RIETI Discussion Paper Series 25-E-102

Health Effects of Retirement Policy Changes: Evidence from Japan

XIE, Mingjia
Liaoning University

YIN, Ting
RIETI

USUI, Emiko
Hitotsubashi University

ZHANG, Yi
Central University of Finance and Economics

RIETI
Research Institute of Economy, Trade & Industry, IAA

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

RIETI Discussion Paper Series 25-E-102
October 2025

Health Effects of Retirement Policy Changes: Evidence from Japan

Mingjia Xie* TingYin†‡ Emiko Usui§ Yi Zhang¶

Abstract

We evaluate the health effects of hypothetical retirement policy changes, accounting for varied individual responses to policy changes and the heterogeneous health impacts of retirement. Using a Policy Relevant Treatment Effect (PRTE) framework with Japanese data, we find a policy's net average health impact depends critically on its scale. Policies which cause marginal downward shifts in retirement rate improve average population health. Conversely, policies which induce large, substantial shifts lead to a net health decline as it faces individuals who stand to gain from retiring to continue working. Our findings highlight the importance of "selection on gains" and suggest that policymakers should favor incremental incentives over broad mandates.

Key words: Retirement; Health; Policy Relevant Treatment Effect; Japan
JEL Classification: I12, J26

The RIETI Discussion Paper Series aims at widely disseminating research results in the form of professional papers, with the goal of stimulating lively discussion. The views expressed in the papers are solely those of the author(s), and neither represent those of the organization(s) to which the author(s) belong(s) nor the Research Institute of Economy, Trade and Industry.

*Liaoning University, Email: xiemingjia@lnu.edu.cn
†Department of International Economics Faculty of Economics, Toyo University, Japan. Email: yin@toyo.jp
‡Research Institute of Economy, Trade and Industry (RIETI), Japan. Email: yin-ting@rieti.go.jp
§Center for Intergenerational Studies, Institute of Economic Research, Hitotsubashi University. Email: usui@ier.hit-u.ac.jp
¶China Center for Human Capital and Labor Market Research, Central University of Finance and Economics, P.R China. Email: yi.zhang@cufe.edu.cn

This study is conducted as a part of the Project "Economic Analysis on the problem of an aging population and a declining birthrate in China and Japan in the COVID-19 pandemic" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This study was supported by JSPS KAKENHI (Grant- in-Aid for Scientific Research (C), Grant Number JP25K05141). The authors gratefully acknowledge this support. We also wish to thank the JSTAR (Japanese Study of Aging and Retirement) team for providing data. JSTAR was conducted by RIETI, Hitotsubashi University, and the University of Tokyo. We wish to thank the support from the Joint Usage and Research Center, Institute of Economic Research, Hitotsubashi University (grant ID: IERPK2312, IERPK2429), and the Program for Innovation Research at Central University of Finance and Economics. We wish to thank Fukao Kyoji, Tomiura Eiichi, Tsuru Kotaro, Sekizawa Yoichi, Ikari Hiroshi, and the participants at the RIETI Seminar (Oct. 2024), for their helpful comments and suggestions. We also thank the RIETI staff for their kind help and cooperation. Any errors are our own. There are no conflicts of interest.

1 Introduction

In an aging society, individuals are increasingly choosing or being compelled to delay retirement. This shift is driven by both policy and technological factors. On the policy front, many governments are implementing reforms aimed at extending working lives, such as raising the statutory retirement age or adjusting pension incentives to encourage later retirement. From a technological perspective, advancements in automation and artificial intelligence have transformed the nature of work, particularly in physically demanding jobs. Technologies that reduce the physical strain of labor make it more feasible for older workers to remain in the workforce longer. Together, these forces are reshaping traditional retirement patterns towards later retirement.

Meanwhile, it well-documented that retirement is an important transition in life, with substantial consequences on health. For some individuals, retirement eliminates work-related stress and increases leisure time enjoyment with positive effects on their well-being and health. For others, retirement is associated with lower income, a loss of daily routines and life purpose, and fewer social contacts. These individuals may perceive retirement as a burden that negatively affects their well-being and health. Thus, while there are good reasons to expect an impact of retirement on individual well-being, the direction of this effect is ex-ante unclear and depends on whether positive or negative aspects of retirement dominate (van Ours, 2022).

Combining the changing retirement patterns with the mixed health effects of retirement, we can anticipate important health consequences arising from policy-induced shifts in retirement behavior. While policies increasingly aim to delay retirement, understanding the resulting health outcomes becomes critical for evaluating their broader social impact. To assess these consequences, two key components must be examined: first, how retirement itself affects health — whether positively or negatively depending on individual circumstances; and second, how individuals respond to specific policy changes that alter their likelihood of retiring. Only by jointly considering these mechanisms can we accurately estimate the

health implications of policies targeting retirement behavior in an aging population.

The evaluation is challenging, as the answer to neither question is clearly predicted. As for the health effect of retirement, a large number of empirical studies have found mixed evidence on the magnitude and direction of the retirement effect. Some studies have estimated positive effects of retirement on health, well-being, and related outcomes (e.g., Charles, 2004; Johnston and Lee, 2009; Eibich, 2015; Kolodziej and García-Gómez, 2019). Other studies found negative effects (e.g. Dave et al., 2008; Rohwedder and Willis, 2010; Bonsang et al., 2012; De Grip et al., 2012; Heller-Sahlgren, 2017; Mazzonna and Peracchi, 2017; Atalay et al., 2019) or no effects (Coe and Zamarro, 2011; Behncke, 2012; Belloni et al., 2016; Fe and Hollingsworth, 2016).[1] To some extent, these inconclusive findings can be attributed to differences in countries, institutional contexts, and chosen identification strategies. As for the response in retirement behavior to policy changes, different sub-populations reacts differently. For example, those who are likely to retire at retirement age or even earlier are affected when retirement age increases. However, those who are reluctant to retire at retirement ages are not affected too much by the policy change.

To answer both questions in an unified framework, we estimate the heterogeneous effects of retirement on health using the Marginal Treatment Effect (MTE) framework that systematically describes the distribution of heterogeneous effects. The MTE framework is introduced by Björklund and Moffitt (1987) and generalized by (Heckman and Vytlacil, 2005, 2001, 1999), which relates the treatment effect (effect on health) to the observed and unobserved characteristics that affect the likelihood of begin retired. Based on the estimated heterogeneous health effects of retirement, we further investigate the health consequences of some specified hypothetical policy changes shifting people's retirement probabilities in given ways. We first find substantial heterogeneity in the effect of retirement on health with respect to both observed and unobserved characteristics determining retirement. Particularly, individuals with lower probability of retirement due to unobservable characteristics suffer more from

[1]For an excellent literature overview on mental health and retirement, see Picchio and van Ours (2019) and van Ours (2022).

retirement in terms of health compared to others, which points to a selection on gains: individuals who are less likely to retire actually suffer more from retirement in terms of health. We also find such pattern in some observed characteristics. For example, women are more likely to retire and they suffer less from being retired than men. Building upon all these heterogeneity, our analysis further shows, with a marginal reduction in the retirement probability due to a policy — so it only affects those who are indifferent between retiring or not (marginal entrants or exiters) — the overall health consequences of such policy is positive. However, this improvement in health no longer holds when expanding this marginal change in retirement probability to larger subpopulations.

This paper contributes to the growing literature that estimates marginal treatment effects in the context of retirement. The finding of this study shows that individuals select themselves into retirement based on the effects of retirement on health. Such selection on gains have been found by the current literature (Carneiro et al., 2011; Heckman et al., 2006; Nybom, 2017; Heckman et al., 2018, e.g.,). We provide evidence of the selection on gains in evaluating the effect of retirement on health, which is consistent to the finding by Xie et al. (2025).[2] This paper also provides new evidence of heterogeneous effects of retirement in Japan by relating the effect to the likelihood of being retired, which is different to the existing literature mainly focusing on heterogeneity with respect to some observables. Moreover, it also extends the discussion from estimating the health effect of retirement to the health effect of policy changes.

The remainder of the paper is organized as follows. Section 2 summarizes the identification strategy of the treatment of interest. Section 3 describes the data. Section 4 describes the estimated heterogeneous effects. Section 5 discusses the health impact of different hypothetical policy changes. Section 6 concludes.

[2]Though uncommon in literature, the reverse selection is found in the migration literature, which find more skilled workers are easier to migrate (Chiquiar and Hanson, 2005; Rooth and Saarela, 2007; McKenzie and Rapoport, 2010).

2 Policy Relevant Treatment Effect

In this study, we explore the Policy Relevant Treatment Effect (PRTE), a key metric grounded in the framework of the Marginal Treatment Effect (MTE). We begin with a concise overview of the MTE and its estimation methods before formally defining the PRTE and its significance for policy evaluation.

2.1 Baseline Model Setup

Let Y_1 be the potential outcome in treated state ($D = 1$) and Y_0 be the potential outcome in untreated state ($D = 0$). The observed outcome (Y) is the realization of one potential outcome:

$$Y = (1 - D)Y_0 + DY_1 \tag{1}$$

The potential outcomes are specified as:

$$Y_j = \mu_j(X) + U_j, \quad j \in \{0, 1\} \tag{2}$$

where μ_j is a state-specific function of the observable X, and U_j is the unobservable which is normalized to $E[U_j|X] = 0$. Equation 2 indicates that the heterogeneity in the treatment effect $Y_1 - Y_0 = \mu_1(X) - \mu_0(X) + U_1 - U_0$ results from both the observed characteristics X and the unobserved characteristics. This specification defines a more flexible heterogeneity than the commonly used specification in which the treatment D is separately additive to all X (homogeneous treatment effect) and the specification in which the interaction terms between D and X are allowed (heterogeneous treatment effect with respect to only the observable). For selection to treatment (defined in this study as being retired), the following latent index

model is used:

$$I_D = \mu_D(Z) - U_D \tag{3}$$

$$D = 1\{I_D > 0\} \tag{4}$$

where μ_D is a function of $Z \equiv \{X, Z_0\}$, and Z_0 is the instrument(s) for D. μ_D represents the gross benefit of receiving treatment, and U_D represents the cost to treatment. In this study, U_D captures not only some unobserved individual characteristics but also some unobserved family background factors that affect retirement decisions. The latter could be even more important because the decision on retirement is heavily affected by various unobserved factors in family.

In the MTE literature, the distribution of U_D is often normalized to uniform distribution on an unit interval. As a consequence, function $\mu_D(Z)$ can be interpreted as the propensity score (the probability of receiving treatment conditional on the unobservable Z), $P(Z) \equiv Pr(D = 1|Z) = Pr(\mu_D(Z) > U_D|Z) = \mu_D(Z)$, where the last equality holds when $U_D \sim U(0, 1)$. Henceforth, the selection equation for treatment is re-defined as

$$D = 1\{P(Z) > U_D\} \tag{5}$$

MTE as a function of X and U_D accesses the heterogeneous treatment effect as follows:

$$\begin{aligned} MTE(x, u) &= E(Y_1 - Y_0|X = x, U_D = u) \\ &= \mu_1(x) - \mu_0(x) + E(U_1 - U_0|X = x, U_D = u) \end{aligned} \tag{6}$$

MTE is the average treatment effect for the individual with observed characteristics $X = x$ and unobserved cost to treatment $U_D = u$ (or the u_{th} quantile of U_D).[3] MTE also allows for the heterogeneity in both the observable X and the unobservable cost to receive treatment

[3]MTE is defined on *Marginal* individuals in receiving treatment because individuals with $U_D = u$ are also ones with $\{P(Z) = u\} \cap \{I_D = 0\}$ (indifferent in receiving treatment with propensity score u).

u. In this study, the MTE summarizes the heterogeneous effects of retirement with respect to the observable (e.g., gender) and the unobservable cost to retire (e.g., preference towards working). As a consequence, we can directly examine the heterogeneous effects with respect to the likelihood of being retired that is described by X and U_D, which is the treatment effect of interest in this study.

We use the local IV approach developed by Heckman (1999; 2001; 2005) to estimate the MTE. A detailed discussion of the method can be found in the Appendix .

2.2 Policy Relevant Treatment Effects

A central question in policy analysis is how individuals' outcomes — such as their health — respond to changes in public policies, particularly when these policies influence behavioral decisions like retirement. The *Policy Relevant Treatment Effect* (PRTE) provides a framework to quantify how average health outcomes change for the subpopulation of individuals whose retirement behavior is altered by a given policy change.

For example, consider a policy reform that raises the statutory retirement age or reduces pension generosity, making retirement less financially attractive. As a result, some individuals who would have chosen to retire under the previous policy may now decide to continue working. The PRTE captures the average health impact of retirement for those individuals who would have retired, but now stay in the labor force due to the policy.

Formally, let S denote retirement status under the baseline policy, and S^* under the new policy. Health outcomes are denoted Y (under the baseline) and Y^* (under the new policy). The PRTE is defined as:

$$\text{PRTE} = \frac{E(Y^*) - E(Y)}{E(S^*) - E(S)}, \tag{7}$$

where the numerator captures the change in average health outcomes, and the denominator captures the net change in retirement status. This effect is localized to those induced to switch their behavior because of the policy.

Importantly, the PRTE depends on a specific, defined policy change, indexed by a parameter α, which could represent a decrease in pension generosity or an increase in the retirement age. Since α can take different values depending on the policy scenario, the PRTE is not a universal treatment effect, but a *policy-dependent* one. Each policy shift corresponds to a different subpopulation being affected, leading to potentially different treatment effects.

The PRTE can be expressed as a weighted average of the *Marginal Treatment Effect* (MTE), which captures how the effect of retirement on health varies across individuals with different observed and unobserved characteristics:

$$\text{PRTE} = \int_0^1 \text{MTE}(u_S) \cdot \omega_{\text{PRTE}}(u_S)\, du_S, \tag{8}$$

where u_S reflects the individual's unobserved resistance to retire, and the weights $\omega_{\text{PRTE}}(u_S)$ reflect how the policy shift affects retirement probabilities across the distribution.

While the PRTE is useful for evaluating specific, discrete changes in policy, it comes with a notable limitation in empirical applications: accurate estimation of PRTE typically requires full support of the propensity score $P(Z)$, i.e., observing individuals with every possible retirement probability between 0 and 1. In practice, especially when instruments or policy variation are limited, this condition often fails, making PRTE difficult or even infeasible to estimate without strong extrapolation.

To address this, analysts often turn to the *Marginal Policy Relevant Treatment Effect* (MPRTE), which focuses on infinitesimal or marginal policy changes and is more tractable with observed data. Rather than requiring full support, the MPRTE relies only on the local variation in the propensity score. It is defined as the limit of a sequence of PRTEs as the size of the policy change α approaches zero:

$$\text{MPRTE} = \lim_{\alpha \to 0} \frac{E(Y^\alpha) - E(Y)}{E(S^\alpha) - E(S)}. \tag{9}$$

A key interpretation of the MPRTE is that it reflects the effect of retirement on health for

individuals who are *just at the margin of changing their behavior*, in other words, those who are indifferent between retiring and not retiring under the current policy. These individuals are the first to respond to a small policy change, such as a minor cut in pension benefits or a small increase in retirement eligibility age. Thus, the MPRTE captures the local treatment effect for the most responsive group, offering valuable insight into the behavioral and welfare consequences of marginal policy adjustments.

In summary, both PRTE and MPRTE are grounded in the idea that treatment effects — in our application, the effect of retirement on health — are heterogeneous and policy-sensitive. While the PRTE offers flexibility in modeling discrete reforms, its empirical implementation is often constrained by data limitations. The MPRTE, on the other hand, provides a focused and empirically efficient approach to evaluating the marginal health consequences of retirement-related policies, especially for those individuals at the edge of behavioral change. For a detailed explanation of the estimation of MTE, please see Appendix.

3 Data and variables

We use data from Japanese Study of Aging and Retirement (JSTAR), which is a biannual panel survey from 2007 to 2013. The survey collects the information of respondents who are aged 50 or above about their basic demographics, employment status, and health outcomes. A national representative sample of households from five cities participated in the first wave of the survey in 2007, and the number of participants increased to more households from 10 cities in 2013. We focus on individuals from all individual-year observations with valid information on all variables of interest in the analysis.

3.1 Outcome: health status

The outcome variables measure the health status of the respondents. We use two measures in this study: self-rated health and health conditions. Self-rated health is reported by

respondents about the feeling of their health. This is a categorical indicator with 5 levels: 1 Not good, 2 Not very good, 3 Average, 4 Fairly good, and 5 Good.[4] The second measure asks whether there is any health condition that interferes the respondents' casual life. This is also a categorical indicator with 4 levels: 1 Has significantly interfered, 2 Has interfered, 3 Has not interfered, 4 Has not interfered at all. Overall, higher values of our measures indicates better health status.

3.2 Retirement status and its instrument

In the survey, the respondent chooses a situation that best describes her current work status: Currently working, Temporarily not working, Not working, or Other. To measure a respondent's retirement status, we compute a binary measure which takes the value 1 if a respondent considers his/herself as not working, and 0 otherwise.[5]

One key argument in the retirement literature is that retirement decisions are endogenous (see e.g. Mazzonna and Peracchi, 2012; Insler, 2014; Mazzonna and Peracchi, 2017). Endogeneity can arise from reverse causality or from unobserved confounders, such as cognitive functioning or health limitations. A common way of dealing with this is to exploit the change in retirement behaviors due to statutory retirement ages in each country's social security scheme (e.g. Rohwedder and Willis, 2010; Coe and Zamarro, 2011; Mazzonna and Peracchi, 2012, 2017). In our application, we construct an instrumental variable measuring the age relative to the statutory retirement age in Japan (60 years old) in years. It is defined as follows:

$$Z \equiv 1\{\text{Age} \geq 60\} \times (\text{Age} - 60) \tag{10}$$

When someone is eligible for statutory retirement or pension, the probability of retirement usually jumps at the statutory retirement age. Besides the "jump", Equation 10 also specifies

[4]The original measure has a opposite meaning so that 5 indicates for Not Good and 1 indicates for Good. We recode it so that both measures used in our analysis have the a consistent interpretation of health status.

[5]There are several definitions of retirement status. Insler (2014) discusses two common definitions of being retired: self-reported retirement status, or not being in paid labor. Both have been used in the literature.

that the likelihood of retirement increases with the number of years a person exceeds the statutory retirement age. We introduce this additional variation to improve the identification of the marginal treatment effect. Furthermore, it helps address concerns that using age 60 as a fixed cutoff may be problematic due to complexities in the pension system, which could make the threshold appear fuzzy.

Our instrument exploits the exogenous variation from the country-level retirement system. As discussed in Gruber and Wise (2009), retirement behavior responds very strongly to incentives set by social security pension systems. Since such policies are defined on the national level and are outside of individual control, they provide credible exogenous variation to individual retirement decisions. It is therefore unlikely that mental health shows discontinuities around retirement eligibility ages that can be attributed to reasons other than retirement.[6]

Table 1 shows the first stage estimation results, i.e., how the instrument (age relative to statutory retirement age) affects retirement status. Column (1) indicates that, using a linear regression model, with one more year above the statutory retirement age, the likelihood to be retired significantly increased by 0.041. There is also little concern on the weak instrument as the Kleibergen-Paap rk LM statistic is 45.1, well above the rule-of-thumb cutoff 10. Since the first stage of the estimation procedure of MTE is based on a logistic regression, we also report the estimated parameters of the first-stage logit model in column (2). To ease the interpretation, we report the marginal effects of the instruments while fixing all other covariates at sample means. The finding is consistent with the results from the linear model that, one more year above the statutory retirement age, the likelihood of being retired increased by 3.5% conditional on all other covariates at their sample means.

3.3 Summary

Table 2 shows the summary statistics of all variables. The average self-rated health is better than "Average", and the average health conditions is better than "Has not interfered".

[6]One concern with the instrument could be that health insurance benefits are correlated with retirement schemes. Since Japanese health insurance benefits are not contingent on age, this is not an issue here.

Table 1: First stage estimation: 2SLS and Marginal effects at means from the logistic regression

	(1)	(2)
Independent variables	Retired	
Age relative to retirement age 60	0.041***	0.035***
	(0.006)	(0.008)
Kp Wald F statistics	45.1	-
χ^2 for test of the excluded instruments	-	17.9
Observations	12,802	12,802

Significance levels: *** 1%, ** 5%, and * 10%. Robust standard errors in parentheses are clustered at individual level. All regressions include covariates: age, age squre, gender, marital status, education level, and survey year fixed effects.

Around 48% of the respondents are retired, and the average years to the statutory retirement age (60) is 6.9 years. For all respondents with slightly more women, the average age is 69, and most of them have got married. Table 2 also shows the summary statistics by retirement status. There are mainly two noticeable differences. First, the retired respondents have relatively worse health status than the working population. Second, the ratio of women and the high-educated in the retired population is higher than the working population.

4 Heterogeneous effects of retirement on health

The MTE investigates the heterogeneous effects in both the observed and the unobserved dimensions. Table 3 shows the estimation results for self-rated health. We can find heterogeneity with respect to some observed dimensions considered in this study. For example, retirement leads to worse health outcome (-0.627) for men compared to women, so the effect of retirement is more detrimental to men than women. As for the unobserved heterogeneity captured by $k(u)$ which is estimated non-parametrically, we present the results in Figure 1. On the X-axis, it is the unobserved resistance to treatment U_D; on the Y-axis, it is the

Table 2: Summary statistics

Variable	All	Retired	
		No	Yes
Outcome variable			
Good self-rated health	3.470	3.684	3.239
	(1.050)	(0.980)	(1.073)
No health condition	3.280	3.459	3.087
	(0.869)	(0.769)	(0.928)
Treatment variable			
Retired	0.481	-	-
	(0.500)	-	-
Instrumental variable			
Years to retirement age 60	6.876	4.628	9.300
	(6.172)	(5.551)	(5.883)
Covariates			
Age	65.953	63.170	68.955
	(7.417)	(7.093)	(6.536)
Male	0.501	0.597	0.397
	(0.500)	(0.490)	(0.490)
Married	0.812	0.836	0.786
	(0.391)	(0.370)	(0.410)
At least high school degree	0.685	0.738	0.627
	(0.465)	(0.439)	(0.484)
Number of observations	12,802	6,642	6,160

Sample average is in number, and the standard deviation is in parenthesis.

estimated effect of retirement on health, i.e., MTE($X = \bar{X}, U_D = u$), where $\bar{X}$ is the sample mean of the covariates. The effect is statistically negatively associated with the unobserved resistance: with relatively low resistance, the treatment can be positive; whereas with relatively large resistance, the treatment is negative. Therefore, the health effect of retirement for individuals who are very likely to retire for various reasons that are unobserved to this study is relatively trivial and even positive; whereas individuals who are very unlikely to be retired due to these unobserved reasons suffer from retirement in terms of their health. For a more concrete understanding of the result, we use an example to illustrate the finding. Suppose that preference towards work is a key component when making retirement decision, and such preference is not observed to us (not included in control variables). Our finding suggests that individuals who have little interest in working (likely to be retired) suffer less negative health impact of retirement than individuals who have strong preference towards working. In summary, when individuals are less likely to retire due to various unobserved reasons, the retirement can be more detrimental to their health. We have similar results when focusing on the other measure for health, i.e., health conditions, as summarized in Table 4 and Figure 2.

5 Health effect from policy changes

Given the estimated MTE summarizing the distribution of the health effect of retirement, we are further interested in the health impact of policies that shift the distribution of the probability of retirement in the population. Before investigating any policy change, we show how treatment effects vary by subpopulations defined by retirement status. Precisely, Table 5 lists four average treatment effects: (1) for the whole population (2) for the retired subpopulation (3) for the non-retired subpopulation (4) complier subpopulation whose retirement status are shifted by the instrumental variable. Retirement leads to worse health outcomes for most subpopulations though they are not statistically significant. Meanwhile, the effect

Table 3: Estimation results for self-rated health

$MTE(x,u) = (\beta_1 - \beta_0)x + k(u)$	Coef.	Std. Err.	P-value
$\boldsymbol{\beta_1 - \beta_0}$			
Age	0.140	0.151	0.353
Age squared	-0.002	0.001	0.169
Male	-0.627*	0.326	0.054
Married	-0.714***	0.126	0.000
Higher education	0.016	0.115	0.891
$\boldsymbol{k(u)}$ (See Figure 1)			
Test of observable heterogeneity			0.000
Test of unobservable heterogeneity			0.998

The estimation include age, age square, gender, marital status, education level, and survey year fixed effects. Bootstrap standard error from 499 replications is clustered at individual level. The null hypothesis of the test of observable heterogeneity is that $\beta_1 - \beta_0 = 0$ are jointly true for all observable. The null hypothesis of the test of unobservable heterogeneity is $k(u) = 0$.

Table 4: Estimation results for health conditions

$MTE(x,u) = (\beta_1 - \beta_0)x + k(u)$	Coef.	Std. Err.	P-value
$\boldsymbol{\beta_1 - \beta_0}$			
Age	-0.011	0.131	0.932
Age squared	0.000	0.001	0.791
Male	-0.699***	0.258	0.007
Married	-0.506***	0.107	0.000
Higher education	0.035	0.097	0.719
$\boldsymbol{k(u)}$ (See Figure 2)			
Test of observable heterogeneity			0.000
Test of unobservable heterogeneity			0.998

The estimation include age, age square, gender, marital status, education level, and survey year fixed effects. Bootstrap standard error from 499 replications is clustered at individual level. The null hypothesis of the test of observable heterogeneity is that $\beta_1 - \beta_0 = 0$ are jointly true for all observable. The null hypothesis of the test of unobservable heterogeneity is $k(u) = 0$.

Figure 1: Estimation results of MTE for self-rated health

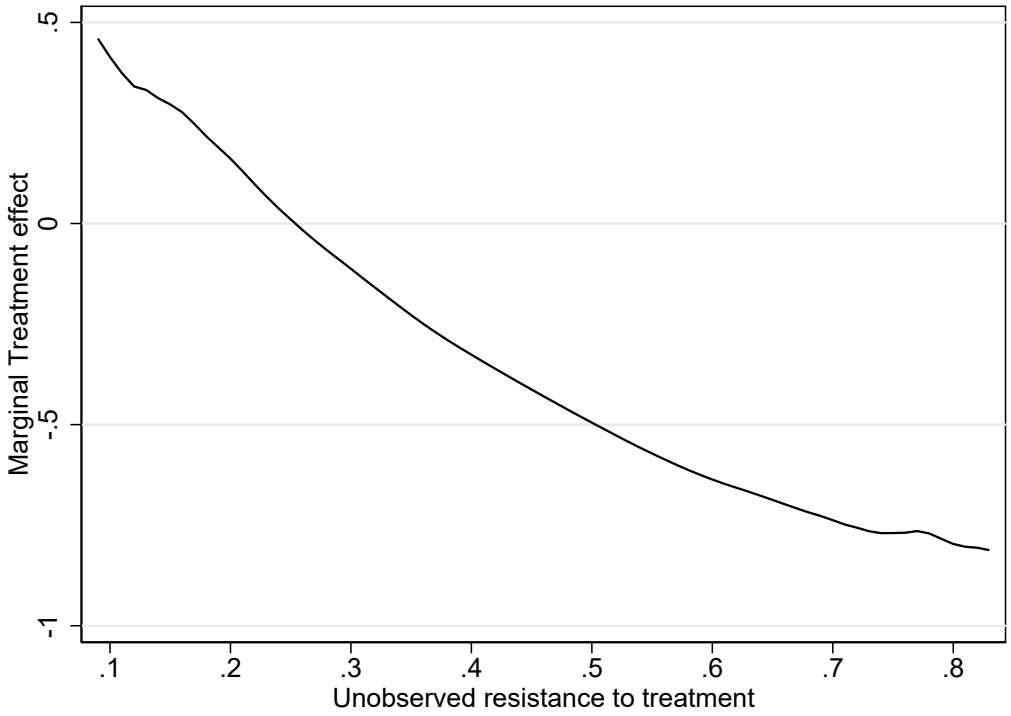

Figure 2: Estimation results of MTE for health conditions

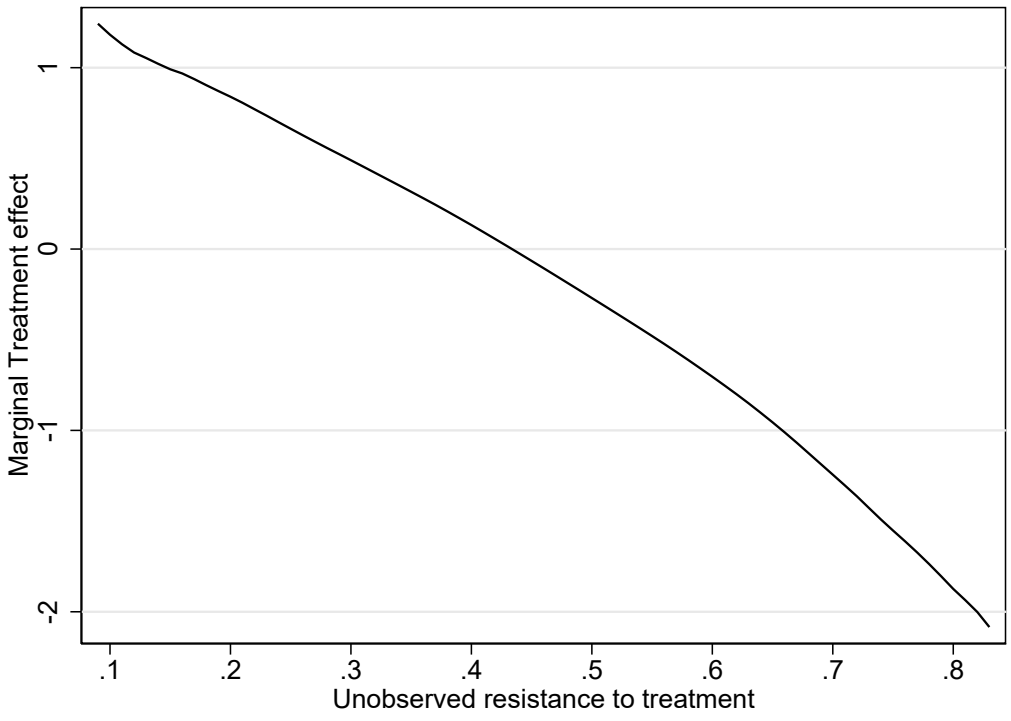

Table 5: Treatment effects of retirement on health by subpopulations defined by retirement status

	Self-rated health	Health conditions
Average Treatment Effect	-0.331	-0.201
	(0.685)	(0.527)
Average Treatment Effect for Treated (retired)	-0.266	0.458
	(0.992)	(0.768)
Average Treatment Effect for Untreated (not retired)	-0.429	-0.906
	(0.804)	(0.675)
LATE	-0.677	-0.745**
	(0.434)	(0.349)
Observations	12,802	12,802

Significance levels: *** 1%, ** 5%, and * 10%. Bootstrap standard error from 499 replications is clustered at individual level. All regressions include covariates: age, age squre, gender, marital status, education level, and survey year fixed effects.

of retirement is less detrimental or even positive for the retired subpopulation compared to the non-retired subpopulation. This indicates for a pattern of self-selection which is in line with the finding from the estimated MTE: those who retired are also those who suffer less or even benefit from retirement in terms of their health.

It can be inferred that any policy that changes the status quo of retirement may leads to worse health outcome based on the self-selection pattern shown in Table 5. However, this inference ignores the heterogeneity within the subpopulations which may lead to unknown directions of the health impact of a given policy change. Therefore, we need to investigate the treatment effects of interest based on all heterogeneity captured by MTE.

We first present the estimation results of MPRTE which evaluates the health effect of a policy that slightly increases the retirement probability. Consider a sequence of policies indexed by a

scalar variable α, with $\alpha = 0$ denoting the baseline, status quo policy. Under a given policy α, the propensity score is P_α which is the fitting probability of retirement, and P_0 corresponds to the propensity score under the baseline policy. Following Carneiro et al. (2011), we consider three types of increments in the retirement probability: (1) a policy that increases the probability of retirement by an amount α, so that $P_\alpha = P_0 + \alpha$. Policies such as cutting public pension payouts or the implementation of robotics that alleviate physically demanding tasks may uniformly decreases people's retirement probability; (2) a policy that changes the probability of retirement by the proportion $(1 + \alpha)$, so that $P_\alpha = P_0(1 + \alpha)$. Some policies disproportionately affect people's retirement behaviors such that some people are affected with larger impact (3) a policy that shifts one of the components of Z, denoted as $Z^{[k]}$, so that $Z_\alpha^{[k]} = Z_0^{[k]} + \alpha$. This type of policy refers to cases when one of the determinants for retirement status, e.g., change in statutory retirement age leading to fewer retired population. It is worthy noting that, though same α is used as the notation, these three types of change in retirement probabilities are driven by different forces that are plausible in different settings. Overall, the estimated MPRTE shows the change in health outcomes in the population when the policy just starts to change (as as retirement probabilities), i.e., $\alpha \to 0$.

Table 6 shows the estimated MPRTE of the two measures for health. The positive coefficients regardless of different types of the estimates indicate that, when a policy decreases the retirement probability for marginal subpopulation who are indifferent between retired or working, we would expect an overall positive impact of the health in the population. In other words, when there is a gradual change in the retirement policy that increases (or decreases) the retirement probability, the first batch of people who are affected would suffer (or benefit) from the change in terms of their health.

The estimation result of the MPRTE reveals that a tiny change in the retirement policy reducing the retirement probability leads to improvement in the overall health in the population. Can we conclude that a policy change always lead to better health when it negatively affects retirement probability? The answer is no because MPRTE focuses only on the subpopulation

who are at margin of retirement — those who are indifferent between retired or not. When retirement probability changes by some non-trivial sizes, a larger subpopulations would be affected, leading the overall health impact to an unknown direction. To understand the health effect of non-trivial changes in retirement probabilities, we specify two types of changes: (1) an universal decrease in the retirement probability by $\alpha \in \{0.1, 0.2, 0.3\}$ so everyone are less likely to retire by same amount in terms of probability (2) only those who are very likely to retire are affected so that retirement probability decreases by $\alpha \in \{0.1, 0.2, 0.3\}$. This type of policy captures the idea that people with higher retirement probability are likely to be affected by the policy change while the other subpopulation are unlikely to retire anyway. We define the subpopulation with higher retirement probability as those with propensity scores larger than 0.5.

As shown in Table 7, when there is an uniform reduction in the retirement probability in the population, the overall health impact is negative which means a worse health outcome of the population. When this reduction is not uniform but only in the subpopulation who has higher (≥ 0.5) probabilities of retirement, the negative impact is even larger. While the second type of policy that aims at preventing people who are likely to retire from retirement, we expect the health consequence is more detrimental than the first type.

6 Conclusion

This paper assesses the heterogeneity in the effects of retirement on health by estimating marginal treatment effects. We find that the effect of retirement on health varies by the likelihood of retirement. Particularly, individuals who have lower probability of retirement due to unobservable determinants of retirement suffer more from retirement in terms of health compared to others. In other words, we find the pattern of selection on health gains in retirement decisions.

We further investigate the health impact of changes in retirement probabilities induced by

Table 6: Estimated Marginal Policy Relevant Treatment Effect

Type of MPRTE	Coefficient	Std.Err.	P-value
Health outcome: self-rated health			
$P_\alpha = P_0 + \alpha$	0.431	0.609	0.479
$P_\alpha = P_0(1+\alpha)$	0.357	0.540	0.509
$Z_\alpha^{[k]} = Z_0^{[k]} + \alpha$	0.479	0.609	0.431
Health outcome: health conditions			
$P_\alpha = P_0 + \alpha$	0.338	0.477	0.479
$P_\alpha = P_0(1+\alpha)$	0.315	0.425	0.458
$Z_\alpha^{[k]} = Z_0^{[k]} + \alpha$	0.619	0.500	0.215

The estimated MPRTE shows the average effect when $\alpha \to 0^-$. Significance levels: *** 1%, ** 5%, and * 10%. Bootstrap standard error from 499 replications is clustered at individual level. All regressions include covariates: age, age squre, gender, marital status, education level, and survey year fixed effects.

Table 7: Estimated Policy Relevant Treatment Effect

	$P_\alpha = P_0 - \alpha$			$P_\alpha = P_0 - 1\{P_0 \geq 0.5\} \cdot \alpha$		
α	0.1	0.2	0.3	0.1	0.2	0.3
Self-rated health	-0.316	-0.295	-0.275	-0.966	-0.889	-0.812
	(0.655)	(0.691)	(0.693)	(0.651)	(0.628)	(0.655)
Health conditions	-0.164	-0.060	0.069	-0.595	-0.362	-0.134
	(0.504)	(0.521)	(0.563)	(0.502)	(0.494)	(0.536)

Significance levels: *** 1%, ** 5%, and * 10%. Bootstrap standard error from 499 replications is clustered at individual level. All regressions include covariates: age, age squre, gender, marital status, education level, and survey year fixed effects.

different policies. The estimation results of the Marginal Policy Relevant Treatment Effect (MPRTE) and the Policy Relevant Treatment Effect (PRTE) reveal nuanced insights into the health impacts of changes in retirement policy. MPRTE focuses on marginal changes in retirement probability and shows that slight increases in retirement (i.e., $\alpha \to 0$) negatively affect health outcomes among those who are indifferent between retiring or not. This suggests that initial policy shifts decreasing retirement likelihood can improve health. To assess broader impacts, PRTE estimates are used for larger, non-marginal changes. When retirement probability is uniformly reduced across the population (by 0.1, 0.2, or 0.3), PRTE results indicate a general decline in health, with even more severe negative health impacts when only individuals with high retirement propensity (propensity score ≥ 0.5) are targeted. These findings suggest that while reducing retirement may theoretically extend working life, it could lead to worse health outcomes, especially for those most inclined to retire.

The study also has the following implication: policymakers should exercise caution with universal measures aimed at increasing labor force participation, such as raising statutory retirement ages or broadly reducing retirement probabilities, as the PRTE estimates indicate these lead to overall worse population health outcomes; the detrimental effects appear particularly pronounced and counterproductive.

References

Atalay, K., G. F. Barrett, and A. Staneva (2019): "The effect of retirement on elderly cognitive functioning," Journal of Health Economics, 66, 37–53.

Behncke, S. (2012): "Does retirement trigger ill health?" Health Economics, 21, 282–300.

Belloni, M., E. Meschi, and G. Pasini (2016): "The effect on mental health of retiring during the economic crisis," Health Economics, 25, 126–140.

Björklund, A. and R. Moffitt (1987): "The estimation of wage gains and welfare gains in self-selection models," The Review of Economics and Statistics, 42–49.

Bonsang, E., S. Adam, and S. Perelman (2012): "Does retirement affect cognitive functioning?" Journal of Health Economics, 31, 490–501.

Brinch, C. N., M. Mogstad, and M. Wiswall (2017): "Beyond LATE with a discrete instrument," Journal of Political Economy, 125, 985–1039.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating marginal returns to education," American Economic Review, 101, 2754–81.

Charles, K. K. (2004): "Is retirement depressing?: Labor force inactivity and psychological well-being in later life," Research in Labor Economics, 23, 269–299.

Chiquiar, D. and G. H. Hanson (2005): "International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States," Journal of political Economy, 113, 239–281.

Coe, N. B. and G. Zamarro (2011): "Retirement effects on health in Europe," Journal of Health Economics, 30, 77–86.

Dave, D., I. Rashad, and J. Spasojevic (2008): "The effects of retirement on physical and mental health outcomes," Southern Economic Journal, 75, 497–523.

De Grip, A., M. Lindeboom, and R. Montizaan (2012): "Shattered dreams: The effects of changing the pension system late in the game," Economic Journal, 122, 1–25.

Eibich, P. (2015): "Understanding the effect of retirement on health: Mechanisms and heterogeneity," Journal of Health Economics, 43, 1–12.

Fe, E. and B. Hollingsworth (2016): "Short- and long-run estimates of the local effects of retirement on health," Journal of the Royal Statistical Society - Series A, 179, 1051–1067.

Gruber, J. and D. A. Wise (2009): Social security programs and retirement around the world: Micro-estimation, University of Chicago Press.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018): "Returns to education: The causal effects of education on earnings, health, and smoking," Journal of Political Economy, 126, S197–S246.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006): "Understanding instrumental variables in models with essential heterogeneity," The Review of Economics and Statistics, 88, 389–432.

Heckman, J. J. and E. Vytlacil (2001): "Policy-relevant treatment effects," American Economic Review, 91, 107–111.

——— (2005): "Structural equations, treatment effects, and econometric policy evaluation 1," Econometrica, 73, 669–738.

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

Heller-Sahlgren, G. (2017): "Retirement blues," Journal of Health Economics, 54, 66–78.

Insler, M. (2014): "The health consequences of retirement," Journal of Human Resources, 49, 195–233.

Johnston, D. W. and W. Lee (2009): "Retiring to the good life? The short-term effects of retirement on health," Economics Letters, 103, 8–11.

Kolodziej, I. W. and P. García-Gómez (2019): "Saved by retirement: Beyond the mean effect on mental health," Social Science & Medicine, 27, 85–97.

Mazzonna, F. and F. Peracchi (2012): "Ageing, cognitive abilities and retirement," European Economic Review, 56, 691–710.

——— (2017): "Unhealthy retirement?" Journal of Human Resources, 52, 128–151.

McKenzie, D. and H. Rapoport (2010): "Self-selection patterns in Mexico-US migration: the role of migration networks," the Review of Economics and Statistics, 92, 811–821.

Nybom, M. (2017): "The distribution of lifetime earnings returns to college," Journal of Labor Economics, 35, 903–952.

Picchio, M. and J. C. van Ours (2019): "Mental health effects of retirement," De Economist, 168, 419–452.

Rohwedder, S. and R. J. Willis (2010): "Mental retirement," Journal of Economic Perspectives, 24, 119–138.

Rooth, D.-O. and J. Saarela (2007): "Selection in migration and return migration: Evidence from micro data," Economics letters, 94, 90–95.

van Ours, J. C. (2022): "How Retirement Affects Mental Health, Cognitive Skills and Mortality; An Overview of Recent Empirical Evidence," De Economist, 1–26.

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

Xie, M., T. Yin, E. Usui, and Y. Zhang (2025): "Heterogeneous Effects of Retirement on Health: Evidence from Japan," RIETI Discussion Paper Series 25-E-002.

Online Appendix

A MTE Identification

One way to identif MTE is the method of local IV developed by Heckman (1999; 2001; 2005). This method identifies MTE as the derivative of the conditional expectation of Y with respect to the propensity score. More precisely, we have

$$\begin{aligned} E(Y|X=x,P(Z)=p) &= \mu_0(x)+p(\mu_1(x)-\mu_0(x)) \\ &\quad + pE(U_1-U_0|X=x,U_D\leq p) \\ &= \mu_0(x)+p(\mu_1(x)-\mu_0(x))+K(x,p) \end{aligned} \tag{A.1}$$

where $K(x,p) \equiv pE(U_1-U_0|X=x,U_D\leq p)$. $K(x,p)$ is a function of X and p that captures heterogeneity along the unobserved cost to treatment U_D. Taking the derivative of Equation A.1 with respect to p and evaluating it at u, we get MTE

$$\begin{aligned} MTE(X=x,U_D=u) &= \frac{\partial E(Y|X=x,P(Z)=p)}{\partial p}|_{p=u} \\ &= \mu_1(x)-\mu_0(x)+k(x,u) \end{aligned} \tag{A.2}$$

where $k(x,u) = E(U_1-U_0|X=x,U_D=u)$. Intuitively, conditioning on $X=x$, when an infinitesimal shift occurs in the propensity score at p (changing the treatment status from untreated state to treated state), the corresponding change in Y is the treatment effect for individuals who have $X=x$ and have p as the propensity score (or unobserved cost), which is exactly MTE. Equation A.2 also indicates that, without further assumptions, we need additional variation conditional on X to identify $\mu_1(x)-\mu_0(x)$ and $k(x,u)$ separately to identify MTE. This additional variation comes from the excluded instrument Z_0, and $MTE(x,p)$ is identified under the following assumption on the instrument.

Assumption 1

$$(U_0, U_1, U_D) \text{ is independent of } Z_0, \text{ conditional on } X$$

The conditional independence assumption requires that the instrument is independent of the unobservable in the outcome equations and the selection equation. The conditional independence between Z and (U_0, U_1, U_D) implies and is also implied by the standard IV assumptions of conditional independence and monotonicity (Vytlacil, 2002).

Besides the assumptions that are required in the literature using Instrumental Variable (IV), there are often more assumptions in estimating MTE. The local IV estimator motivated by Equation A.2 indicates that the support of the propensity score P conditional on X determines the support of the unobserved cost U_D in MTE. Therefore, substantial variation in P conditional on X (which solely comes from the excluded instrument Z_0) is needed to identify $MTE(x, u)$ on a wide range of $U_D \in [0, 1]$. For this reason, additional assumptions are usually required, e.g., at least one of the instruments is continuous, which makes it possible to have a full support in MTE. However, it can be challenging to find proper continuous instrument(s) with sufficient variation conditional on observed covariates in many em pirical studies, including this study. In the case of discrete instrumental variables, alternative approaches include restricting the specifications in the model and specifying a less flexible relation among random variables[7] Following Brinch et al. (2017), we impose the second assumption as follows:

Assumption 2

$$E(Y_j|U_D, X = x) = \mu_j(x) + E(U_j|U_D), \quad j \in \{0, 1\}$$

Assumption 2 specifies a more restrictive version of Equation 2 because it implies that the observable and the unobservable contribute to the potential outcome in a substitute manner.

[7]See a more detailed discussion in Brinch et al. (2017).

consequently, MTE in Equation 6 can be written as

$$MTE(x,u) = \mu_1(x) - \mu_0(x) + E(U_1 - U_0|U_D = u) \tag{A.3}$$

Equation A.3 implies that $MTE(x,u)$ can be identified over the support of u, which is determined by the support of the estimated propensity score P, unconditional on X. Therefore, Assumption 2 makes the discrete instrumental variable feasible in identifying MTE.

After imposing Assumption 2, the treatment effect is still allowed to vary by X and U_D but not by the interaction between the two, and it is weaker than the additive separability assumption between D and X, which is commonly used in empirical analysis such as a linear specification $Y = \alpha D + \beta X + U$. Furthermore, Assumption 2 is implied by (but does not imply) the full independence assumption about random variables, i.e., $(Z, X \perp U_0, U_1, U_D)$ which is assumed in some applied works estimating MTE. Assumption 2 holds when there is no endogenous variable in X in the outcome (health) equation, which is also required in many applied works like the standard IV estimation approach.

Under Assumption 1 and 2, we have:

$$E(Y|X = x, P(Z) = p) = \mu_0(x) + p(\mu_1(x) - \mu_0(x)) + K(p) \tag{A.4}$$

and thus,

$$MTE(x,p) = \frac{\partial E(Y|P(Z) = p, X = x)}{\partial p} = \mu_1(x) - \mu_0(x) + k(p) \tag{A.5}$$

where $K(p) = pE(U_1 - U_0|U_D \leq p)$ and $k(p) = E(U_1 - U_0|U_D = p)$.

B MTE Estimation

For ease of interpretation, we illustrate the idea of estimation procedure with a parametric approach. However, to make our estimates as flexible as possible, we adopt a semi-parametric approach in our estimates.

Equation A.5 suggests the following estimation procedures: We start by estimating the propensity score $\widehat{P}(Z)$ based on Equation 4 using a probability model such as probit or logit model. We then make assumptions about the functional form of the unknown function μ_1, μ_0 and $K(p)$. With these assumed functional forms, we estimate $\widehat{\mu_0}$, $\widehat{\mu_1 - \mu_0}$ and $\widehat{K}(p)$ separately based on the equation $E(Y|X = x, P = p)$ in Equation A.5. Last, we calculate MTE by taking the derivative with respect to p.

In the main specification, the propensity score P is estimated from the the logistic regression. Both μ_0 and μ_1 are specified to be linear: $\mu_0(x) = \beta_0 x$ and $\mu_1(x) = \beta_1 x$. Thereby, the conditional expectation of Y is written as:

$$E(Y|X = x, P(Z) = p) = x\beta_0 + x(\beta_1 - \beta_0)p + K(p) \tag{B.6}$$

Furthermore, $K(p)$ is specified as a polynomial function of p with order 2 in the main specification. Note that MTE is then a linear formula in p as follows:

$$MTE(x, u) = x(\beta_1 - \beta_0) + \gamma u \tag{B.7}$$

$(\beta_1 - \beta_0)$ captures the heterogeneous treatment effects with respect to the observable characteristics X, while γ corresponds to the heterogeneous treatment effects with respect to the unobserved cost to treatment. A negative γ indicates that the treatment effect is larger for those who are more likely to be selected to treatment because of lower unobserved cost to treatment, which is in line with the prediction of the Roy Model, namely selection on gains. On the contrary, a positive γ indicates the reverse selection on gains, i.e., individuals who are less likely to receive the treatment due to the higher unobserved cost are with larger treatment effects.

To make our estimates as flexible as possible, we adopt a semi-parametric estimation approach. We first obtain the estimated $\hat{p}$ from a logistic regression. we then use local polynomial (second order) regressions of Y, X, and $X \times \hat{p}$ on $\hat{p}$ to get residuals e_Y, e_X, and $e_{X \times p}$.

With these residuals, we estimate the following equation using regression and

$$e_Y = e_X\beta_0 + e_{X\times p}(\beta_1 - \beta_0) + \epsilon \tag{B.8}$$

construct residual $\tilde{Y} = Y - X\hat{\beta}_0 - X(\widehat{\beta_1 - \beta_0})\hat{p}$ where $\hat{\beta}_0$ and $(\widehat{\beta_1 - \beta_0})$ are estimated coefficients from above. Furthremore, we use the local polynomial (second order) regression of $\tilde{Y}$ on $\hat{p}$, saving level $\widehat{K(p)}$ and slope $\widehat{k(p)} = \widehat{K'(p)}$. Finally, we have $\widehat{MTE(x,u)} = x(\widehat{\beta_1 - \beta_0}) + \widehat{k(p)}$. In the nonparametric regressions above, the bandwidths are chosen by rule-of-thumb using a polynomial of order 4, and Gaussian kernels are used.