job-related characteristics (e.g., hours of work, control over work schedule, full time or part-time job, commuting time). Schalembier et al. (2019), using Flemish data and logistic regression, emphasize that relative income, rather than absolute income, influences underemployment. Girtz (2021) uses Australian data and multinomial logistic regression to show how the interaction of gender and occupation would influence the probability of overemployment and underemployment.

Some studies bring insights from other methods. For instance, Bell and Blanchflower (2021) construct a better measure of underemployment and compare it over time to investigate the role of Great Recession on underemployment across the United States and European countries. Yamaguchi (2010) provides a theoretical examination of how employers' monopsony power would lead to workers' overemployment. Lo (2023) develops a search and matching model to show how bilateral bargaining between employees and employers would influence overwork and underwork.

Fewer studies focus on older adults, who are in fact more susceptible to the negative consequences of work hour mismatch. Among this small literature a major focus is on why older workers would like to prolong work life. For example, through interviews and thematic analyses, work characteristics, skills and knowledge, financial factors, social factors, health, and purposefulness are found to influence older adults' decisions to work beyond retirement age (e.g., Reynolds et al. 2012; Sewdas et al. 2017). Quantitative analyses find that flexible work arrangement, age discrimination, disability, economic difficulties, social resources, and local job opportunities also matter for working longer. (e.g., Bandara 2017; Van Solinge and Henkens 2017; Choi et al. 2018; Zitikytė 2020).

Very few studies directly investigate the factors influencing older workers' overemployment or underemployment. Charles and Decicca (2007) and Gielen (2009) theoretically model the source of older workers' over(under)employment as employer-imposed work hours constraints, possibly coming from the organizational or technological constraints and/or employers' market power. Bell and Rutherford (2013) find that type of employment matters, that is, older self-employed workers in the UK are more likely to be over- or underemployed than employees. Silver et al. (2019) finds that Canadian workers' over- and underemployment in pre-retirement years are associated with age, health, household income, work hours, relation with coworker and supervisors, control of one's work schedule, and financial strain. They also document heterogeneous effect of determinants among men and women. More evidence worldwide is highly needed to deepen the understanding of factors influencing older workers' overemployment and underemployment.

We add to this line of research by identifying and analyzing the determinants of older workers' overemployment and underemployment in Japan. Japan has provided a highly relevant context, as it is aging fast with prevalent over- and under-employment among older workers (Usui et al. 2016). Given that the factors driving work hour mismatch identified in previous literature are rather wide-ranging yet context-specific, we innovatively use a data-driven approach to find out important determinants for Japanese older workers. More specifically, we use machine learning methods (Random Forest, XG-Boost, and Lasso) to predict the incidence of overemployment and underemployment with a wide range of candidate features (determinants). By cross-validation we choose the best model and the most important features suggested. We then quantify the impact of these chosen determinants on over (under)employment incidence with logistic regressions for the whole sample and for subgroups.

Based on microdata of Japanese Studies of Ageing and Retirement (JSTAR), we find that overemployment is associated with better economic conditions (higher education, more liquid assets), worse health (more chronic diseases, less generous medical insurance), less family support (fewer children, shorter spousal work hours), and unfavorable job characteristics (physically demanding and stressful jobs, dissatisfaction with pay, long actual working hours). Whereas, age, less disposable income, shorter current work hours, temporary nature of the job (wage calculated by hours or days), and low job and pay satisfaction are predictive to underemployment for Japanese older workers. We further use K-means clustering and show that there exist different latent groups within the overemployed and underemployed. For each subgroup, reasons for being work hour mismatched can be highly heterogeneous.

We also find substantial heterogeneity by age groups, gender, and employment status. For younger workers, bad health and less generous medical insurance seem to drive them to report overemployment, while for older workers, high work pressure and low satisfaction with pay are the main drivers of their overemployment. For the younger, underemployment seems to be related with temporary employment (paid by hours) and dissatisfaction with pay, whereas for the older, individual financial stress and dissatisfaction with pay tend to be important to underemployment, implying room for policies encouraging healthy older adults with financial constraints to supply more labor.

For female workers, their overemployment can be alleviated by better spousal pension, improving satisfaction of pay, learning new skills, and involving less physical labor. Women with fewer children are more likely to report overemployment possibly due to less support received from children. Female workers are more likely to report underemployment with lower spousal income, and more unstable jobs (wage calculated by day or hour), pointing to the direction along which policies can be made to increase female labor supply.

Comparing self-employed workers and salaried workers, the self-employed tend to report overemployed for high job pressure while salaried workers report overemployment often for health reasons. Dissatisfaction with pay and unstable jobs are the main reasons for the self-employed to report underemployment, while the salaried tend to report underemployment when their current working hours are not long enough and their satisfaction with job and their level of liquid assets are low. It seems that healthy salaried workers have potentials to prolong working time, consistent with the message drawn from Usui et al. (2016).

We contribute to the literature in following aspects: first, we provide the first thorough investigation into determinants of over- and underemployment incidence for older workers in Japan. Japan has been going through a fast-ageing process (Usui et al. 2017). No matter for its goal of healthy aging, or for a more effective utilization of older population's human capital, it is of utter importance to understand what drives older Japanese workers to work more or less. Our findings unravel the complications that there exist groups with different interests and pursuits of people within the overemployed or underemployed workers. This is crucial for policy makers when revising retirement policies or labor market activation policies. One-size-fits-all policy is not desirable as it may exacerbate the financial difficulties or health problems that some workers have already faced.

Second, we contribute to a deeper understanding of determinants of over- and underemployment for certain policy-relevant subpopulations. For example, women and the self-employed, their work hour mismatches are far from being well understood. Our findings help depict who might conform to the policies incentivizing working hour extension or reduction within a certain subpopulation, and help policy-makers to pin down the obstacles that prevents certain older workers from working longer or shorter.

Third, we make the first step to link the literature of variable selection with machine learning methods to finding out the determinants of over (under)employment. Compared to the context-specific and pre-assumed choice of determinants, we allow for a data-driven method to select relevant determinants in our context. The "let-the-data-speak" approach could avoid the omission of important relevant determinants. This integration of machine learning techniques may provide insights for future research investigating determinants of over- (under-) employment, or even determinants of a broader range of subjects.

The remainder of this paper is organized as follows. Section 2 describes the data and methods. Section 3 presents the empirical results. Section 4 discusses the policy implications and concludes.

## 2. Data and methods

### *2.1 Data*

We employ individual-level pooled data from Japanese Studies of Ageing and Retirement (JSTAR) in 2007, 2009, 2011, and 2013. JSTAR is a panel survey for elderly people aged 50 and above in Japan. It collected various information, including the economic, social, and health conditions of elderly people. To unearth underlying factors of underemployment and over-employment in Japan, we restrict our sample to those who are employed. In addition, we use Multiple Imputation by Chained Equations (MICE) for missing values in the candidate determinants. The resulting sample consists of 8811 observations.

We draw on Usui et al. (2016) for the definition of underemployment and over-employment. For underemployment, we define a binary variable, which equals 1 if respondents cannot increase paid work hours but are willing to increase given the earnings increased in the same proportion, otherwise equals 0. Similarly, over-employment indicators will equal 1 if the respondents cannot decrease paid work hours but are willing to decrease given the earnings decreased in the same proportion, otherwise equal 0.

We include four dimensions of information, including demographics, family information, health-related conditions, and employment-related information, as potential factors driving underemployment and over-employment. For demographics, we include age, age squared, gender, marriage status, education level, income, spousal information, etc. For family information, family structure, providing informal caregiving to parents/parents-in-law, and household assets and debts are included. We also investigate whether health-related conditions are the potential determinants of underemployment and over-employment, such as health status, healthcare utilization, health behaviors, and health insurance. Moreover, we explore if employment-related conditions are relevant potential determinants, such as work type, weekly work hours, ways of wage calculation, and subjective attitudes related to work. The detailed potential determinants of underemployment and over-employment can be found in Table A.1 in Online Appendix A.

### *2.2 Methods*

#### *2.2.1 Methodology framework*

The methodological process of the present paper consists of five steps. The first step, as introduced in Section 2.1, is the data preparation. Next, we employ machine learning (ML) methods, i.e., random forest, XG-Boost, and Lasso to predict over-(under-)employment from a high dimension of candidate variables, and use cross-validation to select the best method to for prediction. We then exploit the selected machine learning method to conduct variable selection. Finally, we employ logistic regression to quantify the association between selected variables and over-(under-) employment.

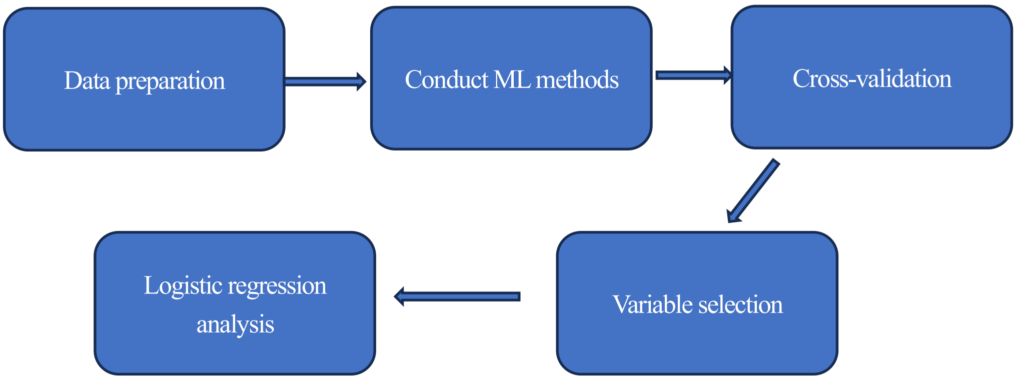



*Figure 1. Methodological framework*

#### *2.2.2 ML methods*

We employ three frequently used ML methods for prediction and variable selection: Random forest, XGBoot, and Lasso (see, for example, James et al. (2013) for detailed descriptions of these methods). The former two methods are tree-based ML methods that are well-known for their flexibility and good predictive performance. Lasso, on the other hand, is a linear regression model with a penalty term, known as a powerful method for variable selection. We briefly introduce the three methods as follows.

Random forest, a supervised ensemble machine learning method working on the concept of bagging, creates a set of classification trees or a series of regression probabilities obtained by the random selection of a set of variables from the variable space and a bootstrap procedure that recurrently selects part of the sample space to fit the model. Random forest can deal with high-dimensional data and non-linear correlation structures with simple parameter setting. It can provide all input variables' importance in terms of Gini or information gain for classification or in terms of mean squared error (MSE) or mean absolute error (MAE) for regression. For a better comparison with Lasso, we do not use random forest classification but random forest regression in this paper.

In contrast with random forest, eXtreme Gradient Boosting (XGBoost) is an implementation of gradient-boosted decision trees (GBDT) designed to be highly efficient, flexible, and portable since XGBoost can provide parallel decision tree boosting. Moreover, XGBoost uses second-order Talor expansion to enhance prediction accuracy and regularized gradient boosting with both L1 and L2 regularization to avoid overfitting. Similar to the random forest, it offers all input variables' importance.

The Least Absolute Shrinkage and Selection Operator (Lasso) is a generalized linear regression method that achieves both feature selection and L1 regularization in order to improve the prediction accuracy and interpretability of the resulting statistical model. Setting a given value of tuning parameter $\lambda$ (the severity of penalty) and shrinking estimates to 0 facilitates Lasso dealing with high-dimension data and performing feature selection. The absolute value of estimates of selected variables indicates variables' importance.

## 3. Data analysis and results

### *3.1 Cross-validation*

Cross-validation iteratively uses different proportions of data as training data and test data, applies the statistical method to train a model with training data, estimates the test error rate with testing data by fitting the model, and acquires the average test error rate to evaluate the performance of the statistical method. Usually, there are two types of cross-validation, Leave-one-out cross-validation (LOOCV) and K-fold cross-validation. Although both types of cross-validation (CV) divide data into training data and testing data, LOOCV only leaves 1 observation as testing data and K-fold leaves a number of observations. K-fold CV divides data into k parts and iteratively uses one part as testing data and the rest of k-1 parts as training data. In this paper, we choose K-fold CV and set k equal to 10 by default. We apply 10-fold CV for each machine learning method, repeat this step 20 times, and calculate the average MSE for each machine learning method. A lower average MSE indicates a better predictive performance.

Table 1. 10-fold cross-validation results (times=20)

| Outcome \ MSE | Random forest | XG-Boost | Lasso |
|---|---|---|---|

| | | | |
|---|---|---|---|
| underemployment | 0.029624 | 0.029677 | 0.030071 |
| overemployment | 0.04361 | 0.043905 | 0.047788 |

As shown in Table 1, we can find that random forest has the best prediction performance no matter for underemployment or overemployment. Thus, we use random forest to select potential determinants (variables) for later analysis.

### *3.2 Variable selection*

We use random forest to conduct feature selection and pick out those variables that their importance is greater than the average of all variables' importance. All selected variables are shown in Table 2.

Table 2. Selected variables from random forest

| Factor group | Variables chosen for Overemployment | Var. chosen for Underemployment |
|---|---|---|
| Demographics | age<br>education years<br># of children<br># of siblings | age<br>education years<br># of children<br># of siblings |
| Income & Assets | income<br>spouse's income<br>liquid assets<br>fix assets (property value)<br>total loan (home mortgage loans + other loans) | income<br>spouse's income<br>liquid assets<br>fix assets (property value)<br>total loan (home mortgage loans + other loans) |
| Health and Healthcare use | # of chronic diseases<br>OOP expenditure of health check<br>Whether have dentist visit<br>OOP expenditure of dentist visit<br>Whether have outpatient visit<br>OOP expenditure of outpatient visit<br>CESD-mild<br>CESD-severe | # of chronic diseases<br>OOP expenditure of health check<br>OOP expenditure of dentist visit<br>OOP expenditure of outpatient visit |
| Medical insurance | Whether have employ medical insurance<br>Whether have other medical insurance | |

| | | |
|---|---|---|
| Employment | Weekly work hours<br>Spouse's weekly work hours | Weekly work hours<br>Spouse's weekly work hours<br>Ways of wage calculation |
| Subjective attitude to work | whether feel time pressure and lots of work<br>Whether involve physical labor or not<br>Whether have hope for promotion<br>Whether have little discretion over work<br>Whether can gain new skills<br>Whether satisfied with current pay | Whether satisfied with current pay<br>Whether satisfied with job |
| Spouse's pension type | Whether spouse can/will get national pension<br>Whether spouse can/will get employ pension | |

*3.3 Regression results*

We employ logistic regression to quantify the impact of selected determinants on overemployment and underemployment, respectively.[1]

In terms of factors driving overemployment, Table 3 shows that those with higher education level, fewer children to raise, and more liquid assets are more likely to feel overemployed, indicating that individuals with more economic resources or family support are inclined to shorten their work hours. Health also matters for overemployment. We find that for those with more chronic diseases, higher probability of outpatient visit, and other medical insurance (refers to less generous medical insurance), they are more likely to be overemployed. Besides, fewer spousal weekly work hours and more own weekly work hours lead to a higher probability of overemployment. This might be due to the "reference dependence effect", i.e., taking spousal working hours as one's own reference point of proper work hours. The job attributes are also important factors affecting overemployment. When a job involves mental pressure or physical labor, it tends to make people feel overemployed. Intuitively, if a worker is satisfied with current pay, he or she is less likely to increase work hours given the same wage rate.

Regarding determinants of underemployment, Table 3 suggests a U-shaped relation between age and the likelihood of underemployment. This means that people initially feel less underemployed as they get

[1] In order to better interpret the results, we add the whole group of dummies even if only one of the dummies is chosen by the selection method. Taking the group of dummies for different ways of wage calculation as an example. Even if only one of this group of dummies, e.g., "dummy for wage calculation by hours", is selected by random forest, we still add all dummies for wage calculation methods into the regression (excluding the reference group).

older, but as they reach 74 years old, they start to feel underemployed and want to work longer as they age further. It is possible that this oldest group might encounter financial difficulties and need more work to make ends meet. Moreover, family's disposable income will decrease the probability of underemployment, such as spouse's income and liquid assets. Surprisingly, having more chronic diseases also lead to a higher chance of underemployment, possibly pointing to the disease-induced financial burden. Consistent with results of overemployment, the more weekly work hours, the more satisfied with the job and current pay, the less likely to feel underemployed. In addition, the ways of wage calculation can predict underemployment well. Those workers whose wage are paid by hours or by days (usually refer to informal or part time jobs) tend to feel underemployed and to be willing to work more hours.

Table 3. Regression results

| | Factors | Overemployment | Underemployment |
|---|---|---|---|
| Demography | age | -0.032 | -0.295*** |
| | age squared | 0.000 | 0.002*** |
| | education years | 0.046*** | -0.000 |
| | # of children | -0.079** | -0.019 |
| | # of siblings | -0.005 | -0.014 |
| Income & Assets | income | -0.012 | 0.018 |
| | spouse's income | 0.009 | -0.030** |
| | liquid assets | 0.033*** | -0.025** |
| | fix assets | -0.001 | -0.004 |
| | total loan | 0.003 | 0.009 |
| Health & healthcare use | OOP expenditure of health check | -0.000 | -0.000 |
| | dentist visit incidence | 0.145 | |
| | OOP expenditure of dentist visit | -0.005 | 0.002 |
| | outpatient visit incidence | 0.180* | |
| | OOP expenditure of outpatient visit | -0.005 | -0.003 |
| | # of chronic diseases | 0.141*** | 0.064** |
| | CESD-mild | 0.036 | |
| | CESD-severe | 0.093 | |

| | | | |
|---|---|---|---|
| Medical insurance (base: nation MI) | employ medical insurance | 0.040 | |
| | other medical insurance | 0.606*** | |
| Employment | weekly work hours | 0.011*** | -0.005** |
| | spouse's weekly work hours | -0.005** | -0.001 |
| | wage calculation: hour | | 0.643*** |
| | wage calculation: daily | | 0.886*** |
| | wage calculation: monthly | | -0.034 |
| | wage calculation: self-employ | | 0.024 |
| Subjective Attitude to work | feel time pressure and lots of work | 0.459*** | |
| | involve physical labor | 0.227*** | |
| | have hope for promotion | 0.027 | |
| | have little discretion over work | -0.042 | |
| | can gain new skills | -0.075 | |
| | satisfied with current pay | -0.196*** | -0.498*** |
| | satisfied with job | | -0.226*** |
| Spouse's type (base: other pension) | Pension national pension incidence | 0.101 | |
| | employ pension incidence | 0.052 | |
| Observations | | 8811 | 8811 |

*3.4 Clustering analysis*

To shed light on whether there exist heterogeneous groups of individuals within the overemployed and underemployed workers, we employ K-means clustering method to identify the latent clusters of workers conditional on being overemployed or underemployed.

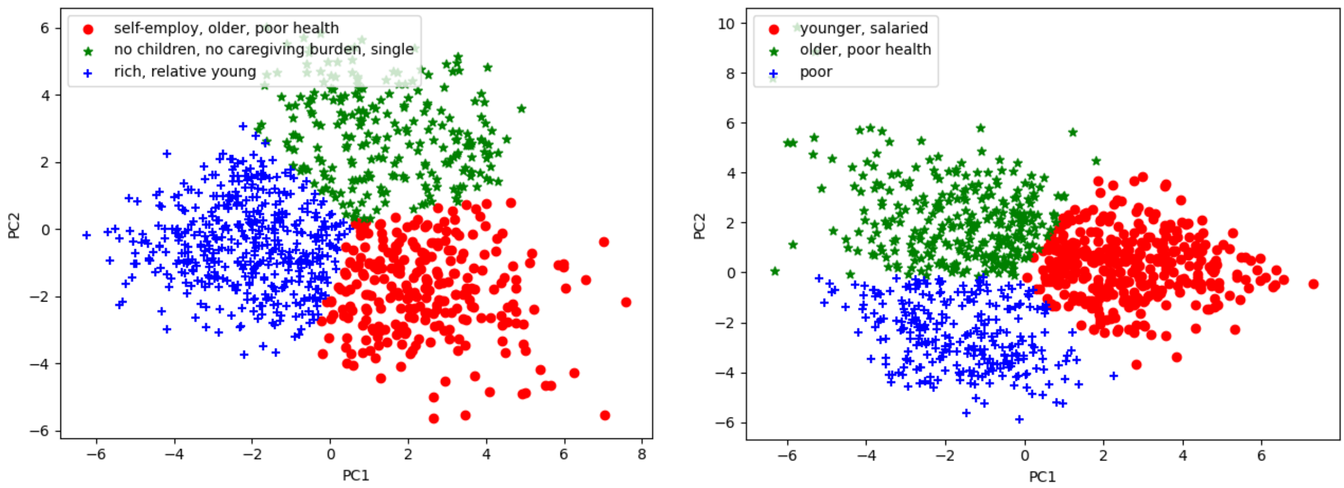
Firstly, we conduct Principal Component Analysis (PCA) on all candidate determinants to acquire the first two principal components (PCs) for the sample of overemployed and the sample of underemployed, respectively.[2] This is to compress high dimensional determinants into two most important dimensions for the overemployed and the underemployed sample, respectively.

Then we interpret the two PCs with their scoring coefficients. The Table A.2 in Online Appendix A shows the scoring coefficients of the two PCs. The absolute values of these coefficients indicate the

[2] Although no matter for overemployment or underemployment 2 PCs only contributes around 5% of variance, respectively, 2 PCs still can give an insight into the factors of overemployment and overemployment.

relative importance of a candidate determinant (variable) to the PC. For example, for the overemployed, PC1 loads heavily on “age”, “wage calculation – self-employed”, “chronic diseases”, “care-giving” related variables. Thus, a higher value of PC1 indicates being older, self-employed, and with poorer health and less care burden. By the same token, PC2 indicates having fewer children, being single and poor. And similar interpretation can be done for the underemployed. PC1 indicates being younger, salaried, healthier, better educated, and with heavier care burden and richer spouses. PC2 refers to being self-employed, rich, and of poor health.

Knowing what the two PCs stand for, we use K-means clustering to find 3 latent groups of individuals along dimensions of two PCs among the overemployed and the underemployed. More specifically, we divide the individuals (each person with two features, PC1 and PC2) into 3 clusters such that the within-cluster variances are minimized. In other words, most similar individuals are put into the same cluster. Finally, we plot the clusters on the PC1-PC2 plane, as shown in Figure 2.




(a) Overemployment (b) Underemployment

Figure 2. K-means clustering results along 2 principal components from PCA

Based on the aforementioned interpretations of PC1s and PC2s, we can have an idea of the latent heterogeneous groups of individuals within the overemployed and the underemployed. For the overemployed (Figure 2a), there are essentially three types of workers: (1) self-employed, old people, with poor health conditions and lower education level; (2) Those who have little family support or responsibilities, such as being single and having fewer children and less need to take care of elderly parents; (3) Those who have much disposable income and are relatively young. Figure 2b shows that underemployed people consist of three groups: (1) those who are younger, salaried, and with heavier care burden; (2) Older people in poor health condition; (3) Older and poor workers.

These heterogeneous clusters suggest that workers are over- or under-employed for different reasons and with different interests. Measures need to take to reduce the heavy workload for the older self-employed workers with poor health. While providing support for workers with little family support may alleviate their overemployment. And it is hard to keep the rich and relatively young group working as they have less financial incentive to do so. For the underemployed, it might be effective to incentivize the younger salaried workers and older poor workers to work more hours, but health support is needed when allowing older workers with poor health to work longer. In any case, a one-size-fits-all policy measure will not work well given mixed incentives and interests among over- (under-) employed workers.

### *3.5 Heterogeneity analysis*

Previous literature (e.g., Bell and Rutherford 2013; Usui et al. 2016; Silver et al. 2019) has documented that men versus women and salaried versus self-employed workers can be heterogeneous in determinants of over-(under-) employment. Unveiling this heterogeneity can facilitate policymakers to identify different socioeconomic subgroups' employment dilemmas and tailor policies to promote full employment or work-life balance for different groups, which can eventually improve social welfare. To shed light on this, we perform the logistic regressions with the afore-chosen variables by age, gender, work type groups for overemployment and underemployment, respectively.[3]

#### *3.5.1 Heterogeneity analysis by age*

Age is an important dimension of heterogeneity. At different ages, statutory (early) retirement and pension incentives may kick in and alter workers' perception of over-(under) employment, whereas health conditions also change as workers grow older, also affecting the over-(under) employment incidence. We divide workers into four age categories based on retirement age and pension age, 60 below (work group), 60-64 (retirement but no pension), 65-69 (retirement and having pension), and 70 & 70 up (the old with pension).

Tables 4 and 5 shows the heterogeneity analysis by age groups for overemployment and underemployment, respectively. The coefficients of age and age squared suggest that workers above age 67 tend to report more overemployment while workers above 57 tend to report more underemployment.[4]

Since companies generally provide re-employment opportunities for people aged 60-64, and pensions

[3] For comparability and interpretability, we assume that the picked determinants remain the same across different subgroups for over-(under-) employment. But we allow for heterogeneous effects of these determinants across subgroups.

[4] Using the coefficients for overemployment in "65-69" group and the coefficients for underemployment in "below 60" group, we can calculate that overemployment and underemployment dip at 67 and 57, respectively.

for better educated people are usually adequate, the higher the education level for people aged 60-64, the more they feel overemployed. In addition, with more chronic diseases workers aged below 65 are more likely to report overemployed as well as underemployed. This is perhaps due to the unavailability of pensions before age 65 and the high healthcare expenditures. The health-related financial stress pushes them to want longer working time to support themselves, but their health condition pulls them back from working more.

Similar to the baseline regression results, more working hours per week generally increases the likelihood of people feeling overemployed and decreases the likelihood of feeling underemployed. However, it is worth mentioning that the 65-69 age group is not sensitive to an increase in the number of hours worked per week, even when the work is stressful or involves physical labor, this group is generally less sensitive than other age groups. In addition, with higher total loans the 65-69 age group significantly increases the likelihood of being underemployed. This means that the 65-69 age group is still a group of people with high employment potential, especially for those under financial pressure. This could provide some rationale for the policy initiative to delay retirement ages.

For people aged above 70, it turns out that the more educated they are, the higher the income they can get from their jobs, the fewer fixed assets they have, the more they feel underemployed and want more work. However, once the work pressure is too high (and mental health deteriorates) and the salary calculation method is relatively stable, they will be less likely to want to work more. If policy makers want to promote labor supply for this group, they need to consider incentives for flexible employment and balance the need for less job pressure and better mental health.

Table 4. Heterogeneity results by age (dependent variable: overemployment)

| | **Factors** | **Overemployment** | | | |
|---|---|---|---|---|---|
| | | 60 below | 60-64 | 65-69 | 70 &70 up |
| Demography | age | -0.956 | -4.807 | -14.928** | -1.707 |
| | age squared | 0.008 | 0.038 | 0.111** | 0.011 |
| | education years | 0.016 | 0.106*** | 0.042 | 0.025 |
| | # of children | -0.066 | -0.008 | -0.132 | -0.237** |
| | # of siblings | -0.016 | 0.009 | 0.018 | -0.051 |
| Income & Assets | income | -0.017 | -0.009 | -0.017 | -0.026 |
| | spouse's income | 0.011 | 0.020 | -0.019 | 0.036 |
| | liquid assets | 0.031** | 0.040* | 0.034 | 0.046 |

| | | | | | |
|---|---|---|---|---|---|
| | fix assets | -0.001 | -0.003 | -0.007 | 0.010 |
| | total loan | 0.002 | 0.002 | -0.003 | 0.016 |
| Health and Healthcare use | OOP expenditure of health check | -0.000 | 0.000 | -0.000 | 0.000 |
| | dentist visit incidence | 0.104 | 0.570 | -0.640 | 0.468 |
| | OOP expenditure of dentist visit | 0.002 | -0.026 | 0.031 | -0.045 |
| | outpatient visit incidence | 0.390*** | -0.196 | 0.228 | -0.062 |
| | OOP expenditure of outpatient visit | -0.008 | 0.011 | -0.006 | -0.011 |
| | # of chronic diseases | 0.173*** | 0.198*** | 0.060 | 0.101 |
| | CESD-mild | -0.121 | 0.048 | 0.124 | 0.494** |
| | CESD-severe | 0.009 | 0.261 | -0.422 | 0.453 |
| Medical insurance (base: nation MI) | employ medical insurance | -0.042 | 0.135 | 0.132 | -0.237 |
| | other medical insurance | 0.723*** | 0.636*** | 0.550*** | 0.177 |
| Employment | weekly work hours | 0.020*** | 0.012*** | 0.002 | 0.021*** |
| | spouse's weekly work hours | -0.003 | -0.011*** | 0.005 | -0.017** |
| | wage calculation: hour | | | | |
| | wage calculation: daily | | | | |
| | wage calculation: monthly | | | | |
| | wage calculation: self-employ | | | | |
| Subjective Attitude to work | feel time pressure and lots of work | 0.468*** | 0.410*** | 0.399** | 0.580** |
| | involve physical labor | 0.105 | 0.477*** | 0.378* | 0.061 |
| | have hope for promotion | 0.139 | 0.051 | -0.185 | -0.266 |
| | have little discretion over work | -0.089 | -0.020 | -0.043 | 0.171 |
| | can gain new skills | -0.092 | -0.261* | 0.228 | 0.052 |
| | satisfied with current pay | -0.113 | -0.245* | -0.386** | -0.232 |
| | satisfied with job | | | | |
| Spouse's Pension type (base: other pension) | national pension incidence | 0.148 | 0.179 | -0.044 | -0.179 |
| | employ pension incidence | 0.015 | 0.240 | -0.025 | -0.538* |
| Observations | | 3,594 | 2,288 | 1,580 | 1,349 |

Table 5. Heterogeneity results by age (dependent variable: underemployment)

| | Factors | Underemployment | | | |
|---|---|---|---|---|---|
| | | 60 below | 60-64 | 65-69 | 70 &70 up |
| Demography | age | -2.434*** | 4.280 | -7.698 | 2.404 |
| | age squared | 0.021*** | -0.035 | 0.057 | -0.016 |
| | education years | -0.015 | -0.035 | 0.031 | 0.079* |
| | # of children | -0.005 | -0.023 | -0.166* | 0.140 |
| | # of siblings | -0.038 | 0.030 | -0.026 | -0.015 |
| Income & Assets | income | 0.017 | -0.028 | 0.035 | 0.069** |
| | spouse's income | -0.018 | -0.004 | -0.076** | -0.045 |
| | liquid assets | -0.004 | -0.039* | -0.041 | -0.043 |
| | fix assets | 0.008 | -0.005 | -0.016 | -0.041** |
| | total loan | 0.005 | 0.004 | 0.037** | -0.010 |
| Health and Healthcare use | OOP expenditure of health check | -0.000 | -0.000 | -0.000 | -0.000 |
| | dentist visit incidence | | | | |
| | OOP expenditure of dentist visit | -0.008 | 0.012 | -0.000 | 0.032 |
| | outpatient visit incidence | | | | |
| | OOP expenditure of outpatient visit | 0.008 | -0.017 | 0.004 | -0.024 |
| | # of chronic diseases | 0.100* | 0.115* | 0.009 | 0.009 |
| | CESD-mild | | | | |
| | CESD-severe | | | | |
| Medical insurance (base: nation MI) | employ medical insurance | | | | |
| | other medical insurance | | | | |
| Employment | weekly work hours | -0.007* | -0.004 | -0.002 | -0.005 |
| | spouse's weekly work hours | -0.003 | -0.002 | 0.005 | 0.003 |
| | wage calculation: hour | 0.527** | 0.857*** | 0.813*** | 0.575* |
| | wage calculation: daily | 1.358*** | 0.968*** | 0.292 | 0.223 |
| | wage calculation: monthly | 0.007 | 0.537* | -0.439 | -0.743* |
| | wage calculation: self-employ | 0.189 | 0.421 | -0.213 | -0.561* |
| Subjective Attitude to work | feel time pressure and lots of work | | | | |
| | involve physical labor | | | | |
| | have hope for promotion | | | | |

| | | | | | |
|---|---|---|---|---|---|
| | have little discretion over work | | | | |
| | can gain new skills | | | | |
| | satisfied with current pay | -0.380*** | -0.427*** | -0.878*** | -0.641*** |
| | satisfied with job | -0.185 | -0.360** | -0.244 | -0.172 |
| Spouse's Pension type | national pension incidence | | | | |
| (base: other pension) | employ pension incidence | | | | |
| Observations | | 3,594 | 2,288 | 1,580 | 1,349 |

*3.5.2 Heterogeneity analysis by gender*

The preference related to labor supply can be different for men and women due to the differential roles in the household. In Japan, female labor participation rate is relatively low in developed countries, partly due to cultural norms and gender roles. While male workers are known for high prevalence of being overemployed. Table 6 investigates the heterogeneity by gender.

More chronic diseases and more liquid assets increase overemployment for both men and women, but rising liquidity only reduces men’s underemployment. Fewer children and more education increase overemployment for women. Previous research generally finds that more children tend to make younger female workers feel overemployed due to harder work-life balance. This contrast probably reveals that children’s role is changing from time investment at earlier life stage into family support in later life.

Men are less likely to feel underemployed as they age. Women are less likely to be underemployed as their husbands earn more, but men are not, probably because of the traditional gender norms. However, when there is more total loan, women are significantly more likely to be underemployed.

An increase in the number of weekly work hours significantly increased the likelihood of both men and women feeling overemployed, but only reduced the likelihood of women feeling underemployed. When men and women's wages are calculated in a relatively flexible manner, i.e., in hours or days, they are both more likely to be underemployed, especially for women. In addition, both men and women are more likely to feel overemployed and want to work fewer hours when they feel stressed at work or their work involves physical labor. This pattern is slightly stronger for women. However, women were significantly less likely to feel overworked if they are able to acquire new skills on the job or they are satisfied with their current salary. And satisfaction with salary would reduce underemployment for both genders.

In summary, due to the cultural norm that Japanese men work outside and women work at home, male workers are less sensitive to work hour mismatch. For women, there is room to further encourage their labor supply especially for those with less family support and higher financial pressure. A combination of decent working hours and wages, the ability to acquire new skills from a job, and the avoidance of excessive physical labor, etc. will all contribute to female workers' work hour match.

Table 6. Heterogeneity results by gender

| | **Factors** | **Overemployment** | | **Underemployment** | |
|---|---|---|---|---|---|
| | | Male | Female | Male | Female |
| Demography | age | -0.020 | -0.082 | -0.350*** | -0.227 |
| | age squared | 0.000 | 0.001 | 0.003*** | 0.001 |
| | education years | 0.022 | 0.105*** | 0.012 | -0.027 |
| | # of children | -0.048 | -0.117** | -0.013 | -0.013 |
| | # of siblings | -0.007 | 0.005 | -0.010 | -0.024 |
| Income & Assets | income | 0.003 | -0.028 | 0.019 | 0.004 |
| | spouse's income | 0.017 | -0.014 | -0.016 | -0.053** |
| | liquid assets | 0.031** | 0.034** | -0.028** | -0.024 |
| | fix assets | -0.012 | 0.012 | -0.005 | -0.001 |
| | total loan | 0.003 | 0.005 | 0.004 | 0.016* |
| Health and Healthcare use | OOP expenditure of health check | -0.000 | -0.000 | -0.000 | -0.000 |
| | dentist visit incidence | -0.330 | 0.824** | | |
| | OOP expenditure of dentist visit | 0.037 | -0.063* | -0.007 | 0.015 |
| | outpatient visit incidence | 0.175 | 0.174 | | |
| | OOP expenditure of outpatient visit | -0.009 | 0.002 | 0.011 | -0.021* |
| | # of chronic diseases | 0.116*** | 0.186*** | 0.046 | 0.082 |
| | CESD-mild | 0.036 | 0.055 | | |
| | CESD-severe | 0.082 | 0.046 | | |
| Medical insurance (base: nation MI) | employ medical insurance | -0.019 | 0.153 | | |
| | other medical insurance | 0.763*** | 0.410*** | | |
| Employment | weekly work hours | 0.005* | 0.022*** | -0.003 | -0.013*** |
| | spouse's weekly work hours | -0.006** | -0.006** | 0.003 | -0.003 |
| | wage calculation: hour | | | 0.919*** | 0.479** |
| | wage calculation: daily | | | 0.760*** | 1.266*** |

| | | | | | |
|---|---|---|---|---|---|
| | wage calculation: monthly | | | 0.029 | -0.158 |
| | wage calculation: self-employ | | | 0.030 | -0.068 |
| Subjective Attitude to work | feel time pressure and lots of work | 0.459*** | 0.464*** | | |
| | involve physical labor | 0.198** | 0.280** | | |
| | have hope for promotion | 0.098 | -0.096 | | |
| | have little discretion over work | -0.064 | 0.043 | | |
| | can gain new skills | 0.040 | -0.287** | | |
| | satisfied with current pay | -0.151 | -0.260** | -0.554*** | -0.393*** |
| | satisfied with job | | | -0.182 | -0.261* |
| Spouse's Pension type (base: other pension) | national pension incidence | 0.030 | 0.282* | | |
| | employ pension incidence | -0.035 | 0.274* | | |
| Observations | | 5,173 | 3,638 | 5,173 | 3,638 |

*3.5.3 Heterogeneity analysis by work type*

Usui et al. (2016) has documented that salaried workers tend to report underemployment and self-employed workers tend to be overemployed. They call for further investigation into the determinants of over-(under-) employment and point out the potential heterogeneity across work types. Table 7 investigates this heterogeneity by wage employment and self-employment.

For the self-employed, the more children and the more total loan they have, the less likely they are to feel overemployed, indicating that financial reasons are important drivers of self-employed workers' willingness to work. Since self-employed people in Japan have to pay for health insurance on their own, when they have more chronic diseases, they are more likely to view themselves as underemployed and want to work more to pay for their health insurance. But for the salaried, having more chronic illnesses only significantly increases overemployment and reduce their work intention. When wages are calculated on a flexible basis, such as on an hourly or daily basis, the self-employed are more likely than the salaried to consider themselves as underemployed.

However, the work hour mismatch of the self-employed is more influenced by their subjective attitude towards work compared to salaried workers. Job stress, physical demandingness, and dissatisfaction about current pay increase overemployment to a larger extent for self-employed workers than for salaried workers.

On the whole, to relieve the self-employed from overemployment, policy makers need to relax their financial constraints and reduce their work pressure. Proving financial support when they face adverse health conditions can also promote self-employed workers' work hour match. Meanwhile, salaried workers with current working hours not long enough, low job satisfaction, and inadequate liquid assets still have potentials to extend work hours and get their human capital fully utilized.

Table 7. Heterogeneity results by work type

| | **Factors** | **Overemployment** | | **Underemployment** | |
|---|---|---|---|---|---|
| | | Self-employed | Salaried | Self-employed | Salaried |
| Demography | age | -0.042 | 0.020 | -0.161 | -0.358*** |
| | age squared | 0.000 | -0.000 | 0.001 | 0.003*** |
| | education years | 0.066** | 0.041** | 0.036 | -0.019 |
| | # of children | -0.200*** | -0.044 | 0.086 | -0.052 |
| | # of siblings | 0.018 | -0.015 | -0.013 | -0.014 |
| Income & Assets | income | 0.024 | -0.031 | -0.004 | 0.024 |
| | spouse's income | 0.012 | 0.008 | -0.036 | -0.020 |
| | liquid assets | 0.042** | 0.033*** | -0.003 | -0.036*** |
| | fix assets | -0.006 | 0.001 | -0.013 | 0.000 |
| | total loan | -0.024* | 0.011* | 0.008 | 0.010 |
| Health and Healthcare use | OOP expenditure of health check | -0.000 | -0.000 | -0.000 | -0.000 |
| | dentist visit incidence | 0.107 | 0.065 | | |
| | OOP expenditure of dentist visit | -0.030 | 0.011 | -0.018 | 0.010 |
| | outpatient visit incidence | 0.041 | 0.192* | | |
| | OOP expenditure of outpatient visit | -0.006 | -0.003 | 0.016 | -0.011 |
| | # of chronic diseases | 0.110 | 0.153*** | 0.116* | 0.044 |
| | CESD-mild | 0.162 | -0.008 | | |
| | CESD-severe | 0.089 | 0.071 | | |
| Medical insurance (base: nation MI) | employ medical insurance | 0.039 | 0.035 | | |
| | other medical insurance | 0.526*** | 0.389*** | | |
| Employment | weekly work hours | 0.017*** | 0.008*** | 0.005 | -0.013*** |
| | spouse's weekly work hours | -0.002 | -0.004* | 0.003 | -0.004 |

| | | | | | |
|---|---|---|---|---|---|
| | wage calculation: hour | | | 1.104*** | 0.507** |
| | wage calculation: daily | | | 1.354*** | 0.779*** |
| | wage calculation: monthly | | | -0.270 | 0.009 |
| | wage calculation: self-employ | | | -0.105 | 0.446 |
| Subjective Attitude to work | feel time pressure and lots of work | 0.723*** | 0.382*** | | |
| | involve physical labor | 0.371** | 0.187** | | |
| | have hope for promotion | -0.034 | 0.025 | | |
| | have little discretion over work | 0.117 | -0.099 | | |
| | can gain new skills | 0.008 | -0.081 | | |
| | satisfied with current pay | -0.268* | -0.206** | -0.711*** | -0.427*** |
| | satisfied with job | | | -0.072 | -0.294*** |
| Spouse's Pension type (base: other pension) | national pension incidence | 0.003 | 0.138 | | |
| | employ pension incidence | 0.209 | -0.022 | | |
| Observations | | 3,032 | 5,779 | 3,032 | 5,779 |

**4. Discussion**

In this work, we focus on the determinants of Japanese older workers' overemployment and underemployment behavior. We use JSTAR 2007, 2009, 2011, and 2013 waves. We set out to use machine learning techniques, namely, "Random Forest", "XG-Boost", and "Lasso" methods, to select important determinants for "overemployment" and for "underemployment". Comparing the cross-validation performance, we choose the "Random Forest" as our preferred prediction method.

We find that demographic characteristics (e.g. age, education, number of children and number of siblings), income & assets (e.g. income, fixed assets, loans), health and healthcare use (e.g. number of chronic diseases, OOP expenditures, CESD), subjective attitude to work, own medical insurance and spousal pension types are important predictors for "overemployment", while demographic characteristics, income & assets, healthcare use, current employment conditions are predictive to "underemployment" among older Japanese workers.

To depict latent classes within the overemployed and underemployed, we further use K-means clustering method and find that "overemployed workers" mainly comprises of three subgroups: "self-employed, older, poor health" people, people with "no children, no caregiving burden, and being single", and "rich, relatively young" people. And "underemployed workers" mainly consists of "younger, salaried", "older, poor health" and "older, poor" individuals. Clearly, reasons for being overemployed and underemployed are highly heterogeneous. For each subgroup, policies need to be tailor-made to properly promote work hour match.

We also find substantial heterogeneity by age groups, employment status, and gender. We find that workers above age 67 with high work pressure and low salary satisfaction are more likely to report overemployment, while workers above 57 with financial stress and salary dissatisfaction tend to feel underemployment. Worse health and high work stress tend to drive male workers' overemployment, while female workers with lower spousal income and more unstable jobs (paid by day or hour) wish to work longer. Financial constraints and job pressure drive self-employed workers' overemployment, while salaried workers would like to work more if their current working hours are not long enough and their job satisfaction and level of liquid assets are low.

Our results show highly heterogeneous determinants of overemployment and underemployment for different subgroups, which imply that policy-makers need heterogeneous ways to incentivize people to work optimal hours from different labor market segments.

*Policy implications*

As a countermeasure against the decline in the working-age population due to the falling birthrate and aging population, the Japanese government has long been trying to promote the employment of the elderly and women. One important take-away from the above findings is that there is much room to promote labor supply at least for 65+ workers facing financial stress and lacking family support, female workers with unstable jobs and low spousal income, and salaried workers not working enough hours. Well targeted policy to encourage working longer would be effective for these subgroups.

We also find that workers whose wage is paid by days or hours tend to report underemployment. These payment methods are usually associated with informal employment (e.g. temporary jobs or part-time jobs). This finding points to the need for further policy efforts to improve salary, benefits, and the disadvantaged labor market position of the non-regular employees.

While making efforts to fully utilize the human capital of current workers, the policy makers also need to carefully balance workers' need for reducing working hours. As indicated by the findings in this study, bad health conditions, high job pressure, less generous medical insurance, long current working hours are all key drivers behind overemployment. Financial support for disease-stricken old workers, better access to generous medical insurance and support for insurance contributions for self-employed workers, and more flexible working arrangement for female workers are plausible policy measures to alleviate the distress from overemployment.

Another heads-up for policy makers is that they need to address underemployment and overemployment integratively. Extending working hours for certain subgroups may inadvertently exacerbate the overemployment issue for other groups. For example, would extending older people's work hours result in loss of employment opportunities for younger people and women? This could be good question for future research. But being aware of heterogeneous incentives and difficulties within the work hour mismatched population, and keeping alert to the potential spillover effects of policy measures across subgroups would be recommended for policy makers.

*Limitations and future directions*

This study makes the first step to employ machine learning methods to analyze determinants of overemployment and underemployment. We choose from the most popular tree-based methods and linear regression models with regularization. Future research could integrate a wider range of methods, e.g., including forward/backward selection methods to select determinants, or employing support vector machine methods to pin down pivotal determinants. More evidence from other countries or institutional contexts are also needed to assess how determinants can differ across countries and cultures. Last but not least, this study focuses on the association between determinants and over-(under-) employment incidence, thus we refrain from any causal interpretation. Future efforts could be made to analyze the causal effects of certain determinants on overemployment and underemployment.

## Reference

Angrave, David, and Andy Charlwood. "What is the relationship between long working hours, over-employment, under-employment and the subjective well-being of workers? Longitudinal evidence from the UK." *Human Relations* 68.9 (2015): 1491-1515.

Asgari, Behrooz, Peter Pickar, and Victoria Garay. "Karoshi and Karou-jisatsu in Japan: causes, statistics and prevention mechanisms." *Asia Pacific Business & Economics Perspectives* 4.2 (2016): 49-72.

Bandara, Samanthi. "What Factors Matter for Choice of Livelihood among the Older Workers in Sri Lanka?." *Indian Journal of Human Development* 11.2 (2017): 210-231.

Bell, David NF, and Alasdair C. Rutherford. "Older workers and working time." *The journal of the economics of ageing* 1 (2013): 28-34.

Bell, David NF, and David G. Blanchflower. "The well-being of the overemployed and the underemployed and the rise in depression in the UK." *Journal of Economic Behavior & Organization* 161 (2019): 180-196.

Bell, David NF, and David G. Blanchflower. "Underemployment in the United States and Europe." *ILR Review* 74.1 (2021): 56-94.

Campbell, Iain, and Brigid van Wanrooy. "Long working hours and working-time preferences: Between desirability and feasibility." *Human Relations* 66.8 (2013): 1131-1155.

Chambers, Mark, Blair Chapman, and Eleanor Rogerson. "Underemployment in the Australian labour market." *RBA Bulletin*, June (2021): 86–99.

Charles, Kerwin Kofi, and Philip DeCicca. "Hours flexibility and retirement." *Economic Inquiry* 45.2 (2007): 251-267.

Choi, Eunhee, et al. "Understanding work enjoyment among older workers: The significance of flexible work options and age discrimination in the workplace." *Journal of gerontological social work* 61.8 (2018): 867-886.

De Moortel, Deborah. "Underemployment, Overemployment, and Mental health." *Handbook of Socioeconomic Determinants of Occupational Health: From Macro-level to Micro-level Evidence*. Cham: Springer International Publishing, 2022. 225-240.

Dvouletý, Ondřej. "Underemployment and overemployment in Central Europe." *Economics and Business Letters* 12.2 (2023): 147-156.

Fernandez, Jacqueline Liza, and Lim Ee Shiang. "Overemployment of Workers in Penang, Malaysia: An Empirical Analysis." *The Journal of Asian Finance, Economics and Business* 4.4 (2017): 17-26.

George, Usha, et al. "Underemployment and life satisfaction: A study of internationally trained engineers in Canada." *Journal of Immigrant & Refugee Studies* 10.4 (2012): 407-425.

Gielen, Anne C. "Working hours flexibility and older workers' labor supply." *Oxford economic papers* 61.2 (2009): 240-274.

Girtz, Britta. "Counting the Hours: The Influence of Gender and Occupation on Work Hour Mismatch Prevalence and Resolution in Australia." *Labor Studies Journal* 46.2 (2021): 101-133.

Golden, Lonnie, and Tesfayi Gebreselassie. "Overemployment mismatches: The preference for fewer work hours." *Monthly Labor Review* 130 (2007): 18.

Hara, Hiromi, and Hiroki Sato. "Work-Life Conflict from the Perspective of the Gap Between the Reality of Working Hours and Hopes--Realizing Work-Life Balance (Japanese)."*Japanese journal of research on household economics* 79 (2008): 72-79

Iwasaki, Kenji, Masaya Takahashi, and Akinori Nakata. "Health problems due to long working hours in Japan: working hours, workers' compensation (Karoshi), and preventive measures." *Industrial health* 44.4 (2006): 537-540.

James, Gareth, et al. *An introduction to statistical learning*. Vol. 112. New York: springer, 2013.

Kaufman, Bruce E. "Adam Smith's economics and the modern minimum wage debate: The large distance separating Kirkcaldy from Chicago." *Journal of Labor Research* 37 (2016): 29-52.

Li, Nan, Huanhuan Liang, Yi Gao, and Dan Wu. "Short-and Long-Term Effects of Underemployment on Workers' Health: Empirical Analysis from the China Labor Force Dynamics Survey." *International Journal of Environmental Research and Public Health* 19.24 (2022): 16695.

Lo, Simon. "Bilateral Bargaining and Working Hour Mismatches." *Available at SSRN 4575810* (2023).

Otterbach, Steffen. "Mismatches between actual and preferred work time: Empirical evidence of hours constraints in 21 countries." *Journal of Consumer Policy* 33 (2010): 143-161.

Otterbach, Steffen, et al. "Working-time regulation, long hours working, overemployment and mental health." *The International Journal of Human Resource Management* 32.22 (2021): 4659-4686.

Pagan, Ricardo. "Impact of working time mismatch on job satisfaction: Evidence for German workers with disabilities." *Journal of Happiness Studies* 18 (2017): 125-149.

Paz, Lourenco. "Does age modulate the impact of import competition on job type?." *Economics and Business Letters* 11.4 (2022): 161-171.

Raykov, Milosh. "Employer Support for Innovative Work and Employees' Job Satisfaction and Job-related Stress." *Journal of occupational health* 56.4 (2014): 244-251.

Reich, Ricarda. "Work hour mismatches and sickness absence and the moderating role of human resource practices: Evidence from Germany." *German Journal of Human Resource Management* (2023): 23970022231193085.

Reynolds, Jeremy, and Lydia Aletraris. "Pursuing preferences: The creation and resolution of work hour mismatches." *American Sociological Review* 71.4 (2006): 618-638.

Reynolds, Jeremy, and Lydia Aletraris. "Work–family conflict, children, and hour mismatches in Australia." *Journal of Family Issues* 28.6 (2007): 749-772.

Reynolds, Jeremy, and Lydia Aletraris. "Mostly mismatched with a chance of settling: Tracking work hour mismatches in the United States." *Work and Occupations* 37.4 (2010): 476-511.

Reynolds, Frances, Alexandra Farrow, and Alison Blank. "“Otherwise it would be nothing but cruises”: exploring the subjective benefits of working beyond 65." *International Journal of Ageing and Later Life* 7.1 (2012): 79-106.

Roh, Youn Hee, et al. "The effects of income and skill utilization on the underemployed's self-esteem, mental health, and life satisfaction." *Journal of Employment Counseling* 51.3 (2014): 125-141.

Ruiz-Quintanilla, S. Antonio, and Rita Claes. "Determinants of underemployment of young adults: A multi-country study." *ILR Review* 49.3 (1996): 424-438.

Schalembier, Benjamin, et al. "How relative income affects work hours preferences." *Applied Economics* 51.51 (2019): 5545-5558.

Sewdas, Ranu, et al. "Why older workers work beyond the retirement age: a qualitative study." *BMC public health* 17.1 (2017): 1-9.

Silver, Michelle Pannor, et al. "Getting the hours you want in the preretirement years: Work hour preferences and mismatch among older Canadian workers." *Work, Aging and Retirement* 5.2 (2019): 175-188.

Usui, Emiko, Satoshi Shimizutani, and Takashi Oshio. "Are Japanese men of pensionable age underemployed or overemployed?." *The Japanese Economic Review* 67.2 (2016): 150-168.

Usui, Emiko, Satoshi Shimizutani, and Takashi Oshio. "Health capacity to work at older ages: Evidence from Japan." *Social security programs and retirement around the world: The capacity to work at older ages*. University of Chicago Press, 2017. 219-241.

Van Solinge, Hanna, and Kène Henkens. "Older workers' emotional reactions to rising retirement age: The case of the Netherlands." *Work, aging and retirement* 3.3 (2017): 273-283.

Wilkins, Roger. "The consequences of underemployment for the underemployed." *Journal of Industrial Relations* 49.2 (2007): 247-275.

Wooden, Mark, Diana Warren, and Robert Drago. "Working time mismatch and subjective well-being." *British Journal of Industrial Relations* 47.1 (2009): 147-179.

Yamaguchi, Kazuo. "Overemployment: The structure, determinants, and countermeasures of involuntarily long work hours (Japanese)." *RIETI Discussion Paper Series* (2008) No. 08-J-051.

Yamaguchi, Kazuo. "Overemployment of full-time workers and work-family conflict (Japanese)." *RIETI Discussion Paper Series* (2010) No. 10-J-008.

Zitikytė, Kristina. "Employment at 55+: do we want to work longer in Lithuania?." *Ekonomika* 99.1 (2020): 50-6

## Online Appendix A: Additional Tables

**Table A.1 Definition of all candidate determinants**

| Variables | | | Note | Variable name |
|---|---|---|---|---|
| **Demographics** | age | | | age |
| | gender | | | gender |
| | education years | | If the respondents did not graduate, his/her education years equaled the education years of the last education level. | eduyear |
| | whether the respondent gets married or not | | | marry |
| | whether the respondent is salaried or not | | | salaried |
| | whether the respondent has any children | | | child |
| | the number of children | | | childnum |
| | the number of siblings | | | num_sibling |
| | whether the respondent owns any house | | | ownhouse |
| | whether the respondent is certified as eligible to receive nursing care | | | can_getcare |
| | whether the respondent is primarily insured | | | primary_insur |
| | whether the respondent has private health insurance | | | privatemi |
| | the type of public health insurance | whether the respondent has a national health insurance | base: the respondent did not have any public health insurance | nationmi |
| | | whether the respondent has an employment health insurance | | employmi |
| | | whether the respondent has other public health insurance | | othermi |
| | the type of pension benefit | whether the respondent has a national pension | base: the respondent did not have any pension | nation_pension |
| | | whether the respondent has an employment pension | | employ_pension |

| | | | | |
|---|---|---|---|---|
| | | whether the respondent has other pensions | | other_pension |
| | the type of the spouse's pension benefit | whether the spouse has a national pension | If the respondent did not have a spouse or his/her spouse did not have any pension, it equals 0. | spnation_pension |
| | | whether the spouse has an employment pension | | spemploy_pension |
| | | whether the spouse has other pensions | | spother_pension |
| | the spouse's age | Whether the spouse is 50 years old or younger | base: the respondent did not have a spouse | spage_50b |
| | | Whether the spouse is between 51 and 60 years old | | spage_50_60 |
| | | Whether the spouse is between 61 and 70 years old | | spage_60_70 |
| | | Whether the spouse is above 70 years old | | spage_70up |
| | the spouse's education | whether the spouse's education is less than lower secondary | base: the respondent did not have a spouse | spedu3_d1 |
| | | whether the spouse's education is upper secondary & vocational training | | spedu3_d2 |
| | | whether the spouse's education is tertiary | | spedu3_d3 |
| **Income & assets** | log(the total income of the respondent + 1) | | | income |
| | log (the total income of the spouse +1) | | | spincome |
| | log (the total liquid assets in the respondent's name +1) | | | tot_liq |
| | log (the total loan + 1) | | The sum of the house loan and other loan | tot_loan |
| | log (the fixed assets of the respondent +1) | | The market value of the owned house | fix_asset |
| **Informal caregiving** | whether the respondent's parents are alive | | | par_living |
| | whether the spouse's parents are alive | | If the respondent did not have a spouse, it equals 0. | sppar_living_d |
| | whether the respondent was involved in caring for parents last year | | | care_par |
| | whether the respondent was involved in caring for the spouse’s parents last year | | If the respondent did not have a spouse or his/her spouse's | care_sppar_d |
| | whether the spouse was involved in caring for the respondent's parents last year | | | spcare_par_d |

| | | | | |
|---|---|---|---|---|
| | whether the spouse was involved in caring for the spouse's parents last year<br>the number of total hours that the spouse's parents need nurse care hours at home<br>the number of total hours that the respondent's parents need nurse care hours at home | | parents are not alive, it equals 0. | spcare_sppar_d<br>par_nurhour<br>sppar_nurhour |
| **Employment-related** | whether the spouse has any work | | If the respondent did not have a spouse, it equals 0. | spwork_d |
| | the work type of spouse | whether the spouse is salaried<br>whether the spouse is self-employed | If the respondent did not have a spouse or the respondent's spouse did not have a job, it equals 0. | spsalaried_d<br>spselfemp_d |
| | the number of the spouse's work hours per week<br>the number of the respondent's work hours per week | | | spworkhour<br>workhour |
| | the type of working hours | whether the respondent's working hours vary each week, but the respondent's work is year-round<br>whether the respondent works during some seasons but not at other time | base: the respondent generally works the same hours every week | wktime_d2<br>wktime_d3 |
| | the way of calculating wage | 1. calculated the salary by hour; 0. otherwise or no work<br>1. calculated the salary by month; 0. otherwise or no work<br>1. calculated the salary by month; 0. otherwise or no work<br>1. self-employed thus no calculating; 0. otherwise or no work<br>1. calculated the salary in other ways; 0. otherwise or no work | | salary_calculate_hour<br>salary_calculate_day<br>salary_calculate_month<br>salary_calculate_self<br>salary_calculate_other |
| | the subjective attitude of the current job | whether the current job involves physical labor<br>whether the hope for promotion is likely to be realized for this job | | job_physical_d<br>job_promotion_d |

| | | | | |
|---|---|---|---|---|
| | | whether can receive appropriate evaluation on this work from co-workers | | job_evaluation_d |
| | | whether the respondent is satisfied with the current job | | job_satisfied_d |
| | | whether has an opportunity to gain new skills in the current job | | job_gainskill_d |
| | | whether colleagues would give advice and help the respondent when he/she has any problems doing work | | job_support_d |
| | | whether the respondent has a lot of work and always feels time pressure in this job | | job_pressure_d |
| | | whether the respondent can decide something on his/her own in this job | | job_decide_d |
| | | whether the respondent is satisfied with the current pay | | job_satispay_d |
| **Health-related** | BMI | whether the BMI index is lower than 18 | base: normal BMI between 18 and 25 | bmi_under |
| | | whether the BMI index is greater than 25 | | bmi_over |
| | CESD scores | whether CESD scores are between 10 and 20 | base: CESD scores are lower than 10 | cesd_mild |
| | | whether CESD scores are greater than 20 | | cesd_severe |
| | self-reported health (dummy variable) | | It equals 1 if the respondent self-resorts good health condition and 0 otherwise. | srh |
| | spouse's self-reported health (dummy variable) | | If the respondent did not have a spouse, it equals 0. | spsrh_d |
| | whether the respondent has any difficulty performing activities of daily living (ADL) | | | ADL_d |
| | whether the spouse has any difficulty performing activities of daily living (ADL) | | | spADL_d |
| | whether the respondent's physical or mental condition interferes with daily life | | | interfer_life |
| | the number of chronic diseases | | | num_chronic |

<table>
<tr><td colspan="2">whether the respondent has any diseases</td><td></td><td>chronic</td></tr>
<tr><td rowspan="19">whether have a certain chronic disease</td><td>heart disease</td><td rowspan="19"></td><td>heart</td></tr>
<tr><td>high blood pressure</td><td>hblood</td></tr>
<tr><td>hyperlipidemia</td><td>hyper</td></tr>
<tr><td>cerebral accident</td><td>cerebral</td></tr>
<tr><td>diabetes</td><td>diabetes</td></tr>
<tr><td>lung disease</td><td>lung</td></tr>
<tr><td>asthma</td><td>asthma</td></tr>
<tr><td>liver disease</td><td>liver</td></tr>
<tr><td>stomach disease</td><td>stomach</td></tr>
<tr><td>arthritis</td><td>arthritis</td></tr>
<tr><td>broken hip</td><td>hip</td></tr>
<tr><td>osteoporosis</td><td>osteo</td></tr>
<tr><td>eye disease</td><td>eye</td></tr>
<tr><td>ear disorder</td><td>ear</td></tr>
<tr><td>bladder disorder</td><td>bladder</td></tr>
<tr><td>Parkinson’s disease</td><td>parkinson</td></tr>
<tr><td>dementia</td><td>dementia</td></tr>
<tr><td>skin disorder</td><td>skin</td></tr>
<tr><td>cancer</td><td>cancer</td></tr>
<tr><td colspan="2">whether the respondent has physical examination or not last year</td><td rowspan="3">If the respondent did not have any health check, it equals 0</td><td>health_check</td></tr>
<tr><td colspan="2">log (the expenditure of health check +1)</td><td>hcheck_fee</td></tr>
<tr><td colspan="2">whether the respondent has seen a dentist in the past year</td><td>dentist</td></tr>
</table>

| | | | | |
|---|---|---|---|---|
| | log (the expenditure of dentist visit +1) | | If the respondent did not have any dentist visits, it equals 0 | dentist_fee |
| | whether the respondent has a doctor visit last year | | | outpatient |
| | log (the expenditure of outpatient care +1) | | It only includes the regular visit fee. If the respondent did not have any outpatient care, it equals 0. | outpatient_fee |
| | whether the respondent has spent one or more nights in the hospital in the past year | | | inpatient |
| | log (the expenditure of inpatient care + 1) | | If the respondent did not have any inpatient care, it equals 0 | inpatient_fee |
| | whether the respondent smokes | | | smoke |
| | drinking frequency | 1. drink 5~6 days per week; 0. otherwise | base: the respondent drinks alcohol daily | drink_d2 |
| | | 1. drink 3~4 days per week; 0. otherwise | | drink_d3 |
| | | 1. drink 1~2 days per week; 0. otherwise | | drink_d4 |
| | | 1. drink several times per month; 0. otherwise | | drink_d5 |
| | | 1. hardly ever or never drink; 0. otherwise | | drink_d6 |
| **Wave** | survey year at 2009 (base: 2007) | | | wave_d2 |
| | survey year at 2011 (base: 2007) | | | wave_d3 |
| | survey year at 2013 (base: 2007) | | | wave_d4 |

**Table A.2 Scoring coefficients of PCAs for overemployment and underemployment**

| | | Overemployment | | Underemployment | |
|---|---|---|---|---|---|
| | **Variables** | PC1 | PC2 | PC1 | PC2 |
| **Demographics** | age | 0.2668 | -0.0818 | -0.2550 | 0.1291 |
| | gender | -0.0719 | 0.0205 | 0.0648 | 0.0763 |
| | eduyear | -0.1387 | 0.0300 | 0.1162 | -0.0519 |
| | marry | -0.1464 | -0.3237 | 0.1740 | 0.2908 |
| | salaried | -0.1708 | 0.1192 | 0.0791 | -0.1711 |
| | child | -0.0382 | -0.1829 | 0.0589 | 0.1812 |
| | childnum | -0.0268 | -0.1819 | 0.0468 | 0.1862 |
| | num_sibling | 0.1123 | -0.0334 | -0.1028 | 0.0956 |
| | ownhouse | -0.0522 | -0.1941 | 0.0636 | 0.1497 |
| | can_getcare | -0.0692 | -0.0051 | 0.1324 | 0.0398 |
| | primary_insur | -0.0738 | 0.1424 | 0.0183 | -0.0650 |
| | privatemi | -0.0904 | -0.0220 | 0.1252 | 0.0404 |
| | nationmi | 0.2554 | -0.0673 | -0.1872 | 0.1219 |
| | employmi | -0.2578 | 0.0502 | 0.1913 | -0.1126 |
| | othermi | -0.0622 | 0.0396 | 0.1358 | -0.0154 |
| | nation_pension | 0.1819 | -0.0767 | -0.0964 | 0.0786 |
| | employ_pension | -0.1575 | 0.0626 | 0.0944 | -0.0562 |
| | other_pension | 0.0684 | 0.0282 | -0.0852 | -0.0650 |
| | spnation_pension | -0.0006 | -0.1388 | 0.0236 | 0.1821 |
| | spemploy_pension | -0.1149 | -0.1626 | 0.1354 | 0.1011 |
| | spother_pension | 0.0112 | -0.0538 | 0.0080 | 0.0049 |
| | spage_50b | -0.1140 | 0.0272 | 0.1251 | -0.0169 |

| | | | | | |
|---|---|---|---|---|---|
| | spage_50_60 | -0.1992 | -0.0862 | 0.2131 | 0.0695 |
| | spage_60_70 | 0.1091 | -0.1339 | -0.0829 | 0.1678 |
| | spage_70up | 0.1300 | -0.1069 | -0.1116 | 0.0889 |
| | spedu3_d1 | 0.0733 | -0.0686 | -0.0369 | 0.1252 |
| | spedu3_d2 | -0.0712 | -0.1391 | 0.0986 | 0.1217 |
| | spedu3_d3 | -0.0922 | -0.0598 | 0.1009 | 0.0442 |
| **Income & assets** | income | -0.0327 | -0.0397 | 0.0105 | 0.0096 |
| | spincome | -0.0776 | -0.2962 | 0.1406 | 0.2288 |
| | tot_liq | -0.0108 | -0.0760 | 0.0437 | 0.0317 |
| | tot_loan | -0.1177 | -0.0355 | 0.0974 | 0.0400 |
| | fix_asset | -0.0531 | -0.1924 | 0.0742 | 0.1352 |
| **Informal caregiving** | par_living | -0.1766 | 0.0300 | 0.1805 | -0.0669 |
| | sppar_living_d | -0.2053 | -0.0984 | 0.2338 | 0.1069 |
| | care_par | -0.1134 | 0.0079 | 0.1506 | -0.0645 |
| | care_sppar_d | -0.1755 | -0.0765 | 0.2107 | 0.0817 |
| | spcare_par_d | -0.1464 | -0.0492 | 0.1808 | 0.0077 |
| | spcare_sppar_d | -0.1491 | -0.0809 | 0.1861 | 0.0665 |
| | par_nurhour | -0.0072 | 0.0169 | 0.0125 | -0.0391 |
| | sppar_nurhour | -0.0266 | -0.0276 | 0.0293 | 0.0435 |
| **Employment-related** | spwork_d | -0.1405 | -0.3006 | 0.1906 | 0.2516 |
| | spsalaried_d | -0.1893 | -0.1869 | 0.1938 | 0.1170 |
| | spselfemp_d | 0.0551 | -0.1628 | 0.0166 | 0.1878 |
| | spworkhour | -0.1094 | -0.2835 | 0.1767 | 0.2309 |
| | workhour | -0.0934 | 0.0580 | 0.1334 | -0.0207 |
| | wktime_d2 | 0.0304 | -0.0013 | -0.0363 | 0.0371 |

| | | | | | |
|---|---|---|---|---|---|
| | wktime_d3 | 0.0877 | -0.0606 | -0.0612 | 0.0738 |
| | salary_calculate_hour | 0.1482 | -0.0721 | -0.0746 | -0.0179 |
| | salary_calculate_day | 0.0768 | 0.0235 | 0.0453 | 0.0107 |
| | salary_calculate_month | -0.1157 | 0.0729 | 0.2056 | -0.0567 |
| | salary_calculate_self | 0.1825 | -0.0699 | 0.0035 | 0.1295 |
| | salary_calculate_other | 0.1125 | -0.0117 | -0.0351 | 0.0678 |
| | job_physical_d | 0.0722 | 0.0320 | -0.0035 | -0.0161 |
| | job_promotion_d | -0.0034 | 0.0258 | 0.0065 | -0.0197 |
| | job_evaluation_d | -0.0374 | -0.0531 | 0.0184 | 0.0382 |
| | job_satisfied_d | -0.0484 | -0.0639 | 0.0115 | 0.0386 |
| | job_gainskill_d | -0.0804 | -0.0089 | 0.0565 | 0.0019 |
| | job_support_d | -0.0603 | -0.0327 | 0.0539 | -0.0073 |
| | job_pressure_d | -0.0606 | 0.0271 | 0.0640 | -0.0436 |
| | job_decide_d | 0.0171 | 0.0185 | 0.0082 | -0.0048 |
| | job_satispay_d | -0.0353 | -0.0423 | 0.0159 | -0.0038 |
| **Health-related** | bmi_under | -0.0054 | 0.0024 | 0.0137 | 0.0008 |
| | bmi_over | 0.0310 | -0.0377 | -0.0296 | 0.0074 |
| | cesd_mild | 0.0483 | 0.0064 | -0.0150 | 0.0075 |
| | cesd_severe | 0.0105 | 0.0239 | -0.0245 | 0.0229 |
| | srh | -0.0732 | 0.0776 | 0.0719 | -0.1361 |
| | spsrh_d | 0.0392 | -0.0660 | -0.0425 | 0.0845 |
| | ADL_d | 0.1165 | -0.0760 | -0.1233 | 0.1238 |
| | spADL_d | -0.0189 | -0.0504 | -0.0103 | 0.0461 |
| | interfer_life | 0.0600 | -0.0638 | -0.0712 | 0.1287 |
| | num_chronic | 0.1260 | -0.1490 | -0.1170 | 0.1927 |

| | | | | | |
|---|---|---|---|---|---|
| | chronic | 0.1024 | -0.1406 | -0.0981 | 0.1504 |
| | heart | 0.0573 | -0.0546 | -0.0603 | 0.0988 |
| | hblood | 0.0839 | -0.0835 | -0.0762 | 0.1029 |
| | hyper | 0.0224 | -0.0547 | 0.0017 | 0.0482 |
| | cerebral | 0.0498 | -0.0496 | -0.0751 | 0.0800 |
| | diabetes | 0.0594 | -0.0784 | -0.0592 | 0.0882 |
| | lung | 0.0103 | -0.0602 | -0.0102 | 0.0021 |
| | asthma | 0.0639 | -0.0170 | 0.0082 | 0.0299 |
| | liver | -0.0214 | -0.0282 | -0.0010 | 0.0382 |
| | stomach | 0.0105 | -0.0160 | 0.0189 | 0.0257 |
| | arthritis | 0.0673 | -0.0536 | -0.0607 | 0.0627 |
| | hip | 0.0061 | -0.0082 | -0.0196 | 0.0220 |
| | osteo | 0.0580 | -0.0763 | -0.0643 | 0.0459 |
| | eye | 0.0600 | -0.0465 | -0.0768 | 0.0788 |
| | ear | 0.0261 | -0.0496 | 0.0039 | 0.0324 |
| | bladder | 0.0518 | -0.0565 | -0.0525 | 0.0823 |
| | parkinson | -0.0071 | -0.0196 | 0.0000 | 0.0000 |
| | dementia | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | skin | 0.0147 | 0.0083 | -0.0101 | 0.0299 |
| | cancer | 0.0188 | -0.0490 | -0.0172 | 0.0547 |
| | health_check | -0.1252 | 0.0133 | 0.0802 | -0.0511 |
| | hcheck_fee | -0.0107 | -0.0080 | 0.0352 | -0.0268 |
| | dentist | -0.0140 | -0.0339 | -0.0220 | 0.0418 |
| | dentist_fee | -0.0174 | -0.0270 | -0.0223 | 0.0390 |
| | outpatient | 0.0473 | -0.1340 | -0.0716 | 0.1110 |

| | | | | | |
|---|---|---|---|---|---|
| | outpatient_fee | 0.0966 | -0.1570 | -0.0858 | 0.1386 |
| | inpatient | 0.0427 | -0.0762 | -0.0683 | 0.1380 |
| | inpatient_fee | 0.0416 | -0.0804 | -0.0689 | 0.1382 |
| | smoke | -0.0523 | 0.0563 | 0.0389 | -0.0330 |
| | drink_d2 | -0.0143 | 0.0521 | 0.0063 | 0.0111 |
| | drink_d3 | -0.0375 | -0.0461 | 0.0195 | -0.0122 |
| | drink_d4 | -0.0147 | 0.0050 | 0.0189 | 0.0163 |
| | drink_d5 | -0.0194 | 0.0063 | 0.0128 | 0.0264 |
| | drink_d6 | 0.0727 | -0.0002 | -0.0685 | -0.0488 |
| | wave_d2 | -0.0163 | 0.0202 | 0.0376 | -0.0032 |
| **Wave** | wave_d3 | 0.0328 | 0.0277 | -0.0919 | -0.0440 |
| | wave_d4 | 0.0525 | -0.0433 | -0.0837 | 0.0135 |

Note: Please refer to Table A.1 for the specific definition of variables.