OSIPP Discussion Paper : DP-2024-E-007

The Effect of Initial Job in Japanese Labor Market

November 8, 2024

Ayaka Nakamura[1]

Specially Appointed Researcher, Graduate School of Economics, Osaka University

Katsuya Takii[2]

Professor, Osaka School of International Public Policy, Osaka University

【JEL Classification】J62

【Keywords】Employment stability; Initial labor market condition; MTE; Assignment model

【Abstract】We investigate the effect of being a regular employee in a job which a worker takes immediately after graduation (the initial job), on subsequent job status. We construct an assignment model that can be estimated by the marginal treatment effect (MTE) framework; the model suggests that the region- and cohort-level unexpected shocks that influences the demand for full time-worker share is a valid instrument under some assumptions. Estimating the MTE, we find that the treatment effect of the initial job is heterogeneous among individuals: male workers who are less likely to obtain regular employment in the initial job enjoy benefits of stable employment; however, the regular initial job does not increase the probability of subsequent regular employment for male workers who are likely to obtain regular employment in the initial job.

1 nakamura.ayaka.econ@osaka-u.ac.jp

2 takii@osipp.osaka-u.ac.jp

The Effect of Initial Job in Japanese Labor Market*

Ayaka Nakamura† and Katsuya Takii‡

November, 2024

Abstract

We investigate the effect of being a regular employee in a job which a worker takes immediately after graduation (the initial job), on subsequent job status. We construct an assignment model that can be estimated by the marginal treatment effect (MTE) framework; the model suggests that the region- and cohort-level unexpected shocks that influences the demand for full time-worker share is a valid instrument under some assumptions. Estimating the MTE, we find that the treatment effect of the initial job is heterogeneous among individuals: male workers who are less likely to obtain regular employment in the initial job enjoy benefits of stable employment; however, the regular initial job does not increase the probability of subsequent regular employment for male workers who are likely to obtain regular employment in the initial job.

1 Introduction

Several studies have found that entering the labor market during a recession is detrimental to a variety of future outcomes, including employment rates and wages.*1 In Japan, studies such

*The authors would like to thank Nobuyoshi Kikuchi, Ayako Kondo, Fumio Ohtake, and Masaru Sasaki, and the participants of Eastern and Western Labor Economics Conference, Kansai Macroeconomics Workshop, Search Theory Workshop, and the workshop on the economics of human resource allocation for useful comments. Financial support from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (Grants-in-Aid for Scientific Research Nos 18H03636 and 24K00265) and Research Grant from Japanese Center for Economic Research are gratefully acknowledged.

†Graduate School of Economics, Osaka University, Email: nakamura.ayaka.econ@osaka-u.ac.jp

‡Osaka School of International Public Policy (OSIPP), Osaka University, Email: takii@osipp.osaka-u.ac.jp

*1 The effects of labor market entry during recessions are wide-ranging and are found to have adverse effects not only on labor market outcomes, but also on family formation (Currie and Schwandt, 2014; Schwandt and as Ohtake and Inoki (1997) and Genda et al. (2010) show that labor market entrants during recessions face lower employment probabilities and wages.

In Japan, at least part of the effect is attributed to the failure to become a regular employee[*2] after working the first job post graduation (hereafter, the initial job). Some studies such as Sakai and Higuchi (2005) find a strong correlation between the employment status at the time of graduation and the subsequent employment status, after which Kondo (2007) and Hamaaki et al. (2013) show that the association can be interpreted as the causal effects of the initial jobs on the employment status by using instrumental variable estimation.[*3]

However, it is known that standard instrumental variable estimators clarify causal effects only for individuals whose treatment status is switched by an instrument.[*4] Therefore, these effects are generally not the same as those for individuals who would take treatment by an intervention or policy of interest.

This research examines the effect of being a regular employee in the initial job on subsequent job status for male workers by taking into account that it is likely to have unobserved treatment heterogeneity. To focus on unobserved heterogeneity in the effects, we construct a non-parametric model of an initial job assignment under a frictional labor market, the reduced form of which is shown to be equivalent to the generalized Roy model. As Heckman and Vytlacil (2007) reveal, the generalized Roy model non-parametrically identifies the marginal treatment effect (hereafter, MTE); our model can identify the MTE of being a regular employee in the initial job on subsequent job status. The MTE framework allows us to analyze the heterogeneity of treatments with respect to the evaluation that workers receive in their initial job using information from the self-selection of individuals. Moreover, it allows us to recover several interesting parameters of interest, in particular, which are the so-called ATE, ATT, and ATUT,[*5] under some assumptions.

Von Wachter, 2020), health (Maclean, 2015; Cutler et al., 2015; Schwandt and Von Wachter, 2020), and crime probability (Bell et al., 2018). The impacts, primarily in the U.S., are summarized in detail by Von Wachter (2020).

[*2]The regular employee of interest here refers to the position called *seisha-in* in Japan. In general, regular employees are 1. employed without a fixed-term contract, 2. working the standard number of hours set by the company (approximately 40 hours per week), and 3. directly employed by the company. Hereafter, we will refer to this position simply as "regular employee".

[*3]Esteban-Pretel et al. (2011) demonstrate that contingent jobs have a lasting impact on individual welfare. In this paper, they construct an on-and-off-the-job search model and conduct structural estimation along with simulations of career paths.

[*4]Innovative works by Imbens and Angrist (1994) and Heckman and Vytlacil (1999) provide an interpretation of the instrumental variable estimator as local average treatment effect.

[*5]The Average Treatment Effect (ATE) represents the average treatment effect for the entire sample, while the Average Treatment Effect on the Treated (ATT) refers to the average treatment effect for individuals who received

Our results find that the effect of being regular worker in the initial job is heterogeneous with respect to unobserved resistance to treatment. The impact of obtaining a regular initial job on securing subsequent regular employment is greater for male workers who have a higher resistance to treatment. As workers with high resistance to treatment are more likely to be evaluated as having lower unobserved abilities during the initial hiring process, this result suggests that securing a regular position as an initial job is more important for workers with lower unobserved abilities.

We recover ATE, ATT, and ATUT as population parameters from the estimated MTE. These population parameters also support our argument. The only parameter that is significantly positive is ATUT. As discussed below, ATUT represents the effect on workers who did not actually become regular employees but would have become regular employees if they had. Therefore, our results suggest that if male workers, who are less likely to be treated, are employed as regular workers in the initial job by chance, they would enjoy more benefit from stable employment. However, ATE and ATT are found to be insignificant. For workers who are average or likely to become regular employees in their initial job, becoming a regular employee in their initial job does not improve their employment stability. That is, their future probability of becoming a regular employee does not change regardless of whether their initial job is regular. Thus, if the goal is to stabilize employment, it is efficient to intervene, especially for workers who have difficulty becoming regular employees in their initial job.

Why do these results emerge? First, we conjecture that for workers who find it more difficult to secure a regular position in their initial job, it is likely that the search costs for finding a new regular job are high, leading them to remain in the regular position they obtained at their initial job. Additionally, in Japan, both institutional and customary factors are said to make it difficult to dismiss regular employees once they are hired. If so, it is expected that such workers benefit significantly from the employment stability gained by becoming regular employees at their initial job.

If this conjecture is correct, when workers who find it more difficult to secure a regular position in their initial job can obtain a regular job at a firm by chance, they are likely to stay in the firm.

the treatment. The Average Treatment Effect on the Untreated (ATUT) represents the average treatment effect for individuals who did not receive the treatment, had they received it. Details are provided in section 3.2.

To explore this hypothesis, we estimate the Marginal Treatment Effect (MTE) for the probability of staying with the initial employer. In this estimation, we find that only ATUT is positively significant, while ATE and ATT are not significant. This evidence supports our conjecture that men who find it difficult to become regular employees in their initial job face high search costs and/or benefit more from their regular position due to protections, such as dismissal regulations, that prevent easy termination.

The contributions of this paper are summarized in the following two aspects. First, we find the heterogeneity of treatment effects with respect to unobserved characteristics of being a regular employee on subsequent job status. Kondo (2007) shows that the failure at market entry has a negative effect on the current probability of regular employment. Hamaaki et al. (2013) show that the effect disappears around 10 years after graduation and that marginal effects are larger for high school graduates than for college graduates. These previous findings show the positive effects of the initial job status on the subsequent job status using the instrumental variable estimation. However, the instrumental variable estimation only reveals the effects for the people influenced by the instrument.[*6] We find that while the IV estimation of the initial job status on the subsequent job status is positive, as suggested by the extant literature, we cannot reject the hypotheses that the ATE and ATT are 0. That is, we cannot interpret the results of the IV estimation are applicable to average people. We rather argue that being regular workers at the initial job is important only for the workers who are less likely to become regular workers without a stroke of luck.

In addition to providing new evidence, our paper also makes a methodological contribution. Heckman (2010) argues that the generalized Roy model can connect structural and the program evaluation approaches. Following the spirit of Heckman (2010), we construct a non-parametric job assignment equilibrium model, the reduced form of which is equivalent to a generalized Roy model. This modeling strategy brings three advantages. First, we propose the candidates of theoretically consistent instruments, which are the residuals of the cohort- and region-level share of regular employees in the initial job after regressing it on local macroeconomic variables. Second, the model clarifies economic conditions under which the instruments satisfy exclusive

[*6] Kondo (2007) also notes that 2SLS estimates the effect on workers who respond well to labor market aggregate shocks and does not claim that there are no unobservable differences in aptitude between those who become regular employees in their initial job and those who do not.

restriction and monotonicity. Third, our theory provides a solid interpretation of the unobserved resistance to treatment in the generalized Roy model: it can be interpreted as the measure of the lack of unobserved ability evaluated by the selection process for having an initial job. These advantages make the assumptions behind our identification strategy transparent and economic interpretation of results easier.

The remaining sections of this paper are organized as follows. In Section 2, we describe the new graduate recruitment system in the Japanese labor market. In Section 3, the MTE framework is described, the parameters identified in this paper are defined, and specific estimation formulas are presented. We construct a non-parametric model of job assignment which can be estimated by the MTE framework, and define treatment, outcome and instrumental variables as well as the MTE in Section 4. In Section 5, we describe the data and explain how we construct our variables. Estimation results are explained in Section 6, where we examine the heterogeneity of the treatment effects by using the MTE framework. We also use additional estimation to interpret the mechanism of the main result. We conclude with a summary of the findings in Section 7.

2 Link Between Education and the Labor Market in Japan

In this section, we describe some of the unique features of the Japanese process of graduating from school and entering the labor market.

One of the characteristics of Japanese employment practices is the lump-sum hiring of new graduates, a system that differs between high school and college graduates.

For high school graduates, the rules and schedule for hiring are clearly defined by the government, major economic organizations, and schools. About 10 months prior to the start of employment, companies announce job offers to the government and high schools. Based on the job postings and the explanations given by the companies when they visit the high schools, high school students narrow down the companies to which they are interested in applying; they then, go through a selection process within the high schools to be screened by the companies and decide on a job about 6 months before they start working.[*7] In the job search for high school graduates,

[*7]The Ministry of Health, Labour, and Welfare has posted a detailed schedule on its website. For those entering the workforce in 2025, Public Employment Security Offices (commonly known as "Hello Work") will

the rule of one-employee-one-firm exists; high school students are not allowed to apply to more than one firm at a time, and if they are accepted by one firm, they must conclude their job search and join that firm.[*8] As for the arrangement between high schools and local governments to place final year students with regional companies, the quality of matching remains high and long-lasting (Genda and Kurosawa, 2001), while Ariga (2007) reports a decline in the quality of matching in the market for new high school graduates after the bursting of the bubble economy.

Although the schedule for university graduates is more relaxed than that for high school graduates, there is a fixed schedule for commencing recruitment activities, which varies from every year. University students generally begin gathering information by attending company information sessions and registering on recruitment websites in the latter half of their junior year.[*9] Firms publish job offers through their own websites, private paths such as recruiting portals, or through public institutions such as Hello Work, and select university students who apply through their own process. Students are allowed to apply to more than one firm, and the selection process begins about ten months prior to employment, with many university students completing their job search six months in advance.[*10]

A large percentage of Japanese firms hire new graduates who enter the company through

---

begin accepting job application forms on June 1, 2024, and companies will start applying to schools and visiting schools on July 1, 2024. On September 5, schools will begin submitting student applications to companies, and on September 16, companies will start screening and making offers of employment.

[*8]This rule is designed to increase the job offer rate for high school students with little knowledge of the job market by having high schools actively intervene in the job search process. The Ministry of Health, Labor, and Welfare reports that the employment rate for high school graduates in 2022 is 99%. However, there are critics of this practice, and some prefectures allow students to apply to more than one company. For more information on the one-person, one-company system, see Tsukioka (2023).

[*9]The Ministry of Health, Labour, and Welfare and the Cabinet Secretariat have notified on their official websites the recruitment system and schedule for college graduates. The schedule for student employment and recruitment activities has been established annually through the following process until 2017: 1. the establishment of "Guidelines for Recruitment and Selection" by the Japan Business Federation, 2. the agreement by the Roundtable Conference on Employment Issues, and 3. requests to economic organizations and others by the Cabinet Secretariat, Ministry of Education, Culture, Sports, Science and Technology, Ministry of Health, Labour, and Welfare, and Ministry of Economy, Trade, and Industry and other relevant ministries and agencies. However, in October 2018, the Japan Federation of Economic Organizations announced its policy not to formulate "Guidelines on Recruitment and Selection" in the future. Since then, to allow students to engage in job-hunting activities with peace of mind while securing study time, and so on, the government has held a liaison meeting of relevant ministries and agencies regarding the schedule of job-hunting and recruitment activities every fiscal year, compiled the "Guidelines Regarding the Schedule of Job-Hunting and Recruitment Activities" for students in their second year of university and others in the relevant year, and determined the schedule of job-hunting and recruitment activities.

[*10]For those entering employment in 2025, publicity and explanation activities by firms will begin on March 1, 2024, and selection activities will begin on June 1 of the same year, with formal offers of employment being made on or after October 1 of the same year. However, students who are determined to have a high level of expertise and ability through internships that meet certain conditions will be able to move into the selection process before June. The percentage of job offers for those entering the workforce in 2024 has remained at 74.8% as of October 1, 2023, 86.0% as of December 1, 2023, and 91.6% as of February 1, 2024.

the process described above, rather than mid-career workers. According to the "Survey on Employment Trends (2019-2022)" by the Ministry of Health, Labour and Welfare, although there is a declining trend, new graduates account for about 21% of the annual number of new hires among general workers*11 overall and is the largest group among college graduates from firms with more than 1,000 employees, amounting to about 33%.*12 This is consistent with the practice of Japanese firms, which have long emphasized training within the firm and hiring people who have accumulated firm-specific human capital for the long term or life. As a result, Japan's labor market tends to be less fluid, and tenure tends to be longer than in other countries. By comparing Japan to European countries, Shikata (2012) argues that Japan has the lowest transitions from non-regular to regular workers, at 30% annually for male workers.*13

In this environment, we can expect the initial job to be important. As firms allocate a larger hiring lot to new graduates, it becomes more difficult for workers who initially could not be regular employees to later obtain regular employment after a certain period following graduation. This is because the cost of catching up with regular employees of the same cohort who have received training is higher.*14

In addition, restrictions on the dismissal of regular employees in Japan tend to be stricter than in other countries (Shikata, 2012), and and once a worker is hired as a regular employee, there must be a socially acceptable reason for dismissing them (Takahashi, 2011).

If these arguments are correct, we can expect that the initial job is likely to be more significant for workers with low ability than for the average worker. However, notably, it is difficult to identify such abilities from observable attributes. Investigating the impacts of initial jobs on the subsequent employment status of workers with different unobserved abilities is the main purpose of this paper. Hence, we believe that our analysis can provide new facts in considering the degree of ease of dismissal in the Japanese labor market.

*11 It refers to workers who are "regular employees," excluding part-time workers.

*12 See Nagano (2012) for the transition from 1990-2010.

*13 Here, non-regular in Europe means temporary worker, which includes fixed-term employment contracts and temporary workers. He also notes a large gender gap existing in Japan, especially in terms of transitions to regular employment within the same company, which are considerably lower for female workers.

*14 See Arellano-Bover (2022) for the impact of recession entry on workers' long-term skill development.

3 MTE Framework

In this section, we explain the MTE framework. First, we briefly review the framework for causal inference. Second, we define parameters of interest estimated in this paper after introducing the generalized Roy model, and discuss why these parameters are important.

3.1 Rubin Causal Model

In this section, we briefly review the framework for causal inference originally proposed by Rubin (1974). This is a model that allows us to clearly define what we mean by causal effects by accepting the concept of potential outcomes and an assignment mechanism. For a simplest case, the observed outcome Y is expressed as follows.

$$Y = Y(0) + D(Y(1) - Y(0)),$$

where $D \in \{0, 1\}$ is a treatment variable, which indicates that the individual receives treatment when D takes the value of 1, and that the individual does not receive treatment when D takes the value of 0. $Y(1)$ and $Y(0)$ are potential outcomes. To clarify the relationship between this causal model and an economic structural model, we explicitly express them as follows:

$$Y(s) = F(\mathbf{X}, \varepsilon(s), s), s \in \{0, 1\},$$

where $\mathbf{X}$ is predetermined observable vectors, $\epsilon(s)$ is an unobservable variable and s is a treatment status.

Using potential outcomes, we can define individual-level treatment effects as follows:

$$\Delta \equiv Y(1) - Y(0).$$

Unfortunately, we cannot identify this individual-level treatment effect from the observed data because we can only observe either $Y(1)$ or $Y(0)$ for the same individual. This is called the fundamental problem of causal inference.

Instead, we consider the average treatment effect, $E[\Delta|\mathbf{X}]$, on the subsample with characteristics $\mathbf{X}$. This can be identified from the data with the following equation:

$$E[\Delta|\mathbf{X}] = E[Y|\mathbf{X}, D = 1] - E[Y|\mathbf{X}, D = 0]$$

if an assignment mechanism of D leads to the following condition:

$$(\varepsilon(1), \varepsilon(0)) \perp D|\mathbf{X}.$$

Again, unfortunately, the latter condition is rarely met in survey data. This is because in most situations, individuals can choose whether to receive treatment according to their own benefits, unless a randomized controlled trial is conducted.

To address this issue, the Roy model explicitly models a treatment assignment mechanism as a result of self-selection.

3.2 Generalized Roy Model and the MTEs

In this section, we describe the generalized Roy model, which adds a selection mechanism to the causal model described in the previous section, and define the parameters that can be identified by it: the MTEs. We also explicitly define the MTE and other parameters of interest that can be recovered from it.

The generalized Roy model assumes the following treatment assignment mechanism:

$$D = I(p(\mathbf{X}, Z) \geq V),$$

where V is uniformly distributed in [0,1], p is the propensity score determined by the characteristics, $\mathbf{X}$, and the value of the instrumental variable, Z. The propensity score, p, represents the probability of being assigned to a treatment. The random variable, V, is commonly referred to as “unobserved resistance to treatment” and is a one-dimensional measure of the extent to which an individual is resistant to treatment. This assignment mechanism implies that an individual is assigned treatment only when their propensity for receiving treatment exceeds their resistance to treatment.

Now we define the MTE as follows (Björklund and Moffitt, 1987; Heckman and Vytlacil, 1999):

$$E[\Delta|\mathbf{X}, V].$$

This is the average treatment effect for individuals with characteristics $\mathbf{X}$ and unobserved resistance to treatment V. As all of the unobserved components are assumed to be summarized by the one-dimensional variable V, identifying the MTE allows us to investigate the treatment effects on the subsample with observed characteristics $\mathbf{X}$ and unobserved characteristics V. Therefore, the MTE allows us to examine treatment effect heterogeneity not only in terms of observed characteristics but also unobserved characteristics.

Heckman and Vytlacil (1999, 2001) show that, within the framework of the generalized Roy model, the local instrumental variable (LIV) identifies the MTE for all p over the common support of $p(\mathbf{X}, Z)$. More concretely, MTE can be identified from observable data by the following equation:

$$E[\Delta|\mathbf{X}, V = p] = \frac{\partial E[Y|\mathbf{X}, p(\mathbf{X}, Z) = p]}{\partial p},$$

where

$$\begin{aligned} E[Y|\mathbf{X}, p(\mathbf{X}, Z) = p] &= E[Y(0)|\mathbf{X}, p(\mathbf{X}, Z) = p] + E[\Delta D|\mathbf{X}, p(\mathbf{X}, Z) = p] \\ &= E[Y(0)|\mathbf{X}] + \int_0^p E[\Delta|\mathbf{X}, V]dV \end{aligned}$$

if

$$(\varepsilon(1), \varepsilon(0), V) \perp Z|\mathbf{X}. \tag{1}$$

As the propensity score is assumed to be influenced by $\mathbf{X}$ and Z, the MTE can also be interpreted as the average treatment effect of a subsample with characteristics $\mathbf{X}$ that is indifferent between receiving the treatment or not when assigned a certain level of $Z = z$.

The MTE itself may not be the parameter of interest, but because it is the treatment effects on the subsample with both observed characteristics and unobserved characteristics, summing MTE with a proper weight may recover our parameter of interest. In fact, Heckman and Vytlacil

(2005) shows that popular parameters of interest can be recovered from MTE.[*15]

Let us, first, define the conventional population-level treatment parameters, which are reported in this paper, and discuss the relation to the MTE.

The first is the average treatment effect (ATE), which is defined as:

$$\Delta^{ATE}(\mathbf{X}) = E[\Delta|\mathbf{X}].$$

ATE represents the average effect over the entire population if everybody were participating in the treatment, or if individuals in the population were randomly assigned to the treatment.

Second, we define the average treatment effect on treated (ATT):

$$\Delta^{ATT}(\mathbf{X}) = E[\Delta|\mathbf{X}, D = 1].$$

ATT represents the average effect for individuals who are currently participating in the treatment.

Finally, we define the average treatment effect on untreated (ATUT):

$$\Delta^{ATUT}(\mathbf{X}) = E[\Delta|\mathbf{X}, D = 0].$$

ATUT represents how the individuals who are currently not participating in the treatment would benefit from the treatment on average if they were participating in the treatment.

Heckman and Vytlacil (2005, 2007) show that treatment parameters defined above can be represented as weighted averages of the MTE:

$$\Delta^k = \int_0^1 \omega_k \Delta^{MTE}(\mathbf{X}, V)dV$$

where ω_k is an appropriate weight for constructing $k \in \{\text{ATE, ATT, ATUT}\}$,[*16] and it can be constructed from observable data.

If we fail to have full common support of p, the exact value for these treatment parameters

[*15]Note that the following treatment parameters would be the same if the treatment effect was homogeneous, or there was no selection into treatment based on gains. In these cases, these treatment parameters also be the same as the IV estimator, thus, as the LATE. In the case when individuals are sorted into treatment based on gains, these treatment parameters would differ each other.

[*16]See Heckman and Vytlacil (2005, 2007), or Cornelissen et al. (2016) for the exact weights.

cannot be estimated. Instead, following Carneiro et al. (2011), we approximate each treatment parameter by the weighted averages of the MTE over the common support:

$$\Delta^k = \int_{V_L}^{V_U} \omega_k \Delta^{MTE}(\mathbf{X}, V) dV$$

where V_L and V_U are the lower and upper bound of the common support, respectively.

4 Model

4.1 Roy Model in a Frictional Labor Market

We develop a non-parametric job assignment model within a frictional labor market for a specific position, which derive the generalized Roy model discussed above as a reduced form. Frictions can arise from factors such as search frictions, regulations such as minimum wage laws and dismissal restrictions, strong union bargaining power, efficiency wages, or other mechanisms. In all models, these frictions create a surplus that is shared between firms and workers. As a result, workers assigned to these positions can obtain rent from the match. We refer to the set of jobs that provide these rents as "good positions." We consider the assignment to a good position as the treatment and the assignment to a bad position—those jobs without such rents—as the control.

Due to the presence of friction, wages cannot equalize the demand and supply of labor. If unknown shocks occur in the supply of good positions, these ex post shocks influence labor demand, creating randomness in the labor market. The model clarifies the conditions under which these random shocks can serve as a source of instrumental variables for our estimation.

Location Choice: Assume that the agent is characterized by predetermined types $\mathbf{T} \in \mathbf{R}^{N_T}$. $B_l(\mathbf{S}^a_\tau)$ is the set of $\mathbf{T}$ with who chooses l as the optimal job-seeking location.

$$B_l(\mathbf{S}^a_\tau) = \left\{ \mathbf{T} | l = \arg\max_{\{\hat{l}\}_{\hat{l}=1}^{L}} W^e_{\hat{l}}(\mathbf{T}, \mathbf{S}^a_\tau) \right\},$$

where $W_l^e(\mathbf{T}, \mathbf{S}_\tau^a) \in \mathbf{R}$ is the expected value of searching for a job at location l and $\mathbf{S}_\tau^a = \left\{\{\mathbf{S}_{l,\tau}\}_{l=1}^L, \tau\right\}$, where $\mathbf{S}_{l,\tau} \in \mathbf{R}^{N_S}$ is the vector of location specific aggregate variables in year τ and τ is a year fixed effect. We assume that $\mathbf{S}_\tau^a$ contains all public information and $\mathbf{T}$ contains all private information that influences location decisions. After choosing location l, the agents decide whether they apply for a good position at location l.

Selection Probability: As we assume that there are frictions to obtain a good position, the wage cannot equate demand and supply of the good positions. Hence, we need to have a mechanism to select workers who are assigned to good positions. When the agent applies for a good position at location l in year τ, we assume that they can be appointed to the good position if

$$I\left[\eta\left(\mathbf{T}, \mathbf{S}_\tau^a, l\right) + \varepsilon_\tau \geq m_{l,\tau}\right] = 1,$$

where $\eta(\mathbf{T}, \mathbf{S}_\tau^a, l) \in \mathbf{R}$ is an evaluation function, $m_{l,\tau} \in \mathbf{R}$ is the minimum score to be appointed to the good position at the location l in year τ, $\varepsilon_\tau \in \mathbf{R}$ is the evaluation errors for the applicants in year τ, which captures the uncertainty under selection process.

Assume that ε_τ has a continuous distribution with $F_l(\varepsilon_\tau)$. Then, the probability of being appointed to a good position is

$$q\left(\mathbf{T}, \mathbf{S}_\tau^a, l, m_{l,\tau}\right) = 1 - F_l\left(m_{l,\tau} - \eta\left(\mathbf{T}, \mathbf{S}_\tau^a, l\right)\right).$$

We assume that workers cannot observe $m_{l,\tau}$ before making application decisions but infer it from the knowledge of the conditional distribution of $m_{l,\tau}$ given $\mathbf{S}_\tau^a$, $Q_{m|\mathbf{S}l}(m_{l,\tau}|\mathbf{S}_\tau^a, l)$. Hence, workers decide whether they apply for a good position based on the following expected probability to be appointed to the good position:

$$q^e\left(\mathbf{T}, \mathbf{S}_\tau^a, l\right) \equiv \int q\left(\mathbf{T}, \mathbf{S}_\tau^a, l, m_{l,\tau}\right) dQ_{m|\mathbf{S}l}\left(m_{l,\tau}|\mathbf{S}_\tau^a, l\right).$$

Application Decision: Both jobs within a good position and jobs within a bad position can be heterogeneous. We assume that the workers who are appointed to a good position are competitively assigned to a job based on their evaluation $\eta(\mathbf{T}, \mathbf{S}_\tau^a, l) + \varepsilon_\tau$. We also assume that if workers apply for a job within a bad position, they are competitively evaluated and assigned to the job. These assumptions imply that both the assignment of jobs within a good position and that within a bad position are determined by $\mathbf{T}$, $\mathbf{S}_{\tau}^{a}$ and the location specific factor l, with some noise. Therefore, the expected values of both a good position and a bad position for workers are expressed by the function of $(\mathbf{T}, \mathbf{S}_{\tau}^{a}, l)$. Let us denote $W_{l}^{1}(\mathbf{T}, \mathbf{S}_{\tau}^{a})$ as the value of applying to a good position; $W_{l}^{0}(\mathbf{T}, \mathbf{S}_{\tau}^{a})$ is the value of applying to a bad position (i.e., value as the control). Then we can express the value of choosing a location l, $W_{l}^{e}(\mathbf{T}, \mathbf{S}_{\tau}^{a})$ by[*17]

$$W_{l}^{e}(\mathbf{T}, \mathbf{S}_{\tau}^{a}) = \max\left\{W_{l}^{1}(\mathbf{T}, \mathbf{S}_{\tau}^{a}), W_{l}^{0}(\mathbf{T}, \mathbf{S}_{\tau}^{a})\right\}.$$

Define the set of types who apply to the good position by $\Omega_{l}(\mathbf{S}_{\tau}^{a})$:

$$\Omega_{l}(\mathbf{S}_{\tau}^{a}) = \left\{\mathbf{T} | W_{l}^{1}(\mathbf{T}, \mathbf{S}_{\tau}^{a}) \geq W_{l}^{0}(\mathbf{T}, \mathbf{S}_{\tau}^{a}), \mathbf{T} \in B_{l}(\mathbf{S}_{\tau}^{a})\right\}.$$

Then when a worker with $\mathbf{T} \in \Omega_{l}(\mathbf{S}_{\tau}^{a})$ applies to a good position, and that with $\mathbf{T} \notin \Omega_{l}(\mathbf{S}_{\tau}^{a})$ does not apply at location l.

Market Equilibrium: After applying to a good position, whether the workers are appointed to it depends on the value of $m_{l,\tau}$. We assume that $m_{l,\tau}$ can be determined by the demand and supply of the good position.

$$N_{l,\tau} = \int q(\mathbf{T}, \mathbf{S}_{\tau}^{a}, l, m_{l,\tau}) I[\mathbf{T} \in \Omega_{l}(\mathbf{S}_{\tau}^{a})] dQ_{T}(\mathbf{T}),$$

where $N_{l,\tau}$ is the number of the supply of a good position. The right hand side is the number of workers who are assigned to the good position. The above equation shows that the minimum score must be adjusted so that the number of the supply of good positions is equal to the number of workers who are assigned to the good positions.

Normalizing both sides of the equation by the number of workers who seek a job at location

[*17]To be specific, we can more explicitly write $W_{l}^{1}(\mathbf{T}, \mathbf{S}_{\tau}^{a})$ as follows:

$$\begin{aligned} W^{1}(\mathbf{T}, \mathbf{S}_{\tau}^{a}) \equiv\ & q^{e}(\mathbf{T}, \mathbf{S}_{\tau}^{a}, l)\left(\tilde{W}_{l}^{1}(\mathbf{T}, \mathbf{S}_{\tau}^{a}) - W_{l}^{0}(\mathbf{T}, \mathbf{S}_{\tau}^{a})\right) \\ & - C_{l}(\mathbf{T}, \mathbf{S}_{\tau}^{a}) + W^{0}(\mathbf{T}, \mathbf{S}_{\tau}^{a}) \end{aligned}$$

where $\tilde{W}_{l}^{1}(\mathbf{T}, \mathbf{S}_{\tau}^{a})$ is the value of having a good position and $C_{l}(\mathbf{T}, \mathbf{S}_{\tau}^{a})$ is the cost of applying to the good position.

l in year τ, $\#B_l(\mathbf{S}^a_\tau) = \int I[\mathbf{T} \in B_l(\mathbf{S}^a_\tau)]\, dQ_T(\mathbf{T})$, the market equilibrium condition can be rewritten as

$$Z_{l,\tau} = \int q(\mathbf{T}, \mathbf{S}^a_\tau, l, m_{l,\tau})\, \omega(\mathbf{T}, \mathbf{S}^a_\tau, l)\, dQ_T(\mathbf{T}),$$

where $\omega(\mathbf{T}, \mathbf{S}^a_\tau, l) = \frac{I[\mathbf{T} \in \Omega_l(\mathbf{S}^a_\tau)]}{\int I[\mathbf{T} \in B_l(\mathbf{S}^a_\tau)]dQ_T(\mathbf{T})}$, and $Z_{l,\tau} = \frac{N_{l,\tau}}{\#B_l(\mathbf{S}^a_\tau)} \in \left[\underline{Z}_l(\mathbf{S}^a_\tau), \bar{Z}_l(\mathbf{S}^a_\tau)\right]$.

As F_l is a continuous distribution, an implicit function theorem suggests that there exists $\mu(Z_{l,\tau} : \mathbf{S}^a_\tau, l)$ such that

$$Z_{l,\tau} = \int p(\mathbf{T}, \mathbf{S}^a_\tau, l, \mu(Z_{l,\tau} : \mathbf{S}^a_\tau, l))\, \omega(\mathbf{T}, \mathbf{S}^a_\tau, l)\, dQ_T(\mathbf{T}) \tag{2}$$

and

$$\mu'(Z_{l,\tau} : \mathbf{S}^a_\tau, l) < 0.$$

That is, an increase in the number of the supply of good positions reduces the minimum score to be appointed to the good positions.

Treatment Variable, Instrument and Propensity Score: Let us assume that $\mathbf{T} = (\mathbf{O}, \mathbf{U})$ where $\mathbf{O} \in \mathbf{R}^{N_O}$ and $\mathbf{U} \in \mathbf{R}^{N_\mathbf{U}}$ are observable and unobservable individual characteristics vectors, respectively. Using the function μ defined by Equation (2), we can demonstrate that our treatment variable, D, which represents the assignment to a good position, is a function of $Z_{l,\tau}$:

$$\begin{aligned} D_\tau &\equiv D(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, Z_{l,\tau}), \\ &\equiv I[\eta(\mathbf{U}, \mathbf{X}_\tau) + \varepsilon_\tau \geq \mu(Z_{l,\tau} : \mathbf{S}^a_\tau, l)]\, I[(\mathbf{O}, \mathbf{U}) \in \Omega_l(\mathbf{S}^a_\tau)]. \end{aligned}$$

where $\mathbf{X}_\tau = (\mathbf{O}, \mathbf{S}^a_\tau, l)$ is the vector of observable variables. If an agent who chooses location l to search the job in year τ receives $D(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, Z_{l,\tau}) = 1$, they are assigned to the good position and if $D(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, Z_{l,\tau}) = 0$, they are not assigned[*18].

[*18]If the supply of good positions is large enough so that all applied workers can be appointed, $\underline{Z}_l(\mathbf{S}^a_\tau) \geq 1$, $I\left[\eta(\mathbf{U}, \mathbf{X}_\tau) + \varepsilon_\tau \geq \mu\left(Z_{l,\tau} : \mathbf{S}^a_\tau, l\right)\right] = 1$ is satisfied for all workers. In this case, there is no friction, and therefore,

$$D\left(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, Z_{l,\tau}\right) = I\left[W^1_l(\mathbf{T}, \mathbf{S}^a_\tau) \geq W^0_l(\mathbf{T}, \mathbf{S}^a_\tau), \mathbf{T} \in B_l(\mathbf{S}^a_\tau)\right].$$

This is equivalent to a standard Roy model. That is, our model extends the standard Roy model into a frictional labor market.

Note that because the worker's location choice depends on $\mathbf{S}_\tau$, but not on $Z_{l,\tau}$,

$$D\left(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, Z_{l,\tau} = z\right) \perp Z_{l,\tau}|\mathbf{X}_\tau$$

where $z_{l,\tau}$ is a realization of $Z_{l,\tau}$. That is, $Z_{l,\tau}$ satisfies a part of the conditions for exclusive restriction. Moreover, for all $z' > z$,

$$D\left(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, Z_{l,\tau} = z'\right) \geq D\left(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, Z_{l,\tau} = z\right), \forall\left(\varepsilon_\tau, \mathbf{U}\right)$$

because $\mu'\left(Z_{l,\tau} : \mathbf{S}^a_\tau, l\right) < 0$. That is, $Z_{l,\tau}$ satisfies the monotonicity conditions discussed in Imbens and Angrist (1994). These properties suggest that $Z_{l,\tau}$ can be a candidate for an appropriate instrument.

Following the arguments in Vytlacil (2002), our appendix proves the following theorem.

Theorem 1 *There exists a random variable* $V_\tau \in [0,1]$ *such that*

$$D\left(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, Z_{l,\tau}\right) = I\left[P\left(\mathbf{X}_\tau, Z_{l,\tau}\right) \geq V_\tau\right]$$

where

$$P\left(\mathbf{X}_\tau, Z_{l,\tau}\right) = \int p\left[\mathbf{U}, \mathbf{X}_\tau, \mu\left(Z_{l,\tau} : \mathbf{S}^a_\tau, l\right)\right] \omega\left(\mathbf{U}, \mathbf{X}_\tau\right) dQ_{\mathbf{U}}\left(\mathbf{U}|\mathbf{O}\right)$$

and

$$V_\tau \perp Z_{l,\tau}|\mathbf{X}_\tau$$

The theorem shows that although the assignment to the good position is influenced by several unobserved variables, comparing the propensity score, $P\left(\mathbf{X}_\tau, Z_{l,\tau}\right)$, and unobserved resistance to treatment, V_τ, is sufficient to determine a treatment decision. If the propensity score is greater than the resistance to the treatment, workers are assigned to a good position; otherwise, they are not. This property is utilized by identifying the MTE.

Different from the standard generalized Roy model, the resistance to treatment, V_τ, is an endogenous variable. Hence, the interpretation of V_τ is less clear than usual. To help us to interpret V_τ, we provide the following theorem.

Theorem 2 *Suppose that for any* $\mathbf{X}_\tau$*, there exist functions* $\Upsilon : \mathbf{R}^{N_\mathbf{U}} \to \mathbf{R}$ *and* $\tilde{\eta} : \mathbf{R} \to \mathbf{R}$ *such that 1.* $\tilde{U} = \Upsilon(\mathbf{U} : \mathbf{X}_\tau)$ *and* $\tilde{\eta}(\Upsilon(\mathbf{U} : \mathbf{X}_\tau), \mathbf{X}_\tau) \equiv \eta(\mathbf{U}, \mathbf{X}_\tau)$ *and 2.* $\tilde{\eta}\left(\tilde{U}, \mathbf{X}_\tau\right)$ *is increase in* $\tilde{U} \in \left\{\tilde{U} | \tilde{U} = \Upsilon(\mathbf{U} : \mathbf{X}_\tau),\ (\mathbf{O}, \mathbf{U}) \in B_l(\mathbf{S}^a_\tau)\right\}$*. Then for any nondecreasing function* $G\left(\tilde{U}\right)$*,*

$$E_{\tilde{U}}\left[G\left(\tilde{U}\right) | \mathbf{X}_\tau, V_\tau = \tilde{v}\right]$$

is nonincreasing in $\tilde{v}$*.*

This theorem implies that if the evaluation function $\tilde{\eta}$ orders the unobservable characteristics vector $\mathbf{U}$ by the one-dimensional variable $\tilde{U}$, V_τ has a negative correlation with $\tilde{U}$. As the appointment to a good position is influenced by the realization of evaluation error, ε_τ, the relationship between V_τ and $\tilde{U}$ is influenced by luck. However, the negative correlation implies that a person with large V_τ is a person whose unobserved characteristics are less likely to be evaluated as competent by firms.

4.2 Outcome

In this section, we model potential outcomes and clarify the conditions under which the candidate instruments satisfy exclusive restriction. Suppose that $t \geq \tau$ is a current year, and $Y_t(s_\tau)$ is a potential outcome of individual workers in year t where $s_\tau \in \{0, 1\}$ is the initial employment status in year τ. Using a switching equation, we can express the observed outcome by

$$Y_t = Y_t(0) + (Y_t(1) - Y_t(0)) D_\tau.$$

We assume that the potential outcome $Y_t(s_\tau)$ is a function of s_τ and the history of the worker at year t, $\mathbf{H}_t$:

$$\begin{aligned} Y_t(s_\tau) &= \Theta(\mathbf{H}_t, s_\tau) \quad (3) \\ \mathbf{H}_t &= \left((\mathbf{S}^a_x, \mathbf{Z}_x, \upsilon_x)^{x=t}_{x=\tau+1}, \mathbf{T}, l\right) \ if\ t \geq \tau + 1 \\ &= \{\mathbf{T}, l\} \ if\ t \leq \tau \end{aligned}$$

where $\mathbf{H}_t$ contains the sequences of the vector of location specific aggregate variables $\{\mathbf{S}_x\}_{x=\tau+1}^{t}$ and the sequences of the vector of shares of good jobs $\{\mathbf{Z}_x\}_{x=\tau+1}^{t}$, where $\mathbf{Z}_x = \{Z_{l,x}\}_{l=1}^{L}$, the sequences of the vector of individual shocks from $\tau+1$ to t, $\upsilon_t \in \mathbf{R}^{N_\upsilon}$, the type of the agent, $\mathbf{T}$, and the location of the initial job search, l. These individual shocks, υ_t, include any shocks that influence human capital accumulation, network formation, changes in employment status and so on.

As our identifying assumption, we need the independence between $Z_{l,\tau}$ and $Y_t(s_\tau)$. To guarantee the independence, we make the following assumptions

$$\begin{aligned} Z_{l,t} &= Z_l\left(\mathbf{S}_t^a, \zeta_{l,t}\right), \forall l, t \qquad (4) \\ \mathbf{S}_{l,t+1} &= \Sigma_l\left(\{\mathbf{S}_{l,t}\}_l^L, \xi_{l,t}\right), \forall l, t \qquad (5) \end{aligned}$$

where $\zeta_{l,t} \in R$, $\xi_{l,t} \in R$ and

$$\left(\{\zeta_{l,t+s}\}_{s\neq 0}, \{\xi_{x,t+s}\}_{\forall x,s}\right) \perp \zeta_{l,t}, \forall t \qquad (6)$$

and υ_t is distributed with

$$F_\upsilon\left[\upsilon_t | \mathbf{S}_t^a, \mathbf{H}_{t-1}, s_\tau\right]$$

Given these assumptions,,we can easily see that

$$\left(Y_t(1), Y_t(0)\right) \perp Z_{l,\tau} | \mathbf{X}_\tau$$

The most important assumption is Equation (6). It guarantees that $\zeta_{l,\tau}$ is independent of $Y_t(s_\tau)$. When unknown shocks, $\zeta_{l,t}$, influence the amount of supply of good positions, the assignment of workers to the position can be influenced by some uncertainty. Equation (6) guarantees the condition under which this random variable $\zeta_{l,t}$ does not directly influence potential outcome.

Define $\epsilon_t(s_\tau)$ by $\epsilon_t(s_\tau) = Y_t(s_\tau) - E\left[Y_t(s_\tau)|\mathbf{X}_\tau\right]$. We can summarize the above argument.

Theorem 3 *Suppose that potential outcome satisfies Equation (3) and also suppose that Equa-*

tions (4), (5), and (6) are satisfied. Then

$$Y_t\left(s_\tau\right) = J\left(\mathbf{X}_\tau, s_\tau\right) + \epsilon_t\left(s_\tau\right),$$

where $J\left(\mathbf{X}_\tau, s_\tau\right) = E\left[Y_t\left(s_\tau\right)|\mathbf{X}_\tau\right]$ *and*

$$\begin{aligned}\left(\epsilon_t\left(1\right), \epsilon_t\left(0\right)\right) &\perp Z_{l,\tau}|\mathbf{X}_\tau, \\ E\left[\epsilon_t\left(1\right)|\mathbf{X}_\tau\right] &= E\left[\epsilon_t\left(0\right)|\mathbf{X}_\tau\right] = 0.\end{aligned}$$

4.3 Identification of MTE and Empirical Strategy

For our estimation, we choose current employment status as our outcome, becoming a regular worker at the initial job as a good position and becoming a non-regular worker at the initial job as a bad position. Then, we define the MTE of becoming a regular worker at the initial job by

$$MTE \equiv E\left[Y_t\left(1\right) - Y_t\left(0\right)|\mathbf{X}_\tau, V_\tau = p\right],$$

where $Y_t(s)$ is the current employment status when the initial job is s, where $s = 1$ means that the initial job is a regular worker and $s = 0$ means that the initial job is a non-regular worker.

As Theorems 1 and 3 satisfy the assumptions of the generalized Roy model in Heckman and Vytlacil (2005), as discussed in subsection 3.2, we can identify the MTE by

$$MTE = \frac{\partial E\left[Y_t|\mathbf{X}_\tau, P\left(\mathbf{X}_\tau, Z_{l,\tau}\right) = p\right]}{\partial p}$$

where

$$\begin{aligned}& E\left[Y_t|\mathbf{X}_\tau, P\left(\mathbf{X}_\tau, Z_{l,\tau}\right) = p\right] \\ =\ & J\left(\mathbf{X}_\tau, 0\right) + \left[J\left(\mathbf{X}_{l,\tau}, 1\right) - J\left(\mathbf{X}_\tau, 0\right)\right]p \\ & + \int_0^p E\left[\epsilon_t\left(1\right) - \epsilon_t\left(0\right)|\mathbf{X}_\tau, V\right]dV\end{aligned}$$

The estimation of the MTE requires the estimation of $E\left[Y_t|\mathbf{X}_\tau, P\left(\mathbf{X}_\tau, Z_{l,\tau}\right) = p\right]$. However, when the number of variables in $\mathbf{X}_\tau$ is large, it is practically infeasible to estimate it without additional assumptions. We assume two standard assumptions for our estimation. The first assumption is additive separability,

$$E\left[\epsilon_t(1) - \epsilon_t(0) | \mathbf{X}_\tau, V_\tau\right] = E\left[\epsilon_t(1) - \epsilon_t(0) | V_\tau\right] = \hat{K}(p),$$

and the second assumption is linearity,

$$\begin{aligned} J(\mathbf{X}_\tau, 0) &= \beta_0 \mathbf{X}_\tau, \\ J(\mathbf{X}_\tau, 1) - J(\mathbf{X}_\tau, 0) &= \beta_1 \mathbf{X}_\tau. \end{aligned}$$

With these two assumptions we can estimate $E\left[Y_t | \mathbf{X}_{l,\tau}, P(\mathbf{X}_\tau, Z_{l,\tau}) = p\right]$ by

$$E\left[Y_t | \mathbf{X}_\tau, P(\mathbf{X}_\tau, Z_{l,\tau}) = p\right] = \hat{\beta}_0 \mathbf{X}_\tau + \hat{\beta}_1 \mathbf{X}_\tau p + \hat{K}(p) \tag{7}$$

where $\hat{\beta}_0$ $\hat{\beta}_1$ and $\hat{K}$ are estimated value of β_0 β_1 and K.

This estimation requires the estimation of the propensity score. We assume that $P(\mathbf{X}_\tau, Z_{l,\tau})$ can be estimated by the following probit estimation

$$P(\mathbf{X}_\tau, Z_{l,\tau}) = \Phi(\Delta_{l,\tau})$$

where Φ is a standard normal distribution and

$$\Delta_{l,\tau} = \delta_0 + \delta_X \mathbf{O} + \delta_s \mathbf{S}_{l,\tau} + \delta_\zeta \hat{\zeta}_{l,t} + \delta_{\zeta X} \hat{\zeta}_{l,t} \mathbf{O} + \delta_{\zeta S} \hat{\zeta}_{l,t} \mathbf{S}_{l\tau} + \delta_l + \delta_\tau.$$

where δ_l and δ_τ are initial location effects and cohort effects, and instrumental variable $\hat{\zeta}_{l,t}$ can be estimated by residuals of the following macro regression.

$$Z_{lt} = \vartheta_0 + G(\mathbf{S}_{l,t}) + \vartheta_l + \vartheta_t + \zeta_{l,t}, \tag{8}$$

where $G(\mathbf{S}_{l,t})$ is a linear function comprising the first-order terms of each macro variable, interaction terms between two macro variables, and interaction terms involving all three macro variables.

5 Data

In this section, we explain the data used and how we construct variables for our estimation. We use the Japanese Panel Study of Employment Dynamics (JPSED) for 2018, which is conducted by Recruit Works Institute (Recruit), and provided by the Social Science Japan Data Archive, Center for Social Research and Data Archives, Institute of Social Science, The University of Tokyo.

The JPSED has been conducted through the Internet every year since 2016. The populations included in the JPSED are men and women aged 15 or older so that its assignment reflects demographics of the Labor Force Survey, which is conducted to understand the status of employment and non-employment in Japan by the Ministry of Internal Affairs and Communications Statistics Bureau. The age and occupational composition in the sample differ from those in the target population. To conduct a representative survey that reflects the population as much as possible, Recruit calculates the required number of respondents by age, gender, employment type, region, and educational background and creates an assignment by setting these numbers as a goal.[*19]

In the 2018 survey, a total of 50,677 valid responses are received. Of these respondents, 40,308 are the respondents since 2016 or 2017. The remaining 10,369 are new respondents who are part of 16,574 participants selected to fill the cells that are insufficient against the assignment. The respondent rates are 79.7% and 62.6% for the continuation and the new samples, respectively.

We use male observations with no more than 15 years of potential years of experience, aged between 18 and 45. This is because we would like to examine the effect of the initial job on a relatively earlier career. This kind of restriction is also employed by Schwandt and Von Wachter (2019).

To the best of our knowledge, the JPSED is the only survey in Japan that can identify the exact graduation year and when and where an individual has obtained an initial job. The extant literature mostly uses the exogenously calculated years from information of educational

[*19] Although the assignment is set to reflect the population, the non-labor population of teens and those aged 65 and over are allocated less than the actual number for analyzing workers in detail. We can obtain the true composition of the population by the weight provided by Recruit.

background as graduation year (e.g., "birth year +6+ years of education" is used as graduation year), and recognizes the graduation year as a year when an individual has obtained the initial job. If an individual delays the enrollment (i.e., redshirting), drops out of the school, repeats a year, or roams, this calculated year includes some noise. Moreover, the literature assumes that an individual obtains the initial job in the region they currently live. If an individual moves from the region during the year they have to obtain the initial job and the survey year; this assumption, however, must include some noise.

Figure 1 represents the transition between the initial and current residence. The vertical axis represents a region where an individual has lived when they has worked at first after graduation, that is, the initial region. The horizontal axis represents a current region. Regarding the Labor Force Survey regional aggregation, we divide 47 prefectures in Japan into 9 regions (see Appendix A for detailed division). The diagonal cells mean that an individual stays at the same region since they have worked there at first after graduation. We can see that over 60% of our sample stays at the same region. However, the remaining individuals have moved between regions from their initial jobs to the present, and it does not seem to be appropriate to recognize their current residence as the initial residence. We can overcome this bias by using the JPSED.

Next, we explain the definition and construction of variables used for estimation in the following paragraphs.

Outcome and Treatment Variables Here, we explain the definition of outcome and treatment variables. Outcome is an indicator that takes the value of one if an individual is currently employed as a regular worker and takes zero otherwise. Treatment is an indicator that takes the value of one if an individual has been employed as a regular worker in the initial job and zero otherwise.

Covariates and Macroeconomic Variables We include dummies of college graduates, dropout, left home, initial job in Tokyo, moved to Tokyo, cohort, and initial region, and local macroeconomic condition at a year prior to the cohort year into the regressions as covariates. College dummy takes the value of one if an individual graduates from a 4-year degree college or more and becomes zero otherwise. Dropout dummy indicates an individual has been dropped out from

Table 1: Descriptive Statistics

	Mean	St.d
Age	31.4883	4.5560
Potential years of experience	9.5750	3.6831
Regular employee (current)	.8131	.3899
Regular employee (initial)	.8709	.3353
Residual	-.0015	.0528
Covariates		
College graduates	.4844	.4998
Dropouts	.1120	.3153
Left home	.3234	.4678
Initial job in Tokyo	.1524	.3594
Moving to Tokyo	.0781	.2683
Region-level macroeconomic condition (normalized by year)		
Log of mean wage	.0856	.0891
Log of per capita GDP	.0981	.1046
Job-opening ratio	.0528	.1731
Observations		4725

The data come from the Japanese Panel Study of Employment Dynamics (JPSED) for 2018. The sample includes individuals with no more than 15 years of potential years of experience, aged between 18 and 45. 4,725 observations are used for the estimations. We have the data of per capita GDP up to 2014, so the observations entering the labor market after 2015 are dropped.

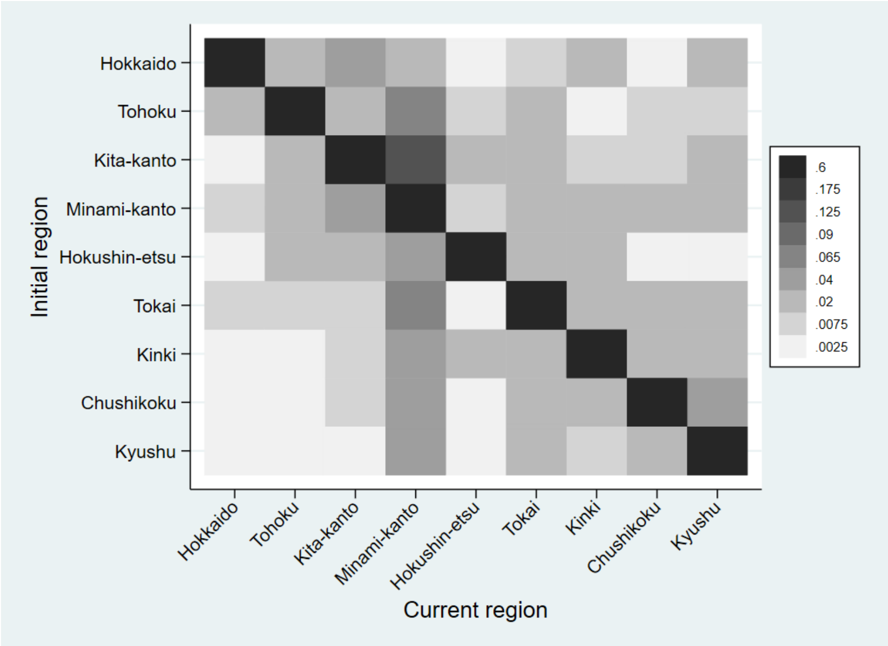


Figure 1: Transition matrix of the initial and current region

These matrices show the region in which individuals in the sample held their initial job and their current region. The vertical region represents the region in which the individual held his or her initial job, and the horizontal region represents the region in which the individual currently resides. The darker the color of a cell, the greater the share of individuals belonging to that cell. Diagonal cells indicate the share of individuals whose region of initial job and current region of residence are the same, while non-diagonal cells indicate the share of individuals whose two regions are different.

school at any one time. Left home dummy indicates an individual whose region of residence at age 15 differs from the region in which they obtain their initial job. Initial job in Tokyo dummy is one if an individual obtains their initial job in Tokyo. Moved to Tokyo dummy is an interaction term of left home dummy and initial job in Tokyo dummy, which indicates that an individual leaves their home and obtains an initial job in Tokyo. Dummies of left home, initial job in Tokyo, and moved to Tokyo are included to reflect some kind of individual's preference for employment. The independence of being able to obtain a job away from one's family and choosing Tokyo, where there are many job opportunities, are thought to correlate with the probability of becoming a regular employee in one's initial job and the probability of remaining as a regular employee thereafter. Cohort dummies represent a year when an individual is employed at first after graduation. Initial region dummies represent residences where an individual has lived when they have worked at first after graduation. As region-level macroeconomic conditions, we use the log of mean wage, log of per capita GDP, and job-opening ratio. These variables are normalized to mean zero by year.

Instrumental Variables Here, we explain the definition of instrumental variable and its construction. We use the residual of share of people who have obtained regular employment in the initial job. This share is constructed based on our sample. We first calculate the share by educational background (4-year degree college graduates or not), initial region, and the year when an individual obtains the initial job[*20]. For the calculated share to avoid including information about the individual, we calculate the share for samples other than the individual. Then, we regress these shares on macroeconomic conditions and cohort and initial region dummies and use these residuals as instruments, which correspond to ζ_t in our model. We call this instrument *Residual*. The interaction terms of the residual and covariates are also included as additional instruments. Table 1 contains the summary statistics of the sample used in the estimation.

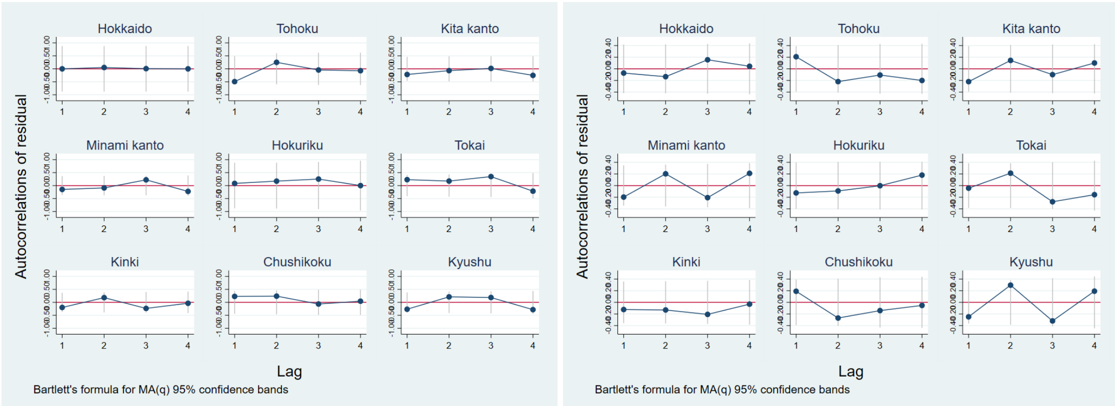
6 Estimation Results

In this section, we describe the estimation results. First, we discuss the absence of serial correlation within instrumental variables, which is a critical assumption for our instruments to be valid. Second, we conduct the standard two-stage least square by using these instrumental variables in order to replicate the results in the previous literature. Third, we show the results from first-stage estimation to measure the MTE, that is, the effect of our instruments on the probability of taking treatment. Finally, we show and discuss our main results from the MTE framework.

6.1 Serial Correlation within the Instrumental Variables

In this section, we discuss the absence of serial correlation within instrumental variables. As shown in Equations (4), (5) and (6), our instrumental variables—that is, the residuals of the share of regular employees in each education-region-cohort cell—have to be serially uncorrelated with each other. To confirm it, we plot the autocorrelations within the residuals as shown in Figure 2. We plot the autocorrelations by education and the initial region, because the instrumental variables are constructed for each education-region-cohort cell. Figure 2 represents

[*20] We eliminate observations if they are in the cell in which the number of observations is less than 10.



College graduates High school graduates

Figure 2: Serial Correlation within the instrumental variables

These figures show the results of testing of autocorrelation in the residuals obtained by equation (8) by education-state cell. The dots show the estimates and gray bars show Bartlett's formula for MA(q) 95% confidence bands.

the results for college graduates and high school graduates. Each dot represents the calculated autocorrelation functions, and each bar represents 95% confidence bands of the autocorrelation based on Bartlett's formula for MA(q) processes. In any education-region cell, we cannot reject the null hypothesis that the autocorrelation function equals zero. Therefore, our instruments seem to be valid in our sample.

6.2 Results from OLS and IV Estimations

Prior to examining the results of the MTE estimation, we first confirm that our data can replicate the findings of previous studies by analyzing the results of OLS and IV estimation.

The OLS and IV rows in Table 2 show the coefficients of the dummy variable indicating whether participant was a regular employee at their initial job when the current job status (regular or not) is regressed on the initial job status and individual characteristics. The OLS row shows the treatment effect obtained from OLS estimation, and the IV row presents the effect from the standard two-stage least square estimation.

In both rows, the coefficient on the dummy variable for regular initial employment is significantly positive, indicating that individuals whose initial job was regular employment tend to remain in regular employment. The values are 0.5831 for the OLS estimate and 0.5700 for the

Table 2: OLS and IV Estimation

	Coefficient of Regular Initial Job
OLS	.5831***
	(.0150)
IV	.5700***
	(.0808)
Observations	4725

These coefficients are the coefficients from regressing current job status on a dummy variable whose initial job was a regular job. The equation also includes individual characteristics and the macro variables of region at the time of employment as covariates. Ordinary least squares and two-stage least squares are used for obtaining OLS and IV estimates, respectively. The instrumental variable used to estimate the two-stage least square is $\zeta_{l,t}$ derived from our model.

IV estimate. These results are consistent with previous literature.

Although our instrument differs from those used in previous studies, the result is qualitatively consistent with the findings of Kondo (2007), with somewhat different coefficients. Kondo (2007) reports that individuals who are regular employees upon labor market entry are about 50% more likely to be regular employees later, and that the OLS and IV estimators are similar (and we cannot reject the null hypothesis that the two coefficient values are equal). Considering that Kondo (2007) and our analysis use different IVs as well as different types and periods of data, our results indicate that the findings of Kondo (2007) are robust regardless of the IVs and data periods.

However, notably, the results from IV estimation only indicate LATE. Hence, these results are only for the effect on workers who were able to become regular employees in their initial jobs due to unknown demand shocks. As MTE can reveal heterogeneity in effects across individuals, we would obtain deeper insights by restoring population parameters such as ATE and ATT. In the following section, we present results using MTE.

6.3 First Stage and Common Support

In this section, we show results from the first stage estimation and common support. We use the probit model for estimating the probability of being a regular employee in the initial job. Table 3 shows the results from the first stage regression. We see that the instruments (jointly) serve as predictors of obtaining a regular initial job, as well as college graduates and dropouts from

Table 3: First Stage Estimation

	Average derivative
Covariates	
College graduates	.0393***
	(.0098)
Dropouts	-.1361***
	(.0186)
Left home	.0505***
	(.0108)
Initial job in Tokyo	.0012
	(.0195)
Moving to Tokyo	-.0325
	(.0308)
Instruments	
Residual	.2912*
	(.1226)
Test for joint significance of instruments: p-value	.0000
Observations	4725

This table reports the average derivatives from a probit regression of being regular employee in the initial job (the dummy variable which takes one if an individual is a regular employee in the initial job) on the set of variables listed in the table and on the region of the initial job and cohort dummies, and the region-level macroeconomic conditions. Interaction terms of an instrument and each covariate listed in the table are also included as additional instruments. Standard errors are in parentheses. At the bottom of the table we present p-values for the test of joint significance of coefficients on the instruments.

school.

Figure 3 represents common support. Blue and white bars represent the density of the propensity scores predicted from the first stage estimation for the treatment and for control groups, respectively. Individuals in our sample have propensity scores above .26; thus, we fail to have full-interval common support, and we can only identify the MTE over the common support obtained from the data when we non-parametrically estimate $K(q)$. Along with it, notably, none of the treatment parameters, such as ATE, ATT and ATUT described in Section 4, can be exactly recovered from our results of MTE estimation. Instead, we approximate these parameters using the MTE estimations over the common support by weighting so that each weight over the common support adds to one as explained in Section 3.2.

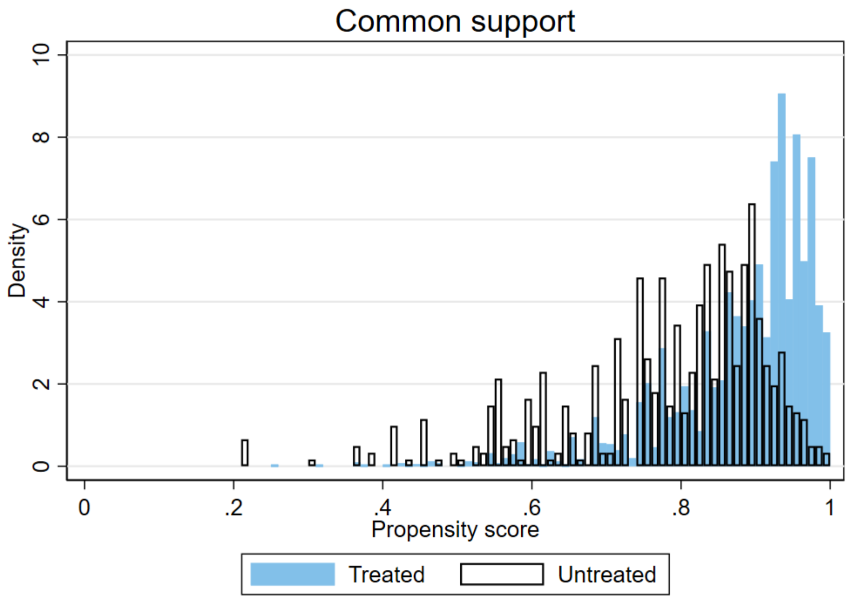


Figure 3: Common Support

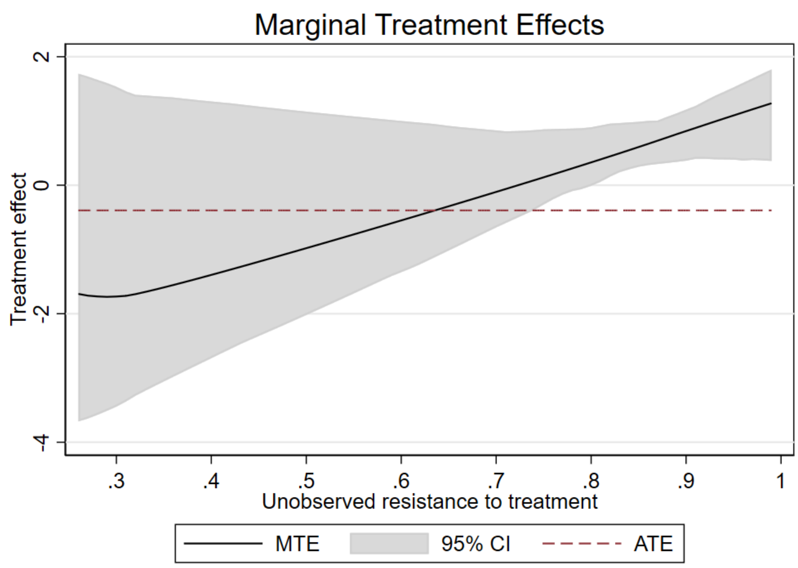
Notes: Propensity score is the estimated probability of being regular employee in the initial job. It is estimated from a probit regression of regular employee in the initial job on dummies of collage graduates, dropout, left home at the initial job, initial job in Tokyo, moving to Tokyo at the initial job, region-level macroeconomic conditions, cohort, and region of the initial job, and their interactions of instruments (see Table 3).

6.4 MTE and Other Treatment Parameters

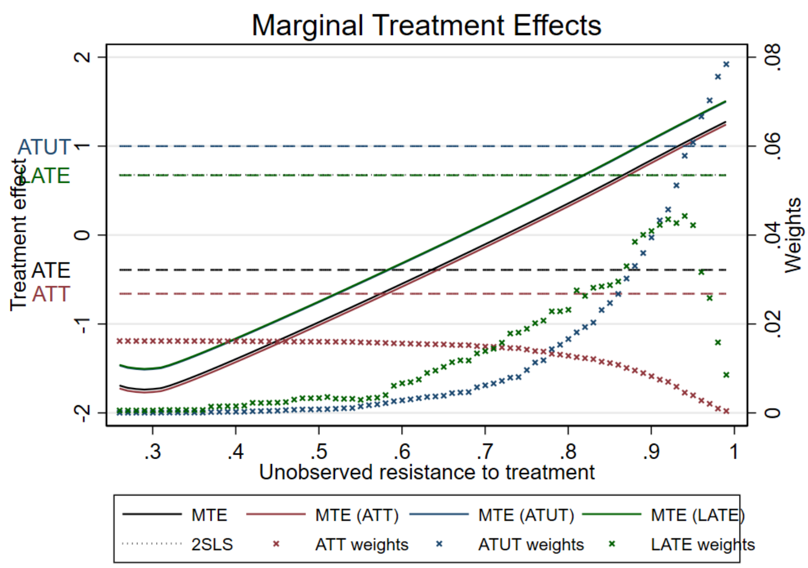
In this section, we show the results from the MTE framework. First, we show the plot of MTE over unobserved resistance to treatment. We also present the heterogeneity of treatment effects by showing popular treatment parameters.

Figure 4 shows the MTE. Panel A of Figure 4 represents the MTEs with respect to each point of resistance to the treatment, denoted by V_τ in Section 4. Therefore, the left side of the plot represents the MTEs for individuals who have relatively small resistance to the treatment, who in turn, are likely to be treated, and the right side of each plot represents the MTEs for individuals who are less likely to be treated.

As shown in Panel A of Figure 4, the MTE of obtaining a regular initial job on the probability of current regular employment is high for male workers with larger resistance to treatment. Thus, male workers are negatively sorted into treatment; that is, male workers who have benefited more from being regular employees in the initial job who are less likely to be employed as regular



Panel A: Marginal treatment effects and their confidence interval



Panel B: Treatment parameters and their weights

Figure 4: MTE Plots

Note: These plots represent the MTE of being regular employee in the initial job on subsequent probability of being regular employee. The upper plot represents parametrically estimated MTEs by solid line with respect to each points of the unobserved resistance to treatment, and 95% confidence interval by gray scale. The bottom plot represents treatment parameters from the MTE by horizontal dotted lines and weights for calculating each treatment parameters by dotted curves.

Table 4: Treatment Parameters

	Treatment effect
LATE	.6721***
	(.1952)
ATE	-.3922
	(.5535)
ATT	-.6598
	(.6819)
ATUT	.9993***
	(.2137)
t-test of ATT v.s. ATUT	.0000
Observations	4725

Note: Bootstrapped standard errors (200 replications) are in parentheses. These treatment parameters are recovered by the marginal treatment effects reported in Panel A of Figure 4. The estimate in LATE row corresponds to the two-stage least square estimator. The estimate in ATE row represents the average effects over whole population if everybody were participating in the treatment, or if individuals were randomly assigned to the treatment. The estimate in ATT row represents the average effects for individuals who are currently participating in the treatment. The estimate in ATUT row represents the average effects when individuals who are currently not participating in the treatment participate in the treatment.

workers.

Theorem 2 shows that V_τ is negatively correlated with $\tilde{U}$ which is a one-dimensional measure of unobserved ability evaluated during the hiring process of an initial job. Therefore, our results suggest that male workers whose unobserved abilities are evaluated as lower during the hiring process may benefit more from obtaining a regular worker position in their initial job.

The Panel B of Figure 4 comprises three components: treatment parameters, the MTE evaluated at mean covariates for constructing each treatment parameter, and the weights. Treatment parameters (i.e., ATE, ATT, ATUT and LATE)[*21] are denoted by a horizontal dotted line. The MTEs evaluated at mean covariates for constructing each treatment parameter are denoted by solid lines. These are the MTEs conditional on mean observed characteristics weighted by appropriate weights for constructing each treatment parameter.[*22] The weights for constructing treatment parameters are denoted by dotted curves, whose color corresponds to the color the of treatment parameters.

Table 4 shows the recovered treatment parameters based on the MTE and weights in the

[*21]Here, we refer to the parameter which corresponds to the two-stage least square (2SLS) estimator as LATE for convenience.

[*22]See Andresen (2018) for the exact way to calculate each treatment parameter and their weights.

right plot of Figure 4. The LATE row shows the standard 2SLS estimator recovered from the MTE. Consistent with the results in the previous section, becoming a regular employee in the initial job has a large impact on the current probability of being a regular employee.

Population parameters provide more insight into this impact. Table 4 shows that ATUT has the largest value and is significantly positive. The next largest value is ATE, followed by ATT . However, these two parameters are insignificant. As ATUT is a treatment effect for those who are not actually treated, it indicates that male workers who do not obtain regular employment in the initial job would have been more likely to be regular employment in the current job if they had obtained regular employment in the initial job. However, for those who are more likely to obtain the regular initial job or who are at mean, we cannot say that being a regular employee at the initial job increase the subsequent probability of being regular employee. These parameters indicate the same pattern of sorting as MTE plots—male workers are negatively sorted into regular employees in the initial job.

We observe the is heterogeneity of effects in male workers. The bottom row of Table 4 shows the p-values of the test that the values of ATT and ATUT are significantly different for male workers. As noted above, ATUT has the largest and most positive significant value for the probability of current regular employment, while ATE and ATT have negative, though not significant, point estimates. We can find the existence of a difference between the values of ATT and ATUT in male workers at the 1% level of significance.

This point has not been emphasized in previous studies. The LATE row of Table 4 is the 2SLS estimate recovered from the MTE. Its value is positively significant. In our model this value represents the effect on individuals who happen to be regular employees in their initial job due to a marginal increase in the unexpected local demand for regular employees. The results of the 2SLS estimation only picks up the effect on individuals whose probability of regular employment of the initial job is affected by luck, and our results show that this result does not hold for all individuals. Our results suggest that for individuals who are less likely to become a regular employee in their initial job, the initial job is important for employment stability; however, conversely, for individuals who are more likely to become a regular employee in their initial job, the initial job has a less important role.

The results discussed thus far are based on non-parametric estimation of $K(q)$ in Equation (7),

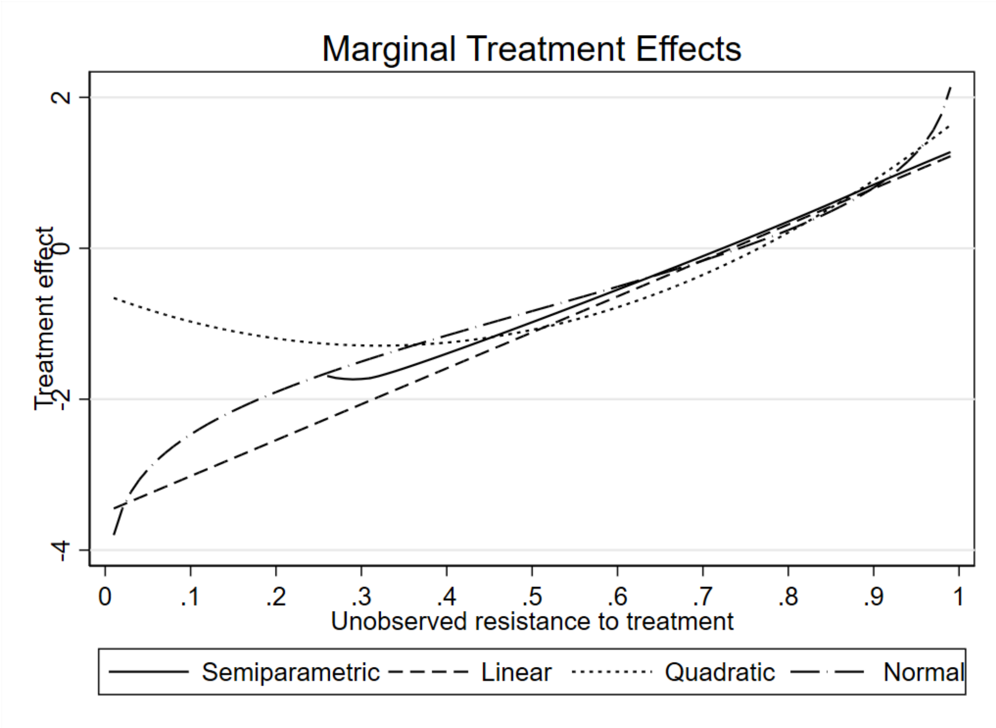


Figure 5: MTE with parametric $K(q)$

Notes: The figure displays MTE curves for the outcome of the probability of regular employee at the current job. The solid line refers to our baseline specification, where we estimate $K(q)$ nonparametrically, implying a semiparametric MTE. The figure also presents three additional MTE based on parametric approach: one curve obtained from jointly normally distributed assumption of $(\epsilon(0), \epsilon(1), V)$ and two curves based on specifications, in which the MTE curves are restricted to linear and quadratic.

but similar trends can be observed when we change the specification of $K(q)$. In Figure (5), the solid line represents our baseline estimation of semi-parametric MTE, that is, the non-parametric estimation of $K(q)$. We add three estimation results that impose constraints on the MTE curve. One curve is obtained from the jointly normally distributed assumption of $(\epsilon(0), \epsilon(1), V)$, and the other two curves are obtained from estimations that impose the assumption that the MTE is linear or quadratic.

Similar to the baseline estimates, we observe an upward slope in MTE in any specification, with male workers who are less likely to be regular employees in their initial job tending to receive more benefits. When parametric assumptions are imposed on $K(p)$, the estimates on the common support closely resemble the MTE estimates obtained through nonparametric estimation. However, notably, the MTE estimates outside the common support can vary significantly depending on how the assumptions are applied to $K(p)$. We contend that the MTE estimates outside the common support, when parametric assumptions are applied to $K(p)$, are unreliable. Therefore, we have decided to focus our discussion on the non-parametric estimates obtained on the common support, as these provide the most conservative insights regarding the treatment parameters.

6.5 Treatment Effect on the Initial Firm Dummy

We find that the MTE is increasing in p. Additionally, as shown in Table 5, we find that while we cannot reject the null hypothesis that ATT = 0, we can reject the null hypothesis that ATUT = 0. We also find that we can reject the null hypothesis that ATT = ATUT. The obvious question is what is the mechanism for it.

Theorem 2 demonstrates the inverse correlation between evaluations of unobserved characteristics at the initial job, $\tilde{U}$, and the resistance to treatment, V_τ, and Panel B in Figure 4 suggests that ATT are constructed with more weight on small V_τ and ATUT are constructed with more weight on large V_τ. Hence, workers who are more likely to become regular employees at their initial job tend to receive better evaluations from employers. Such workers are expected to be able to transition relatively easily to other regular employee positions, even if they do not become regular employees in their initial jobs. For these workers, whether their initial job is as a regular employee is not a significant concern in terms of their future prospects for becoming a regular employee; conversely, the benefits of securing a regular position in their initial job are likely to be minimal.

However, workers who find it challenging to become regular employees in their initial job tend to receive lower evaluations from employers. If they do not secure a regular position at their initial job, they face difficulties in finding new regular employee positions thereafter. For these workers, the benefits of having become a regular employee in their initial job are likely to be substantial.

This is because such workers are expected to face high search costs when looking for new regular employee positions. Additionally, in the Japanese labor market, whether institutionally or conventionally, it is often said to be difficult to dismiss regular employees once they are hired. Our findings are consistent with this opinion.

Table 5: Treatment parameters of Staying in the Initial Firm

	Treatment effect
LATE	.3507***
	(.1963)
ATE	.3106
	(.3256)
ATT	.3039
	(.4023)
ATUT	.3471*
	(.1963)
Observations	4725

Note: Bootstrapped standard errors (200 replications) are in parentheses. These treatment parameters are recovered by the marginal treatment effects from the regressing the dummy variable of staying the initial firm using the main specification. The estimate inLATE row corresponds to the two-stage least square estimator. The estimate in ATE row represents the average effects over whole population if everybody were participating in the treatment, or if individuals were randomly assigned to the treatment. The estimate in ATT row represents the average effects for individuals who are currently participating in the treatment. The estimates in ATUT row represents the average effects when individuals who are currently not participating in the treatment participate in the treatment.

If these arguments are correct, workers who received favorable evaluations in their initial job are likely to not have remained at that company, whereas workers who received poor evaluations are more likely to stay with the company.

To confirm whether this prediction is correct, we investigate an alternative outcome, that is, the initial firm dummy. This variable takes the value of one when the individual is currently affiliated with the same company as in their initial job, and zero otherwise.

Table 5 presents the treatment parameters restored based on the estimated MTE. First, the LATE row, which is the 2SLS estimation recovered by the MTE, is significant at approximately 0.35, indicating that being fortuitously employed as a regular employee in the initial job due to an unexpected demand shock increases the probability of remaining with the initial firm.

None of the parameters differ significantly from the LATE estimates, and all point estimates are positive. However, only ATUT takes on a significant value. This suggests that being a regular employee in the initial job increases the probability of remaining with the initial firm for male workers who find it difficult to secure regular employment in their initial job, while it does not impact the retention rate for other men.

These results are consistent with our argument that there are difficulties in dismissing regular employees once hired due to factors such as dismissal regulations or customary practices and/or that workers who find it difficult to secure regular employment in their initial job face high search costs in finding new regular positions.

7 Conclusion

This paper uses data on male workers in Japan to investigate the effect of being a regular employee in one's initial job on one's subsequent employment status. To investigate the heterogeneity in the treatment effects, we employ the MTE framework.

First, we construct a non-parametric model of initial job assignments, which can be estimated from the MTE framework. Our model proposes the unexpected demand shock is a candidate of instruments on which our identification strategy relies. Moreover, our model clarifies the conditions under which the instruments satisfy exclusion restriction and monotonicity. We confirm that the estimated IV is valid by checking the clarified condition.

From the MTE framework estimation, we find that heterogeneity exists across individuals in the treatment effect of having a regular initial job. We obtain the upward MTE curve, that is, being a regular employee in the initial job significantly increases the probability of being a regular employee in the subsequent job only for male workers who are less likely to be regular employees in the initial job. That is, they are negatively sorted into regular employment in the initial job. We argue that this pattern of sorting is likely to occur when a firm has a difficulty in dismissing workers with low ability and/or when such workers face high search costs. As it is relatively difficult to obtain regular employment, and because once a worker becomes a regular employee, they cannot be easily dismissed for various reasons, workers who have normally find it difficult to become regular employees seem to enjoy benefit from becoming regular employees if they are fortunate enough to become regular employees in their initial job.

It is the advantage of the MTE framework to be able to reveal these heterogeneities in the treatment effects. The extant literature has not prioritized this perspective. From the perspective of employment stability, this paper finds that we cannot apply previous findings to the entire economy. It may not be a big issue for competent workers to obtain a regular worker position as an initial job. The policy should focus support for workers who are less competent if employment stability is the main policy target. We hope that our analysis helps to understand the mechanism of employment and to politically intervene more effective manner.

References

Andresen, Martin Eckhoff, "Exploring marginal treatment effects: Flexible estimation using Stata," *The Stata Journal*, 2018, *18* (1), 118–158.

Arellano-Bover, Jaime, "The effect of labor market conditions at entry on workers' long-term skills," *Review of Economics and Statistics*, 2022, *104* (5), 1028–1045.

Ariga, Ken, "Kosotusya no Rodo Shhijyo (Market for High School Graduates in Japan)," *Keizai-Seido No Bunseki to Sekkei (Analyses and Design of Japanese Economic Institutions)*, 2007, pp. 227–263.

Bell, Brian, Anna Bindler, and Stephen Machin, "Crime scars: recessions and the making of career criminals," *Review of Economics and Statistics*, 2018, *100* (3), 392–404.

Björklund, Anders and Robert Moffitt, "The estimation of wage gains and welfare gains in self-selection models," *The Review of Economics and Statistics*, 1987, pp. 42–49.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil, "Estimating marginal returns to education," *American Economic Review*, 2011, *101* (6), 2754–81.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg, "From LATE to MTE: Alternative methods for the evaluation of policy interventions," *Labour Economics*, 2016, *41*, 47–60.

Currie, Janet and Hannes Schwandt, "Short-and long-term effects of unemployment on fertility," *Proceedings of the National Academy of Sciences*, 2014, *111* (41), 14734–14739.

Cutler, David M, Wei Huang, and Adriana Lleras-Muney, "When does education matter? The protective effect of education for cohorts graduating in bad times," *Social Science & Medicine*, 2015, *127*, 63–73.

Esteban-Pretel, Julen, Ryo Nakajima, and Ryuichi Tanaka, "Are contingent jobs dead ends or stepping stones to regular jobs? Evidence from a structural estimation," *Labour Economics*, 2011, *18* (4), 513–526.

Genda, Yuji and Masako Kurosawa, "Transition from school to work in Japan," *Journal of the Japanese and International Economies*, 2001, *15* (4), 465–488.

_ , Ayako Kondo, and Souichi Ohta, "Long-term effects of a recession at labor market entry in Japan and the United States," *Journal of Human resources*, 2010, *45* (1), 157–196.

Hamaaki, Junya, Masahiro Hori, Saeko Maeda, and Keiko Murata, "How does the first job matter for an individual's career life in Japan?," *Journal of the Japanese and International Economies*, 2013, *29*, 154–169.

Heckman, James J, "Building bridges between structural and program evaluation approaches to evaluating policy," *Journal of Economic literature*, 2010, *48* (2), 356–398.

_ and Edward J Vytlacil, "Local instrumental variables and latent variable models for identifying and bounding treatment effects," *Proceedings of the national Academy of Sciences*, 1999, *96* (8), 4730–4734.

_ and _ , "Instrumental variables, selection models, and tight bounds on the average treatment effect," in "Econometric Evaluation of Labour Market Policies," Springer, 2001, pp. 1–15.

_ and _ , "Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments," *Handbook of econometrics*, 2007, *6*, 4875–5143.

_ and Edward Vytlacil, "Structural equations, treatment effects, and econometric policy evaluation 1," *Econometrica*, 2005, *73* (3), 669–738.

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

Kondo, Ayako, "Does the first job really matter? regiondependency in employment status in Japan," *Journal of the Japanese and International Economies*, 2007, *21* (3), 379–402.

Maclean, Johanna Catherine, "The lasting effects of leaving school in an economic downturn on alcohol use," *ILR Review*, 2015, *68* (1), 120–152.

Nagano, Hitoshi, "Kigyou no Jinzai Saiyou no Doukou (Trends in Corporate Recruitment)," *JILPT*, 2012, *619*, 22–28.

Ohtake, Fumio and Takenori Inoki, "Rodo-shijo ni Okeru Sedai Koka (Cohort effects in the labor market)," *Gendai macro keizai bunseki—tenkanki no nihon keizai*, 1997, pp. 297–320.

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

Sakai, Tadashi and Yoshio Higuchi, "Freeter no Sonogo (future outcomes of freeters)," *Monthly Journal of Japan Institute of Labor*, 2005, *535*, 29–41.

Schwandt, Hannes and Till M Von Wachter, "Socioeconomic decline and death: Midlife impacts of graduating in a recession," Technical Report, National Bureau of Economic Research 2020.

_ **and Till Von Wachter**, "Unlucky Cohorts: Estimating the Long-term Effects of Entering the Labor Market in a Recession in Large Cross-sectional Data Sets," *Journal of Labor Economics*, 2019, *37* (S1), S161–S198.

Shikata, Masato, "Is temporary work " dead end " in Japan? Labor market regulation and transition to regular employment," *Japan Labor Review*, 2012, *9* (3), 59–79.

Takahashi, Kenji, "Study of Dismissals: A Critical Examination of Deregulation and Dismissal Jurisprudence (Kaiko no kenkyu: Kisei kanwa to kaiko houri no hihan teki kousatsu)," 2011.

Tsukioka, Tadashi, "Shinki Koukou Sotsugyousha no Roudou Shijou ni Okeru Shusyoku Kankou–Hitori Issha-sei to Fukusuu Oubo-sei no Hikaku (Job Search System in the labor market for New High School Graduates One-Person, One-Company System vs Multiple-Company Application System)," *Studies in commerce*, 2023, *59*, 159–178.

Vytlacil, Edward, "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 2002, *70* (1), 331–341.

Wachter, Till Von, "The persistent effects of initial labor market conditions for young adults and their sources," *Journal of Economic Perspectives*, 2020, *34* (4), 168–194.

Appendix A Division of Region

Hokkaido Hokkaido

Tohoku Aomori, Iwate, Miyagi, Akita, Yamagata

Kita Kanto Ibaraki, Tochigi, Gunma, Yamanashi, Nagano

Minami Kanto Saitama, Chiba, Tokyo, Kanagawa

Hokuriku Niigata, Toyama, Ishikawa, Fukui

Tokai Gifu, Shizuoka, Aichi, Mie

Kinki Shiga, Kyoto, Osaka, Hyogo, Nara, Wakayama

Chu-Shikoku Tottori, Shimane, Okayama, Hiroshima, Yamaguchi, Tokushima, Kagawa, Ehime, Kochi

Kyushu Fukuoka, Saga, Nagasaki, Kumamoto, Oita, Miyazaki, Kagoshima, Okinawa

Appendix B Proof of the Theorem in Our Model

The Proof of Theorem 1: Following the arguments in Vytlacil (2002) we can construct V_τ by the following procedure:

1. Define the set of Never taker, $NV(\mathbf{X}_\tau)$, Always taker, $AL(\mathbf{X}_\tau)$, and Compliers, $CO(\mathbf{X}_\tau)$,

$$NV(\mathbf{X}_\tau) = \left\{(\varepsilon_\tau, \mathbf{U}) : D\left(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, \bar{Z}(\mathbf{S}_\tau)\right) = 0\right\} \tag{A-1}$$

$$AL(\mathbf{X}_\tau) = \left\{(\varepsilon_\tau, \mathbf{U}) : D\left(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, \underline{Z}(\mathbf{S}_\tau)\right) = 1\right\} \tag{A-2}$$

$$CO(\mathbf{X}_\tau) = \left\{\begin{array}{c} (\varepsilon_\tau, \mathbf{U}) : \\ D\left(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, \bar{Z}(\mathbf{S}_\tau)\right) = 1 \\ D\left(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, \underline{Z}(\mathbf{S}_\tau)\right) = 0 \end{array}\right\} \tag{A-3}$$

2. Construct V_τ as follows.

$$\begin{aligned} V_\tau &= 1 \; iff \;\; (\varepsilon_\tau, \mathbf{U}) \in NE(\mathbf{X}_\tau) \\ &= 0 \; iff \;\; (\varepsilon_\tau, \mathbf{U}) \in AL(\mathbf{X}_\tau) \\ &= \inf_{z \in \vartheta_1(\varepsilon_\tau, \mathbf{U})} P(\mathbf{X}_\tau, z) \;\; iff \;\; (\varepsilon_\tau, \mathbf{U}) \in CO(\mathbf{X}_\tau) \end{aligned} \tag{A-4}$$

where

$$P(\mathbf{X}_\tau, Z_{l,\tau}) = \int q[\mathbf{U}, \mathbf{X}_\tau, \mu(Z_{l,\tau} : \mathbf{S}_\tau, l)]\, \omega(\mathbf{U}, \mathbf{X}_\tau)\, dQ_{\mathbf{U}}(\mathbf{U}|\mathbf{O})$$

and

$$\vartheta_1(\varepsilon_\tau, \mathbf{U}) = \left\{ z \in \left[\underline{Z}(\mathbf{S}_\tau), \bar{Z}(\mathbf{S}_\tau) \right] : D(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, Z_{l,\tau}) = 1 \right\}$$

Following the same argument as Vytlacil (2002), it can be shown that

$$D(\varepsilon_\tau, \mathbf{U}|\mathbf{X}_\tau, Z_{l,\tau}) = I[P(\mathbf{X}_\tau, Z_{l,\tau}) \geq V_\tau]$$

and

$$V_\tau \perp Z_{l,\tau}|\mathbf{X}_\tau$$

The Proof of the Theorem 2: As the theory of first order stochastic dominance, it is enough to show that $\Pr\left(\tilde{U} \leq u|\mathbf{X}_\tau, V_\tau = \tilde{v}\right)$ is increasing in $\tilde{v}$. As the always taker (equation (A-2)) is always $V_\tau = 0$ and the never taker (equation (A-1)) is always $V_\tau = 1$ by the construction of V_τ, $E_U\left[G\left(\tilde{U}\right)|\mathbf{X}_\tau, V_\tau = \tilde{v}\right]$ must be nonincreasing in $\tilde{v}$ by definition. Hence, it is enough to focus on the complier (equation (A-3)). Suppose that complier exists and take $(\varepsilon^*, \mathbf{U}^*) \in CO(\mathbf{X}_\tau)$. Then using equation (A-4), we can construct the corresponding $v^* \in (0,1)$ such that

$$v^* = \inf_{z \in \vartheta_1(\varepsilon^*, \mathbf{U}^*)} P(\mathbf{X}_\tau, z)$$

where

$$\vartheta_1(\varepsilon^*, \mathbf{U}^*) = \left\{ z \in \left[\underline{Z}(\mathbf{S}_\tau), \bar{Z}(\mathbf{S}_\tau) \right] : D(\varepsilon^*, \mathbf{U}^*|\mathbf{X}_\tau, z) = 1 \right\}$$

Let $z^* \equiv \arg\inf_{z\in\vartheta_1(\varepsilon^*,\mathbf{U}^*)} P(\mathbf{X}_\tau, z)$. Because $\frac{\partial P(\mathbf{X}_\tau,z)}{\partial z} > 0$, $z^* = \arg\inf_{z\in\vartheta_1(\varepsilon^*,\mathbf{U}^*)} z$. Let $\tilde{U}^* = \Upsilon(\mathbf{U}^* : \mathbf{X}_\tau)$. At z^*, the applicant with $(\varepsilon^*, \mathbf{U}^*)$ must be the border to be selected or not. That is, z^* must satisfy

$$\varepsilon^* = \mu(z^* : \mathbf{S}_\tau, l) - \tilde{\eta}\left(\tilde{U}^*, \mathbf{X}_\tau\right).$$

Because $(\varepsilon^*, \mathbf{U}^*) \in CO(\mathbf{X}_\tau)$, $\mu\left(\bar{Z}(\mathbf{S}_\tau) : \mathbf{S}_\tau, l\right) \le \tilde{\eta}\left(\tilde{U}^*, \mathbf{X}_\tau\right) + \varepsilon^*$, and $\mu\left(\underline{Z}(\mathbf{S}_\tau) : \mathbf{S}_\tau, l\right) > \tilde{\eta}\left(\tilde{U}^*, \mathbf{X}_\tau\right) + \varepsilon^*$. Moreover, $\mu'(z : \mathbf{S}_\tau, l) < 0$ means that there exists a unique z^* in $\left[\underline{Z}(\mathbf{S}_\tau), \bar{Z}(\mathbf{S}_\tau)\right]$ and, by the definition of z^*,

$$v^* = P(\mathbf{X}_\tau, z^*).$$

Because $\frac{\partial P(\mathbf{X}_\tau,z)}{\partial z} > 0$, there exists P^{-1}

$$\begin{aligned} z^* &= P^{-1}(v^* : \mathbf{X}_\tau), \\ 0 &< \frac{\partial P^{-1}(v, \mathbf{X}_\tau)}{\partial v}. \end{aligned}$$

Hence, for this v^*, $\left(\varepsilon^*, \tilde{U}^*\right) \in \Gamma(v^* : \mathbf{X}_\tau)$ where

$$\Gamma(v^* : \mathbf{X}_\tau) \equiv \left\{\left(\varepsilon, \tilde{U}\right) | \varepsilon = \mu\left(P^{-1}(v^* : \mathbf{X}_\tau.) : \mathbf{S}_\tau, l\right) - \tilde{\eta}\left(\tilde{U}, \mathbf{X}_\tau\right)\right\}.$$

Take any $\tilde{u} \ge \tilde{U}$ and $\left(\varepsilon, \tilde{U}\right) \in \Gamma(v^* : \mathbf{X}_\tau)$. Then

$$\begin{aligned} \varepsilon &= \mu\left(P^{-1}(v^* : \mathbf{X}_\tau) : \mathbf{S}_\tau, l\right) - \tilde{\eta}\left(\tilde{U}, \mathbf{X}_\tau\right) \\ &\ge \mu\left(P^{-1}(v^* : \mathbf{X}_\tau) : \mathbf{S}_\tau, l\right) - \tilde{\eta}(\tilde{u}, \mathbf{X}_\tau) \end{aligned}$$

Hence

$$\begin{aligned} &\left\{\left(\varepsilon, \tilde{U}\right) \in \Gamma(v^* : \mathbf{X}_\tau) | \tilde{U} \le \tilde{u}\right\} \\ = &\left\{\left(\varepsilon, \tilde{U}\right) \in \Gamma(v^* : \mathbf{X}_\tau) | \varepsilon \ge \mu\left(P^{-1}(v^* : \mathbf{X}_\tau) : \mathbf{S}_\tau, l\right) - \tilde{\eta}(\tilde{u}, \mathbf{X}_\tau)\right\} \end{aligned}$$

Let us define

$$C\tilde{O}(\mathbf{X}_\tau) = \left\{\left(\varepsilon, \tilde{U}\right) | \tilde{U} = \Upsilon(\mathbf{U} : \mathbf{X}_\tau), \ (\varepsilon, \mathbf{U}) \in CO(\mathbf{X}_\tau)\right\}.$$

Then

$$
\begin{aligned}
& \Pr\left(\tilde{U} \leq \tilde{u} | \mathbf{X}_\tau, V_{n.\tau} = v^*\right) \\
= & \int \int_{\mu\left(P^{-1}(v^*:\mathbf{X}_\tau.):\mathbf{S}_\tau,l\right)-\tilde{\eta}(\tilde{u},\mathbf{X}_\tau)} I\left[\left(\varepsilon, \tilde{U}\right) \in C\tilde{O}\left(\mathbf{X}_\tau\right)\right] dF_l\left(\varepsilon_\tau\right) dQ_{\tilde{U}}\left(\tilde{U}|\mathbf{O}\right)
\end{aligned}
$$

Therefore, $\Pr\left(\tilde{U} \leq \tilde{u} | \mathbf{X}_\tau, V_{n.\tau} = v^*\right)$ is strictly increase in v^*. The desired result is immediate.