June 2019

Empirical Essays on Causal Effects of Job Training Programs: Evidence From Korea

Seokchae Hwang
Syracuse University

ABSTRACT

This dissertation comprises two chapters that are related to the research topics in Korean job training programs using administrative data. The first chapter examines the effectiveness of the programs on the probability of re-employment and on wages. This observational study tackles multiple biases by adopting: (1) a propensity score matching approach for selection into treatment, and (2) a principal stratification framework for selection into employment when looking at wages. This chapter finds positive effects on employability (0.026), but negative effects on wages (-8.4%), both of which are statistically significant. We conduct sensitivity analysis to violations of our main identifying assumption that confirms our results are robust to unobserved confounding. The second chapter analyzes the heterogeneous treatment effects of the programs on employability using a recent causal forest estimator, which is a machine learning technique. This chapter finds that almost a third of trainees (31.0%) is likely to experience negative effects, in spite of a positive and significant average treatment effect on the treated (0.029). We illustrate distinctive characteristics of the two groups that are affected positively and negatively. Based on these characteristics, we suggest some alternative assignment rules and find that several of the suggested assignment rules outperform the current one.

Empirical Essays on Causal Effect of Job Training Programs: Evidence from Korea

by

Seokchae Hwang

B.S., Seoul National University, 2004

M.A., Syracuse University, 2016

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics.

Syracuse University

June 2019

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor Alfonso Flores-Lagunes for his insightful advice and enormous support throughout the process. His professional guidance nurtured my enthusiasm, and his positive attitude encouraged my research during my humble years in Syracuse. I am hugely appreciative to my passionate co-author, Chung Choe, for his constant support and invaluable suggestions. I am also very grateful to my dissertation defense committee: Hugo Jales, Menxiao Liu, William C. Horrace, Yoonseok Lee, and the chair Leonard M. Lopoo. I am deeply grateful for their invaluable advice and suggestions on my work. I also thank the weekly labor group meeting, hosted by Alfonso Flores-Lagunes, Carmen Carrión-Flores, and Hugo Jales, for providing opportunities to discuss and get constructive feedback. A special gratitude goes out to Stanley Ziemba, Edward Godwin, and other staff at the Maxwell IT department, for providing computing resource for my dissertation. I also thank the other faculty members and the staff of the Economics department and the Center for Policy Research for their support.

Personally, I thank Hoang Pham, Jonathan Presler, Xiuming Dong, and my other fellow graduate students, and friends who have provided words of encouragement and fruitful discussions at this completion. I am grateful to my parents, parents-in-law, and my brother who have supported my dreams and appraised another chapter of my life. Last but most importantly, I am extremely indebted to my beloved wife, Shinae, who has provided unconditional support and encouragement by standing on my side. Without her, I could not finish my long journey of the Ph.D. program. I would dedicate my work to her with my heartfelt gratitude.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Job training programs (hereafter JTPs) have been recognized as an important active labor market policy (ALMP) throughout the world (Heckman et al., 1999; Card et al., 2010; Kluve, 2010). In South Korea, it has been highlighted as one of the major policies since the economic crisis in 1997, which raised the unemployment rate abruptly from 2.6% in 1997 to 7.0% in 1998 (Choe et al., 2015). The fragile economic climate brought extensive industrial restructuring including large scale layoffs, outsourcing, and mergers and acquisitions (M&A) (Chopra et al., 2001). To accommodate the sudden changes in the labor market, the Korean government expanded JTPs on a large scale.

JTPs played a meaningful role to deal with the high unemployment rate, as expected to increase human capital investment to cope with the labor demand.[1] Moreover, the Korean government renovated JTPs in 2008-2011, to improve effectiveness in response to the labor market shock from another financial crisis in 2008. More specifically, three important changes were made to the policy. First, the government introduces a voucher program, which motivates training agencies to focus more on trainees who play a main role

[1] The "Job Training Stimulation Act", which took effect as of 1997, states that the primary objectives of JTPs is to develop trainees' capabilities in order to help them to be employed and eventually reduce labor market mismatch.

in subsidizing process. Consequently, the agencies are likely to focus more on trainees' demand, so that the agencies are likely to supply better quality JTPs to attract trainees. Second, a licensing system is adopted which improves market accessibility of training agencies. Agencies can freely open and operate JTPs if they satisfy certain conditions to be a legitimate providers, such as facilities, instructors and financial status, whereas previously only a limited number of agencies could operate classes under the government's permission. This change stimulates competition among the agencies so as to provide programs with better quality. Third, displaced workers are eligible to receive both UI benefits and JTPs simultaneously, while they previously were able to receive either UI benefits or JTPs at a time, even though they were eligible for both. The effectiveness of JTPs becomes more tractable, as the two benefits do not compete with each other.

The government's budget for JTPs steadily increased from KRW 41.4 billion in 2008 to KRW 546 billion in 2017, and correspondingly the number of participants increased from 94 thousand to 220 thousand (Ministry of Employment and Labor, 2009; 2018). Despite its growing importance, the effectiveness of JTPs is still under debate in South Korea. Specifically, some studies suggest that JTPs are likely to increase employability of trainees (e.g. Yoo and Lee, 2008), while other studies argue that less job-search time of trainees due to JTPs ("lock-in" effect) may impede the effectiveness of JTPs in the short run (e.g., Choi and Kim, 2012). The effects of training programs on wages are more ambiguous, as the previous literature shows diverse results (e.g., Yoo and Lee, 2008; Chae and Kim, 2004). More importantly, despite the 2011 substantial reform, no research has paid attention to the effectiveness of the reformed JTPs.

Against this backdrop, the primary focus of this dissertation is to estimate the effectiveness of JTPs after the 2011 reform. We use rich administrative data which combines three different sources. We provide evidence whether JTPs improve employability and wages by comparing the effects of the reformed JTPs with the previous studies. This extends our understandings of the effectiveness of the reformed JTPs in South Korea. To the best of our knowledge, this is one of the first attempts to analyze the reformed JTPs. The analysis can contribute to the empirical literature by providing not only evidence on the after-the-reform effectiveness but also policy prescriptions.

More specifically, the first chapter, which I co-authored with Alfonso Flores-Lagunes, Chung Choe, and Sangjun Lee, addresses two common biases in observational studies. To deal with one bias, selection into treatment, we use the popular propensity score matching. Even after the matching, another bias may arise when estimating causal effects on wages because wages are observed depending on employment status; the wage distribution is truncated by non-employment. We use the principal stratification framework suggested in Frangakis and Rubin (2002) to tackle selection into employment. This framework analyzes the effects on wages within subgroups defined by potential employment status (e.g., Zhang et al., 2009; Frumento et al., 2012). In addition, as our inference relies on the unconfoundedness assumption, we implement sensitivity analysis to gauge the robustness of our results to the existence of unobserved confounders (Schwartz et al., 2012; Bia et al., 2018).

Our findings suggest that the reformed JTPs have a positive and statistically significantly effect (0.026) on the probability of re-employment in accordance with their intention. However, the results show a negative and statistically significant effect (-8.4 percent) on wages. We conjecture that the negative effect may result from the wrong assignment.

This chapter contributes to the literature in three important ways. First, this study is, to the best of our knowledge, the first attempt to analyze the effectiveness of the reformed JTPs in South Korea, and provide consistent results by successfully controlling for multiple selection biases. Second, we define four different types of individuals by their potential employability, and illustrate their characteristics which may provide evidence of our results. Third, we conduct sensitivity analysis to gauge robustness of our results to potential unobserved measures affecting selection..

The second chapter, which I co-authored with Alfonso Flores-Lagunes and Chung Choe, analyzes heterogeneous treatment effects of Korean JTPs. Previous studies have been interested in traditional estimators such as average treatment effect (ATE) and ATE on treated (ATT), which can only assess the general tendency of policy effects. We extend our inference to analyze heterogeneous treatment effects which can explain variations across targets. In general, analysis of heterogeneous treatment effects needs to be carefully addressed due to multiple testing problems (List et al., 2016). Research in observational studies typically investigates interaction terms under regression analysis or conduct subgroup analyses set by their prior information (Athey and Imbens, 2017). However, these approaches cannot examine all the important heterogeneity, and the analysis heavily relies on researchers' prior information. Consequently, it is hard to implement systematic analysis which can investigate unexpected important heterogeneity.

In order to address these difficulties, machine learning techniques have received considerable attention over the past few decades. In particular, in the program evaluation literature, the use of machine learning techniques has remarkable potentialities because they show excellent performance in "prediction" (e.g., Varian, 2014; Mullainathan and Spiess, 2017). These methods are primarily based on data-driven approaches, thus they were treated as "black-box" models at the early stage due to lack of theoretical background. However, recent research suggests several promising methods which provide valid statistical inference while examining effect heterogeneity (e.g., Imai and Ratkovic, 2013; Belloni et al., 2014, 2017; Nie and Wager, 2017; Wager and Athey, 2018; Athey et al., 2019). Based upon these theoretical progress, empirical studies have followed to shed light on heterogeneous treatment effects (e.g., Tian et al., 2014; Bertrand et al., 2017; Davis and Heller 2017, 2018; Andini et al., 2018; Kleinberg et al., 2018; Knaus et al., 2018; Strittmatter, 2018).

We use the most recent causal forests suggested by Athey et al. (2019) among others, which is considered as one of the best estimators for effect heterogeneity (Knaus et al., 2018). The common concerns including spurious error and multiple testing problems can be addressed by the 'honest' process, which uses two mutually exclusive subsamples that are randomly divided: one is for identifying subgroups which shares similar characteristics, and the other is for estimating treatment effects. The independent implementation of the two tasks contributes provably valid statistical inference, that is, the forest estimator is consistent and asymptotically normal.

We sketch the distribution of treatment effects, and find a considerable proportion (31%) of trainees has a negative effect, despite of positive and statistically significant effect (0.029). We systematically examine heterogeneous treatment effects to uncover the most influential characteristics, and find that positive effects are observed for those are more likely old, female, to have 2-year college or bachelor's degree, work experience and longer tenure, separated due to firm-oriented reasons (e.g., layoff) or employment expiration (e.g., retirement), and to have shorter inactive duration.

Based on our findings, we suggest several hypothetical assignment rules and find most of the rules outperform the current one, for example, assignment on those who have longer tenure (0.112), on those who are separated due to firm-oriented reasons (0.074), employment expiration reasons (0.057), and those who are the elderly (0.074). We also show that the current voluntary assignment rule may suffer from adverse selection in the sense that it is worse than random assignment rule (0.041).

This chapter contributes to the literature in three ways. First, this research provides one of the early empirical studies using the most recent causal forest estimator. Second, this study systematically examines heterogeneous treatment effects, which reveals that a substantial amount of individuals can experience negative treatment effects, although the average treatment effect is significantly positive. Third, this study uncovers substantial characteristics that have large impacts on effectiveness of JTPs, which provide evidence of better treatment assignment rules for policymakers.

Chapter 2

Estimating Average Treatment Effects of Korean Job Training Programs

2.1 Introduction

In numerous countries, job training programs (hereafter, JTPs) have been widely recognized as a major part of active labor market policies over the past decade. In South Korea, JTPs have been powerful tools for dealing with the rapidly worsening unemployment rate since the economic crisis of 1997 (Choe et al., 2015). To address another crisis in 2008, the necessity for re-training labor force has raised the government budget from KRW 41.4 billion ($\approx$ USD 37 million) in 2008 to KRW 546 billion ($\approx$ USD 488 million) in 2017, and correspondingly the participants have increased from 94 thousand in 2008 to 220 thousand in 2017. The importance of JTPs has grown further since the major reform in 2011 that aims to improve their effectiveness by introducing more demand-oriented policies.[1] Many scholarly works have studied Korean JTPs for the displaced workers, but there is no consensus about their impact on labor market performance. Moreover, despite the substantial change in 2011, few studies have addressed the effectiveness of the reformed JTPs. Thus, from both policy and economic perspectives, the need for analyzing their effectiveness increases, which assesses the effectiveness of the programs and provides policy implications.

Evaluation in the selection-on-observables setting is usually complicated due to multiple potential selection biases. One potential bias occurs when individuals are not randomly assigned into treatment, which usually occurs in observational studies. Another bias may arise from non-random selection into employment: wage information cannot be observed for all individuals but only for those who are employed. Since individuals self-select into finding and holding employment, such selection is likely not independent of having received training, and creates a post-treatment bias. Each bias escalate complication of estimating causal effects.

[1] The government budget for JTPs increased 129.4% during 6 years after the reform (2011-17), while the budget had increased 71.2% for the previous 6 years (2005-11). We will revisit this institutional change in the next section.

One of the popular methods to deal with selection into treatment is propensity score matching, suggested in Rosenbaum and Rubin (1983a). Many empirical researches use this method in many fields including economics, social science and biomedical science (Imbens and Rubin, 2015). To tackle selection into employment, namely post-treatment bias, Heckman's selection model is widely used in the econometrics literature (e.g., Heckman, 1979; Ahn and Powell, 1993; Vella, 1998). The selection model however relies on exclusion restriction assumption for point identification and disregards the existence of subgroups having a negative effect, i.e., defier, under the monotonicity. Recently, Frangakis and Rubin (2002) suggest the principal stratification framework, which defines mutually exclusive principal strata using the combination of joint potential (intermediate) outcome. Under the framework, causal interpretation can be possible within well-distributed subsamples (principal strata). Empirical studies are followed using the framework to estimate average treatment effects in a randomized experiment (e.g., Zhang et al., 2009; Frumento et al., 2012). More recently, Bia et al. (2018) extends its application to a selection-on-observables setting, which takes into account both biases mentioned above.

This study contributes to extend upon the above studies to analyze the effects of JTPs on employability and wages using observational data from South Korea. Our identification strategy relies on a two-stage process to deal with both selection biases as in Bia et al. (2018). However, we have some distinctive features relative to the previous studies. We specifically implement propensity score matching to deal with selection into treatment, given the large size of the control group in our administrative data. We also use the estimated linearized propensity score as a regressor to achieve computationally tractable estimation by reducing the dimensionality of covariates during our estimation process. Finally, unlike Bia et al. (2018), we do not need to employ an auxiliary outcome variable to undertake statistical inference.

Our analysis contributes to the literature in at least three important ways. First, this study is, to the best of our knowledge, the first attempt to analyze the effectiveness of the reformed JTPs in South Korea, and extends upon empirical evidence of the JTPs' impacts both on employability and wages in South Korea. Second, we provide consistent results by successfully controlling for multiple selection biases. We find evidence of a positive effect on the probability of re-employment, and a negative effect on wages for individuals who are employed regardless of treatment assignment, which are both statistically significant. Third, we define four different types of individuals by their potential employability (principal strata), and sketch their characteristics to provide some implications for policymakers. Methodologically, we employ propensity score matching and the principal stratification framework sequentially to deal with multiple selection biases. We use the estimated linearized propensity score as a regressor in our estimation, and find that our results are consistent under stochastic dominance assumptions. We also conduct sensitivity analysis to gauge robustness of our results.

The remainder of this paper is organized as follows. The next section provides some background on Korean JTPs, and reviews the literature in South Korea. In Section 2.3, we describe the data used and define our sample. In Section 2.4, we explain the theoretical framework and key assumptions used to identify causal effects. In Section 2.5, we explain our estimation method. Section 2.6 presents our results and interpretation. In Section 2.7, we implement sensitivity analysis to formally gauge the robustness of our findings. The final section provides a conclusion and policy implications.

2.2 Background

2.2.1 Korean Job Training Programs

Prior to the Reform in 2011

The main purposes of JTPs in South Korea are (1) to re-train displaced workers,[2] (2) to upgrade the skills of incumbent workers, and (3) to improve employability of new entrants to labor market (Ministry of Employment and Labor, 2018). In this paper, we focus on the JTPs for displaced workers which have the largest proportion of Korean JTPs. The government's budget for them constitutes more than half of total amount of the budget for JTPs. More specifically, in 2017, KRW 546 billion was used to fund the JTPs for displaced workers, amounting to approximately 52.4% of total budget for JTPs.[3]

Since JTPs are financed by the central government, the government's intervention seems reasonable.[4] However, the government intervened not only to promote job-matching, but also to determine details of operating JTPs, which may give rise to inefficiency. For example, training agencies needed a permit from the government to operate JTPs as legitimate providers, and were allowed to open limited classes with trainee quota under the 'government workforce plan'. This top-down decision making process were likely to be inflexible on the trainees' demand, and gave rise to malfunction in three main ways.

First, limited discretion were allowed for training agencies to operate their own JTPs. The agencies had less motivation to improve quality of their programs, and they

[2]In this study, we define "displaced workers" as persons 17 years of age and over who lost and left jobs involuntarily because of their personal (changing jobs, family issues, disease or injury, disciplinary dismissal, or miscellaneous), firm-oriented (layoff, shutdown, exploitation), or expiration issues (retirement, contract expiration).

[3]Ministry of Employment and Labor, Overview of the Ministry Budget Report in 2018, South Korea [white paper]

[4]A few of the programs were occasionally financed by the local government. For example, some special programs for farmers, fisheries, North Korean refugees, and workers in the key industries designated by the government were supported by the local governments given their regional and geographical interests. Je-ju (province) also operated its own JTPs separately due to regional and administrative reasons.

were likely to be passive to react toward trainees' needs. Trainees also experienced limitations. Since the limited programs could not fully take into account their preference, some trainees had difficulties to develop their human capital by JTPs. This could lead JTPs less preferable.

Second, training agencies' incentive structure was likely distorted by the fund flow of training subsidies. Since the government supported JTPs by delivering subsidies directly to the agencies, trainees could not play any role in the process. The weak role of trainees allowed the agencies to engage more with the government—a primary funding source—than with trainees for securing more fund, and hindered the agencies from improving the quality of programs. Therefore the program quality was likely to remain relatively low.

Lastly, displaced workers could not receive both unemployment insurance (UI) benefits and training subsidies simultaneously, even if they were eligible for both; they were allowed to receive only one at a time. As many of the displaced workers preferred to receive UI benefits first because of the magnitude and fewer conditions, JTPs were likely to be treated as extra subsidies for those who had used up their UI benefits and remained still unemployed.

The Reformed JTPs under the New Scheme after 2011

Having noted the above problems, the Korean government introduced a new subsidy package called the 'Individual Training Account' system, and a few substantial reforms became effective nationwide as of 2011.[5] First, the government introduces a voucher program instead of the direct subsidizing system. A voucher amounts annually up to KRW 2 million is supplied to trainees. Using this voucher, trainees can take any JTPs from training agencies, and the agencies will be reimbursed from the government. The voucher draws the agencies' attention to trainees who become key players for their funding, and

[5] This system was adopted in some regions in 2008 as a pilot project and expanded nationwide from 2011.

expect to improve JTPs' quality for appealing to trainees. To prevent moral hazard, trainees are charged 20-50% copayment of their training expenses.[6]

Second, the government issues a license for legitimate training agencies instead of a permit with trainee quota. The new system allows any institutions' free entry into the market for JTPs unless they cannot meet the minimum requirement such as financial stability, facilities and instructors. More agencies are allowed to operate JTPs with larger discretion, and are expected to open various JTPs with better quality. This change also improve trainee's accessibility to more programs by motivating the agencies to provide more competitive JTPs that fit to trainees' needs. Consequently, this change is expected not only to cultivate autonomy for agencies but also to provide JTPs with better quality to trainees.

Lastly, displaced workers are able to receive both UI benefits and training subsidies simultaneously. The JTPs become more relevant to their purpose which aims to develop trainees' ability, as those two benefits do not compete with each other. Also, the effectiveness of JTPs becomes more tractable in the sense that training subsidies are no longer treated as a second chance after UI benefits.

Consequently, once a worker is separated from employment due to some reasons[7] they need to register at Worknet to be eligible for the governmental support including training subsidies or UI benefits. During the registration process, they are asked to file their information at Worknet, which can be used as a background information or as a resume to apply for a job within Worknet. After the registration, a worker is required to meet with a caseworker at a local office, and the caseworker will issue a voucher to an eligible worker.

[6]The Korean government also implements a special subsidy program called "employment success package" for targeted groups, such as low-income families, for which the cost of training was subsidized up to KRW 3 million and exempt from out-of-pocket payments (Ministry of Employment and Labor, 2014). Since this program is operated separately, we do not include this group in our analysis.

[7]We classify the reasons into three categories. Personal reasons: changing jobs, disease/injury, and disciplinary dismissal. Firm-oriented reasons: temporary closure, shutdown, and layoff. Employment expiration reasons: retirement, contract expiration, and project completion.

2.2.2 Literature

Most of the previous studies analyze the effects of JTPs prior to the 2011 reform. Under the former scheme, estimating the causal effect of JTPs seems more complicate because the choice between UI benefits and JTPs is likely endogenous. For this reason, Lee and Lee (2005) and Yoo and Lee (2008) define their main interest as the relative effect of JTPs versus UI benefits instead of effect of JTPs taker versus JTPs non-taker; that is, they shed light on preference of government programs between JTPs and UI benefits.

More specifically, Lee and Lee (2005) conclude that JTPs are less effective than UI benefits as JTPs have lengthened the unemployment duration of Korean women, in the connected administrative data of the UI file and the job training file from 1999-2000. Yoo and Lee (2008) also compare the effectiveness between the two subsidized groups using similar administrative data as in Lee and Lee (2005) but more recent one from 2002. They conduct simple logit regressions, and find that JTPs are more effective than UI benefits by increasing re-employment probability by around 13%. Choi and Kim (2012) try to disentangle effects of the two government supports, by defining their sample in terms of application status for UI benefits; they analyze the causal effect of JTPs for those (1) UI benefit applicants and (2) non-applicants. They employ a similar dataset as in Lee and Lee (2005) and Yoo and Lee (2008), but a more recent one from 2007. They use propensity score matching estimators to address selection into treatment, and find that the average treatment effect on treated on the probability of re-employment at 12 months of their job separation is insignificant, which is almost zero. The estimated effect is observed as negative because trainees are "locked-in" training programs in the short run, which is likely observed due to their inclusion of training duration into job search period for trainees. However the effect becomes zero at 12 months and gradually increase approximately up to 8%p at 18 months. To address endogeneity of training participation, they conduct additional analysis using the fixed effects model with a panel dataset, called Korea Labor and Income Panel Study (KLIPS) data between 2005 and 2008. They find the long term effects become smaller than those effects from the matching estimator, and statistically insignificant. Their estimation results can be arguable in the sense that the UI applicants may contain both UI benefit recipients and non-recipients.

Compared to the literature that investigates JTPs' impact on the probability of re-employment, less studies shed light on the JTPs' causal effects on wages. Chae and Kim (2004) examine wage effects of JTPs using a similar administrative dataset of UI data and training files from 1999-2002, while restricting their samples within the UI recipients to estimate the JTPs' causal effect. They use Heckman's two-stage estimation to control for selection into employment, and find that the JTP takers received 2.8-7.4% less in wages than the control group. However, Yoo and Kang (2010) find that JTPs in Korea have a statistically significantly positive effect at 2.6-4.7% of the average monthly wages, using data from Current Population Survey Panel between 2007 and 2009. They define the sample within wage workers and examine wage differences before and after the programs with a fixed effects model. They conduct another analysis using propensity score matching estimator to deal with selection into treatment, and find larger positive effects around 7.6-9.8%, which is also statistically significant.

The mixed evidence in the above studies results from the use of different dataset and time period. Even among studies using the similar administrative data, the different definition of the control group and different identification strategies result in arguably different results. Our analysis addresses the two common selection biases using rich administrative data that is the most recent available one from 2012-2014. We compare our results with the previous research to indirectly gauge the effectiveness of the 2011 reform in terms of employability and wages.

2.3 Data

2.3.1 Administrative data

Under the Ministry of Labor and Employment in South Korea, the Korea Employment Information Service (KEIS) provided three sources of administrative data: (1) the human resource development (HRD) file, (2) the unemployment insurance (UI) file, and (3) the Worknet dataset.[8] The HRD file contains information on trainees such as demographics and start/end date of training. The UI file contains the last workplace related profiles of those eligible for UI benefits; details include (1) date of employment and separation (2) reasons for separation, (3) occupation type, (4) firm size, (5) location, and (6) initial monthly wage when re-employed. These two data sources are similar to those used in earlier studies (e.g., Lee and Lee, 2005, 2009; Choi and Kim, 2012; Choe et al., 2015).

However, our data have two advantages. First, ours are constructed after the 2011 reform, so that we can estimate effect of the reformed JTPs. Second, we exploit an additional dataset called "Worknet" which provides more information about the displaced workers; such as educational attainment, work experience within a given industry, the registration date at Worknet, certificate and other personal characteristics. Thus, our dataset is richer than the previous studies by including more covariates from the additional dataset.

[8]The Worknet, a portal website, was developed by the KEIS in 1998 to provide public job related information to job seekers. As of the 2011 reform, the target has expanded to firms in the private sector, so as to provide more comprehensive job related information including both public and private sectors. Displaced workers are required to register at Worknet to receive UI benefits or training subsidies from the government, which also encourage employers to participate at Worknet.

2.3.2 Sample definition

Our data focuses mainly on displaced workers who are eligible for UI benefits in South Korea. The sample includes those who lost and left their jobs involuntarily and show their intention to find a job by registering at Worknet, regardless of participation in JTPs.

The HRD file contains JTP participants—who start and finish their JTPs within 2013. These participants comprise the treatment group ($W_i = 1$) with a sample of 58,512 observations. Using the UI file and Worknet dataset, we construct the control group ($W_i = 0$) who is separated and registers at Worknet within 2013 but takes no JTPs. Our control group consists of a sample of 306,923 observations.

Employment status (S_i) indicates whether individuals are employed within 12 months since the start date of job-search, and this definition has two advantages comparing to that in the previous studies. One is the use of 12-month duration. The 12-month duration is common criteria to measure short-term effects (Card et al., 2010), and it has an advantage to control for seasonal recruitment than shorter duration such as 3-month or 6-month. The other is using different start date of job-search between the treatment and control group.[9] For the treated, it starts from the end of JTPs, which implies that training duration is not considered as job-search duration. This seems plausible because trainees are likely to focus on programs during training duration in the sense that trainees cannot be subsidized the training expenses unless they finish the programs in good attendance. Also, our definition can define causal effects of JTPs better by improving comparability between the two treatment arms since the length of JTPs up to 6 months may decrease job-search opportunity for trainees. For the control group, the 12-month duration starts from the date of registration at Worknet. The duration between the date of employment separation and registration at Worknet, say "inactive duration", is excluded from job-search period because we regard the duration as discouraged period in which the displaced workers are

[9] These definitions are similar with Choe et al. (2015).

not looking for a job. This also seems plausible in two ways. First, this is in line with the objectives of Worknet (job-matching) because the government does not treat them as job seekers eligible to receive any benefits until they register at Worknet. Also, since trainees need to register at Worknet in order to take JTPs, the similar rules at least should be applied for the control. In that sense, our duration of job-search conservatively increases comparability between the two treatment arms than the previous studies which measures both duration from separation.

Wage information (Y_i) indicates the monthly wage in the first month of re-employment.[10] We discard any outliers whose wage records are below the 1^{st} or above the 99^{th} percentile of wages among the employed (4,462 observations dropped). After dropping these observations, the size of our samples is 360,973 which consists of 58,132 individuals for the treated, and 302,841 individuals for the control.

The list of covariates (X_i) in our dataset is reported in Table 2.1. The set of covariates is chosen to cover observed characteristics of individuals, which serves as important controls for confounding factors in our specification. Our covariates include demographic (age, gender, disability), personal information (educational attainment, residential region, inactive duration, and work experience within a given industry), and the last workplace related information (tenure, occupation, industry, reasons for separation, location, firm size in persons). The "inactive duration", between the date of separation and registration at Worknet, is a variable created by ourselves, which is used as a proxy for how actively an individual wants re-employment. Since we postulate the registration date for the start of job-search, this variable implies that the earlier registrants are likely to escape from inactive status and start their job-search earlier than the others. The continuous age variable is converted into categorical dummies because in general the outcomes does not

[10]Hourly wages are used as a more accurate measure of human capital accumulation. However, our date does not contain hours worked to define the wage rate hourly. Thus, our monthly wage measure may be affected if the hours worked vary significantly among workers. We will discuss in a later section how our results may be affected by not having acces to hours worked.

have a linear relationship with age, but concave. We drop some observations that contains binary variables with fewer than 1% of participants to avoid multicollinearity.[11] Finally, our sample size for the analysis is 355,071 (treated: 57,164, control: 297,907).

2.3.3 Summary statistics

Table 2.1 shows descriptive statistics for the full list of covariates and outcome variables (employment status and monthly wage) in our sample defined by treatment status. The last column from Table 2.1 reports the standardized differences[12] to assess balance property between the treated and control. The use of standardized difference is desirable because we can avoid to underestimate differences unlike the t-test and other statistical tests, when we have a large sample. While there is no consensus on the criteria for imbalance, we use a rule of thumb suggested by Rosenbaum and Rubin (1985); the standardized differences less than 20 between the two groups are regarded as negligible. Table 2.1 suggests that a large number of covariates shows significant imbalance between the two treatment arms: gender (22.8), work experience within a given industry (60.1), age—less than 30 (35.7) and 60 or more (33.0), educational attainment—elementary school or less (26.6), and reasons for separation—personal miscellaneous (46.5) and exploitation (34.9).[13] More specifically, in the treatment group, the proportion of males is 37% which is significantly less than in the control group (48%). The treatment group (36.62 year-old) is generally younger than the control group (42.34 year-old): the treatment group are more likely 29 year-old or less (0.33 vs. 0.18), but less likely 60 year-old or more (0.04 vs. 0.13). The treatment group are less likely to have elementary school diploma or less (0.01 vs. 0.06), and middle school diploma

[11] The number of observations (percentage) of the dropped variables: Residential and Workplace region (Jeju: 2,951 (0.8%)), Industry (farming and fishing: 889 (0.25%), energy industry: 2,062 (0.58%))

[12] $d = \frac{100 \times |\bar{x}_t - \bar{x}_c|}{\sqrt{1/2 \times (\hat{s}_t^2 + \hat{s}_c^2)}}$, d: standardized difference, $\bar{x}_g$: sample mean of g, $\hat{s}_g$: sample standard deviation of g

[13] "Exploitation" indicates that any following violations occur for longer than 2 months within a year before employment separation: (1) employers urge workers to work overtime without extra payment, or employers pay less than (2) wages in the contract, (3) the minimum wage, and (4) 70% of wages during temporary closure of a firm.

(0.03 vs. 0.07). The treatment group is also likely to have less work experience within a given industry (0.39 vs. 0.67). The most common reason for separation is exploitation of a firm for the control group (0.40), personal miscellaneous issues for the treatment group (0.47). Tenure, inactive duration, residential region, last workplace location, industry, and the ex-firm size (in persons) are similar between the two treatment arms to some extent, while trainees are slightly more likely to have longer inactive duration (0.55 year vs. 0.44 year) and shorter tenure (1.68 year vs. 2.22 year) than the control group, and less likely to have worked as a manual laborer (0.10 vs. 0.14). Lastly, trainees are likely to earn significantly lower wage (KRW 1,456 $\times$ 10^3 vs. 1,582 $\times$ 10^3), but trainees have slightly better employability than the control (0.63 vs. 0.59) which is not significant.

Table 2.1: Summary Statistics of our Administrative Data

	Control (T_i=0)		Treated (T_i=1)		Std. Diff
Variable	Mean	Std. Dev.	Mean	Std. Dev.	.
Gender (male=1)	0.48	(0.50)	0.37	(0.48)	22.8
Not disabled	0.98	(0.13)	0.99	(0.11)	4.6
Inactive duration (years)	0.44	(0.58)	0.55	(0.64)	18.6
Work experience (within a given industry)	0.67	(0.47)	0.39	(0.49)	60.1
Tenure (years)	2.22	(3.88)	1.68	(3.10)	15.5
Age (years)	42.34	(12.89)	36.62	(11.02)	47.7
$\leqslant$ 29	0.18	(0.39)	0.33	(0.47)	35.7
30s	0.30	(0.46)	0.32	(0.47)	2.8
40s	0.21	(0.41)	0.20	(0.40)	3.9
50s	0.17	(0.38)	0.11	(0.31)	17.0
$\geqslant$ 60	0.13	(0.34)	0.04	(0.20)	33.0
Educational Attainment					
Elementary school or less	0.06	(0.25)	0.01	(0.12)	26.6
Middle school	0.07	(0.26)	0.03	(0.17)	19.3
High school	0.35	(0.48)	0.35	(0.48)	0.3
College	0.23	(0.42)	0.29	(0.45)	14.6
Bachelor's	0.26	(0.44)	0.30	(0.46)	8.9
Graduate or more	0.03	(0.16)	0.02	(0.13)	5.9
Residential Region					
Seoul	0.22	(0.41)	0.21	(0.41)	1.1
Gyeong-in	0.33	(0.47)	0.32	(0.47)	0.7
Gang-won	0.03	(0.16)	0.02	(0.14)	3.5
Chung-cheong	0.10	(0.29)	0.08	(0.28)	4.3
Jeolla	0.08	(0.28)	0.10	(0.30)	4.7
Gyeong-sang	0.25	(0.43)	0.26	(0.44)	2.6
Workplace Region					
Seoul	0.35	(0.48)	0.38	(0.49)	7.6
Gyeong-in	0.25	(0.44)	0.23	(0.42)	5.1

(Cont'd on next page)

Gang-won	0.02	(0.15)	0.02	(0.13)	3.3
Chung-cheong	0.09	(0.28)	0.07	(0.26)	4.2
Jeolla	0.07	(0.26)	0.08	(0.26)	1.5
Gyeong-sang	0.22	(0.41)	0.22	(0.41)	0.5
Industry					
Manufacturing	0.22	(0.42)	0.22	(0.41)	0.6
Construction	0.04	(0.20)	0.04	(0.19)	1.6
Wholesale/retail	0.17	(0.38)	0.17	(0.38)	0.3
Lodging/restaurant	0.05	(0.22)	0.05	(0.22)	1.9
Broadcasting/publishing	0.05	(0.22)	0.06	(0.23)	2.1
Finance/insurance	0.02	(0.14)	0.02	(0.15)	1.3
Real estate	0.02	(0.15)	0.02	(0.14)	1.4
Tech service	0.05	(0.22)	0.05	(0.22)	0.6
Facility management	0.13	(0.34)	0.13	(0.33)	0.1
Public administration	0.10	(0.29)	0.09	(0.28)	3.2
Social service	0.15	(0.36)	0.16	(0.36)	1.9
Ex-firm size (persons)					
$\leqslant$10	0.35	(0.48)	0.32	(0.47)	5.7
10-29	0.20	(0.40)	0.18	(0.38)	5.8
30-99	0.16	(0.37)	0.16	(0.37)	0.9
100-299	0.11	(0.31)	0.11	(0.32)	0.5
$\geqslant$300	0.18	(0.39)	0.23	(0.42)	12.6
Ex-occupation					
Manager	0.13	(0.34)	0.12	(0.33)	2.8
Professional/engineer	0.26	(0.44)	0.26	(0.44)	0.4
Supervisor/Clerk	0.18	(0.38)	0.22	(0.41)	9.1
Merchandiser	0.12	(0.33)	0.15	(0.36)	8.7
Operators	0.16	(0.36)	0.15	(0.35)	3.3
Farmer/Fishery	0.01	(0.10)	0.00	(0.06)	7.1
Manual Laborer	0.14	(0.35)	0.10	(0.30)	12.6
Reason for Separation					
Changing jobs	0.04	(0.19)	0.06	(0.24)	12.3

(Cont'd on next page)

Family issue	0.02	(0.12)	0.03	(0.16)	7.0
Disease/injury	0.02	(0.13)	0.02	(0.15)	4.9
Disciplinary Dismissal	0.03	(0.17)	0.02	(0.14)	6.6
Miscellaneous (personal)	0.25	(0.43)	0.47	(0.50)	46.5
Shutdown	0.04	(0.20)	0.02	(0.14)	12.4
Exploitation	0.40	(0.49)	0.24	(0.42)	34.9
Retirement	0.02	(0.12)	0.01	(0.09)	7.6
Contract Expiration	0.20	(0.40)	0.14	(0.34)	16.5
Monthly wage (KRW, $\times 10^3$)	1,582.4	(679.3)	1,455.9	(538.3)	22.1
Employed within 12 months	0.59	(0.49)	0.63	(0.48)	8.0
Observations	297,907		57,164		

2.4 Econometric Methodology

2.4.1 Basic Framework and Key Assumptions

We adopt the potential outcomes framework to identify our parameters of interest (Rubin, 1974). Throughout the paper, uppercase letters indicate random variables and lowercase letters indicate the realization of the random variables. The treatment assignment denoted as W_i is an indicator variable: 1 if participating in a JTP, and 0 otherwise. Let $Y_i(W_i)$ denote the potential outcomes—monthly wage—for an individual i in terms of treatment assignment W_i. S_i is a binary indicator which is equal to 1 if the individual is employed, and 0 otherwise. Since it is likely affected by the treatment, we denote the potential employment status by $S_i(W_i)$. Note that the potential values of employment determine the observability of the monthly wage outcome.

In our context, a first complication arises because of voluntary enrollment in JTPs, which may give rise to self-selection bias into programs (W_i). We will address this issue by adopting unconfoundedness (or selection-on-observables) assumption. The second complication arises from the observability of wages (Y_i) on employment status (S_i). That is, since S_i as an intermediate outcome variable it is likely affected by W_i, implying the non-random observability of Y_i. To address this, we will rely on principal stratification, which will be described below.

We start with the following identifying assumptions:

(Assumption 1) Stable Unit Treatment Value Assumption (SUTVA):

$$Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$$

(Assumption 2) Unconfoundedness: $(Y_i(1), Y_i(0)) \perp W_i | X_i$

(Assumption 3) Overlap: $0 < P(W_i = 1 | X_i) < 1$

The first assumption implies that all outcome variables of each individual are mutually independent (Rubin, 1978). This assumption holds when treatment assignment of an individual does not affect potential outcomes of others. This assumption is violated if there are general equilibrium effects due to the training program, as in that situation the potential outcomes of an individual depend on the potential outcomes of other individuals. It may also be violated if there are peer effects at work within the training program. In the first case, JTPs are not likely to create general equilibrium effects because total number of trainees is only 0.03% of the total job vacancy in 2013 (a period that corresponds with our data) which was 1.76 million. For the second case, we only point out that we are not aware of any peer or network effects among the trainees. Thus, even though those effects may exist, they may be unimportant and so we disregard them.

The second assumption is unconfoundedness, which implies that the distribution of the potential outcomes is independent from the treatment assignment conditional on the observed pre-treatment covariates. The unconfoundedness assumption implies that there are no unobserved confounders that jointly affect treatment and potential outcomes, conditional on observable factors (X). Under this assumption, we are able to estimate causal effects by comparing outcomes after conditioning on those observed factors. This assumption is untestable and has to be argued in substantive grounds. In our context, we rely on the rich administrative data at our disposal to argue that we have access to all variables that determine selection into the training program and that are simultaneously related to the potential outcomes. Indeed, we employ all available covariates in our linked administrative data. Furthermore, since this assumption is not testable, we will check the plausibility of the assumption through sensitivity analysis in the spirit of Rosenbaum and Rubin (1983b) in section 2.6.

Third, we impose the overlap assumption which implies that the probabilities of being treated are bounded away from zero or one. Intuitively, it requires that we are able to find a suitable comparison unit for individual i on the common support X. This assumption can be assessed by comparing the distribution of the estimated propensity score between the two treatment arms (Rosenbaum and Rubin, 1983).

2.4.2 Propensity Score Matching

The large imbalance between the two treatment arms documented in the previous section implies the likely existence of selection into JTPs. Our identification assumptions postulate that selection into JTPs is based on observable covariates. Therefore, we need to make sure that we undertake comparisons of the outcomes (employment and monthly wages) conditional on the covariates available to us. However, conditioning on all the available covariates (and functions of them) may be intractable due to the high dimensionality of the resulting vector X. Thus, we use the propensity score, defined as the probability of assigning treatment W_i (i.e., taking JTPs) given the observed covariates X_i, $e(X_i) = P(W_i = 1|X_i)$ (Rosenbaum and Rubin, 1983a), to flexibly condition on all covariates. As we will explain below, using the propensity score will also aid estimate our model from a computational perspective.

Since the propensity score is generally not known in observational studies, it needs to be estimated (for details, see Imbens and Rubin, 2015). A logit or a probit are the most popular estimators of the propensity score, among others (e.g., Dehejia and Wahba 1999; Gerfin and Lechner, 2002; Mercatanti and Li, 2017). Some alternatives have been proposed recently, including machine learning approaches like classification and regression trees (Breiman et al., 1984) and generalized boosted models (McCaffrey et al., 2004).

We use a standard logit regression to estimate the propensity score. For its specification, we adopt the forward stepwise selection method as suggested in Imbens and Rubin (2015). The method adds a covariate at each stage that gives the highest explanatory power by testing all possible variables X_i; we test covariates up to quadratic terms, and the process is repeated until no covariates are further selected.[14] While implementing the method, we exclude some highly correlated variables (i.e., larger than ± 0.99) to avoid multicollinearity, and finally choose 234 of relevant covariates (39 linear and 195 interaction terms).

Once the mode for the propensity score is estimated, we compute, for each individual in the sample, the estimated propensity score and its odds ratio—linearized propensity score ($l(x) = ln\frac{e(x)}{1-e(x)}$)—using the selected covariates. Given the estimated linearized propensity score, we conduct a one-to-one nearest neighbor matching algorithm without replacement to construct a matched control group. We do this since the pool of control group individuals is large relative to the pool of treated individuals. After the matching, the size of the matched control group will be the same as that of the treated group: 57,164 respectively. In section 2.6 we will show that this procedure results in acceptable balance of covariates between the two treatment arms, suggesting that this procedure is successful in creating comparison groups that will allow leveraging the

[14] We use the thresholds of t-statistic suggested in Imbens and Rubin (2015) as a rule of thumb: the threshold is 2.74 for linear terms, 1.00 for quadratic terms, respectively.

unconfoundedness assumption for estimation of causal effects.

2.4.3 Principal Stratification

The second empirical complication to deal with is the observability of monthly wages. This outcome variable is not observed for individuals that are not employed, and since employment status likely depends on whether an individual enrolled in JTPs, the "truncation" mechanism is non-random.[15] As a result, it is necessary to "control" for the endogenous employment status when estimating the effect of JTPs on the monthly wage. To address this issue, we use the principal stratification framework proposed by Frangakis and Rubin (2002). This framework is related to the more familiar framework of "compliance types" in instrumental variables estimation in Angrist, Imbens and Rubin (1996).

The principal strata are defined as the joint values of the potential employment status in terms of treatment assignment $(S_i(0), S_i(1))$. Our empirical setting gives rise to four mutually exclusive principal strata. The key feature of principal strata is that, for individuals within principal strata, their employment status is affected in the same way by the treatment, and thus comparisons of the monthly wage within a given strata yield causal effects. Note, however, that principal strata are latent subpopulations since only one of the potential values of employment are observed for each individual. This will create an estimation complication since individuals will need to be classified into each of the four strata to subsequently obtain the causal effect of JTPs on the monthly wage.

The four principal strata that arise in our framework are the following.

- EE: $i = \{i|S_i(1) = 1, S_i(0) = 1\}$, those who will be employed irrespective of the treatment assignment;

[15] In the biostatistics literature, since the outcome cannot be observed for those who died, this is called "truncation by death" (Zhang and Rubin, 2003).

- EN: $i = \{i|S_i(1) = 1, S_i(0) = 0\}$, those who will be employed only under treatment, not under control;
- NE: $i = \{i|S_i(1) = 0, S_i(0) = 1\}$, those who will be employed only under control, not under treatment; and
- NN: $i = \{i|S_i(1) = 0, S_i(0) = 0\}$, those who will not be employed irrespective of the treatment assignment.

One important aspect to note is that there is only one principal strata above, the EE for which the monthly wage is observed under the two treatment arms. In other words, the monthly wage is observed under both participating and not participating in JTPs only for individuals that are always employed irrespective of treatment assignment. For the remaining three principal strata the monthly wage is unobserved at least in one of the treatment arms. The implication is that, to identify the effect of JTP on the population or on principal strata other than the EE requires assumptions to allow imputation of those unobserved monthly wages. In order to avoid additional assumptions to extrapolate, the principal stratification approach concentrates on identification and estimation of the treatment effects for the EE strata (Zhang et al., 2008; Zhang et al., 2009; Frumento et al., 2012; Blanco et al., 2013; Bia et al., 2018). These effects are also known as principal strata effects.

2.4.4 Parameters of Interest and Identification

Under the aforementioned framework, our parameters of interest are average treatment effects on the treated (ATT), which evaluate effectiveness of JTPs for those who undergo training. In general, the ATT is useful to gauge the 'actual' policy effects, especially in observational studies such as ours.

The first parameter is the ATT on the probability of re-employment (ATT_{emp}).

Using the prior framework and denoting by π_g the proportion of individuals in principal strata g, the ATT_{emp} can be defined as the difference between the proportion of EN and NE:

$$\begin{aligned} ATT_{emp} &= \mathbb{E}(S_i(1) - S_i(0)|W_i = 1, X_i = x) \\ &= \pi_{EN} - \pi_{NE}. \end{aligned}$$

The first line is the definition of the ATT, and the second equality holds under the unconfoundedness assumption. The proportions π_g can be estimated under the unconfoundedness assumption after controlling for selection into treatment by using the propensity score matching, which matches only the treated individuals. It is also necessary to impute the strata membership for each individual. The details of the estimation procedure are presented in the following section.

The next parameter is the ATT on wages for the EE stratum ($ATT_{w,EE}$). This parameter can be expressed as the average wage differences between the two treatment arms, conditional on covariates and being in the treated group and the EE stratum:

$$\begin{aligned} ATT_{w,EE} &= E(Y_i(1) - Y_i(0)|W_i = 1, X_i = x, G_i = EE) \\ &= w_{EE,1} - w_{EE,0} \end{aligned}$$

where $w_{g,t}$ represents the average wage over individual i within strata g under the treatment status t. The first line is the definition of the ATT for the EE stratum, and the second equality holds under the unconfoundedness assumption. The average wages $w_{EE,1}$ and $w_{EE,0}$ can also be estimated following the unconfoundedness assumption (after propensity score matching) and a method for imputing the strata membership for each individual.

2.5 Maximum Likelihood Estimation

As we discussed earlier, we cannot observe the principal strata to which an individual belongs because $S_i(1)$ and $S_i(0)$ cannot be observed simultaneously. Define the observable groups by the joint values of the treatment (W_i) and the observed employment status(S_i^{obs}) as follows:

- O(1,1) $= \{i|W_i = 1, S_i^{obs} = 1\}$, those who are treated and employed;
- O(1,0) $= \{i|W_i = 1, S_i^{obs} = 0\}$, those who are treated but not employed;
- O(0,1) $= \{i|W_i = 0, S_i^{obs} = 1\}$, those who are not treated but employed; and
- O(0,0) $= \{i|W_i = 0, S_i^{obs} = 0\}$, those who are not treated and employed.

Each observable group is a mixture of two principal strata. Specifically, both EE and EN are observed in O(1,1), both EE and NE are observed in O(1,0), EN and NN are observed in O(0,1), and NE and NN are observed in O(0,0). Note that wages can only be observed in observed groups O(1,1) and O(0,1), and thus only for EE in both treatment arms.

Given the observability of only latent mixture distributions in the observed groups above, we need a systematic way to separate the wage distributions of two principal strata. When a model formulation involves latent data, the EM algorithm can separate two distributions by iteratively evaluating parameters while estimating parameters by maximum likelihood estimation (Dempster et al., 1977). The advantage of the EM algorithm is that membership of principal strata need not be explicitly known, because they are estimated during the iteration process.

In the context of our observational study, it is important to flexibly condition on the available covariates in the implementation of the EM algorithm. Unfortunately, prior studies explicitly conditioning on a large set of covariates report that the EM algorithm can become unstable or exhibit multimodality in the implied likelihood function.[16] For this reason, we use the estimated linearized propensity score as a single regressor instead of all the covariates. We find that by reparameterizing X to $l(X)$ the implied likelihood function from the EM algorithm is well behaved.[17]

To construct the maximum likelihood function for the EM algorithm, we need to impose some parametric assumptions. First, we assume a multinomial logistic model for the membership to principal strata conditional on covariates as in Zhang et al. (2008):

$$\pi_{g:i} = P\big[G_i = g\big] = \frac{exp\big[\gamma_g \cdot l(X_i)\big]}{\sum_{g'} exp\big[\gamma_{g'} \cdot l(X_i)\big]} \tag{2.1}$$

where $G_i = g$ and g' denote membership to principal strata $g, g' \in \{EE, EN, NE, NN\}$, γ_g are the parameters of the model, and NN is chosen as the omitted stratum (i.e., $\gamma_{NN} = 0$), without loss of generality.

In turn, we assume a log-normal model for the conditional wage distribution given covariates and the principal strata:

$$\begin{cases} \text{O(1,1): } log(Y_i(1)) \sim N(\mu_{g,1}, \sigma^2_{g,1}), \quad \text{where } \ g \in \{EE, EN\} \\ \text{O(0,1): } log(Y_i(0)) \sim N(\mu_{g,0}, \sigma^2_{g,0}), \quad \text{where } \ g \in \{EE, NE\} \end{cases}$$

[16] Multimodality refers to a problem of the likelihood function in which several modes exist, making it difficult to find the global optima.

[17] In our estimation results reported below, we find that the same numerical optimum of the likelihood function is found after attempting over 50 random starting values. Conversely, if we include all the covariates in estimation, the algorithm has problems converging to an optimum.

where $\mu_{g,t} = \alpha_{g,t} + l(x)\beta_{g,t}$; and $\alpha_{g,t}$, $\beta_{g,t}$ and $\sigma_{g,t}$ are parameters of the model, $t = 0, 1$.

In addition to the above parametric assumptions, we further impose stochastic dominance assumptions that facilitate the separation of the two strata distributions in each observable group. The assumptions imply that the wage distribution for EE stochastically dominates the one for EN in the treatment group, and that the wage distribution for EE stochastically dominates the one for NE in the control group. Formally, $P(Y_{EE}(1) \leqslant w) < P(Y_{EN}(1) \leqslant w)$ and $P(Y_{EE}(0) \leqslant w') < P(Y_{NE}(0) \leqslant w')$ for any w, w'.[18] These assumptions can be advocated by positive selection into employment, as individuals with higher capabilities are more likely to be employed (e.g., Blundell et al., 2007; Lechner and Melly, 2010; Blanco et al., 2013; Bia et al., 2018).

We now construct the likelihood function based on the above assumptions to estimate the treatment effects of interest. The likelihood function consists of multiplications of mixture models containing two principal strata. More specifically, the mixture distributions in each observable group O(1,1) and O(0,1) are convex combinations of probability density functions of principal strata. The distribution of O(1,1) consists of two log-normal distributions of EE and EN. Correspondingly, the distribution of O(0,1) includes two log-normal distributions of EE and NE. Given the mixture nature of the likelihood function, it is often the case that it exhibits multiple modes, which is ameliorated with the stochastic dominance assumptions and our use of a single regressor in the linearized propensity score.

The likelihood function for the model can be expressed as follows:

[18] To implement these stochastic dominance assumptions, we follow Zhang et al. (2009) and set the coefficient β_g and variances σ_g^2 to be the same in each observable group, but the constant terms α_g in the equations of {EE,T=1} and {EE,T=0} are set to be greater than those of {EN,T=1} and {NE,T=0}, respectively. That is, $\beta_{EE,1} = \beta_{EN,1}$, $\beta_{EE,0} = \beta_{NE,0}$, $\sigma^2_{EE,1} = \sigma^2_{EN,1}$, $\sigma^2_{EE,0} = \sigma^2_{NE,0}$, but $\alpha_{EE,1} > \alpha_{EE,0}$, $\alpha_{EE,0} > \alpha_{NE,0}$.

$$\mathcal{L}(\theta|T, S^{obs}, Y^{obs}, X) \propto \prod_{i \in O(1,1)} \Big[\pi_{EE,i} \cdot N\big(\mu_{EE,1}, \sigma^2_{EE,1}\big) + \pi_{EN,i} \cdot N\big(\mu_{EN,1}, \sigma^2_{EN,1}\big)\Big]$$
$$\times \prod_{i \in O(0,1)} \Big[\pi_{EE,i} \cdot N\big(\mu_{EE,0}, \sigma^2_{EE,0}\big) + \pi_{NE,i} \cdot N\big(\mu_{NE,0}, \sigma^2_{NE,0}\big)\Big]$$
$$\times \prod_{i \in O(1,0)} \Big[\pi_{NE,i} + \pi_{NN,i}\Big] \times \prod_{i \in O(0,0)} \Big[\pi_{EN,i} + \pi_{NN,i}\Big]$$

As mentioned before, a feature of the likelihood function results from the unobservable membership of individuals to principal strata π_g. The likelihood function can be maximized by the EM algorithm with respect to the parameter $\theta = \{\alpha, \beta, \gamma, \sigma\}$ as follows. First, the algorithm starts with a randomly selected initial parameter, θ_0. Second, in the expectation (E) step, the conditional probabilities of the principal strata are estimated by using θ_0. Third, in the maximization (M) step, using the estimated conditional probabilities obtained from the E-step, parameter θ can be estimated by maximum likelihood estimation. Afterwards, the estimated parameter $\hat{\theta}$ is plugged back into the E-step to update conditional probabilities and, in the M-step, re-estimate parameters with the updated conditional probabilities. These E- and M-steps are iterated until convergence. The threshold for convergence is set as 10^{-5}, and the maximum number of iterations as 1,000.

Given the estimated parameters $\hat{\theta}$, the estimated membership of principal strata g, denoted by $\hat{\pi}_g$, can be obtained by averaging the probability of individual i being in principal strata g ($\pi_{g:i}$) with respect to i:

$$\hat{\pi}_g = \frac{1}{N}\sum_{i=1}^{N}\hat{\pi}_{g:i}, \qquad \text{where } g \in \{EE, EN, NE, NN\}$$

Note that the summation of π_g with respect to g is always equal to one, due to the mutual exclusiveness of principal strata ($\sum_g \pi_{g:i} = 1$ for all i). The average wages are estimated by plugging in the estimated parameters $\hat{\theta}$ in the following equations that use the parametric model for the wages:

$$\hat{w}_{g,t} = \frac{\sum_i \hat{\pi}_{g:i} exp(\hat{\mu}_{g,t} + \frac{1}{2}\hat{\sigma}^2_{g,t})}{\sum_i \hat{\pi}_{g:i}}$$

where $\hat{\mu}_{g,t} = \hat{\alpha}_{g,t} + \hat{l}(X_i) \cdot \hat{\beta}_{g,t}$, and $(g,t) \in \{(EE,1),(EN,1),(EE,0),(NE,0)\}$.

2.6 Results

Figure 2.1 reports the covariate balances by displaying the standardized differences before (marked as $\blacklozenge$) and after the matching (marked as $\times$). We confirm that the covariate imbalances between the two treatment arms become almost negligible after the matching, as all the standardized differences after the matching are aligned around zero, while some of those before the matching show large magnitude. More specifically, the standardized differences only range up to 0.9 (e.g., Resident: Seoul, Industry: wholesale/retail) after matching, whereas before matching ranging up to 60.1 (e.g., work experience).

We also find the estimated linearized propensity score between the two treatment arms are well-balanced in the same support, as can be seen in Figure 2.2, which suggests that the overlap condition largely holds.

Figure 2.1: Standardized Differences of all the Covariates

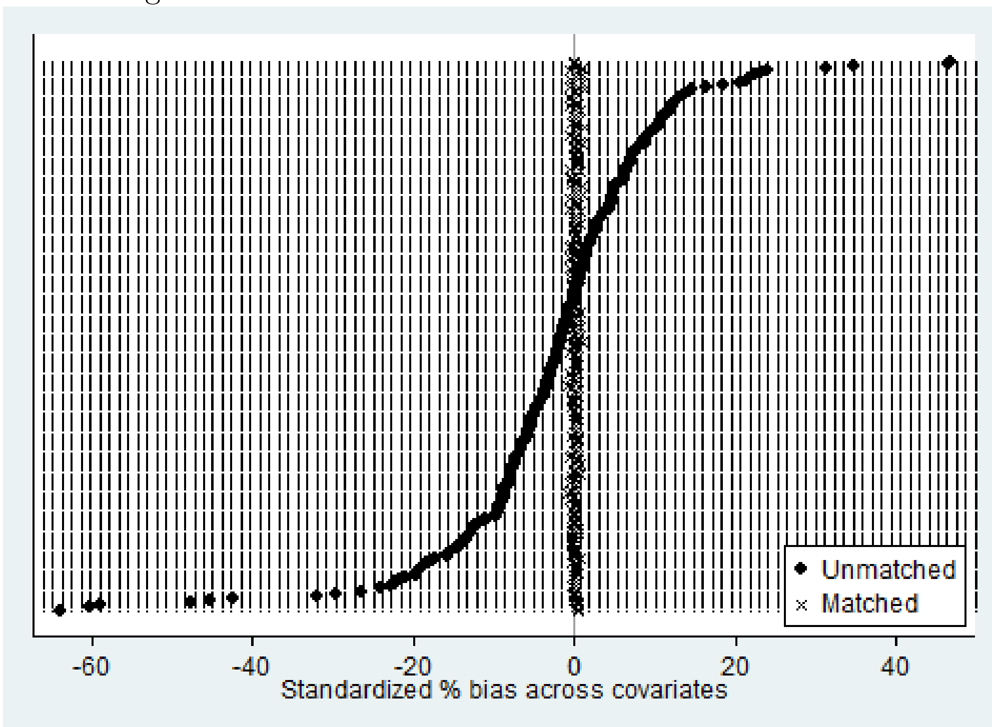
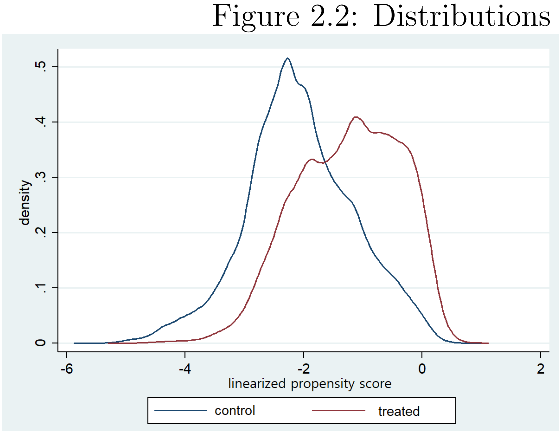
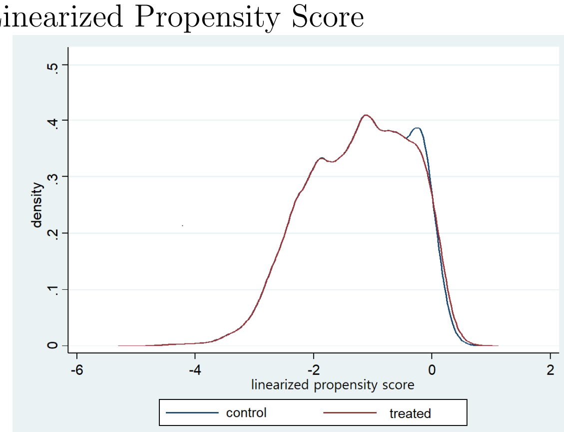


Figure 2.2: Distributions of Linearized Propensity Score



(a) Before Matching



(b) After Matching

As a preliminary analysis, we apply simple naive OLS regressions to the matched sample. The results of the OLS regression indicate that the JTPs have a negative effect on wage by KRW -33,060 (std. error 3,299) which is statistically significant. However, these results are likely biased because selection into employment is not taken into account in the regression. That is, we expect the magnitude of negative effect will be greater than the OLS estimation after controlling for selection bias, if there exists positive selection into employment.

Table 2.2 presents the ATT estimates on the re-employment probability and wages, along with the estimated proportion of each principal strata ($\hat{\pi}_g$) and the estimated average wage.[19] For the probability of re-employment ($\hat{\pi}_{EN} - \hat{\pi}_{NE}$), the ATT (left panel of Table 2.2) is estimated as 0.026 (std. err. 0.005), which is positive and statistically significant. In terms of positive effectiveness of JTPs, our results are in line with those from previous studies (e.g., Yoo and Lee, 2008; Choi and Kim, 2012) and the stated primary objective of JTPs in South Korea.[20]

The estimated proportions of principal strata are estimated as 0.066 in EE, 0.396 in EN, 0.370 in NE, and 0.169 in NN, respectively. The proportion of EE (0.066) is comprised of individuals always employed regardless of JTPs, which is the subsample we focus on to estimate the ATT on wages. We will discuss about the proportion of EE later in this section. The proportion of EN (0.396) contains individuals that would be employed under training but not under no participation in training. On the other hand, the proportion of NE (0.370) is comprised of individuals who would be unemployed if they participate in training but employed if they do not participate in it. For these individuals, JTPs have a negative impact on employability within 12 months of the start of job-search. One possibility is that NE remained unemployed because they increase their reservation wage

[19]We repeatedly ran our Matlab code 50 times with randomly selected initial parameters, and find similar results, which suggests that the one found is a global optimum.

[20]Worker's Capacity Building Act. Article 1. (2018) "The purpose of this Act is to contribute economic development by promoting and stabilizing the employment ..."

Table 2.2: Estimated Results

Proportions in Each Principal Strata		Estimated Average Wage ($\times$ 10^3 KRW)	
$\hat{ATT}_{emp}$	0.026 (0.005)	$\hat{ATT}_{w,EE}$	-205.3 (5.67)
$\hat{\pi}_{EE}$	0.066 (0.003)	$EE, W_i=1$	2,243.7 (3.83)
$\hat{\pi}_{EN}$	0.396 (0.003)	$EE, W_i=0$	2,449.1 (4.18)
$\hat{\pi}_{NE}$	0.370 (0.004)	$EN, W_i=1$	1,299.8 (2.50)
$\hat{\pi}_{NN}$	0.169 (0.006)	$NE, W_i=0$	1,315.8 (2.59)

after taking JTPs.

The right panel of Table 2.2 reports that the estimated ATT on wages for the *EE* stratum is KRW -205.3 $\times 10^3$ (std. error 5.67 $\times 10^3$), which is statistically significant.[21] Although this result may appear counterintuitive, it is in line with Chae and Kim (2004), which documented a negative effect on wages from the Korean JTPs before the 2011 reform. The estimated average wage of *EE* under treatment is KRW 2,243.7 $\times 10^3$ (std. error 3.83 $\times 10^3$), and that under control is 2,449.1 $\times 10^3$ (std. error 4.18 $\times 10^3$). Meanwhile, the estimated average wage of the treated in *EN* is KRW 1,299.8 $\times 10^3$ (std. error 2.50 $\times 10^3$), and that of the control in *NE* is KRW 1,315.8 $\times 10^3$ (std. error 2.59 $\times 10^3$). Interestingly, the wages of *EE* in both treatment arms are twice the size of the average wages for the other strata, *EN* and *NE*. Despite different parameters of interest between the analyses, it is worth noting that the negative impact of JTPs is more negative relative to the naive estimate from an OLS regression, -KRW 33 $\times 10^3$.

The proportion of *EE* (0.066) seems quite small[22] relative to the size of this stratum in other studies from different countries (0.36-0.51) such as Zhang et al. (2009), Frumento et al. (2012) and Bia et al. (2018), while the wages of this group are twice as large as those

[21]This result is obtained under the assumption that the treatment does not change the hours worked. Alternatively, if the hours worked differ by treatment status but this difference is accounted for by the pre-treatment covariates, then the lack of acces to hours worked do not bias our estimates.

[22]Despite the small proportion of *EE* (0.066), we have enough sample size in the group for our estimation (n_{EE}=7,541).

of the other strata. The large size difference may result from the use of different dataset among studies. To gain more insights about our EE stratum, for whom we estimate the ATT on wages, we analyze the average pre-training characteristics of the group.

Table 2.3 reports the average characteristics for the principal strata EE and NN, which are obtained by taking average of covariates across the corresponding principal strata, $E(X_i|G_i = g) = \frac{1}{N}\sum_{i=1}^{N} \pi_{g:i} X_i$.[23] We can observe in the table that there is a large number of statistically different characteristics between the EE and NN strata. More specifically, the EE group is more likely to be male (0.49 vs. 0.35), older (40.98 year-old vs. 36.42 year-old), to have shorter inactive duration (0.38 year vs. 0.54 year), and to have work experience at a given industry (0.67 vs. 0.38). They are more likely to be separated due to mainly firm-oriented reasons such as shutdown or exploitation of a firm, however less likely to be separated due to personal miscellaneous reasons. We now can illustrate a stereotype of EE: they are individuals that, on average, earn twice more than others, were born in the early 1970's and thus are older than others, have considerable work experience in a given industry, were separated from their prior job due to firm-oriented/expiration reasons, and appear motivated to find another job (they have shorter inactive duration).

Based upon the analysis of average pre-treatment characteristics, we conjecture that higher wages of EE are associated with higher seniority, higher work experience in a given industry, the reasons for separation from the prior job, and their motivation to actively seek re-employment.

Turning back to the negative effects of JTPs on the wages of EE individuals, one possibility is that, for these more-experienced individuals, a marginal increase in human capital from JTPs does not significantly contribute to raise their wages. Moreover, it seems that the time invested in the JTPs for these individuals results in a statistically significant

[23] For the following comparison of the average characteristics between EE and others, we choose NN that has the smallest sample size among them. This comparison provides more conservative t-statistics than the other strata. The average characteristics among the strata NN, NE, and EN observe very similar average characteristics.

Table 2.3: Average Characteristics across Principal Strata

Covariates	EE		NN		
	mean	std.err.	mean	std.err.	t-stat.
Gender (male=1)	0.49	(0.50)	0.35	(0.48)	19.53***
Tenure (years)	1.96	(3.42)	1.70	(3.11)	0.35
Inactive Duration (years)	0.38	(0.55)	0.54	(0.63)	-34.40***
Work Experience at a given industry	0.67	(0.47)	0.38	(0.48)	5.60***
Age (years)	40.98	(12.34)	36.42	(11.05)	19.53***
$\leqslant$ 29	0.21	(0.41)	0.35	(0.48)	-1.28
30s	0.31	(0.46)	0.31	(0.46)	-0.68
50s	0.16	(0.37)	0.11	(0.31)	0.85
$\geqslant$ 60	0.10	(0.30)	0.04	(0.19)	1.39
Educational Attainment					
Elementary school or less	0.04	(0.20)	0.01	(0.11)	0.80
Middle school	0.06	(0.25)	0.03	(0.17)	0.78
High school	0.35	(0.48)	0.34	(0.47)	-0.37
College	0.25	(0.43)	0.30	(0.46)	-0.39
Bachelor's	0.28	(0.45)	0.31	(0.46)	-0.34
Residential Region					
Chung-cheong	0.10	(0.29)	0.08	(0.27)	0.37
Jeolla	0.08	(0.28)	0.10	(0.30)	-0.48
Workplace Location					
Seoul	0.34	(0.47)	0.39	(0.49)	-0.75
Gyeong-in	0.26	(0.44)	0.23	(0.42)	0.46
Chung-cheong	0.09	(0.28)	0.07	(0.26)	0.35
Industry					
Manufacturing	0.23	(0.42)	0.21	(0.41)	0.48
Wholesale/retail	0.16	(0.37)	0.17	(0.38)	-0.36
Finance/insurance	0.02	(0.14)	0.02	(0.15)	-0.30
Ex-firm size (persons)					
$\leqslant$10	0.33	(0.47)	0.32	(0.47)	-0.26
10-29	0.20	(0.40)	0.18	(0.38)	0.64
30-99	0.17	(0.37)	0.16	(0.36)	0.44
$\geqslant$300	0.18	(0.39)	0.24	(0.42)	-0.92
Ex-occupation					
Supervisor/Clerk	0.18	(0.39)	0.22	(0.42)	-0.62
Merchandiser	0.12	(0.32)	0.16	(0.36)	-0.89
Operator	0.16	(0.37)	0.14	(0.35)	0.45
Manual Laborer	0.14	(0.34)	0.09	(0.29)	1.08
Reason for Separation					
Family issues	0.01	(0.12)	0.03	(0.16)	-0.72
Disease/Injury	0.02	(0.12)	0.02	(0.15)	-0.39
Miscellaneous (personal)	0.27	(0.44)	0.46	(0.50)	-3.13***
Shutdown	0.04	(0.19)	0.02	(0.13)	0.46
Exploitation	0.38	(0.49)	0.24	(0.43)	2.36**
Retirement	0.01	(0.11)	0.01	(0.09)	0.10
Contract Expiration	0.20	(0.40)	0.14	(0.34)	1.47

, and * indicate that mean difference is statistically significant at the 95, and 99% level, respectively.

wage penalty of about 8 percent.

We conjecture that the type of individuals belonging to our estimated EE stratum have mistakenly enrolled in JTPs, given that these are individuals that would be employed regardless of JTP enrollment (i.e., experience no effect on employment) and experience a monthly wage penalty 12 months after starting their job-search. The following aspects may, at least partially, help explain this phenomenon. First, the reformed JTPs reduced the opportunity costs for all individuals to take training. Thus, for this small EE stratum, it appears that it resulted on miscalculation about the marginal benefit of such JTPs. Second, the more-experienced individuals in EE may have participated in JTPs even though the programs may not be relevant to their career. This is consistent with an a priori lack of concern about employability. Indeed, Yoon (2014) reports that 26.3% of the JTPs enrollment is related to classes that seem more like a hobby, such as cook, barista, pâtissier, beautician, and esthetician. In this case, individuals in the EE stratum may actually derive consumption value from this type of JTPs.

Third, the supply of appropriate (advanced) training programs for more-experienced individuals may be insufficient due to the following reasons: (1) the more experienced individuals are not the main target of the government intervention that focuses primarily on higher employability, (2) training agencies are less likely to provide advanced programs to target the more experienced individuals that represent a small proportion of trainees, since those programs may not be profitable, and (3) there may be scarcity of capable certified instructors (Yoon et al., 2017) to offer those advanced programs. Fourth, the time spent by EE individuals taking JTPs may reduce their opportunities to accumulate firm-specific human capital through employment, which may be the channel through which the untrained EE obtain higher wages. In our context, as in Table 2.3, those in EE are likely to have more work-experience and longer tenure, which can be a reason for higher wage by accumulated human capital by on-the-job trainings. Finally, it is also possible that participation in unrelated JTPs (like those described in Yoon, 2014) would be a negative signalling effect to firms, whom may expect JTP trainees to undergo generic programs. There is a possibility that training agencies may play a role on the effectiveness of the JTPs through systematic practices. However, we are not able to explore this possibility due to lack of information about the training agency attended by trainees.

2.7 Sensitivity Analysis

As discussed earlier, our identification under the potential outcome framework relies heavily on the potentially strong assumption of unconfoundedness. Since this assumption is not testable, we conduct sensitivity analysis in the spirit of Rosenbaum and Rubin (1983b) to gauge the robustness of our results to existence of unobserved confounders which violate the key assumption. Through this assumption, the potential outcomes are independent of treatment assignment conditional on covariates. The conditional independence of potential outcomes implies that there are no unobserved confounders that result in selection into treatment. It also implies that the conditional probabilities of being in a given principal strata across the two treatment arms are the same: $Pr(G_i|W_i = 1, l(X_i)) = Pr(G_i|W_i = 0, l(X_i))$. Once we introduce unobserved counfounders to gauge sensitivity, those conditional probabilities across the two treatment arms are no longer the same, $Pr(G_i|W_i = 1, l(X_i)) \neq Pr(G_i|W_i = 0, l(X_i))$, which is a consequence of a violation of unconfoundedness. The key insight of our sensitivity analysis is to consider plausible values of the unobserved confounders we introduced, and if the results under these violations of unconfoundedness are similar to the baseline results (under its validity), then we can conclude that the baseline results are robust to such violations.

We consider sensitivity parameters ξ_g (where $g = EE, EN, NE$) that represent unobserved factors that affect both treatment and employment status simultaneously for each principal strata, and re-estimate the effects of JTPs to gauge impacts of the unobserved confounders on our results (following the same method outlined in section 2.5). It should be note that unobserved confounders for wages need not to be considered separately in our analysis because the consequences of unobserved confounders on wages cannot be disentangled from the treatment effects without exclusion restriction assumption (Schwartz et al., 2012; Bia et al., 2018). Thus, our sensitivity parameters address both effects simultaneously. In the following discussion, for simplicity, we approximate the conditional strata probabilities on $l(x)$ (the linearized propensity score) with the unconditional strata probabilities: $P(G_i = g|W_i = w, l(X_i) = l(x)) = P(G_i = g|W_i = w)$.

We now describe the plausible values chosen for the sensitivity parameters, ξ_g, where $g \in \{EE, EN, NE\}$. First, we discuss about the signs of sensitivity parameters, the ξ_{EE} is expected to decrease the probability of being in EE under treatment, relative to that of being in EE under control, $P(G_i = EE|W_i = 1) < P(G_i = EE|W_i = 0)$. We interpret the negative ξ_{EE} as an unobserved parameter that discourage individuals to take JTPs or encourage them to be employed without JTPs. As we discussed earlier, a strong preference to take (firm-specific) on-the-job training at work can be an example in the sense that it may increase their human capital for raising wages more than JTPs. Also, less professional instructors and negative signalling effect of JTPs are more likely to discourage them not to take JTPs, which implies negative sign of the sensitivity parameters. The ξ_{EN} can be interpreted as an unobservable parameter that increases the likelihood of taking JTPs, that is, $P(G_i = EN|W_i = 1) > P(G_i = EN|W_i = 0)$. JTPs are likely to have positive effects for EN in the sense that they are the main beneficiary of the current JTPs. Therefore, for example, those in EN who are highly motivated to participate in JTPs are more likely employed. ξ_{NE} is also assumed to increase the probability of being in NE for the treatment group relative to the control, $P(G_i = NE|W_i = 1) > P(G_i = NE|W_i = 0)$. This may be associated with their confidence that their human capital can be increased through JTPs, or willingness to increase their reservation wages, which is preferable for NE.

Next, to set plausible magnitude of the sensitivity parameters, ξ_g, our departure encompasses the difference of conditional probability by treatment arms, Δ_g, which is defined as $\Delta_g = P(G_i = g|W_i = 1) - P(G_i = g|W_i = 0)$. The values of Δ_g that we consider are $\Delta_{EE} \in \{0, -0.025, -0.05\}$, $\Delta_{EN} \in \{0, 0.15, 0.3\}$, $\Delta_{NE} \in \{0, 0.15, 0.3\}$. We obtain conditional probabilities of principal strata $P(G_i = g|W_i = w)$ by adding or subtracting $\Delta_g/2$ to the corresponding membership of principal strata $P(G_i = g) = \pi_g$ from our main results. We can gauge relative importance of the departures from the unconfoundedness by refering the following percentage of deviation relative to the corresponding proportion of principal strata (π_g): $\Delta_{EE} \in \{0, -38\%, -76\%\}$, $\Delta_{EN} \in \{0, 38\%, 76\%\}$, $\Delta_{NE} \in \{0, 41\%, 81\%\}$, respectively.

A list of values of Δ_g delineates 27 different scenarios: combinations of three different Δ_g. We rewrite our parametric form by adding sensitivity parameters ξ_g. Our baseline scenario is consistent with the validity of the unconfoundedness assumption: $\{\Delta_{EE}, \Delta_{EN}, \Delta_{NE}\} = \{0, 0, 0\}$, which means that there are no unobserved confounders; that is, the sensitivity parameters ξ_g equal to zero. A large absolute value of ξ_g implies that influential confounders exist which give rise to imbalance of the conditional probabilities of principal strata between the two treatment arms. As in the earlier section, we implement the maximum likelihood estimation, using the following modified conditional probabilities of principal strata with sensitivity parameters ξ_g.

$$P(G_i = g|W_i = w, l(X_i) = l(x)) = \frac{exp\big[l(x) \cdot \gamma_g + w \cdot \xi_g\big]}{\sum_{g'} exp\big[l(x) \cdot \gamma_{g'} + w \cdot \xi_{g'}\big]}$$

where $g, g' = \{EE, EN, NE, NN\}$, NN is a base stratum.

Table 2.4 through 2.6 reports the results of our sensitivity analysis. Each table presents the estimated parameters such as the estimated membership of principal strata, average wages, and causal effects on wages and the probability of re-employment along with their standard errors by the combination of sensitivity parameters of Δ_{EN} and Δ_{NE} which correspond to the 9 different scenarios given Δ_{EE}. Despite the sensitivity parameters, our sensitivity analysis are fairly robust to deviations from our baseline. More specifically, the average treatment effects on the probability of re-employment in all the scenarios are around 0.026, because π_{EN} and π_{NE} are generally similar to the baseline. Even when they show different results, they tend to deviate about the same amount and in the same direction, at the cost of π_{NN}. The π_{EE} is quite stable at 0.066. The other probabilities, such as π_{EN}, π_{NE}, and π_{NN}, deviate less than 0.02, 0.02, and 0.04, respectively. The estimated standard error are also stable across all the scenarios.

Table 2.7 reports the maximum differences along with their standard errors; where even the largest difference appears to be reasonably small. Specifically, the ATT estimates on the probability of re-employment range between 0.025 to 0.028, which means [-3.85%, 7.69%] of differences from our main result. The differences are not statistically significant at the 95% confidence level. The estimated effects on wages range between [-210.9$\times 10^3$, -188.7$\times 10^3$], which means [-2.73%, 8.10%] of differences at most from the main result KRW -205.3 $\times 10^3$. The minimum wage difference (-210.9$\times 10^3$) is not statistically significant at the 95% confidence level, but the maximum difference (-188.7$\times 10^3$) is statistically significant at the 95% confidence level, but not significant at the 99% level. Therefore, based on the sensitivity analysis, we argue that our results under unconfoundedness in Table 2.2 are fairly robust to departures from the key assumption, and bring more confidence to those results.

Table 2.4: Sensitivity Analysis Estimates for $\Delta_{EE} = 0$

	$\hat{ATT}_{emp}$	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{ATT}_{w,EE}$	$\hat{w}_{EE1}$	$\hat{w}_{EE0}$	$\hat{w}_{EN1}$	$\hat{w}_{NE0}$
$\Delta_{EN} = 0$	0.026	0.066	0.404	0.378	0.152	-196.6	2,249.2	2,445.9	1,299.9	1,315.8
$\Delta_{NE} = 0$	(0.005)	(0.003)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0$	0.026	0.066	0.393	0.367	0.174	-209.4	2,247.4	2,456.8	1,299.8	1,315.7
$\Delta_{NE} = 0.15$	(0.005)	(0.003)	(0.003)	(0.004)	(0.005)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0$	0.026	0.066	0.401	0.375	0.158	-204.1	2,248.5	2,452.6	1,299.9	1,315.7
$\Delta_{NE} = 0.3$	(0.004)	(0.003)	(0.003)	(0.004)	(0.005)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.15$	0.028	0.066	0.416	0.389	0.129	-188.7	2,258.0	2,446.7	1,300.1	1,315.7
$\Delta_{NE} = 0$	(0.005)	(0.003)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.15$	0.027	0.066	0.421	0.394	0.119	-202.3	2,247.9	2,450.2	1,299.9	1,315.7
$\Delta_{NE} = 0.15$	(0.005)	(0.003)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.15$	0.027	0.066	0.410	0.383	0.141	-195.0	2,252.1	2,447.1	1,299.9	1,315.8
$\Delta_{NE} = 0.3$	(0.005)	(0.003)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.3$	0.027	0.066	0.408	0.381	0.145	-199.8	2,252.2	2,452.0	1,299.9	1,315.7
$\Delta_{NE} = 0$	(0.005)	(0.003)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.3$	0.026	0.066	0.406	0.380	0.147	-200.8	2,249.4	2,450.1	1,299.8	1,315.8
$\Delta_{NE} = 0.15$	(0.005)	(0.003)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.3$	0.026	0.066	0.395	0.369	0.170	-210.9	2,246.4	2,457.3	1,299.8	1,315.7
$\Delta_{NE} = 0.3$	(0.004)	(0.003)	(0.002)	(0.003)	(0.005)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)

$\Delta_{EE} = P(G_i = EE|T_i = 1) - P(G_i = EE|T_i = 0)$

$\Delta_{EN} = P(G_i = EN|T_i = 1) - P(G_i = EN|T_i = 0)$

$\Delta_{NE} = P(G_i = NE|T_i = 1) - P(G_i = NE|T_i = 0)$

Table 2.5: Sensitivity Analysis Estimates for $\Delta_{EE} = -0.025$

	$\hat{ATT}_{emp}$	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{ATT}_{w,EE}$	$\hat{w}_{EE1}$	$\hat{w}_{EE0}$	$\hat{w}_{EN1}$	$\hat{w}_{NE0}$
$\Delta_{EN} = 0$	0.025	0.066	0.397	0.372	0.165	-208.2	2,250.1	2,458.3	1,299.7	1,315.8
$\Delta_{NE} = 0$	(0.005)	(0.004)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0$	0.027	0.066	0.409	0.383	0.142	-198.1	2,253.8	2,451.9	1,299.9	1,315.7
$\Delta_{NE} = 0.15$	(0.005)	(0.004)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0$	0.027	0.066	0.421	0.393	0.120	-207.9	2,243.8	2,451.7	1,299.9	1,315.7
$\Delta_{NE} = 0.3$	(0.005)	(0.004)	(0.003)	(0.004)	(0.007)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.15$	0.026	0.066	0.405	0.379	0.150	-199.6	2,256.0	2,455.6	1,299.8	1,315.8
$\Delta_{NE} = 0$	(0.005)	(0.004)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.15$	0.026	0.066	0.396	0.370	0.168	-199.5	2,251.9	2,451.4	1,299.9	1,315.7
$\Delta_{NE} = 0.15$	(0.005)	(0.004)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.15$	0.027	0.066	0.407	0.380	0.147	-206.7	2,243.0	2,449.6	1,299.9	1,315.7
$\Delta_{NE} = 0.3$	(0.005)	(0.004)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.3$	0.027	0.066	0.410	0.382	0.142	-195.9	2,254.0	2,449.9	1,300.1	1,315.7
$\Delta_{NE} = 0$	(0.005)	(0.004)	(0.003)	(0.004)	(0.007)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.3$	0.027	0.066	0.412	0.385	0.138	-207.4	2,249.8	2,457.2	1,299.9	1,315.7
$\Delta_{NE} = 0.15$	(0.005)	(0.004)	(0.003)	(0.004)	(0.007)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.3$	0.026	0.066	0.405	0.379	0.150	-209.6	2,246.2	2,455.9	1,299.7	1,315.8
$\Delta_{NE} = 0.3$	(0.004)	(0.004)	(0.003)	(0.003)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)

$\Delta_{EE} = P(G_i = EE|T_i = 1) - P(G_i = EE|T_i = 0)$

$\Delta_{EN} = P(G_i = EN|T_i = 1) - P(G_i = EN|T_i = 0)$

$\Delta_{NE} = P(G_i = NE|T_i = 1) - P(G_i = NE|T_i = 0)$

Table 2.6: Sensitivity Analysis Estimates for $\Delta_{EE} = -0.05$

	$\hat{ATT}_{emp}$	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{ATT}_{w,EE}$	$\hat{w}_{EE1}$	$\hat{w}_{EE0}$	$\hat{w}_{EN1}$	$\hat{w}_{NE0}$
$\Delta_{EN} = 0$	0.027	0.066	0.418	0.391	0.126	-196.7	2,253.5	2,450.2	1,299.9	1,315.8
$\Delta_{NE} = 0$	(0.005)	(0.005)	(0.003)	(0.004)	(0.007)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0$	0.027	0.066	0.406	0.379	0.149	-202.2	2,251.3	2,453.5	1,299.9	1,315.7
$\Delta_{NE} = 0.15$	(0.005)	(0.005)	(0.003)	(0.004)	(0.007)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0$	0.026	0.066	0.391	0.366	0.177	-204.3	2,251.8	2,456.1	1,299.9	1,315.7
$\Delta_{NE} = 0.3$	(0.004)	(0.004)	(0.003)	(0.003)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.15$	0.026	0.066	0.403	0.377	0.153	-200.2	2,250.6	2,450.7	1,299.7	1,315.8
$\Delta_{NE} = 0$	(0.005)	(0.005)	(0.003)	(0.004)	(0.007)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.15$	0.026	0.066	0.390	0.365	0.179	-198.9	2,254.3	2,453.3	1,299.9	1,315.7
$\Delta_{NE} = 0.15$	(0.004)	(0.005)	(0.003)	(0.003)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.15$	0.026	0.066	0.398	0.372	0.163	-206.2	2,250.2	2,456.3	1,299.8	1,315.7
$\Delta_{NE} = 0.3$	(0.004)	(0.004)	(0.003)	(0.003)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.3$	0.025	0.066	0.394	0.369	0.171	-206.8	2,252.5	2,459.3	1,299.7	1,315.8
$\Delta_{NE} = 0$	(0.004)	(0.005)	(0.002)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.3$	0.027	0.066	0.403	0.376	0.155	-193.7	2,255.7	2,449.4	1,300.0	1,315.7
$\Delta_{NE} = 0.15$	(0.004)	(0.004)	(0.002)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)
$\Delta_{EN} = 0.3$	0.027	0.066	0.421	0.394	0.119	-193.8	2,257.5	2,451.3	1,300.0	1,315.7
$\Delta_{NE} = 0.3$	(0.004)	(0.004)	(0.003)	(0.004)	(0.006)	(5.7)	(3.8)	(4.2)	(2.5)	(2.6)

$\Delta_{EE} = P(G_i = EE|T_i = 1) - P(G_i = EE|T_i = 0)$

$\Delta_{EN} = P(G_i = EN|T_i = 1) - P(G_i = EN|T_i = 0)$

$\Delta_{NE} = P(G_i = NE|T_i = 1) - P(G_i = NE|T_i = 0)$

Table 2.7: Range of Estimated Results from Sensitivity Analysis

	$\hat{ATT}_{emp}$	$(\Delta_{EE}, \Delta_{EN}, \Delta_{NE})$	$\hat{ATT}_{w,EE}$	$(\Delta_{EE}, \Delta_{EN}, \Delta_{NE})$
min.	0.025 (0.005)	(-0.025, 0, 0)	-210.9 (5.7)	(0, 0.3, 0.3)
max.	0.028 (0.005)	(0, 0.15, 0)	-188.7 (5.7)	(0, 0.15, 0)

2.8 Conclusion

This study evaluated the effectiveness of Korean job training programs (JTPs) in regards to two factors—probability of re-employment and wages, measured within 12 months from their start of job-search. Having the advantage of a large sample and rich set of covariates, we apply propensity score matching to tackle selection into JTPs. In addition, we employ the principal stratification framework to deal with selection into employment and identify the average treatment effect on wages for those employed irrespective of taking JTPs (EE). We employ the estimated linearized propensity score as a single covariate in our estimation, which allow us to have computationally tractable estimation results. We conduct the maximum likelihood estimation using the EM algorithm under the latent mixture model. The bimodal mixture groups are separately defined under the stochastic dominance assumption. Furthermore, we implement a sensitivity analysis to assess the plausibility of our results under the existence of fairly large unobserved confounders that allow deviations from our results.

Our findings indicate that participation in JTPs is likely to increase the probability of re-employment by 0.026 which is economically and statistically significant. By contrast, we do not find evidence that JTPs increase wages for EE, rather, we do find that JTPs have a negative effect on monthly wages by 8.4% (KRW -205.3$\times$ 10^3) when re-employed. This amounts approximately USD 18, which is also economically and statistically significant. The estimated effects are found to be robust to considerable departures from our key identifying assumption which are obtained through sensitivity analysis.

From a policy perspective, our results show that Korean JTPs are likely to aid displaced workers by improving their likelihood of re-employment, which corresponds to the primary objective of the Korean government as stated in the law. Meanwhile, JTPs are less likely to be helpful for increasing wages for the particular group of individuals that would be employed regardless of enrollment in JTPs, which implies that JTPs do not seem to provide enough relevant human capital to raise their wages. Given that those individuals for whom the wage effect is estimated possess more experience, we conjecture that the current JTPs are programmed for more vulnerable groups with lower wages. It can be argued that JTPs mainly focus on providing human capital for employability. In this context, more-experienced and senior individuals should not be preferred to take the current JTPs given both perspectives of trainees—to avoid negative effects—and of the government—not to waste government resources. Human capital accumulation to increase wages may need more (advanced) programs with better quality and/or more appropriate assignment of programs. The former can be addressed by using several policy tools, such as cultivating more qualified instructors and caseworkers, limiting licenses for the agencies with low quality programs, and encouraging to open advanced programs. The latter, better assignment, can be done by adopting assignment rules using some distinctive characteristics of subgroups instead of the current voluntary system. However, as the assignment rules likely have a more sophisticated justification, that remains a subject for future research.

Chapter 3

Better Assignment Rules using Heterogeneous Treatment Effects

3.1 Introduction

The importance of job training programs (hearafter JTPs) in South Korea has been emphasized since the 1997 economic crisis (Choe et al., 2015). The JTPs have served as a major active labor market policy dealing with the unemployment, and the participants and the total budget of the programs has considerably expanded over the past decades.

In South Korea, extensive literature have studied on JTPs to gauge the effectiveness (e.g., Lee and Lee, 2005; Yoo and Lee, 2008; Choi and Kim, 2012; Choe et al., 2015). The main interests in the previous studies were conventional estimators such as the average treatment effects (ATE) and the average treatment effects on treated (ATT). Some previous studies extend their application to analyze effect heterogeneity using subgroup analysis for college graduates, or the unemployment insurance (UI) applicants (e.g. Yoo and Lee, 2008; Choi and Kim, 2012), However, the subgroup studies do not widely examine the heterogeneous treatment effects across subgroups.

The traditional approaches to estimate effect heterogeneity are reasonable if researchers have perceived the heterogeneity prior to the analysis and they design the research carefully (Casey et al., 2012). Heterogeneity investigation for all possible subgroups to examine heterogeneous treatment effects may encounter the multiple hypothesis testing problem which may lead to imprecise estimates and invalid hypothesis tests (Knaus et al., 2018).

Recently, in the policy evaluation literature, machine learning algorithms provide systematic methods to investigate effect heterogeneity (see Athey and Imbens, 2017). These methodologies have gained more popularity, and they provide intuition for causal inference by showing great performance in "predictions" (Varian, 2014; Athey, 2015; Mullainathan and Spiess, 2017). Recent research suggests promising methods to estimate effect heterogeneity, and to achieve better statistical properties using: Lasso (e.g., Imai and

Ratkovic, 2013; Knaus, Lechner, and Strittmatter, 2017), decision tree (e.g., Su et al., 2009; Athey and Imbens 2016), random forests (e.g., Wager and Athey, 2018; Lechner, 2018; Oprescu, Syrgkanis and Wu, 2018; Athey, Tibshirani and Wager, 2019), Bayesian regression tree (Hahn et al, 2017), and ensemble methods (e.g., Nie and Wager, 2017). Profound empirical studies have also appeared in the literature (e.g., McBride and Nichols, 2015; Chalfin et al., 2016; Bertrand et al., 2017; Davis and Heller, 2017a, 2017b; Knaus, Lechner and Strittmatter, 2017; Andini et al., 2018; Kleinberg et al., 2018; Strittmatter, 2018).

Knaus et al. (2018) recently compare proposed machine learning estimators by using Monte Carlo simulation. They conclude that there exists no single best estimator, but some estimators, including the causal forest estimator in Athey et al. (2019), perform well.[1] In this paper, we use the causal forest method suggested in Athey et al. (2019) to examine effect heterogeneity of Korean JTPs using rich administrative data. The causal forest uses a non-parametric estimator that adopts the basic structure of the conventional random forests (for an overview, see Hastie, Tibshirani and Friedman, 2009). They further suggest using weighted moment equations which is more precise and computationally efficient. Intuitively, their forest-based weights are similar to a kernel weighting function, but avoid the curse of dimensionality by using the forest algorithm.

This paper contributes to the literature in three ways. Our study is one of the earliest empirical studies using the most recent causal forest estimator. We extend its application into a selection-on-observable setting, while other empirical studies are conducted based on a randomly assigned treatment. This is the first study using machine learning techniques to analyze Korean JTPs and to investigate effect heterogeneity, to the best of our knowledge. This study finds large variations of predicted impacts across trainees and across various subgroups. We extend our understanding of the different characteristics of trainees' between benefiting and harmed groups. Based on our findings, we present

[1] The better estimators are 1) Causal Forest with local centering (Athey et al., 2019), 2) Double machine learning estimator with random forest (Chernozhukov et al., 2017), 3) Modified Covariate Method with efficiency augmentation (Tian et al., 2014), and 4) R-learning (Nie and Wager, 2017)

possible candidates of treatment assignment rules using important observed characteristics, and assess how much impacts vary across the different hypothetical assignment rules. We find that some easy-to-implement assignment rules outperform the current one.

The remainder of this paper is organized as follows. The next section provides some background and literature on Korean JTPs. In Section 3.3, we review our methodologies to identify the effect heterogeneity. In Section 3.4, we describe the data used and define the sample. Section 3.5 presents the estimation results and interpretations. The final section provides a conclusion and policy implications.

3.2 Background

3.2.1 Institutional Background

In South Korea, large-scale job training programs were introduced since the 1997 economic crisis that increased the unemployment rate from 2.6% in 1997 to 7.0% in 1998. Since the enactment of the Job Training Stimulation Act in 1997, JTPs are generally recognized as a major part of active labor market policies over the past two decades. The 2017 participants in JTPs (220 thousand) have become more than twice as many as those in 2008 (94 thousand), and the 2017 budget (KRW 546B) has increased more than ten times than that in 2008 (KRW 41.4B).

The main purposes of JTPs are (1) to re-train displaced workers,[2] (2) to upgrade the skills of incumbent workers, and (3) to improve employability of new entrants to labor market (Ministry of Employment and Labor, 2018). In this paper, we focus on the JTPs for displaced workers which comprises the largest proportion of Korean JTPs (Ministry of

[2]In this study, we define 'displaced workers' as persons 17 years of age and over who lost and left jobs involuntarily because of their personal issues (changing jobs, family issues, disease or injury, disciplinary dismissal, or miscellaneous), firm-oriented issues (layoff, shutdown, exploitation, or miscellaneous), or expiration (retirement, contract expiration, project completion).

Employment and Labor, 2018b).[3]

It should be noted that the Korean government has introduced a few substantial reforms that became effective nationwide as of 2011.[4] First, the government introduces a voucher program instead of the direct subsidizing of training agencies. An annual voucher up to KRW 2 million is supplied for a trainee, and he/she can take any JTPs from training agencies using the voucher. Once a trainee takes a JTP at a training agency, the agency can be reimbursed for training expense by the government. Thus, the voucher program draws the agencies' attention to trainees, and expect to improve JTPs' quality for appealing to trainees. Note that, trainees are charged 20-50% copayment of their training expenses, to prevent moral hazard.[5] Second, the government issues a license for legitimate training agencies instead of a permit with a limited trainee quota. The new system allows any agencies' free entry into the market for JTPs unless they cannot meet the minimum requirement such as facilities, instructors and financial stability. More agencies are allowed to operate JTPs, and are expected to open various JTPs with their discretion. This change improve trainee's accessibility to more programs by stimulating the agencies to provide more competitive JTPs that fit to trainees' needs. Consequently, this change is expected to cultivate agencies' autonomy and to provide better JTPs. Third, displaced workers are able to receive both UI benefits and training subsidies simultaneously. Previously, since trainees were not allowed to receive both benefits simultaneously, UI benefits are usually preferred due to the large amount and less conditions. However, after the reform, the JTPs become more relevant to their purpose which is to develop trainees' human capital, as training subsidies were no longer treated as a second chance after the UI benefit. As those two benefits do not compete with each other, the effectiveness of JTPs becomes more tractable.

[3] The total government's budget for these workers constitutes more than half of the total amount of budget for the JTP (e.g., 52.4% in 2017).

[4] This system was adopted in some regions in 2008 as a pilot project and expanded nationwide from 2011.

[5] The Korean government implemented a special subsidy program called "employment success package" for targeted groups, such as low-income families, for which the cost of training was subsidized up to KRW 3 million and exempt from out-of-pocket payments (Ministry of Employment and Labor, 2014).

Consequently, once a worker is separated from employment due to some reasons[6] they need to register at Worknet to be eligible for the governmental support including training subsidies or UI benefits. During the registration process, they are asked to file their information at Worknet, which can be used as a background information or as a resume to apply for a job within Worknet. After the registration, a worker is required to meet with a caseworker at a local office, and the caseworker will issue a voucher to an eligible worker.

3.2.2 Literature

Most previous studies in South Korea analyze the effects of the former JTPs prior to the 2011 reform when the choice of displaced workers is limited to choose either JTPs or UI benefits. To measure the causal effect of JTPs, the self-selected benefit results in complication. For this reason, Lee and Lee (2005) and Yoo and Lee (2008) define their main interests as the relative effect of JTPs versus UI benefits instead of effect of JTPs taker versus JTPs non-taker; that is, they gauge preference of government programs between JTPs and UI benefits.

On the other hand, the main targets for estimating parameters are population (ATE), trainees (ATT), or some designated subgroups such as women (Lee and Lee, 2005), college graduates (e.g. Yoo and Lee, 2008) and the UI benefits applicants (Choi and Kim, 2012). More specifically, Lee and Lee (2005) focus on the effects on women, and find that JTPs are generally less effective than UI benefits because JTPs have lengthened the unemployment duration of Korean women. Their duration analysis uses the connected administrative data of the UI file and the job training file from 1999-2000. Yoo and Lee (2008) also compare the effectiveness of the government supports between the two policies, JTPs and UI benefits, using the similar administrative data as in Lee and Lee (2005) but

[6] We classify the reasons into three categories. Personal reasons: changing jobs, disease/injury, and disciplinary dismissal. Firm-oriented reasons: temporary closure, shutdown, and layoff. Employment expiration reasons: retirement, contract expiration, and project completion.

more recent one from 2002. They use simple logit regressions, and find that JTPs show better employability (around 13%) than UI benefits. They conduct another analysis which compares the effectiveness of JTPs for college graduates. To construct the comparison group, they employ another dataset from the Korean Chamber of Commerce and Industry (KCCI), which contains the information of JTPs held by the KCCI (i.e., private JTPs). Using the similar methodology, they find that the private JTPs shows better employability (13.2%) than the government-sponsored JTPs.

Choi and Kim (2012) attempt to disentangle the two effects of government policies by defining their sample using the UI benefits application status—the UI applicants and non-applicants. They conduct subgroup analysis, and find that the effects on employability in both subgroups show similar trends, which demonstrate a 'J-curve' shape, while the UI applicants generally have larger effects than the non-applicants. They show that JTPs have a negative effect in the short term due to their "lock-in" the training (van Ours, 2004); which is likely to happen because they include training period in their job-search duration. However, the estimated effects on trainees become zero at 12 months, and gradually increase approximately up to 8%p after 18 months of their job separation.

The mixed evidence in the above studies likely results from the use of different datasets, subgroups, variables, time period, the different definition of the control group, and different identification strategies. Also, some subgroup analyses in the previous literature do not fully explain the effect heterogeneity across the samples, but focus on a specific group based upon the prior information. That is, a more general investigation of the effect heterogeneity has not been implemented yet. Thus, our analysis focuses on examining effect heterogeneity in a more systematic way. The goal is to uncover extensive effect heterogeneity across subsamples using rich administrative data that is similar to the previous studies but the most recently available from 2012-2014, which is extracted after the 2011 reform.

3.3 Methodology

3.3.1 Potential Outcome Framework

We adopt the potential outcome framework for our discussion (Rubin, 1974). Let W_i denote the treatment assignment for individual i, as a binary indicator: 1 if treated, 0 otherwise. Let $Y_i(W_i)$ denote the potential employment status under treatment status W_i for individual i, which is also a binary indicator; 1 if re-employed, 0 otherwise. Let X_i denote covariates which illustrate characteristics of individual i. The treatment effect for each individual i can be described by subtracting two potential outcome, $\theta_i = Y_i(1) - Y_i(0)$. However, the fundamental problem arises because both potential outcomes are not observable simultaneously. Thus, causal effect in general is defined as the average treatment effect, $\tau = \mathbb{E}(\theta_i) = \mathbb{E}(Y_i(1) - Y_i(0))$, which can be identified under the following key assumptions:

- (Stable Unit Treatment Value Assumption, SUTVA) $Y_i = Y_i(1)W_i + Y_i(0)(1 - W_i)$, for all i
- (Unconfoundedness) $(Y(1), Y(0)) \perp W_i | X_i$
- (Overlap) $0 < P(W_i = 1|X_i) < 1$
- (Exogeneity) $X_i(W_i = 1) = X_i(W_i = 0)$, where X_i is a candidate for assignment rules

The SUTVA implies that all outcome variables of each individual are mutually independent (Rubin, 1978). This assumption is violated if there are general equilibrium effects due to the training program, as in that situation the potential outcomes of an individual depend on the potential outcomes of other individuals. It may also be violated if there are peer effects at work within the training program. In the first case, JTPs are not likely to create general equilibrium effects because total number of trainees is only 0.03% of the total job vacancy in 2013 (a period that corresponds with our data) which was 1.76

million. For the second case, we only point out that we are not aware of any peer or network effects among the trainees. Thus, even though those effects may exist, they may be unimportant and so we disregard them.

The unconfoundedness assumption implies that there are no unobserved confounders that jointly affect treatment and potential outcomes, conditional on observable factors (X). Under this assumption, we are able to estimate causal effects by comparing outcomes after conditioning on those observed factors. This assumption is untestable and has to be argued in substantive grounds. In our context, we rely on the rich administrative data at our disposal to argue that we have access to all variables that determine selection into the training program and that are simultaneously related to the potential outcomes. Indeed, we employ all available covariates in our linked administrative data.

The overlap assumption implies that the probabilities of being treated are bounded away from zero or one. Intuitively, it requires that we are able to find a suitable comparison unit for individual i on the common support X. This assumption can be assessed by comparing the distribution of the estimated propensity score between the two treatment arms (Rosenbaum and Rubin, 1983a).

Lastly, the exogeneity assumption indicates that treatment status cannot affect individuals' characteristics. That is, the characteristics of our samples will be the same regardless of the treatment assignment. If not, the effects of assignment can be distorted as trainees easily change their observed characteristics. Thus, those covariates that can be simply switched are not desirable for the assignment rules even if we consider the covariates in our estimation.

Our main parameter of interest is the average treatment effect on treated (ATT), which illustrates the 'actual' policy impact on a designated group of trainees. To address effect heterogeneity, we define the ATT in several ways according to its aggregation level: from the finest (individual level) to the coarsest (population level). These can be defined by conditioning on different combinations of the covariates. First, the finest parameter mentioned above, the individualized treatment effect on the treated (ITET), can be defined as the difference of outcomes between a treated unit and a corresponding control, measuring the mean difference in outcome between the two treatment arms for an individual i with characteristics x: $\theta_i = Y_i(1) - Y_i(0) = \mathbb{E}(Y_i(1) - Y_i(0)|W_i = 1, X = x)$.[7] This finest definition allows us to overview the distribution of treatment effects, and illustrate how their characteristics are associated with policy effectiveness.

The coarsest definition, the average treatment effect on the treated (ATT), is defined as the average of the differences in outcomes between the treatment and control groups, conditional on being treated: $\tau = \mathbb{E}(\theta_i) = \mathbb{E}(Y_i(1) - Y_i(0)|W_i = 1)$. This parameter can capture the average effects on the treated population, and used to gauge usefulness of JTPs in general.

Finally, we can define conditional average treatment effects on the treated (CATTs) which are coarser than the ITET since they condition on a few covariates: $\tau(x) = \mathbb{E}(Y_i(1) - Y_i(0)|X_i^1 = x, W_i = 1)$, for $X^1 \subset X$. This parameter narrows down the target to a designated group of trainees within the same support of covariates x. Estimation of these CATTs allows examining how treatment affects vary across individuals' characteristics x. These CATTs also give important information to policymakers about the effectiveness of the training program for different types of individuals.

[7] Strittmatter (2018) and Lechner (2019) have a slightly different definition of this concept by acknowledging explicitly that one can only define the treatment effect on individuals conditional on the size of the support of the observed covariates X^{obs} (where $X^{obs} \subset X$). They call it the Individualized Average Treatment Effect (IATE). To the extent that the individualized treatment effect (ITE) can be identified only if we observe all the covariates affecting heterogeneity, the variance of the expectation of the true ITE is likely smaller than that of IATE, conditional.

3.3.2 Causal Forest approach

Overview

The forest algorithm we use is practically based on the conventional forest estimator in Breiman (2001), thus, we first illustrate the properties of the conventional forest and move toward our methods.

To construct the forest estimator, we will take a look at a tree estimator first, which constitutes the main element of the forest. A (binary) tree estimator is constructed by recursively partitioning sample into two subsamples, called nodes, that share the same support of covariates x.[8] Each partition investigates all the possible splits of covariates to find the best split, which maximizes in-sample goodness-of-fit within subsamples. The goodness-of-fit is measured by several methods such as the sum of mean squared error of the observed outcomes within the subsamples. After finishing the partitioning process, we obtain the (predicted) outcomes within subsamples. Despite the advantage of interpretability, a tree estimator has major drawbacks: (1) high sensitivity and (2) high variance. More specifically, the former implies that a small change of setting, such as a different covariate selection or criteria for goodness-of-fit, may lead to very different results. This is because the estimator partitions the sample sharply along a value of x. The latter drawback implies that recursive partitioning tends to split nodes to the minimum level, leading to maximum goodness-of-fit. However, this overfitting tendency results in high variance. To overcome these drawbacks, the random forest estimator suggested by Breiman (2001) aggregates many trees to reduce its dependency on a tree. The random forest is an ensemble method that averages over many results from tree estimations, which are

[8] In the machine learning literature, the subsample is defined as a node which contains an individual and its neighborhood that shares similar covariates x. We use the definition of three different nodes: (1) children node, (2) parent node, and (3) terminal node. The parent node is a precedent node that has children nodes. The children nodes are descendant nodes which come from the parent node, generally two children nodes are split from a parent node in a binary decision tree. The terminal node is defined as nodes without any splits, that is, they do not have any children nodes. The terminal node should contain more than the minimum number of observations, which is pre-determined by researchers.

estimated from randomly drawn subsamples. This aggregation process to reduce variance is called 'bagging (bootstrap+aggregating)'.

Wager and Athey (2018) first introduce a causal forest estimator and show that it is consistent and asymptotically normal under some conditions.[9] Their forests estimator also follows the conventional process of forest-based estimator, which is based on tree estimators that have recursive partitioned nodes, and a forest estimator is the average of many tree estimators. The key contribution of the causal forest is extending parameters of interest from outcomes to causal effects by adopting an 'honest' process that splits a sample into two mutually exclusive subsamples: (1) for partitioning trees, and (2) for estimating causal effects, respectively (Athey and Imbens, 2016).[10] That is, using a first half, the partitioning process is implemented to define terminal nodes that shares common support of x, while, using the other half, they estimate causal effects within a terminal node defined by trees. The use of independent subsamples can decorrelate the two processes so as to avoid spurious error. Intuitively, one of the algorithm suggested in Wager and Athey (2018), the Procedure 2, is similar to the propensity weighting estimator, in which the weights are defined by tree splitting.

In this study, we use the most recent causal forest algorithm in Athey et al. (2019) which outperforms the previous causal forest estimators. This method is generalized random forest that identifies the treatment effects using local moment conditions. This forest estimator is similar to the aforementioned forest estimator (Procedure 1 of the causal forest in Wager and Athey (2018)), but has a few advantages. First, this estimator is more precise than the other forests estimators, as the potential bias can be alleviated by adopting the Neyman's orthogonality condition (Neyman, 1979). The orthogonality

[9]The assumptions are as follows: (1) samples are independently identically distributed, (2) Lipschitz continuity of the outcome variable conditional on the covariates, (3) the subsamples for growing trees should increase slower than the number of samples.

[10]The typical random forest estimator sets aside a validation subsample, which is used for testing the quality of the model. However, in the causal forest approach, since the treatment effect cannot be validated due to the counterfactual, a validation subsample is not needed.

condition exploits residualized regressions which partialing out the effect of confounders from outcomes and treatment in the spirit of Frisch-Waugh-Lovell theorem. This orthogonalization may reduce bias and gives more precise estimations in the presence of confounding factors. Second, the recursive partitioning process uses a gradient approximation instead of greedy search for the optimal partition at x, which is computationally more efficient. Third, it preserves the statistical properties which is consistent and asymptotically normal. We implement the causal forest using the R package *grf* (Athey et al., 2019).

Local GMM Estimation

Suppose that we have n samples independently drawn from the population. Each observation has a covariate vector (characteristics) $X_i \in \mathcal{X}$, treatment status $W_i \in \{0,1\}$ and outcome (employment status) $Y_i \in \{0,1\}$. Our goal is to estimate our parameter of interest $\tau(x)$ which is defined via a local moment condition. The moment equation in the machine learning literature is usually defined as a scoring function which is amenable to using gradient-based optimization (e.g., Chernozhukov et al., 2018). $\tau_0(x)$ is the unique solution with respect to $\tau(x)$ of:

$$\mathbb{E}\big[\psi_{\tau(x)}(Y_i, W_i | X_i = x)\big] = 0$$

where $\psi(\cdot)$ is a moment equation, $\tau(x)$ is conditional treatment effect at x.

We adopt a local orthogonalization; that is, we use the Neyman-orthogonality moment condition inspired by Neyman (1979). Intuitively, by regressing out the effect of covariates (X_i) on the treatment (W_i) and outcome (Y_i), the condition alleviate concerns of having confounding factors that may affect both treatment and outcome. The similar marginal regression of Robinson (1988) is often used in recent machine learning algorithms to provide more precise estimators (e.g., Belloni et al., 2014, 2017; Chernozhukov et al., 2017; Nie and Wager, 2018; Athey et al., 2019).

$$\mathbb{E}\big[\psi_{\tau(x)}(\tilde{Y}_i, \tilde{W}_i | X_i = x)\big] = 0$$

where $\tilde{Y}_i = Y_i - E(Y_i|X_i = x)$ and $\tilde{W}_i = W_i - E(W_i|X_i = x)$, $\mathbb{E}(Y_i|X_i = x)$ is the mean outcome conditional on X_i, and $\mathbb{E}(W_i|X_i)$ is propensity score.

Next, we construct the weighted moment condition using the weights $\alpha_i(x)$. The weights $\alpha_i(x)$ are defined as the relevance of individual i to fitting $\tau(\cdot)$ at x. We fit our parameter $\tau(x)$ using an empirical version of the weighted moment condition. Thus, we find $\hat{\tau}(x)$ which solves:

$$\sum_{i=1}^{n} \alpha_i(x)\psi_{\tau(x)}(Y_i, W_i) = 0$$

where $\alpha_i(x)$ is the weights for individual i.

The forests estimator $\hat{\tau}(x)$ is consistent and asymptotically normal under the following assumptions (Athey et al., 2019): (1) samples are independently identically distributed, (2) the moment condition is Lipschitz continuous in x, (3) the moment condition is twice differentiable in $\tau(x)$ when x is fixed. (4) regularity condition, (5) the moment function $\psi(\cdot)$ is convex (or at least a negative subgradient of a convex function). (6) the subsamples for growing trees should be smaller and increase slower than the number of samples.

Construction of the weights $\alpha_i(x)$

Our causal forest follows the similar recursive partitioning process by minimizing the mean squared error of predictions, as other forest estimators grow trees. However, while the other forests use a partitioning process to define sharp terminal nodes in which conditional average treatment effects are estimated, this forest uses the partitioning process to construct the weights for defining the relevance of each individual i; that is, in our setting, the $\tau(x)$ are not identified at each stage of partitioning because $\tau(x)$ is only identified through the weighted moment condition, which needs fully constructed weights. As the weights and estimation are connected by reciprocal necessity, the optimization may be computationally expensive and the outcome may not be numerically attainable for every stage. Thus, during the partitioning process, this forest uses an approximate outcome, called pseudo outcome (denoted by ρ_i), which is the gradient-based approximation using parameters from the precedent sample (parent node), instead of the descendent subsamples (children nodes) themselves. The approximate estimates, $\tilde{\tau}_{c:j}$ (for $j = 1, 2$), can be obtained by averaging the estimates from the parent node (L_p) and the average of the pseudo outcomes in a child node j ($L_{c:j}$) as follows:

$$\tilde{\tau}_{c:j} = \hat{\tau}_p + \frac{1}{n_{c:j}} \sum_{\{i: X_i \in L_{c:j}\}} \rho_i \tag{3.1}$$

$$= \hat{\tau}_p - \frac{1}{n_{c:j}} \sum_{\{i: X_i \in L_{c:j}\}} \Big(\frac{1}{n_p} \sum_{\{i: X_i \in L_p\}} \nabla \psi_{\hat{\tau}_p(x)}(Y_i, W_i) \Big)^{-1} \psi_{\hat{\tau}_p(x)}(Y_i, W_i) \tag{3.2}$$

where L_p denotes a parent node, L_c denotes a child node, and j ($j = 1, 2$) is an index for the children nodes partitioned from a parent node. n_p is a number of observations in the parent node L_p, $n_{c:j}$ is a number of observations in the child node $L_{c:j}$, and ∇ denotes the gradient of a function. This implies that the estimates in children nodes $\tilde{\tau}_{c:j}$ can be optimized by moving in the direction of largest differences from the estimate in the parent node. The pseudo outcome ρ_i can be treated as numerical approximation by the Newton-Raphson method using parameters from the parent node $\hat{\tau}_p(x)$ as (3.2). The use of the pseudo outcome substantially reduces computational burden.

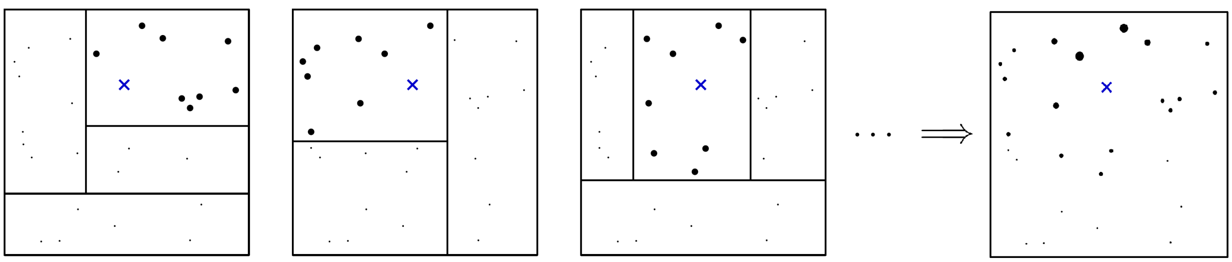
Having the pseudo outcome ρ_i from a node, we can recursively find the best split which minimize in sample goodness-of-fit, the sum of mean squared pseudo outcomes $(\tilde{\Delta}_c(L_{c:1}, L_{c:2}))$ within each children node $L_{c:1}$ and $L_{c:1}$:

$$x \in argmin_x \, \tilde{\Delta}_c(L_{c:1}, L_{c:2})$$
$$\text{where } \tilde{\Delta}_c(L_{c:1}, L_{c:2}) = \sum_{j=1}^{2} \frac{1}{n_{c:j}} \Big(\sum_{\{i: X_i \in L_{c:j}\}} \rho_i \Big)^2$$

Note that this partitioning process is to find x that partition our observations to have similar covariates. This recursive partitioning processes are repeated until some criteria are met such as the minimum number of individuals in a node, which is predetermined by a researcher. After finishing the partitions, we obtain a weight $\alpha_{i:b}(x)$ (b: an index of a tree) for a tree. The weight is 1 if an individual i shares the same support of x (falls into the same terminal node), 0 otherwise. Every individual within a terminal node of a tree share the same weight. This weight can be normalized as $\alpha_{i:b}(x) = \frac{\mathbf{1}(\{X_i \in L_b(x)\})}{|L_b(x)|}$, where $L_b(x)$ is the set of samples in the node x. As with the 'bagging' process used in other forests, we grow sufficiently a large number of trees (B) to obtain many weights, which are aggregated to construct an adaptive weight $\alpha_i(x) = \frac{1}{B} \sum_{b=1}^{B} \alpha_{i:b}(x)$. This adaptive weighting matrix captures the frequency of individuals sharing the same support of x with individual i.

Figure 3.1 displays how the weights are constructed. The left three subfigures indicate weights $\alpha_{i:b}(x)$ obtained from a tree, respectively. In each subfigure, individual i (marked as $\times$) falls in a terminal node defined by the two dimensions of x. Subsamples that fall into the same terminal node (marked as $\bullet$) have the same weight with an individual i, which implies their outcome will affect the estimate of the treatment effect of i. Note that other observations in different terminal nodes from the individual i (marked as $\cdot$) have zero weight. Since each tree is grown by using randomly drawn subsamples, the terminal nodes may be defined by different combinations of x. After growing many trees, we can obtain an adaptive weight for individual i by aggregating them together. The rightmost subfigure demonstrates the adaptive weight of individual i, which is the average of many weights from trees.

Figure 3.1: Construction of the Weights (source: Athey et al. (2019))



Using the adaptive weight $\alpha_i(x)$ defined by covariates x, we solve the moment equation to obtain predicted outcomes for individuals, i.e., the conditional average treatment effects (CATEs) conditonal on characteristics. Correspondingly, the distribution of conditional average treatment effects on the treated (CATTs) can also be obtained by considering only individuals who take JTPs.

We further improve our inference using a bias reduction estimator. Instead of simple aggregation of $\hat{\tau}(x)$, we use a doubly robust (DR) estimator using augmented inverse propensity weighting estimation to obtain a better estimate of the ATT ($\hat{\tau}^1$) as in Robins et al. (1994). The DR estimator reduces bias by employing an inverse propensity weighting estimator along with the baseline results from the causal forest. Intuitively, if either the propensity score ($e(X_i)$) or the mean outcome ($\mu(X_i)$) is correctly specified, the DR estimator will be unbiased.

We need to estimate tuning parameters, $e(X_i)$ and $\mu_0(X_i)$ before the DR estimation. Chernozhukov et al. (2018) use machine learning techniques to estimate those nuisance parameters and show that this DR estimator is consistent and semi-parametrically efficient. The form of the DR estimator for the ATT is as follows:

$$ATT = \hat{\tau}^1_{DR} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_t/n}\Big\{\big(W_i - \frac{\hat{e}(X_i)}{1-\hat{e}(X_i)}(1-W_i)\big)(Y_i - \hat{\mu}_0(X_i))\Big\}$$

where n_t/n is the proportion of the treated, $\hat{e}(X_i) = \mathbb{E}(W_i|X_i)$ is the estimated propensity score, $\hat{\mu}_0(X_i) = \mathbb{E}(Y_i(0)|W_i = 1)$ is the estimated (counterfactual) mean outcome.

Application to our Case: Estimation of Effect Heterogeneity

As we discussed earlier, our causal forest is one special case of the generalized random forest. Our non-parametric estimator of ATE can be simply summarized as the mean of the observed outcome under the aforementioned assumptions. We obtain our estimated parameter $\hat{\tau}$ from the following equation:

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}(Y_i(1)W_i - Y_i(0)(1-W_i))$$

Next, using the partitioning process as we previously described, we construct the weights $\alpha(x)$, which define an adaptive local neighborhood within the same support of x. The partition occurs based on the pseudo outcome ρ, a gradient approximation. We use the moment equation with residualized variables of Y_i and W_i; the use of locally centered variables $\tilde{Y}_i$ and $\tilde{W}_i$ under the orthogonality condition also contributes to have more precise results by eliminating confounding effects. In our context, the the approximate estimates $\tilde{\tau}_{c:j}$ (for $j = 1, 2$), can be obtained as:

$$\tilde{\tau}_{c:j} = \hat{\tau}_p + \frac{1}{n_{c:j}} \sum_{\{i:X_i \in L_{c:j}\}} \rho_i \tag{3.3}$$

$$= \hat{\tau}_p - \frac{1}{n_{c:j}} \sum_{\{i:X_i \in L_{c:j}\}} \Big(\frac{1}{n_p} \sum_{\{i:X_i \in L_p\}} \nabla \psi_{\hat{\tau}_p(x)}(Y_i, W_i) \Big)^{-1} \psi_{\hat{\tau}_p(x)}(Y_i, W_i) \tag{3.4}$$

$$= \hat{\tau}_p - \frac{1}{n_{c:j}} \sum_{\{i:X_i \in L_{c:j}\}} \frac{[(Y_i - \bar{Y}_p) - \hat{\tau}_p(x) \cdot (W_i - \bar{W}_p)](W_i - \bar{W}_p)}{Var_p(W_i)} \tag{3.5}$$

where j $(j = 1, 2)$ is an index for the children node $L_{c:j}$ partitioned from a parent node L_p, $\bar{W}_p = E(W_{-i}|X = x_p)$ and $\bar{Y}_p = E(Y_{-i}|X = x_p)$ are propensity score and average outcome from a parent node, respectively, which can be obtained by a leave-one-out estimation. x_p is the support of the parent node L_p, $Var_p(W_i)$ is variance of W_i at a parent node L_p.

We use the orthogonalized treatment and outcome in (3.5), and find x which minimizes mean squared error of pseudo outcome within children nodes. We finally construct a weight for a single tree, $\hat{\alpha}_{i:b}(x)$, using randomly drawn subsamples. The same processes are repeated to construct a large number of the weights $\hat{\alpha}_{i:b}(x)$, the weights are aggregated over many trees as the 'bagging' process, $\hat{\alpha}_i(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{\alpha}_{i:b}(x)$.

Having the weights $\hat{\alpha}_i(x)$ using the procedure, we can obtain our parameter via a non-parametric weighted estimator. This parameter can be defined as conditional average treatment effect (CATE) by employing the weight $\hat{\alpha}_i(x)$.

$$\hat{\tau}(x) = \sum_{i=1}^{n} \hat{\alpha}_i(x) \big[Y_i(1) W_i + Y_i(0)(1 - W_i) \big]$$

In our context, the assumptions for the causal forest hold. More specifically, our samples are supposed to be independently identically distributed under our setting. The conditional moment functions $\mathbb{E}(W_i|X_i = x)$, $\mathbb{E}(Y_i|X_i = x)$, and $\mathbb{E}(W_i Y_i|X_i = x)$ are all Lipschitz continuous in x, The orthogonal moment condition we employed above is twice differentiable in $\tau(x)$ when x is fixed. The regularity condition and the convexity of the moment function hold by the definition of the moment condition.

Researchers need to determine several tuning parameters for the causal forest algorithm such as (1) number of trees, (2) the minimum node size,[11] (3) number of covariates to be considered during each split in a tree, (4) proportion of sample for each tree and (5) proportion of sample for an honest process. Since more trees are preferable to construct our forests, we obtain 5,000 tree estimators to reduce dependency on a single tree. We choose the minimum node size equals to at least two, which means recursive partitioning process continues until a node size has at least one treated and one untreated individuals. The minimum node size has a trade-off between bias and variance; that is, the smaller node size defines our predicted results more finer ways which may reduce bias, but increase variance, and vice versa. We choose the smaller node size, since a lager number of trees can offset the increasing variance. In addition, the number of randomly chosen covariates used to grow a tree is set to 50. This number is mainly related to runtime of the algorithm, thus we choose a number which is sufficiently large but not too big. Finally, we randomly split the sample in half for constructing a weight from each tree, and partition into half again for an honest process, which is the most general and popular split. Consequently, a randomly chosen one-fourth of the sample is used for weight construction and another one-fourth is used for each tree estimation.

3.4 Data

3.4.1 Administrative data

The Korea Employment Information Service (KEIS) under the Ministry of Labor and Employment in South Korea provided three different sources of administrative data: (1)

[11] The minimum node size restricts the number of the treated (control) in a node.

the human resource development (HRD) file, (2) the unemployment insurance (UI) file, and (3) the Worknet dataset.[12] These three datasets are connected to each other by an individual's identification number. Our dataset is analogous to that used in Hwang et al. (2019). The HRD file contains information on trainees such as demographics and start/end dates of training. The UI file contains the last workplace related profiles of those eligible for UI benefits. Details include (1) date of employment and separation (2) reasons for separation, (3) occupation type, (4) firm size, (5) location, and (6) initial monthly wage when re-employed. These two data sources are similar to those used in earlier studies (e.g., Lee and Lee, 2005, 2009; Choi and Kim, 2012; Choe et al., 2015).

However, our data have two advantages comparing to that in the previous studies. First, it is constructed after the 2011 reform, so that the estimation of the effects of JTPs is more tractable as the receipt of the both benefits, UI benefits and JTPs, became possible.[13] Second, we additionally employ "Worknet" dataset, which provides more information about individuals such as the registration date at Worknet, educational attainment, work experience within a given industry, certificate and other personal characteristics. Thus, our dataset is richer than the previous studies in that it includes more covariates from the additional dataset.

3.4.2 Sample definition

Our data focuses on displaced workers who are eligible for UI benefits in South Korea. The sample includes those who lost and left their jobs involuntarily and show their intention to find a job by registering at Worknet, regardless of participation in JTPs.

[12]The Worknet, a portal website, was developed by the KEIS in 1998 to provide public job related information to job seekers. The target has expanded to the private sector in 2011, so as to provide more comprehensive job related information including both public and private sectors. Displaced workers are required to register on Worknet to receive UI benefits or training subsidies from the government, which also encourage employers to participate in the Worknet.

[13]Prior to the 2011 reform, a displaced worker needs to choose either JTPs or the UI benefits even both benefits are available.

The HRD file contains participants of JTPs—who start and finish their JTPs within 2013. These participants comprise the treatment group ($T_i = 1$) with 58,512 observations. Using the UI file and Worknet dataset, we construct the control group ($T_i = 0$) who becomes separated and registers at Worknet within 2013—the same period with the treatment group to control their start date of job-search—but do not undergo any JTPs. Our control group consists of 306,923 observations.

Employment status (S_i) indicates whether individuals are employed within 12 months since the start date of job-search. Our definition has two advantages comparing to that in the previous studies. One is the use of 12-month duration. In general, the 12-month duration is common criteria to measure short-term effects (Card et al., 2010), and it has an advantage than shorter duration, such as 3-month or 6-month, to control for seasonal recruitment. The other is the use of different start date of job-search (12-month window) between treatment and control group. For treatment group, it starts from the end of JTPs, and for control group, from the date of registration at Worknet.[14] Our definition for trainees implies that training duration is not included in job-search because trainees are likely to focus on programs during training duration in the sense that the training expenses cannot be supported by the government unless they finish the programs in good attendance. Also, this definition can improve better causal interpretation of the program effects by controlling for trainees' opportunity for job-search, since the length of JTPs up to 6 months may decrease trainees' opportunity for job-searching. Our definition for the control group implies that the duration between the date of separation and registration at Worknet, i.e. "inactive duration", is excluded because we regard the duration as discouraged period which the displaced workers are inactive to look for a job. This seems plausible in two regards. First, this is in line with the objectives of Worknet (job-matching) because the government does not treat them as job seekers eligible to receive any benefits until they register at Worknet. Also, since trainees need to register at Worknet in order to

[14]These definitions are similar with Choe et al. (2015) and Hwang et al. (2019).

take JTPs, the similar rules need to be applied for the control; they need to register at Worknet in order to start job-search. In that sense, our duration of job search increase the comparability between the two treatment arms than the previous studies which measures both duration from separation.

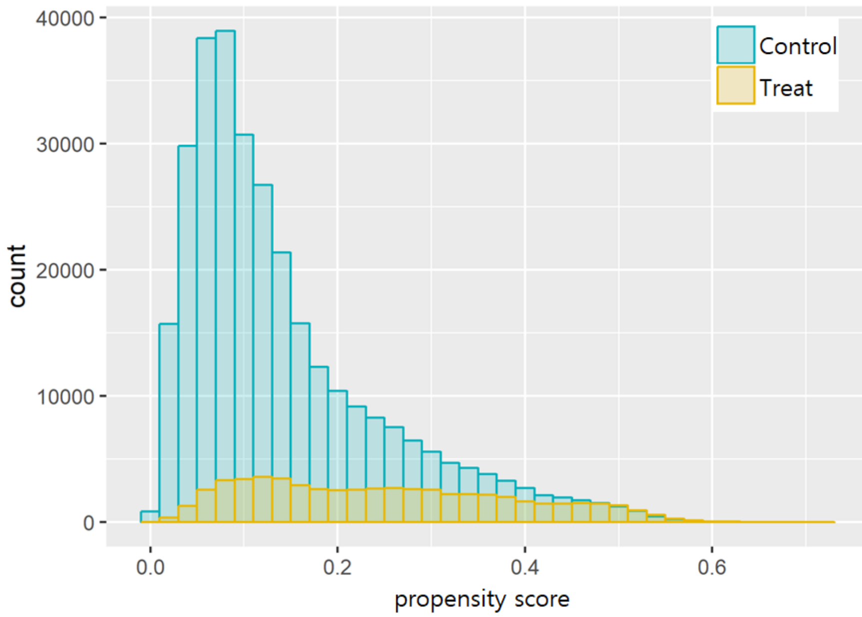
The set of covariates is chosen to cover observed characteristics of individuals, which serves as important controls for confounding factors in our specification. Our covariates span demographic (age, gender, and disability), personal characteristics (educational attainment, residential region, duration between their separation and registration at Worknet, and resume completion rate), and the last workplace related information (tenure, work experience at a given industry, occupation, industry, reasons for separation,[15] location, firm size in persons, and month of separation.

The duration between the date of separation and the date of registration at Worknet, namely 'inactive duration', is a variable created by ourselves, which is used as a proxy for how actively an individual wants re-employment. Since we assume the registration date is the start of job-search, this variable implies that the earlier registrants are likely to escape from inactive period and start their job-search earlier than the others.

The 'resume completion rate' is obtained from the Worknet dataset. A displaced worker is required to fill out their resume when they register at Worknet. This resume can be used as background information for counseling with a caseworker, or as a resume to apply for a job through Worknet. Thus, less blanks in their resume are likely favorable for either taking JTPs or finding a job. Filling in supplementary information such as certificates and detailed work experiences can make the completion rate exceed 100%.

[15]The reasons for separation documented by the last workplace are categorized into three parts: (1) personal reasons (changing jobs, family issue, disease/injury, disciplinary dismissal, and personal miscellaneous), (2) firm-oriented reasons (shutdown, layoff, exploitation, and firm miscellaneous), (3) employment expiration (retirement, contract expiration, project completion)

Figure 3.2: Assess Overlap between the Two Treatment Arms



3.4.3 Summary statistics

We first assess covariate balance between the two treatment arms by using propensity score. We estimate propensity score using a simple logistic regression with linear terms of covariates because moderate misspecification may be allowed in our doubly robust estimation. Figure 3.2 presents histograms of the estimated propensity scores for the treated and for the control groups respectively. The histogram shows that there exist some observations which do not overlap in the common support of the estimated propensity score. We trim observations from both tails which are greater than the maximum value of the estimated propensity score for the control, and which are smaller than the minimum estimated propensity score of the treated. The final size of our samples is 365,179 which consists of 58,490 individuals for the treated, and 306,689 individuals for the control.

Table 1[16] reports descriptive statistics of the treated and control along with their

[16]This table shows some important characteristics, and the full list of covariates can be found in the Appendix.

standardized differences.[17] The differences less than 20 between the two groups can be regarded as negligible by rule of thumb (Rosenbaum and Rubin, 1985). A large number of covariates shows significant imbalance between the two treatment arms: age (48.3), gender (23.3), work experience within a given industry (59.8), educational attainment—elementary school (24.7), and miscellaneous reasons for separation—personal (46.6) and firm-oriented (34.2). More specifically, the treatment group (36.71 year-old) was generally younger than the control group (42.53 year-old), and the proportion of males is 37% which is significantly less than in the control group (48%). The treatment group are less likely to be middle school graduates or below, more likely to have 2-year college or bachelor's diploma. The treatment group has less work experience within a given industry (0.39 vs. 0.67), and less tenure in the last workplace (1.69 year vs. 2.24 year). While the most common reason for separation is firm-oriented miscellaneous for the control group (0.38), it is personal miscellaneous issues for the treatment group (0.46). Trainees are likely to have longer inactive duration than the control group (28.65 weeks vs. 22.66 weeks). The residential and last workplace regions, industry, the ex-firm size (in persons), and the month of separation are quite similar between the two treatment arms: trainees are slightly less likely to have worked as a manual laborer (0.10 vs. 0.14) and more likely to work at the large($\geqslant$1000) firm.

Note that the imbalance between the two treatment arms may not affect our analysis as we discussed earlier. The causal forest estimators are quite robust in the presence of confounders that may affect treatment and outcome simultaneously in our setting: (1) our forest works similar to the weighting estimator defined by terminal nodes and (2) we conduct partialing out regressions using locally-centered variables. Thus, we implement the causal forest method using the above sample with the final size of 365,179.[18]

[17] $d = \frac{100\times|\bar{x}_t - \bar{x}_c|}{\sqrt{1/2\times(\hat{s}_t^2+\hat{s}_c^2)}}$, d: standardized difference, $\bar{x}_g$: sample mean of g, $\hat{s}_g$: sample standard deviation of g.

[18] For robustness check, we construct the weighting matrix using a matched control group constructed by one to one propensity score matching, and obtain CATTs. These estimates do not show any substantial differences from the ones presented here.

Table 3.1: Descriptive Statistics

Variable	Treated (W_i=1) Mean	Treated (W_i=1) Std. Dev.	Control (W_i=0) Mean	Control (W_i=0) Std. Dev.	Std. Diff.
Age (years)	36.71	14.97	42.53	12.76	48.30
Gender (male=1)	0.37	0.48	0.48	0.50	23.30
Disabled	0.03	0.33	0.04	0.41	3.92
Tenure (years)	1.69	4.09	2.24	3.78	15.49
Work Experience (within a given industry)	0.39	0.49	0.67	0.47	59.80
Registration at Worknet (weeks)	28.65	37.81	22.66	30.15	18.75
Resume completion rate (%)	91.96	20.98	90.85	17.88	6.33
Educational Attainment					
None	0.00	0.06	0.01	0.11	10.89
Elementary school	0.01	0.10	0.05	0.22	24.66
Middle school	0.03	0.17	0.07	0.26	19.13
2-year College	0.29	0.45	0.22	0.42	14.89
Bachelor's	0.30	0.46	0.26	0.44	8.78
Graduate	0.02	0.13	0.03	0.16	6.16
Residential Region					
Seoul	0.21	0.41	0.22	0.41	1.15
Daegu	0.06	0.24	0.05	0.21	6.06
Gwangju	0.04	0.20	0.03	0.17	6.76
Gyeong-gi	0.25	0.43	0.26	0.44	2.80
Workplace Region					
Seoul	0.38	0.48	0.34	0.47	7.43
Gwangju	0.03	0.16	0.02	0.14	4.10
Gyeong-gi	0.19	0.39	0.21	0.41	4.93
Industry					
Lodging/Restaurant	0.05	0.22	0.05	0.21	2.07
Public Admin.	0.04	0.19	0.04	0.20	3.33
Social service	0.11	0.32	0.11	0.31	2.31
Ex-firm size (persons)					
<5	0.21	0.40	0.22	0.41	2.58
5-9	0.12	0.32	0.13	0.34	4.33
10-29	0.18	0.38	0.20	0.40	5.51
500-999	0.05	0.22	0.05	0.21	3.09
⩾1000	0.14	0.34	0.10	0.30	12.40
Ex-occupation					
Supervisor/Clerk	0.22	0.41	0.18	0.38	9.18
Merchandiser	0.15	0.36	0.12	0.33	8.95
Farmer/Fishery	0.00	0.07	0.01	0.10	6.92
Manual Laborer	0.10	0.30	0.14	0.35	12.80
Reason for unemployment					
Personal: Changing jobs	0.06	0.25	0.04	0.19	12.58
Personal: Family issue	0.03	0.16	0.02	0.12	7.17
Personal: Disease/Injury	0.02	0.15	0.02	0.13	4.97
Personal: Misc.	0.46	0.50	0.25	0.43	46.58
Firm: Shutdown	0.02	0.13	0.03	0.18	10.90
Firm: Layoff	0.02	0.14	0.03	0.17	6.43
Firm: Misc.	0.23	0.42	0.38	0.49	34.23
Expiration: Retirement	0.01	0.09	0.02	0.13	7.82
Expiration: Contract expiration	0.14	0.34	0.20	0.40	17.03
Expiration: Project Completion	0.00	0.06	0.01	0.09	5.92
Employment status	0.63	0.48	0.59	0.49	7.30
Observations	58,490		306,689		

3.5 Results and Interpretation

3.5.1 Distribution of Treatment Effects on Trainees

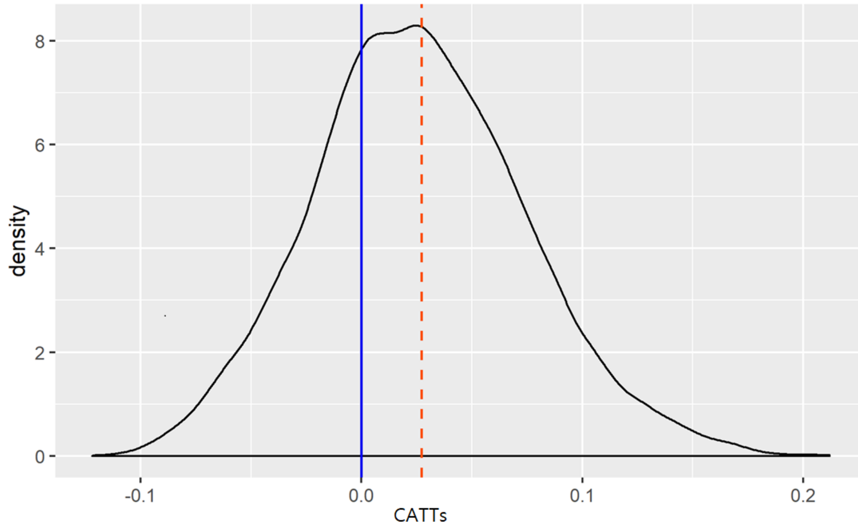
Our results using the doubly robust estimator suggests that the ATT on the probability of re-employment ($\hat{\tau}^1_{DR}$) is a statistically significant 0.029 (std. error 0.003). This result implies that trainees in Korean JTPs on average experience positive effects on employability. The positive result is consistent with the previous studies (e.g. Yoo and Lee, 2008; Hwang et al., 2019).

Figure 3.3 displays the distribution of the CATTs, $\hat{\tau}^1(x)$. The x-axis indicates the conditional treatment effect of trainees and the y-axis indicates density. We observe that around 31.0% of individuals have negative effects on employability, while the majority of trainees (69.0%) experience positive effects; that is, approximately a third of trainees seem not benefiting from JTPs.[19] In order to investigate the groups that do not benefit from training (in terms of employability), we analyze their characteristics.

Table 3.2 reports the average characteristics of the groups that are defined by the sign of the CATTs, $\hat{\tau}^1(x)$. We observe that the two groups have several differences in average characteristics that are statistically significant. More specifically, the group with positive effects is older (37.10 vs. 35.83 year-old), more likely to be female (0.32 vs. 0.48), to have longer tenure (2.13 vs. 0.71 years), work experience within a given industry (0.43 vs. 0.30), and shorter inactive duration (26.95 vs. 32.42 weeks). They are also more likely to have a bachelor's degree (0.32 vs. 0.26) or 2-year college degree (0.30 vs. 0.27), and less likely to have a high school diploma (0.32 vs. 0.41). They are more likely to live in Seoul

[19] Note that, if we were to consider any out-of-pocket expenses of training, the proportion of negatively affected individuals would increase.

Figure 3.3: The Distribution of the Treatment Effects of Trainees



(0.22 vs. 0.19) and Gyeong-gi (0.26 vs. 0.22), and to have worked in Seoul (0.40 vs. 0.33). They are also less likely to have worked in manufacturing (0.19 vs. 0.28) as manual laborer (0.07 vs. 0.17), but more likely to have worked in financial/insurance (0.03 vs. 0.01), technological service (0.05 vs. 0.04), wholesale/retail services (0.15 vs. 0.12), or as professional/engineer (0.29 vs. 0.20). They are less likely to be separated due to personal miscellaneous reasons (0.36 vs. 0.69), but more likely to be separated due to firm-oriented reasons such as miscellaneous (0.29 vs. 0.09), and due to employment expiration such as contract expiration (0.16 vs. 0.09). They also have a lower propensity to take JTPs (0.24 vs. 0.28) even though they experience positive effects, which may suggest the existence of some adverse selection into taking on JTPs.

3.5.2 Heterogeneous Treatment Effects

In this section, we investigate heterogeneous treatment effects on the treated by different characteristics of trainees. Using some distinctive characteristics, we categorize the samples into several groups. For example, continuous characteristics such as age, tenure, inactive

Table 3.2: Some Distinctive Average Characteristics by the Sign of the Treatment Effect

	$\hat{\tau}_i > 0$		$\hat{\tau}_i \leqslant 0$		t-stat.
Variable	Mean	Std. Dev.	Mean	Std. Dev.	
Age (years)	37.10	14.83	35.83	15.37	13.07***
Gender (male=1)	0.32	0.47	0.48	0.50	-36.87***
Tenure (years)	2.13	4.10	0.71	10.40	68.82***
Work Experience within a given industry	0.43	0.49	0.30	0.46	30.50***
Inactive Duration (weeks)	26.95	29.70	32.42	27.35	-17.82***
Resume Completion Rate (%)	92.41	19.37	90.96	19.15	9.58***
Educational Attainment					
High school	0.32	0.46	0.41	0.49	-22.67***
2-year College	0.30	0.46	0.27	0.44	8.05***
Bachelor's	0.32	0.47	0.26	0.44	16.32***
Residential Region					
Seoul	0.22	0.42	0.19	0.39	10.01***
Gyeong-gi	0.26	0.44	0.22	0.42	9.39***
Workplace Region					
Seoul	0.40	0.49	0.33	0.47	15.71***
Gyeong-buk	0.04	0.19	0.05	0.22	-7.78***
Gyeong-nam	0.05	0.21	0.06	0.24	-8.04***
Industry					
Manufacturing	0.19	0.39	0.28	0.45	-24.85***
Wholesale/Retail	0.15	0.36	0.12	0.33	8.58***
Finance/Insurance	0.03	0.16	0.01	0.11	11.16***
Tech. service	0.05	0.23	0.04	0.19	9.83***
Educational service	0.05	0.23	0.04	0.19	8.28***
Social service	0.12	0.32	0.10	0.30	8.44***
Ex-firm size (persons)					
150-299	0.06	0.23	0.08	0.28	-11.14***
⩾1000	0.15	0.36	0.12	0.32	10.98***
Ex-occupation					
Professional/Engineer	0.29	0.45	0.20	0.40	24.20***
Supervisor/Clerk	0.23	0.42	0.18	0.39	13.05***
Operator	0.13	0.33	0.18	0.39	-16.79***
Manual laborer	0.07	0.26	0.17	0.37	-31.59***
Reason for Separation					
Personal: Family issue	0.03	0.18	0.01	0.11	17.62***
Personal: Disease/Injury	0.03	0.16	0.02	0.12	9.71***
Personal: Miscellaneous	0.36	0.48	0.69	0.46	-79.56***
Firm: Shutdown	0.02	0.14	0.01	0.10	9.58***
Firm: Layoff	0.02	0.15	0.01	0.10	13.20***
Firm: Exploitation	0.01	0.10	0.00	0.06	8.83***
Firm: Miscellaneous	0.29	0.45	0.09	0.28	66.43***
Expiration: Retirement	0.01	0.10	0.00	0.03	17.20***
Expiration: Contract Expiration	0.16	0.36	0.09	0.29	21.65***
Propensity score	0.24	0.14	0.28	0.13	-31.11***
Observations	40,338		18,152		

*** indicate that the estimates are statistically significant at the 99% level.

duration, and resume completion rate, are divided into deciles. This partition rule has the advantage of having the same number of trainees within every bin.

Figure 3.4 reports CATTs by the decile of each continuous variable. Specifically, Figure 3.4(a) shows the effects by age decile, which reports that most trainees experience significantly positive effects at the 95% confidence level, while the magnitudes do not show any strong trends across deciles. It is worth to taking a look at the second (27-29), the ninth and tenth ($\geqslant$55) deciles, which show larger positive effects (0.048-0.066) than other deciles. The second decile whose age is between 27 and 29 may be relevant to those who finished their bachelor's degrees but with scant work experience, or who have less than college diploma but some have some work experience. For the last two deciles, corresponding to ages above 55, large positive effects are observed from the elderly as we observed in the previous section. They could be individuals that are close or have retired, making it probable that the motivation for taking and experiencing larger effects from training are related to the loss of financial stability close to or after retirement.

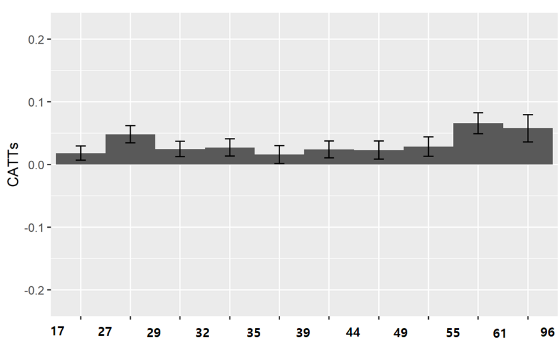
Figure 3.4(b) presents CATTs by deciles of the inactive duration variable. Throughout the deciles, positive effects are observed which are individually statistically significant at the 95% confidence level, except for the last decile (more than 507 days of inactive duration). The larger positive effects are shown from the third to the fifth deciles, corresponding to those who are inactive for 12-53 days after their separation. The magnitude of positive effects gradually decreases as the inactive duration becomes longer. The training effect in the last decile appears significantly smaller than in the others; this suggests that longer inactive duration than 507 dyas are unlikely to contribute to the effectiveness of JTPs.

Figure 3.4(c) shows CATTs by the resume completion rate at Worknet. The effects are similar across the deciles, which is positive and statistically significant at the 95% confidence level, except in the first two deciles. It seems that those who completed less than

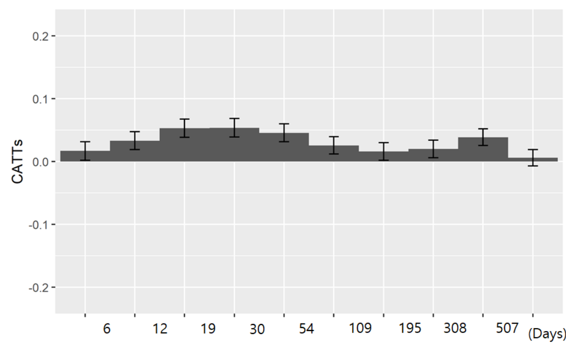
75% of their resume at Worknet do not show statistically significant effects from training.

Finally, Figure 3.4(d) shows CATTs over the tenure decile. We observe a strong positive relationship between tenure and the magnitude of the CATTs, which indicates that those with longer tenure are likely to experience larger positive treatment effects from training. More specifically, those who worked less than 17 weeks in their prior job are likely to experience statistically significant negative effects from training, while those who worked more than 27 weeks are more likely to have positive and statistically significant effects. The largest positive effect (0.099) is observed for the last tenure decile—who have more than 256 weeks of tenure in their last job. Interestingly, if we categorize the tenure deciles into four groups, the group-specific average CATTs are statistically different from each other at the 95% confidence level. The groups are: (1) less than 0.5 year (<17 weeks, 1^{st}-3^{rd} decile), (2) 0.5-1 year (17-52 weeks, 3^{rd}-5^{th} decile), (3) 1-2.7 years (52-138 weeks, 6^{th}-8^{th} decile), and (4) 2.7 years or more ($\geqslant$138 weeks, 9^{th}-10^{th} decile).

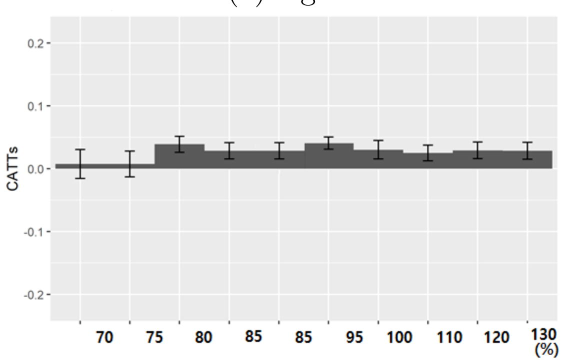
Figure 3.5 reports CATTs by some important categorical variables. Figure 3.5(a) shows CATT by educational attainment. We observe that all the educational groups experience significantly positive effects except, the most educated group (graduate level). The largest positive effects are observed in the group with lower educational attainment, such as no education (0.078), elementary school graduates (0.066), and middle school graduates (0.042), but these effects are not statistically different from each other. Those who have a bachelor's diploma also experience a moderate positive effect (0.041), which is statistically significantly larger than that of high school graduates and 2-year college graduates. The most educated group has the smallest positive effects (0.015), which is statistically insignificant. It seems plausible that low educational attainers are more likely to be employed after learning basic skills through JTPs, because many JTPs aim at re-employment in lower-skilled occupations such as clerk, office assistant and assistant cook (Yoon, 2014).



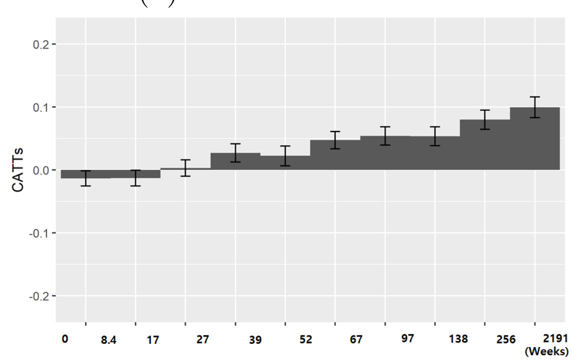
(a) Age



(b) Inactive Duration



(c) Resume Completion Rate



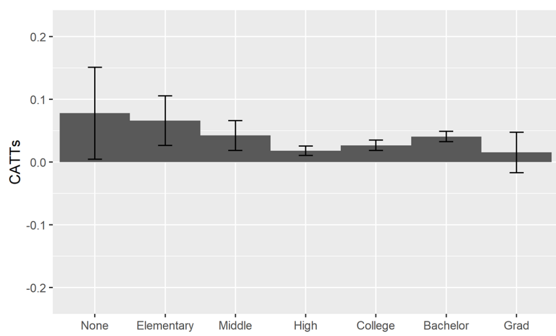
(d) Tenure

Figure 3.4: CATT by Decile of Each Continuous Variable

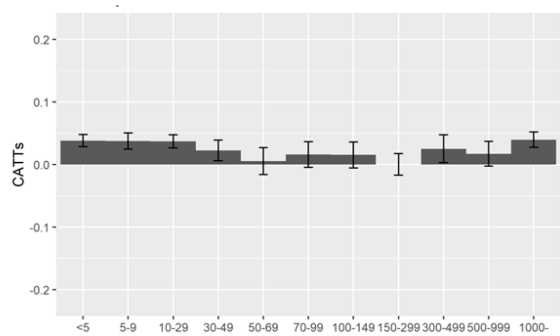
Figure 3.5(b) reports CATTs by deciles of the firm size (in persons) of the last workplace. A slightly U-shaped pattern is observed, with those who have worked at the smaller (49 or less persons) or the larger (300-499 and 1,000 or more) firms experiencing positive and statistically significant effects (0.037-0.040), while the others at mid-sized firms generally show statistically insignificant effects.

Figure 3.5(c) reports CATTs by the month of separation. Large differences across the month of separation are not observed, except for those who are separated in March, April and July show positive and statistically significant effects, while those who are separated in the other months show statistically insignificant results. In general, those who become separated in the first half of the year (January to June) seem more likely to have greater positive effects than those separated during the other half (0.022-0.052 vs. 0.002-0.031). And those separated during the last three months of the year (from October to December) show small positive or negative effects that are statistically insignificant.

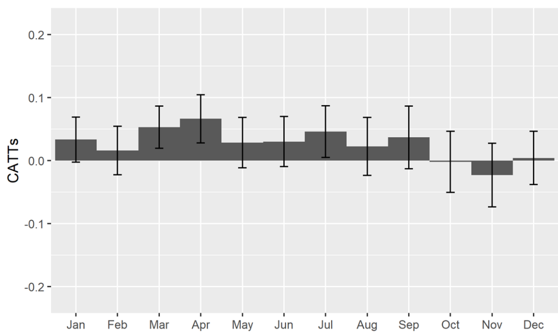
Figure 3.5(d) shows that the CATTs by the reasons for employment separation. Despite the fact that all trainees are involuntarily separated, the trainees whose separation is related to personal reasons, such as changing jobs, disease/injury, disciplinary dismissal and other miscellaneous, do not benefit from JTPs (they show a precisely estimated zero effect). In contrast, trainees who have any of the other two reasons for employment separation experience large and statistically significant positive effects: (1) firm-oriented reasons (0.076) such as layoff and firm's miscellaneous issues, and (2) employment expiration (0.053) such as retirement and contract expiration. It is possible that trainees who are involuntarily separated through personal reasons cannot eliminate those personal reasons (e.g., injury/disease or disciplinary dismissal) and thus they may not be immediately re-employed after undergoing the JTPs, which results in their lower effectiveness.



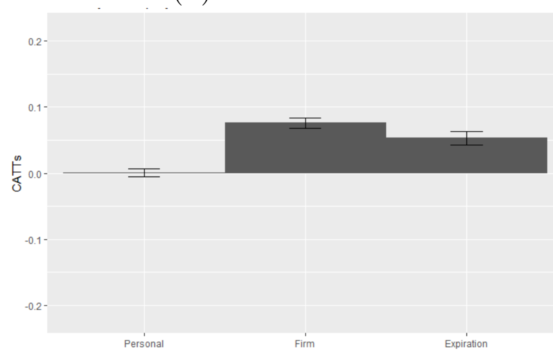
(a) Educational Attainment



(b) ex-Firm Size



(c) Month of Separation



(d) Reasons for Separation

Figure 3.5: CATTs by some Categorical Variables

3.5.3 Hypothetical Assignment Rules

Using the distinctive features from our results, we hypothetically re-assign treatment using practical rules defined by characteristics of individuals, which are obtained from Table 3.2 and the CATTs in the previous section; positive effects are observed for those who are more likely to have lower educational attainment and longer tenure, more likely to be separated due to firm-oriented or expiration reasons, shorter inactive duration. In addition, we also postulate some assignment rules for the socially-disadvantaged groups, such as the elderly, disabled, female, and those who have worked at small-size firm (less than 10). Note that easily changeable characteristics may not be a good candidate for assignment rules since these rules may affect to change trainees' behavior.

When simulating these rules, we set the constant number of trainees (58,490) across hypothetical assignment rules. The fixed number of participants is desirable for better comparability. For simplicity, we assume that there is no searching cost for trainees and agencies, and every trainees are equally subsidized by the government.[20] We impose different assignment rules for the type of characteristics. More specifically, for continuous characteristics, such as age, tenure, and inactive duration, we assign JTPs to the top or bottom of 58,490 individuals, and estimate (hypothetical) CATTs using the DR estimator. Meanwhile, for categorical covariates, we assign JTPs to the subgroup which shows higher CATTs from our analysis. To have the fixed number of trainees, we randomly draw samples from the target group with replacement, and estimate the CATTs and their standard errors using bootstrap. For example, for the rule 'lower education', we assign treatment to those who have middle school diploma or less. The number of target group is 45,661 which is less than our treated individuals (58,490), so we randomly draw 58,490 individuals from the target group (lower educational attainer) with replacement, and use them as a hypothetical treatment group. We iterate this process 500 times using the bootstrap to obtain estimates of our parameters of interest and their corresponding standard errors.

We document the following twelve assignment rules: (1) assignment of displaced workers with the highest CATTs ('best case'), (2) assignment of displaced workers with the lowest CATTs ('worst case'), (3) random assignment, (4) assignment of displaced workers with longer tenure ('longer tenure'), (5) assignment of displaced workers due to firm-oriented reasons ('firm-oriented'), (6) assignment of displaced workers due to employment expiration reasons ('expiration'), (7) assignment of displaced workers with shorter inactive duration ('early registrant'), (8) assignment of displaced workers with middle school diploma or less ('lower education'), (9) assignment of displaced workers who have worked at a small firm, less than 10 persons ('smaller firm'), (10) assignment of the disabled ('disabled'), (11) assignment of the elderly ('elderly'), and (12) assignment of

[20]Training subsidies are allocated based on the number of trainees, hours of training programs, and type of programs. Since we do not have information of the type of JTPs in our dataset, we assume homogeneous JTPs, which has no difference of the cost. In fact, the government pre-determines 251 types of programs' subsidies ranging roughly from KRW 5,000 to 8,000 per hour per trainee.

female (‘female’).

Table 3.3 reports the CATTs under the hypothetical assignment rules for JTPs. The ‘worst’ and the ‘best’ cases of the CATTs are -0.033 and 0.125 respectively, which indicate the lower and upper bounds of the CATTs. It implies that the current JTPs can improve employability up to 0.125 with the oracle assignment rule. The results from any other rules would be bounded away from these results. Note that these bounds are obtained by simple average of individual treatment effects from the top/bottom. The CATT estimates by the ‘random assignment’ rule is 0.041, which is greater than our baseline (0.029). This implies that the current voluntary enrollment system of JTPs may suffer from adverse selection of trainees. This is consistent with Yoon (2014), which argues that around 26% of trainees are likely to take on some programs which are less relevant to re-employment, such as programs for cook, barista, pâtissier, beautician, and esthetician; these programs show lower employment rates (less than 30%) than the average (32.8%). Meanwhile, the CATT estimates by ‘longer tenure’ rule, which indicates assignment on those with longer tenure, shows the best performance of employability (0.112) among our assignment rules, which shows statistically significantly better performance than the other rules. It is probable that those who spend longer time in a firm may be preferred in the labor market, or they are more likely to have knowledge about which JTP is more beneficial for re-employment. The CATT estimates by ‘firm-oriented’ is 0.074, which is the second best among our potential assignment rules, and the CATT by ‘expiration’ is 0.057 which is also better than the current rules, while the CATT of ‘personal reasons’ is insignificant (0.000). Thus, the reason for separation can classify who should be treated effectively. The CATT estimates by ‘early registrant’ is 0.015, which is positive and statistically significant at the 95% confidence level, but it is worse than the current rule. As we observed in Figure 3.4(b), those who registered between 12-30 days (0.053) may show better results than the current one, however, the use of this rule is not recommended because the displaced workers’ can easily change their registration date regardless of their intention to work, which violate the exogeneity assumption.

We also observe potential assignment rules focusing on some socially disadvantaged groups, and find that the CATTs of those rules are positive and statistically significant at the 95% confidence level. More specifically, the CATT for 'lower education' is 0.044, for 'smaller firm(<10)' is 0.054, of 'elderly' is 0.074, of 'female' is 0.061, and of 'disabled' is 0.054, and we find that these rules are better than the current one, which is statistically significant. However, these rules may not be appropriate or practical for ethical or political reasons. It however should be noted that policymakers can refer these results to use for policy design purpose, such as to improve equality (e.g. develop more programs for the disadvantage groups).

Table 3.3: Potential Treatment Assignment Rules

Assignment Rules	CATT	Std.error	% Change*
ATT (baseline)	0.029	(0.002)	-
Highest CATTs	0.125	-	325.5
Lowest CATTs	-0.033	-	-212.7
Random Assignment	0.041	-	40.8
Longest Tenure	0.112	(0.007)	282.4
Reason:Firm-oriented	0.074	(0.000)	151.9
Reason:Expiration	0.057	(0.000)	94.1
Lower Educational Attainment	0.044	(0.000)	50.2
Firm Size (in persons) <10	0.054	(0.000)	82.8
Earlier Registration at Worknet	0.015	(0.006)	-49.9
Disabled	0.054	(0.000)	84.8
Elderly	0.074	(0.009)	153.9
Female	0.061	(0.005)	108.9

* The percentage change measures relative to our baseline result (ATT).

3.6 Discussion and Conclusion

We estimated heterogeneous treatment effects of Korean JTPs using the causal forest in a selection on observables setting. We find that JTPs have positive effects on trainees, 0.029, which is statistically significant. However, a large proportion of trainees, around 31%, experience negative effects, while the rest 69% of trainees has positive effects. We illustrate their characteristics in terms of the sign of the effects, and compare conditional average treatment effects by using important characteristics. We find that some characteristics are highly associated with positive treatment effects: those who are likely to have longer tenure, lower educational attainment, be separated due to firm-oriented issues or employment expiration. Some socially disadvantage groups such as elderly and female also show positive effects.

Given the results, we examine some hypothetical treatment assignment rules and their expected treatment effects, and find some assignment rules outperform the current one. Specifically, assignment on those with longer tenure is the most effective (0.112) on employability, and assignment on those separated due to firm-oriented reasons such as layoff (0.074), due to employment expiration such as retirement (0.057), on the elderly (0.074) also shows large and statistically significant effects. Therefore, we can suggest that the adoption of some assignment rules can improve employability of JTPs.

Some policy suggestions are still hypothetical in the sense that we estimate the effects using bootstrap. Also, those suggested assignment rules do not reach at the (ideal) upper bounds, which implies there still exist potential room for further improvement. Future work will find tailored assignment rules that can maximize welfare as in Kitagawa and Tetenov (2018) and Athey and Wager (2018), and get closer to the upper bound using the the effect heterogeneity.

Chapter 4

Conclusion

This dissertation evaluated the effectiveness of Korean job training programs (JTPs) in various manners. Both chapters estimate the average treatment effects on the treated (ATT) on the probability of re-employment. The results from the dissertation persistently show that JTPs after the 2011 reform are likely to help the displaced workers to be re-employed: the ATT estimates in the first chapter is 0.026, and that in the second chapter is 0.029, each of which is statistically significant.

The first chapter additionally investigates the ATT on wages. Our analysis addresses two common biases by: (1) propensity score matching to tackle selection into JTPs; (2) principal stratification framework to deal with selection into employment. We conduct maximum likelihood estimation under stochastic dominance assumptions, and find that JTPs have negative effects on those who are always employed regardless of the JTPs to the amount of 8.4% (KRW -205.3$\times 10^3$), despite the positive effects on employability.

From a policy perspective, the results of this chapter corresponds to one of the primary stated objectives that aims to lower unemployment rate. However, the negative effects on wages seems not to be desirable. We conjecture that the negative effects result from wrong assignments of JTPs. Interestingly, this argument can be supported by the results from the second chapter.

The second chapter focuses on heterogeneous treatment effects on the probability of re-employment by identifying our parameter in finer ways. This chapter systematically investigates heterogeneous treatment effects by using the causal forest method under the selection on observables setting. Our causal forest performs valid inference by avoiding spurious error using the 'honest' process and mitigating bias by adopting the Neyman's orthogonality condition in the spirit of Robinson (1988). We additionally use a doubly robust estimator to obtain more precise results (Robins et al., 1994; Chernozhukov et al., 2017, 2018).

This chapter finds that the distribution of treatment effects demonstrates that around 31% of trainees experiences a negative effect, despite the positive ATT (0.029). We uncover distinctive characteristics between trainees with positive and negative effects; trainees who have positive effects are likely old, female, to have longer tenure in their work, and become separated due to firm-oriented reasons or employment expiration.

We postulate some hypothetical assignment rules, and find that the current assignment rule (0.029) seems to have worse employability than (simulated) random assignment (0.041), which implies adverse selection. We also find that some hypothetical assignment rules outperform the current one. For example, assignment on those who have longer tenure is the most effective (0.112) among others, and assignment on those who are separated due to firm-oriented reasons such as layoff (0.074) and employment expiration such as retirement (0.057). Some assignment rules for the disadvantaged groups such as female, the elderly, and those who have worked in a firm with fewer than 10 employees also perform better than the current one.

Based on our findings, we argue that the current JTPs can be improved by adopting program assignment rules. The current assignment mechanism of the reformed JTPs highly relies on trainees' choice, which suffers from self-selection bias into JTPs due to the weak role of caseworkers and lowered opportunity cost for taking JTPs. Thus, policymakers can consider two policy tools to improve the effectiveness: (1) adoption of some assignment rules, which prioritize some characteristics such as tenure and reasons for separation, and (2) intensive counseling with a caseworker, which can alleviate adverse selection.

Although we systematically examine heterogeneous treatment effects and find some assignment rules that outperform the current one, the rules are still bounded away from the upper bound that is the maximum output based on oracle assignment of JTPs. Future work will be examining the best assignment rules that may combine a few rules simultaneously as in Kitagawa and Tetenov (2018) and Athey and Wager (2018).

Chapter 5

Appendices

5.1 Log-likelihood function for our analysis

$$
\begin{aligned}
&log\ \mathcal{L}(\theta; G_i, Y_i^{obs}, l(X_i)) \\
&\quad \propto \sum_{i \in O(1,1)} \mathbb{1}(G_i = EE)\big[log(\pi_{EE,i}) - \frac{1}{2}log(\sigma^2_{EE,1}) - \frac{[log(Y_i^{obs}) - \alpha_{EE,1} - l(X_i)\beta_{EE,1}]^2}{2\sigma^2_{EE,1}}\big] \\
&\quad + \sum_{i \in O(1,1)} \mathbb{1}(G_i = EN)\big[log(\pi_{EN,i}) - \frac{1}{2}log(\sigma^2_{EN,1}) - \frac{[log(Y_i^{obs}) - \alpha_{EN,1} - l(X_i)\beta_{EN,1}]^2}{2\sigma^2_{EN,1}}\big] \\
&\quad + \sum_{i \in O(0,1)} \mathbb{1}(G_i = EE)\big[log(\pi_{EE,i}) - \frac{1}{2}log(\sigma^2_{EE,0}) - \frac{[log(Y_i^{obs}) - \alpha_{EE,0} - l(X_i)\beta_{EE,0}]^2}{2\sigma^2_{EE,0}}\big] \\
&\quad + \sum_{i \in O(0,1)} \mathbb{1}(G_i = NE)\big[log(\pi_{NE,i}) - \frac{1}{2}log(\sigma^2_{NE,0}) - \frac{[log(Y_i^{obs}) - \alpha_{NE,0} - l(X_i)\beta_{NE,0}]^2}{2\sigma^2_{NE,0}}\big] \\
&\quad + \sum_{i \in O(1,0)} [\mathbb{1}(G_i = NE)log(\pi_{NE,i}) + \mathbb{1}(G_i = NN)log(\pi_{NN,i})] \\
&\quad + \sum_{i \in O(0,0)} [\mathbb{1}(G_i = EN)log(\pi_{EN,i}) + \mathbb{1}(G_i = NN)log(\pi_{NN,i})]
\end{aligned}
$$

5.2 EM algorithm

The EM algorithm in the chapter 2 follows the subsequent steps.

1. E-step

In O(1,1),

$$\mathbb{P}(G_i = EE) = \frac{\pi_{EE,i}N(\mu_{EE,1}, \sigma^2_{EE,1})}{\pi_{EE,i}N(\mu_{EE,1}, \sigma^2_{EE,1}) + \pi_{EN,i}N(\mu_{EN,1}, \sigma^2_{EN,1})}$$

$$\mathbb{P}(G_i = EN) = \frac{\pi_{EN,i}N(\mu_{EN,1}, \sigma^2_{EN,1})}{\pi_{EE,i}N(\mu_{EE,1}, \sigma^2_{EE,1}) + \pi_{EN,i}N(\mu_{EN,1}, \sigma^2_{EN,1})}$$

In O(0,1),

$$\mathbb{P}(G_i = EE) = \frac{\pi_{EE,i}N(\mu_{EE,0}, \sigma^2_{EE,0})}{\pi_{EE,i}N(\mu_{EE,0}, \sigma^2_{EE,0}) + \pi_{NE,i}N(\mu_{NE,0}, \sigma^2_{NE,0})}$$

$$\mathbb{P}(G_i = NE) = \frac{\pi_{NE,i}N(\mu_{NE,0}, \sigma^2_{NE,0})}{\pi_{EE,i}N(\mu_{EE,0}, \sigma^2_{EE,0}) + \pi_{NE,i}N(\mu_{NE,0}, \sigma^2_{NE,0})}$$

In O(1,0),

$$\mathbb{P}(G_i = NE) = \frac{\pi_{NE,i}}{\pi_{NE,i} + \pi_{NN,i}}, \qquad \mathbb{P}(G_i = NN) = \frac{\pi_{NN,i}}{\pi_{NE,i} + \pi_{NN,i}}$$

In O(0,0),

$$\mathbb{P}(G_i = EN) = \frac{\pi_{EN,i}}{\pi_{EN,i} + \pi_{NN,i}}, \qquad \mathbb{P}(G_i = NN) = \frac{\pi_{NN,i}}{\pi_{EN,i} + \pi_{NN,i}}$$

2. M-step

$$
\begin{aligned}
& log\ \mathcal{L}(\theta; G_i, Y_i^{obs}, l(X_i)) \\
& \quad \propto \sum_{i \in O(1,1)} \mathbb{P}(G_i = EE)\big[log(\pi_{EE,i}) - \frac{1}{2} log(\sigma^2_{EE,1}) - \frac{[log(Y_i^{obs}) - \alpha_{EE,1} - l(X_i)\beta_{EE,1}]^2}{2\sigma^2_{EE,1}}\big] \\
& \quad + \sum_{i \in O(1,1)} \mathbb{P}(G_i = EN)\big[log(\pi_{EN,i}) - \frac{1}{2} log(\sigma^2_{EN,1}) - \frac{[log(Y_i^{obs}) - \alpha_{EN,1} - l(X_i)\beta_{EN,1}]^2}{2\sigma^2_{EN,1}}\big] \\
& \quad + \sum_{i \in O(0,1)} \mathbb{P}(G_i = EE)\big[log(\pi_{EE,i}) - \frac{1}{2} log(\sigma^2_{EE,0}) - \frac{[log(Y_i^{obs}) - \alpha_{EE,0} - l(X_i)\beta_{EE,0}]^2}{2\sigma^2_{EE,0}}\big] \\
& \quad + \sum_{i \in O(0,1)} \mathbb{P}(G_i = NE)\big[log(\pi_{NE,i}) - \frac{1}{2} log(\sigma^2_{NE,0}) - \frac{[log(Y_i^{obs}) - \alpha_{NE,0} - l(X_i)\beta_{NE,0}]^2}{2\sigma^2_{NE,0}}\big] \\
& \quad + \sum_{i \in O(1,0)} [\mathbb{P}(G_i = NE) log(\pi_{NE,i}) + \mathbb{P}(G_i = NN) log(\pi_{NN,i})] \\
& \quad + \sum_{i \in O(0,0)} [\mathbb{P}(G_i = EN) log(\pi_{EN,i}) + \mathbb{P}(G_i = NN) log(\pi_{NN,i})]
\end{aligned}
$$

5.3 The full list of covariates

1. Continuous variables

age (years), tenure (years), inactive duration (weeks), resume completion rate (%)

2. Binary variables

gender, disability status, work experience in a given industry,

3. Categorical variables

Educational Attainment

none, elementary school diploma, middle school diploma, high school diploma, 2-year college diploma, bachelor's degree, graduates or more

Residential Region

Seoul, Busan, Daegu, Incheon, Gwangju, Daejeon, Ulsan, Sejong, Gyeong-gi, Gang-won, Chung-buk, Chung-nam, Jeon-buk, Jeon-nam, Gyeong-buk, Gyeong-nam, Je-ju

Workplace Region

Seoul, Busan, Daegu, Incheon, Gwangju, Daejeon, Ulsan, Sejong, Gyeong-gi, Gang-won, Chung-buk, Chung-nam, Jeon-buk, Jeon-nam, Gyeong-buk, Gyeong-nam, Je-ju

Industry

agriculture, manufacturing, energy, construction, wholesale/retail, transportation, lodging/restaurant, publishing/broadcasting, finance/insurance, real estate, tech. service, facility management, public admin., educational service, social service, leisure related, private service

Ex-firm size (persons)

<5, 5-9, 10-29, 30-49, 50-69, 70-99, 100-149, 150-299, 300-499, 500-999, $\geqslant 1000$

Ex-occupation

manager, professional/engineer, supervisor/clerk, merchandiser, operator, farmer/fishery, manual laborer

Reason for Separation

(Personal) changing jobs, family issues, disease/injury, disciplinary dismissal, miscellaneous

(Firm-oriented) shutdown, layoff, exploitation, miscellaneous

(Employment Expiration) retirement, contract expiration, project completion

Month of Separation

January, February, March, April, May, June, July, August, September, October, November, December

Chapter 6

Bibliography

Ahn, H., Powell, J. L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1-2), 3-29.

Andini, M., Ciani, E., de Blasio, G., D'Ignazio, A., Salvestrini, V. (2018). Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior Organization*, 156 , 86–102.

Angrist, J. D., Imbens, G. W., Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434), 444-455.

Athey, S., (2015). Machine Learning and Causal Inference for Policy Evaluation. *In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 5–6. ACM.

Athey, S., Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 2016 Jul 5; 113(27): 7353–7360.

Athey, S., Imbens, G. W. (2017). The Econometrics of Randomized Experiments a. *In Handbook of Economic Field Experiments*, Vol. 1, pp. 73–140. Elsevier.

Athey, S., Wager, S. (2018). Efficient Policy Learning (No. 1702.02896).

Athey, S., Tibshirani, J., Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2), 1148-1178.

Belloni, A., Chernozhukov, V., Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2), 29-50.

Belloni, A., Chernozhukov, V., Fernández-Val, I., Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1), 233-298.

Bertrand, M., Crepon, B., Marguerie, A., Premand, P. (2017). Contemporaneous and Post-Program Impacts of a Public Works Program: Evidence from Côte d'Ivoire. (English). Washington, D.C.: *World Bank Group.*

Bia, M., Flores-Lagunes, A., Mercatanti, A. (2018). Evaluation of Language Training Programs using Principal Stratication: The Case of Luxembourg. Available at: ftp.iza.org/dp11973.pdf

Blanco, G., Flores, C. A., Flores-Lagunes, A. (2013). The effects of Job Corps training on wages of adolescents and young adults. *American Economic Review*, 103(3), 418-22.

Blundell, R., Gosling, A., Ichimura, H., Meghir, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, 75(2), 323-363.

Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A. (1984). Classification and regression trees. *CRC press.*

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Casey, K., Glennerster, R., Miguel, E. (2012). Reshaping institutions: Evidence on aid impacts using a preanalysis plan. The Quarterly *Journal of Economics*, 127(4), 1755-1812.

Chae, C., Kim, M. (2004). Analysis the Effectiveness of the Job Training Programs for the

Unemployed, *Korea Research Institute Vocational Education and Training* (In Korean)

Chalfin, A. , Danieli, O. , Hillis, A. , Jelveh, Z. , Luca, M. , Ludwig, J. , Mullainathan, S. , 2016. Productivity and selection of human capital with machine learning. *American Economic Review* 106 (5), 124–127 .

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. (2017). Double/Debiased/Neyman machine learning of treatment effects. *American Economic Review Papers and Proceedings*, 107 (5), 261–265.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J. (2018). Double/Debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21 (1), C1-C68.

Choe, C., Flores-Lagunes, A., Lee, S. J. (2015). Do dropouts with longer training exposure benefit from training programs? Korean evidence employing methods for continuous treatments. *Empirical Economics*, 48(2), 849-881.

Choi, H. J., Kim, J. (2012). Effects of public job training programmes on the employment outcome of displaced workers: results of a matching analysis, a fixed effects model and an instrumental variable approach using Korean data. *Pacific Economic Review*, 17(4), 559-581.

Chopra, A., Kang, K., Karasulu, M., Liang, H., Ma, H., and Richards, A., "From Crisis to Recovery in Korea: Strategy, Achievements, and Lessons", IMF Working paper, *International Monetary Fund*, 2001.

Das, M., Newey, W. K., Vella, F. (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies*, 70(1), 33-58.

Davis, J., Heller, S. B. (2017). Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs. *American Economic Review*, 107(5), 546-50.

Davis, J., Heller, S. B. (2017). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs (No. w23443). *National Bureau of Economic Research*.

Dehejia, R. H., Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, 84(1), 151-161.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1-38.

Frangakis, C. E., Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1), 21-29.

Frumento, P., Mealli, F., Pacini, B., Rubin, D. B. (2012). Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. *Journal of the American Statistical Association*, 107(498), 450-466.

Gerfin, M., Lechner, M. (2002). A microeconometric evaluation of the active labour market policy in Switzerland. *The Economic Journal*, 112(482), 854-893.

Hahn, P. R., Murray, J. S., Carvalho, C. M. (2017). Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects.

Hastie, T., Tibshirani, R., Friedman, J. (2009). The elements of statistical learning: prediction, inference and data mining. Springer-Verlag, New York.

Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161. doi:10.2307/1912352

Hwang, S., Flores-Lagunes, A., Choe, C., Lee, S. (2019). Estimating treatment effects of Korean job training programs using the administrative data.

Imai, K., Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1), 443-470.

Imbens, G. W., Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-475.

Imbens, G. W., Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. *Cambridge University Press.*

Kitagawa, T., Tetenov, A. (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2), 591-616.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., Mullainathan, S., 2018. Human decisions and machine predictions. *Quarterly Journal of Economics* 133 (1), 237–293 .

Knaus, M., Lechner, M., Strittmatter, A. (2017). Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach. *arXiv preprint* arXiv:1709.10279.

Knaus, M., Lechner, M., Strittmatter, A. (2018). Machine Learning Estimation of Heterogeneous Causal Effects: Empirical Monte Carlo Evidence.

Lechner, M., Melly, B. (2010). Partial identification of wage effects of training programs.

Lechner, M. (2019). Modified Causal Forests for Estimating Heterogeneous Causal Effects.

Lee, M. J., Lee, S. J. (2005). Analysis of job training effects on Korean women. *Journal of Applied Econometrics*, 20(4), 549-562.

Lee, M. J., Lee, S. J. (2009). Sensitivity analysis of job training effects on reemployment for Korean women. *Empirical Economics*, 36(1), 81-107.

McBride, L., Nichols, A. (2015). Improved poverty targeting through machine learning: An application to the USAID Poverty Assessment Tools. Retrieved April.

McCaffrey, D. F., Ridgeway, G., Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological* *methods*, 9(4), 403.

McLachlan, G.J.; Peel, D. (2000). Finite Mixture Models. *Wiley.* ISBN 978-0-471-00626-8.

Mercatanti, A., Li, F. (2017). Do debit cards decrease cash demand?: causal inference and sensitivity analysis using principal stratification. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(4), 759-776.

Ministry of Employment and Labor (2014), The 2013 Employment and Labor Policy in Korea, South Korea [whitepaper]

Ministry of Employment and Labor. (2018a). The Annual Labor Review in 2018, South Korea [whitepaper]

Ministry of Employment and Labor. (2018b). Overview of the department budget report in 2018, South Korea [white paper]

Mullainathan, S., Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.

Neyman, J. (1979). C(α) tests and their use. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 41(1/2), 1-21.

Nie, X., Wager, S. (2017). Quasi-oracle estimation of heterogeneous treatment effects.

OECD (2013). Back to Work: Korea: Improving the re-employment prospects of displaced workers, *OECD Publishing.* http://dx.doi.org/10.1787/9789264189225-en

Oprescu, M., Syrgkanis, V., Wu, Z. S. (2018). Orthogonal random forest for heterogeneous treatment effect estimation. *arXiv preprint* arXiv:1806.03467.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, 931-954.

Rosenbaum, P. R., Rubin, D. B. (1983a). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Rosenbaum, P. R., Rubin, D. B. (1983b). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, 212-218.

Rosenbaum, P. R., Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5), 688.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, 34-58.

Schwartz, S., Li, F., Reiter, J. P. (2012). Sensitivity analysis for unmeasured confounding in principal stratification settings with binary variables. *Statistics in medicine*, 31(10), 949-962.

Strittmatter, A. (2018). What is the Value Added by using Causal Machine Learning Methods in a Welfare Experiment Evaluation?. *arXiv preprint arXiv:1812.06533*.

Su, X., C.L, Tsai, H. Wang, D.M. Nickerson, B. Li (2009). Subgroup Analysis via Recursive Partitioning, *Journal of Machine Learning Research*, 10, 141-158.

Tian, L., A.A. Alizadeh, A.J. Gentles, R. Tibshirani (2014). A Simple Method for Estimating Interactions Between a Treatment and a Large Number of Covariates, *Journal of the American Statistical Association*, 109 (508), 1517-1532.

Wager, S., Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523), 1228–1242.

Van Ours, J. C. (2004). The locking-in effect of subsidized jobs. *Journal of Comparative Economics*, 32(1), 37-55.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3-28.

Yoo, G., Kang, C. (2010) The impacts of vocational training on earnings in Korea: evidence from the economically active population survey. *KDI Journal of Econonimic Policy* 32:29–53 (In Korean)

Yoo, G., Lee, C. (2008). The estimation of the effectiveness of vocational training for the unemployed. *Korean Journal of Labour Economics*, 31(1), 59-103. (In Korean)

Yoon, H., (2014). Policy suggestions of job training programs for the unemployed. *Korean Development Institutes Focus, 45, 1-8* (in Korean)

Yoon, K., Kim, W., Bae, K., Choi, M., Son, D., Woo, H. (2017). Training performance analysis of vocational training teachers. *Korea University of Technology and Education* (annual report, in Korean)

Zhang, J. L., Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics*, 28(4), 353-368.

Zhang, J. L., Rubin, D. B., Mealli, F. (2008). Evaluating the effects of job training programs on wages through principal stratification. *In Modelling and Evaluating Treatment Effects in Econometrics*, 117-145.

Zhang, J. L., Rubin, D. B., Mealli, F. (2009). Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association*, 104(485), 166-176.

VITA

Seokchae Hwang

Department of Economics	Email: shwang01@syr.edu
Syracuse University	Mobile: +1 202 888 5271
110 Eggers Hall	
Syracuse, NY 13244	

Education

Syracuse University	
Ph.D., Economics	June 2019 (expected)
M.A. in Public Administration	August 2016
Seoul National University	
B.S. in Civil, Urban, Geo-system Engineering	March 2006

Fields of Interest

Program Evaluation, Labor Economics, Applied Microeconomics, Machine Learning, Big Data

Working Papers

Estimating Average Treatment Effects of Korean Job Training Programs using Administrative Data
Better Assignment Rules using Heterogeneous Treatment Effects: The Case in South Korea

In Progress

Optimal Assignment Rules under effect heterogeneity: The Case in South Korea

Employment

Teaching Assistant	
Microeconomics for Business (Prof. D. Laing)	Fall 2015,
Mathematical Economics (Prof. A. Tobias)	Fall 2017; Spring 2018,19
Game Theory (Prof. K. Buzard)	Fall 2018
Intermediate Macroeconomics (Prof. P. Howe)	Spring - Fall 2016
Economics Ideas and Issues (Prof. M. Ali)	Spring 2017
Government Service	
Deputy Director, Ministry of Finance and Economy	Apr 2007 - Present

Awards and Honors

Best Graduate Poster Award, New England Statistical Society (NESS), NextGen: Data Science Day	Oct 2018
Syracuse University Graduate Assistantship	July 2015 - May 2019
Maxwell School Summer Fellowship	2016 - 2018
Fullbright Scholarship	July 2014 - Aug 2015
Seoul National University Scholarship	March 1998 - Feb 2004

Languages and Skills

English (Fluent), Korean (Native), French (Beginner)
Matlab, R, Stata, LaTeX, MS Office