RIETI Discussion Paper Series 15-E-090

Extensions of Rubin's Causal Model for a Latent-Class Treatment Variable: An analysis of the effects of employers' work-life balance policies on women's income attainment in Japan

YAMAGUCHI Kazuo
RIETI

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

RIETI Discussion Paper Series 15-E-090
July 2015

Extensions of Rubin's Causal Model for a Latent-Class Treatment Variable: An analysis of the effects of employers' work-life balance policies on women's income attainment in Japan[1]

YAMAGUCHI Kazuo
The University of Chicago / RIETI

Abstract

Combining inverse-probability weighting based on propensity scores and a semiparametric outcome model with a latent-class variable as an intervening variable, this paper introduces extensions of Rubin's causal model for the case where the treatment variable is a latent-class variable with indicators. Although the paper first introduces a method for the analysis of cross-sectional survey data, some extensions of the method for panel survey data analysis are also described. The method is especially useful when we have a set of mutually related categorical variables to characterize a specific latent characteristic of social contexts such as firms, schools, or neighborhoods, and when the latent characteristic is hypothesized to affect an individual-level outcome and we need to control for selection bias of people in different social contexts.

An application, which is based on data for employees and employers collected in 2009 by the Research Institute of Economy, Trade and Industry in Japan, focuses on the effect of employers' work-life balance policies on female regular white-collar employees' income. Six dichotomous indicators of policies are employed. Variables tested for possible confounding factors include individual human-capital and labor-hour variables and some firm-level exogenous variables. The analytical results show that although a certain portion of the effects of employer's work-life balance policies on income are explained as a result of such a selection bias that firms of larger size have a higher prevalence rate of those policies and at the same time higher average income for employees, the positive effects of those policies on income among female employees still remain significant after the elimination of the selection bias.

Keywords: Rubin's causal model, Latent-class variable, Firms' work-life balance policies, Income attainment of women

JEL classification: C14, J31, J71

[1] This paper was written as a part of the author's study of "Women's Economic Empowerment, Low Fertility Rate, and Work-Life Balance" at the Research Institute of Economy, Trade and Industry, in Japan.

1. INTRODUCTION

This paper introduces extensions of Rubin's causal model (RCM) when the treatment variable is a latent-class variable with indicators. The inverse probability (IP) weighting based on propensity-score estimates introduced by Rubin (1985) and extended for the marginal structural model by Robins et al. (2000) has been employed as a general tool for eliminating selection bias in the treatment variable. It has also been employed by Abadie (2005) to elaborate the difference-in-difference (DID) estimator of the treatment effect with panel survey data. The method introduced in this paper will be especially useful when we have data on two levels, the individual level and the level of social contexts, such as firms, schools, or neighborhoods, and we have a set of mutually related categorical indicators of a particular aspect of social contexts and are concerned with the causal effect of that aspect on individual outcomes. A key methodological issue here is an effective control for selection bias of people in distinct social contexts. Like the RCM, the method introduced in this paper relies on a semiparametric control for selection bias by the use of IP weights.

The use of IP weighting rather than a control for confounding covariates by a regression has the advantage of allowing endogeneity of confounding covariates in addition to the semiparametric treatment for the outcome regression. Unlike the standard RCM, however, the method with a latent-class treatment variable simultaneously identifies latent classes as the states of a treatment variable and estimates the propensity score for selection into the states of the treatment variable.

Below, we (1) review the assumption of the RCM and the modification of it for the case with a latent-class treatment variable, (2) describe the method and models for the analysis of cross-sectional survey data, including the method for the estimation of the treatment effects and their standard errors, and diagnostic analyses for examining the adequacy of adjustments made by IP weighting, (3) describe some extensions of the method for panel survey data analysis, and (4) present an application of the method to the analysis of the effects of employers' work-life balance policies on income among female regular white-collar employees in Japan.

2. METHOD

2.1. A Review of the Assumptions of the RCM and Modifications of Them When the Treatment Variable Is a Latent-Class Variable

The RCM for cross-sectional data analysis makes the following set of assumptions (Rosenbaum and Rubin 1983, 1984; Morgan and Winship 2007):

(A1) The stable unit treatment value assumption (SUTVA) holds. The RCM makes a distinction between *treatment* and observed *treatment assignment* X that distinguishes between the treatment group, for which $X = 1$, and the control group, for which $X = 0$. For treatment, the RCM conceives for every person i both Y_{1i}, which indicates the outcome with treatment, and Y_{0i}, which indicates the outcome without treatment. Y_{1i} is observed only for the treatment group, and Y_{0i} is observed only for the control group. Hence, the treatment effect for person i, $Y_{1i} - Y_{0i}$, is always unobserved. The SUTVA justifies such a conception of Y_{1i} and Y_{0i} by assuming that the pair of outcomes $\{Y_{1i}, Y_{0i}\}$ for each person i does not depend on who among others in the population are assigned to the treatment group.[2] The observed outcome, Y_{obs}, can be expressed as $Y_{obs} = XY_1 + (1-X)Y_0$, and the average treatment effect is given as $E(Y_1) - E(Y_0)$.

[2] The dependence of individual potential outcomes on the characteristics of others under the treatment does not violate the SUTVA if that dependence can be characterized as a confounding

(A2) The treatment effects are heterogeneous and vary with persons. This leads to the dependence of the average treatment effect in the RCM on the specification of the population, such as the total population, the population of the treated, or the population of the untreated.

(A3) The strong ignorability of treatment assignment (SITA), $(Y_{1,}Y_0) \perp X \mid \mathbf{V}$, holds such that controlling for *observed confounding variables* **V** that affect both the outcome and the treatment assignment *X*, conditional independence between a pair of potential outcomes and the treatment assignment holds. This assumption implies that no *unobserved* confounding variables exist.

(A4) The confounding covariates **V** can be endogenous; that is, they may not be independent of the unobserved determinant of the outcome.

Under those assumptions, we obtain

$$\begin{aligned} E(Y_x) &= \int E(Y_x \mid \mathbf{v}) f(\mathbf{v}) d\mathbf{v} = \int E(Y_x \mid X = x, \mathbf{v}) f(\mathbf{v}) d\mathbf{v} \text{(from the SITA assumption)} \\ &= \int E(Y_{obs} \mid X = x, \mathbf{v}) f(\mathbf{v}) d\mathbf{v} = \int \omega_x(\mathbf{v}) E(Y_{obs} \mid X = x, \mathbf{v}) f(\mathbf{v} \mid X = x) d\mathbf{v}, \text{for } x = 0,1, \end{aligned} \quad (1)$$

where $f(.)$ is the probability density function, and

$$\omega_x(\mathbf{v}) \equiv \frac{f(\mathbf{v})}{f(\mathbf{v} \mid X = x)} = \frac{p(X = x)}{p(X = x \mid \mathbf{v})}, \quad (2)$$

where lowercase *p* indicates the probability measure. The RCM estimate of the average treatment effect $E(Y_1) - E(Y_0)$ is obtained by using a consistent estimate of $p(X = x \mid \mathbf{v})$ by a logit or probit regression. Here and below, capital *L*, *X*, **X**, and **V**, where the boldface letters denote a set,

---

covariate, as in the case of the racial composition of students in schools that may differ between schools under treatment and those under no treatment. This variable may be correlated with individuals' treatment assignments and also may affect individual outcomes over and above the effect of individuals' race.

indicate variables, and lower case *l*, *x*, **x**, and **v** indicate their values. The expression of $p(X = x)$, for example, is often abridged as $p(x)$.

When the treatment variable is a latent-class variable, we can regard it as a variable of the treatment assignment. Although we do not observe this assignment directly, we can obtain a set of probabilities each of which indicates the probability that a person has been assigned to an alternative treatment group. No categories of the latent-class variable represent the control group without treatment, however. Hence, the causal analysis here is not concerned with an assessment of the treatment effect by the comparison of outcomes with treatment to those without treatment, but is concerned with an assessment of differences in the outcome among alternative treatments. Thus, in the case in which latent-class variable L is dichotomous and takes a value of 0 or 1, we denote by Y_{1i} the potential outcome under treatment 1, and by Y_{0i} the potential outcome under treatment 0, and $Y_{1i} - Y_{0i}$ indicates the difference in the outcome between the two alternative treatments. The SUTVA, therefore, needs to be modified such that a set of outcomes under alternative treatments for each subject is independent of the probabilities of other people's assignment to the alternative treatments. For a dichotomous latent-class variable L, $Y_{obs} = LY_1 + (1-L)Y_0$ still holds as in the RCM, even though neither Y_1 nor Y_0 is observed directly for any subject, and the average difference in the treatment effects between treatment 1 and treatment 0 is given as $E(Y_1) - E(Y_0)$.

We basically retain assumptions A2, A3, and A4. As for assumption A2, however, we will be concerned with differences in the outcomes among alternative treatments for the total population—except for an extension of Abadie's generalized DID for a latent-class treatment variable for the reason explained later. As for the SITA assumption, A3, we need to replace the observed treatment variable X with the latent-class variable L, so that $(Y_{1,}Y_0) \perp L \mid \mathbf{V}$ holds when L is dichotomous—and extend it to $(Y_{1,}Y_2,...,Y_m) \perp L \mid \mathbf{V}$ if L has m categories. Note, however, that for the analysis of panel survey data, the strong ignorability assumption is not necessary and will be replaced by a weaker assumption explained in a later section.

2.2. The Latent-Class Model and Treatment Effects

Although the model can be easily extended for the case of a latent-class variable having more than two categories, let us assume for simplicity of description that indicator variables reveal two latent classes. Thus, we assume a model for the conditional joint distribution of the latent-class variable L and its indicators **X**, for a given set of covariates **V**, such that

$$P(l, \mathbf{x}_i \mid \mathbf{v}_i) = P(l \mid \mathbf{v}_i) \prod_{j=1}^{J} P(x_{ij} \mid l, \mathbf{v}_i) \text{ , for } l = 0,1, \qquad (3)$$

where J variables **X** are indicators of latent-class variable L and satisfy *local independence* for each given state of the latent-class variable, and **V** is a set of confounding covariates. Here and below, capital P indicates a measure of expected probability from a given model, and lowercase p indicates a measure of probability not based on a model, including the proportion of people in the population.

We employ the "three-step estimation" (Bakk et al. 2013), with the application of a latent-class model as the first step, the assignment of a set of expected probabilities for being in alternative latent classes, such as in the *super population* that cross-classify population and latent classes, as the second step, and the estimation of the treatment effects for a given latent-class variable as the treatment variable as the third step. Generally, a one-step estimation, that is, a simultaneous estimation of parameters for the latent-class model and the outcome equation, not only requires a parametric specification of the outcome equation, but would also adjust the characteristics of the latent-class variable *L* to maximize the prediction of *Y*, but such a dependence of the characteristics of the treatment variable on the outcome variable is not desirable for causal analysis. It is also noteworthy that although the indictors **X** of the latent variable may have some unique effects on the outcome, controlling for *L*, we do not control their effects on the outcome in the estimation of the average treatment effects. This is because we consider the indicator variables **X** to be causal descendants of *L*, as is implicitly assumed in the local independence assumption of the latent-class model, and a control for the effects of causal descendants of *L* is inadequate in assessing the causal effect of *L* (Pearl 2009).

From equation (1), each person *i*'s probability of being a member of each latent class *l* is given by $P_{li} \equiv P(l \mid \mathbf{x}_i, \mathbf{v}_i) = P(l, \mathbf{x}_i \mid \mathbf{v}_i) / \left(P(1, \mathbf{x}_i \mid \mathbf{v}_i) + P(0, \mathbf{x}_i \mid \mathbf{v}_i) \right)$ for l = 0,1. In other words, each person *i* contributes P_{li} "observations" to the superpopulation that represent the set of combinations of persons in the population and latent classes—though randomness in P_{li} needs to be taken into account in the estimation of the standard error of the treatment effects.

Then, since $E(Y_1 \mid L = 1, \mathbf{v}) = E(Y_{obs} \mid L = 1, \mathbf{v}) = E_{P_1}(Y_{obs} \mid \mathbf{v})$ and $E(Y_0 \mid L = 0, \mathbf{v}) = E(Y_{obs} \mid L = 0, \mathbf{v}) = E_{P_0}(Y_{obs} \mid \mathbf{v})$ hold, where E_{P_1} and E_{P_0} indicate weighted means weighted by P_{1i} and P_{0i}, respectively, for person *i*, we obtain under the SITA assumption,

$$\begin{aligned} E(Y_l) &= \int E(Y_l \mid \mathbf{v}) f(\mathbf{v}) d\mathbf{v} = \int E(Y_l \mid L = l, \mathbf{v}) f(\mathbf{v}) d\mathbf{v} \text{ (from the SITA assumption)} \\ &= \int E(Y_{obs} \mid L = l, \mathbf{v}) f(\mathbf{v}) d\mathbf{v} = \int \omega_l(\mathbf{v}) E(Y_{obs} \mid L = l, \mathbf{v}) f(\mathbf{v} \mid L = l) d\mathbf{v} \\ &= \int \omega_l(\mathbf{v}) E_{P_l}(Y_{obs} \mid \mathbf{v}) f(\mathbf{v} \mid L = l) d\mathbf{v}, \text{ for } l = 0,1, \end{aligned} \qquad (4)$$

and

$$\omega_l(\mathbf{v}) \equiv \frac{f(\mathbf{v})}{f(\mathbf{v} \mid L = l)} = \frac{p(L = l)}{p(L = l \mid \mathbf{v})}. \qquad (5)$$

Hence, as in the standard RCM, the weights are inversely proportional to propensity scores. Equations (4) and (5) differ from equations (1) and (2), however, because the estimates of both $p(L=l)$ and $p(L=l\,|\,\mathbf{v})$ as well as the estimated P_{li} depend on the latent-class model. Accordingly, equations (4) and (5) indicate that $E(Y_l)$, the expected mean outcome under treatment *l*, is obtained as a *doubly weighted* average of Y_{obs} weighted for person *i* by (a) the IP weight for latent class *l*, $\omega_l(\mathbf{v}_i)$, and (b) the probability of being in latent class *l*, P_{li}, for each of *l* = 0 and 1.

For equation (3), we consider special cases rather than a general case. We first consider a model that is expressed as $P(l,\mathbf{x}_i\,|\,\mathbf{v}_i)=P(l\,|\,\mathbf{v}_i)\prod_{j=1}^{J}P(x_{ij}\,|\,l)$, where covariates **V** affect only the latent-class variable. We refer to it as a type-I model below, and since this model satisfies $P(\mathbf{x}\,|\,l,\mathbf{v})=P(\mathbf{x}\,|\,l)$, it assumes $\mathbf{V}\perp\mathbf{X}\,|\,L$. Since this is a strong assumption, we need to consider relaxing it. However, a very general dependence of **X** on **V** can make the substantive meaning of latent classes inconsistent across different covariates' values. Generally, the dependence of each indicator variable on the latent-class variable and covariates is expressed by the multinomial logit equation. I consider as an extension for the type-I model an additive-effects model that permits a subset of **V**, which we denote by **V***, to affect the intercepts of $\log\left(P(X_{ij}=k\,|\,l,\mathbf{v}_i^*)/P(X_{ij}=1\,|\,l,\mathbf{v}_i^*)\right)$ *without any interaction effects of covariates and the latent-class variable on the indicator variables,* and refer to this model as a type-II model below. This model has the same pattern and extent of association between the indicator variables and the latent-class variable, expressed by the set of log odds ratios between the two, and thereby retains the same substantive meaning for the latent-class variable across covariate states. Even if interaction effects of the latent-class variable and covariates on the indicator variables exist, it should not be a problem to employ type-I and type-II models as long as revealed characteristics of latent classes are stable in their patterns of responses to indicators regardless of a control for covariates. This is because we are concerned here primarily with the causal effect of a a meaningful summary measure of the set of indicators on the outcome. The merit in the use of the type-II model compared with the type-I model is that the former characterizes the selection process into alternative latent-class states more accurately. On the other hand, if an introduction of covariate effects on the latent-class variable changes the characteristics of latent classes because of the instability of latent-class estimates, the data for those indicators are not suitable for the use of the method introduced in this paper.

2.3. The Estimation of the Treatment Effect

From the latent-class model of equation (3), the estimate of the IP weights is given as

$$\frac{\hat{p}(l)}{\hat{p}(l \mid \mathbf{v})} = \frac{P(l)}{\sum_{\mathbf{x}} P(l, \mathbf{x} \mid \mathbf{v})} = \frac{P(l)}{P(l \mid \mathbf{v})\left(\sum_{\mathbf{x}} P(\mathbf{x} \mid l, \mathbf{v})\right)} = \frac{P(l)}{P(l \mid \mathbf{v})}. \tag{6}$$

If $\hat{p}(l \mid \mathbf{v})$ is a consistent estimate of $p(l \mid \mathbf{v})$, then

$$\begin{aligned} \operatorname{plim}_{N \to \infty} \frac{1}{N} \sum_i \frac{\hat{p}(l)}{\hat{p}(l \mid \mathbf{v}_i)} &= \int_{\mathbf{V}} \frac{p(l) f(\mathbf{v} \mid l)}{p(l \mid \mathbf{v})} d\mathbf{v} \\ &= \int_{\mathbf{V}} \frac{f(\mathbf{v}, l)}{f(\mathbf{v}, l) / f(\mathbf{v})} d\mathbf{v} = \int_{\mathbf{V}} f(\mathbf{v}) d\mathbf{v} = 1, \end{aligned} \tag{7}$$

and, therefore, the IP weights have an asymptotic mean of 1. However, it is more efficient to employ the ratio estimator, so that the average weights sum to $NP(l)$ for each l in the sample. Hence, we will employ the adjusted IP weights that are defined as

$$w_{li} \equiv NP(l) \frac{P(l)}{P(l \mid \mathbf{v}_i)} \Big/ \left\{ \sum_{k=1}^{N} P_{lk} \frac{P(l)}{P(l \mid \mathbf{v}_i)} \right\} = \frac{NP(l)}{P(l \mid \mathbf{v}_i)} \Big/ \left\{ \sum_{k=1}^{N} \frac{P_{lk}}{P(l \mid \mathbf{v}_i)} \right\}, \tag{8}$$

where $P_{lk} \equiv P(l \mid \mathbf{x}_k, \mathbf{v}_k) = P(l, \mathbf{x}_k \mid \mathbf{v}_k) / \left(\sum_{l=0}^{1} P(l, \mathbf{x}_k \mid \mathbf{v}_k) \right)$ and is the expected probability that a person *k* will be in latent class *l* based on the model of equation (3).

It follows that we obtain the mean of Y_1 and that of Y_0, respectively, by the ratio estimator as

$$\hat{E}(Y_1) = E_{w1}(Y_{obs} \mid L = 1) = \frac{\sum_{i=1}^{N} P_{1i} w_{1i} y_i}{\sum_{i=1}^{N} P_{1i} w_{1i}} = \sum_{i=1}^{N} w_{1i}^{*} y_i \ , \qquad (9)$$

where $w_{1i}^{*} \equiv P_{1i} w_{1i} \Big/ \sum_{k=1}^{N} P_{1k} w_{1k}$ and E_{w1} indicate the weighted mean with weights $\{w_{1i}\}$, and

$$\hat{E}(Y_0) = E_{w0}(Y_{obs} \mid L = 0) = \frac{\sum_{i=1}^{N} P_{0i} w_{0i} y_i}{\sum_{i=1}^{N} P_{0i} w_{0i}} = \sum_{i=1}^{N} w_{0i}^{*} y_i \ , \qquad (10)$$

where $w_{0i}^{*} \equiv P_{0i} w_{0i} \Big/ \sum_{k=1}^{N} P_{0k} w_{0k}$ and E_{w0} indicate the weighted mean with weights $\{w_{0i}\}$.

An extension of the method for doubly robust estimation (Bang and Robins, 2005), for which the treatment effect becomes consistent if either the outcome regression or the equation for the estimation of the propensity score, but not necessarily both, is correct, is possible by assuming the following regression model for the outcome combined with the latent-class model of equation (3):

$$Y_i = \alpha + \beta l_i + \phi(\mathbf{v}_i, \boldsymbol{\gamma}) + \varepsilon_i \ , \qquad (11)$$

where $\phi(\mathbf{v}_i, \boldsymbol{\gamma})$ is a parametric linear function of **V**, $\boldsymbol{\gamma}$ indicates parameters included in the function, and the error term is assumed to be random and independent of *L* and **V**. We then apply equation (11) in the third step of the three-step estimation to the data weighted $P_{li} w_{li}$ to obtain the doubly robust estimate by using the fact that a consistent estimate for parameter β satisfies the minimization of the following expected weighted sum-squared errors:

$$E(SSE)=\sum_i \sum_{l=1}^{2} P_{li} w_{li} \left(y_i - \alpha - \beta l_i - \phi(\mathbf{v}_i, \boldsymbol{\gamma}) \right)^2 . \qquad (12)$$

Note, however, that while this doubly robust estimation is useful to check the consistency of its estimate for the treatment effect with the RCM estimate, equation (11) assumes that the treatment effect is homogenous and covariates **V** are exogenous, neither of which is assumed in the RCM. Hence we employ the doubly robust estimation only for a diagnostic purpose for checking the consistency of its estimate with the RCM estimate.

2.4. The M-Estimation of the Standard Error of the Average Treatment Effect

By treating weights $\{\mathbf{w}_1^*, \mathbf{w}_0^*\}^T$ in equations (9) and (10) as given and fixed, and assuming that errors are independent across observations and $V(Y_i)$ is homogenous, we obtain the weighted least squares estimate for the standard error of the treatment effect.[3] However, in addition to making strong assumptions that are not made in the RCM, this method is likely to generate a downward bias in the estimate of the standard error, as does its use for the three-step estimation for the latent-class model (Bakk et al. 2013). In this paper, I employ the M-estimation method, which treats weights as random variables and does not assume either independent errors or the homogeneity of $V(Y_i)$. Although some methods for the adjustment of standard errors for the latent-class model with the three-step estimation have been developed (Bakk et al. 2013), we do not employ any of those methods, because we also need to take into account as sources of variability in the estimate of the treatment effects not only variability in the estimates of latent classes but also variability in the estimates of propensity scores. Note also that since the RCM assumes unspecified heterogeneity in the treatment effects, any method for the estimation of the

[3] When weights are given and fixed and if can assume that $V(Y_i)$ is homogenous, we then obtain $V(E(Y_1)-E(Y_0))=\left\{\sum_{i=1}^{N}(w_{1i}^*-w_{0i}^*)^2\right\}V(Y)$, where the estimate of *V(Y)* is simply a sample variance of *Y*.

standard error, such as the maximum likelihood method, that requires a parametric outcome model cannot be employed.

In the M-estimation (Huber 1964, 1967; Wooldridge 2002), we assume an M-function, $m(\mathbf{y},\boldsymbol{\theta}_0)$ which satisfies—for a set of independent observations **y** and a set of *n* parameter values $\boldsymbol{\theta}_0$—a set of *n* equations

$$\int m(\mathbf{y},\boldsymbol{\theta}_0)dF(\mathbf{y})=\mathbf{0} \tag{13}$$

for the population distribution $F(\mathbf{y})$ of **y**. If $\boldsymbol{\theta}_0$ is the unique solution for the set of equations, then the sample estimate of $\boldsymbol{\theta}_0$, $\hat{\boldsymbol{\theta}}$, that satisfies $\sum_{i=1}^{N} m(\mathbf{y}_i,\hat{\boldsymbol{\theta}})=\mathbf{0}$ for the sample is consistent and asymptotically normally distributed (Huber 1964, 1967), and $\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ has an asymptotic variance $\mathbf{V}(\boldsymbol{\theta}_0)$ given by

$$\mathbf{V}(\boldsymbol{\theta}_0)=\mathbf{A}(\boldsymbol{\theta}_0)^{-1}\mathbf{B}(\boldsymbol{\theta}_0)(\mathbf{A}(\boldsymbol{\theta}_0)^{-1})^T, \tag{14}$$

where $\mathbf{A}(\boldsymbol{\theta}_0)\equiv E\left(-\frac{\partial}{\partial\boldsymbol{\theta}^T}m(\mathbf{y},\boldsymbol{\theta}_0)\right)$, whose sample estimate is $\mathbf{A}(\mathbf{y},\hat{\boldsymbol{\theta}})=-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\boldsymbol{\theta}^T}m(\mathbf{y}_i,\hat{\boldsymbol{\theta}})$,

and $\mathbf{B}(\boldsymbol{\theta}_0)\equiv E\left(m(\mathbf{y},\boldsymbol{\theta}_0)m(\mathbf{y},\boldsymbol{\theta}_0)^T\right)$, whose sample estimate is $\mathbf{B}(\mathbf{y},\hat{\boldsymbol{\theta}})\equiv\frac{1}{N}\sum_{i=1}^{N}m(\mathbf{y}_i,\hat{\boldsymbol{\theta}})m(\mathbf{y}_i,\hat{\boldsymbol{\theta}})^T$.

An M-estimator with $\boldsymbol{\theta}_0=(E(Y_1),E(Y_0))^T$ can be obtained by specifying the M-function as follows by using equation (9) and (10):

$$m(\mathbf{y},\mathbf{w}_1^*,\mathbf{w}_0^*,E(Y_0),E(Y_1))=\begin{pmatrix} w_1^*(y-E(Y_1)) \\ w_0^*(y-E(Y_0)) \end{pmatrix}. \tag{15}$$

It follows that

$$\mathbf{A}(\boldsymbol{\theta}_0) = E\begin{pmatrix} w_1^* & 0 \\ 0 & w_0^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \text{ and}$$

$$\mathbf{V}(\boldsymbol{\theta}_0) = \mathbf{B}(\boldsymbol{\theta}_0) = E\begin{pmatrix} (w_1^*)^2(y - E(Y_1))^2 & w_1^* w_0^*(y - E(Y_1))(y - E(Y_0)) \\ w_1^* w_0^*(y - E(Y_1))(y - E(Y_0)) & (w_0^*)^2(y - E(Y_0))^2 \end{pmatrix}.$$

We then obtain an M-estimation for the variance of $E(Y_1) - E(Y_0)$ as

$$\frac{1}{N^2}\sum_{i=1}^{N}\left\{(w_{1i}^*)^2(y_i - E(Y_1))^2 - 2w_{1i}^* w_{0i}^*(y_i - E(Y_1))(y_i - E(Y_0)) + (w_{0i}^*)^2(y_i - E(Y_0))^2\right\} \quad (16)$$

However, this estimator does not take into account the fact that two sets of weights $\{\mathbf{w}_1^*, \mathbf{w}_0^*\}$ are estimated from the latent-class model and the equation for the estimation of propensity score and are not observed. Hence, a revised M-function with $\boldsymbol{\theta}_0 = \{E(Y_1), E(Y_0), \boldsymbol{\theta}_1)^T$ can be defined as

$$m(\mathbf{y}, \mathbf{v}, \mathbf{x}, \mathbf{w}_1^*, \mathbf{w}_0^*; \boldsymbol{\theta}_1, E(Y_0), E(Y_1)) = \begin{pmatrix} w_1^*(\boldsymbol{\theta}_1)(y - E(Y_1)) \\ w_0^*(\boldsymbol{\theta}_1)(y - E(Y_0)) \\ \dfrac{\partial}{\partial \boldsymbol{\theta}_1}\log L(\mathbf{x}, \mathbf{v} \mid \boldsymbol{\theta}_1) \end{pmatrix}, \quad (17)$$

where $L(\mathbf{x}, \mathbf{v} \mid \boldsymbol{\theta}_1)$ is the likelihood function of the latent-class model, and $\boldsymbol{\theta}_1$ is the set of parameters of the model. This revised method takes into account variability associated with the estimates of propensity scores and the estimates of the set of probabilities that each subject will be being in alternative latent classes more accurately, because it takes into account the presence of parametric constraints in the estimation.

For a doubly robust estimation with a regression extension, we can obtain the M-function by using the fact that parameter estimates are obtained by setting the first-order derivatives of the expected sum-squared errors of equation (12), which becomes 0 with respect to each parameter, as additional components of the M-function.[4]

2.5. The Extension to the Latent-Class Variable Having Three or More Categories

The extension of the method described above for cases where the latent-class treatment variable has three or more categories is straightforward, because the equations given above are unconstrained by the categories of *L* and can be generalized very easily for a case where *L* has three or more categories. The standard error of the difference between two alternative treatments can be calculated for each pair of treatments separately for the simplified M-estimation method, but an entire set of the covariance of $E(Y_l)$, l = 1,…, *m*, needs to be estimated for the revised method that includes the likelihood function of the latent-class model in the M-function.

2.6. Diagnostic Analyses

This paper employed two diagnostic methods to check the adequacy of the construction of the propensity scores. One is a test of statistical independence after the IP weight adjustment between the latent-class variable and each covariate. This criterion is necessary but not sufficient, because the independence of the marginal distribution of each covariate from the latent-class variable does not guarantee the independence of the joint distribution of all covariates from the latent-class variable.

---

[4] An executable program that can calculate both the simplified and revised estimates of standard errors by the M-estimation is available from the author by request with a user manual. The program requires the results from the latent-class models, including parameter estimates and conditional probabilities, which are available by the use of LEM or Latent Gold, as well as the individual-level data for the dependent variables and covariates as input data. However, the program is currently developed only for models with a set of dichotomous indicator variables for the revised method of M-estimation—although models with indicators having three or more categories are allowed for the use of a simplified method of M-estimation without including the likelihood function of the latent-class model in the M-function.

The second method is to test for reasonable agreement in the treatment effect between the RCM estimate, based on the pair of equations (9) and (10), and the doubly robust estimate, based on the outcome regression equation (11) for the IP-weighted data. If the use of the IP weights attains statistical independence of L and $\mathbf{V}$, the results should be nearly identical.

It is noteworthy that the common area of support for the propensity score is not an issue when the latent-class variable is used as the treatment variable. In the standard application of the propensity-score matching and weighting methods, the issue of the common area of support arises when there are observations for one of the states of the treatment variable but not for other states for a particular range of propensity scores. In the case where the treatment variable is a latent-class variable, however, we have "observations" for each treatment state for each observed sample subject—unless $P(l \mid \mathbf{x}, \mathbf{v})$ takes a value of 0 for a particular state of L. Since covariates themselves are *balancing scores* (Rosenbaum and Rubin 1983), we almost always have corresponding "observations" of covariate states across the states of the treatment variable. This is a significant merit of the method that employs a latent-class variable as the treatment variable.

2.7. Some Extensions of the Method for the Analysis of Panel Survey Data

I describe below only methods where we have a single time measurement for the treatment variable. A method with more than a single time measurement for the latent-class treatment variable will require the modeling of latent transitions (Collins and Lanza 2009) among treatment states but is not described here. One simple extension is that when we have a measurement of outcomes at times s and t where $t > s$, and a measurement of the indicators of the latent-class treatment variable between times s and t, we can include Y_s in covariates $\mathbf{V}$. Although we still make a strong ignorability assumption, $(Y_{1t,} Y_{0t}) \perp L \mid \mathbf{V}$, one major potential cause of endogeneity of the treatment variable, namely, the effects of Y_s on L, is removed in this panel data method.

Extensions of the DID and Abadie's (2005) generalized DID for the case with a latent-class treatment variable are also possible but require an additional minor assumption, as explained below. Generally, the use of the DID assumes that every subject is under the control state of no treatment at time *s*, and only subjects in the treatment group get treated by a particular later time *t*. Let *X* be the variable of treatment assignment that takes a value of 1 for the treatment group, and a value of 0 for the control group. We will use 1 and 0 as subscripts to *Y* to denote treatment and no treatment, respectively. It follows that

$$\begin{aligned} \text{DID} &\equiv E(Y_{obs,t} - Y_{obs,s} \mid X = 1) - E(Y_{obs,t} - Y_{obs,s} \mid X = 0) \\ &= E(Y_{1t} - Y_{0s} \mid X = 1) - E(Y_{0t} - Y_{0s} \mid X = 0). \end{aligned} \tag{18}$$

Note that the DID is not affected by any *unobserved time-constant confounder* of *X* and *Y*. On the other hand, the average treatment effect for the treated (ATT) is defined as

$$\text{ATT} = E(Y_{1t} - Y_{0t} \mid X = 1)\,. \tag{19}$$

It follows from equations (18) and (19) that DID = ATT holds if the following condition holds:

$$E(Y_{0t} - Y_{0s} \mid X = 1) = E(Y_{0t} - Y_{0s} \mid X = 0)\,. \tag{20}$$

Equation (20) implies that if the time effect on *Y under no treatment* is independent of the treatment assignment, DID = ATT holds. However, since this condition is strong, Abadie (2005) replaces the condition with the following weaker condition:

$$E(Y_{0t} - Y_{0s} \mid X = 1, \mathbf{v}) = E(Y_{0t} - Y_{0s} \mid X = 0, \mathbf{v})\,, \tag{21}$$

where **V** are covariates that affect both $Y_{0t} - Y_{0s}$ and *X* and include time-varying covariates of *Y* evaluated at time *s* and covariates with time-varying effects on *Y*. Note that time-constant covariates with time-constant effects are not components of **V**, because they cannot affect

$Y_{0t} - Y_{0s}$. Equation (21) is satisfied when $(Y_{0t} - Y_{0s}) \perp X \mid \mathbf{V}$, which is a much weaker ignorability condition than the SITA condition, holds. Then we can employ the propensity-score weights that make the distribution of **V** for the control group equal to that of the treatment group, that is,

$$\omega_{01}(\mathbf{v}) \equiv \frac{p(\mathbf{v} \mid X = 1)}{p(\mathbf{v} \mid X = 0)} = \frac{p(X = 0)p(X = 1 \mid \mathbf{v})}{p(X = 1)p(X = 0 \mid \mathbf{v})}, \tag{22}$$

and calculate the DID with those weights for the control group and constant weight 1 for the treatment group. Then we obtain the DID estimate of the ATT under the weaker assumption of $(Y_{0t} - Y_{0s}) \perp X \mid \mathbf{V}$. This is Abadie's generalized DID method.

An extension of the DID and the generalized DID for a latent-class treatment variable is possible for data that satisfy the following conditions: (1) all subjects were in the state of no treatment at time *s*, (2) a set of dichotomous indicators **X** of the latent-class variable indicate an occurrence ($X_j = 1$) versus nonoccurrence ($X_j = 0$) of treatment element *j* and they were applied in various combinations between times *s* and *t* for subjects, and (3) there is a particular latent class, which we specify as $L = 1$, for which $P(X_j = 1 \mid L = 1)$ is very small for every indicator *j*.

Condition (3) is usually satisfied when there is a significant proportion of sample subjects for which **X** = 0 holds for all indicators at time *t*. Although the illustrative application presented below focuses on the analysis of a set of dichotomous work-life balance policies of employers on income among regularly employed women with cross-sectional survey data, the data requirement above may apply if we analyze data of subjects employed in firms that did not start any such work-life balance policies at time *s*. The difference from the DID or the generalized DID is that (1) the treatment variable is a latent-class variable, and (2) no categories of the latent-class variable exactly correspond to the control state of no treatment at time *t*.

For a latent-class variable L we define the *latent* DID, abbreviated as L-DID, for a pair of $L = l$ and $L = 1$ as

$$\text{L-DID}(l) \equiv E(Y_{lt} - Y_{0s} \mid L = l) - E(Y_{1t} - Y_{0s} \mid L = 1) = E_{wl}(Y_{obs,t} - Y_{obs,s}) - E_{w1}(Y_{obs,t} - Y_{obs,s}), \quad (23)$$

where subscript 0 implies "no treatment," and E_{wl} denotes the weighted mean with the probability that each person i is in latent class l, P_{li}, as weights.

The treatment effect of receiving treatment l compared with receiving treatment 1 among persons in latent class l, which we refer to as the latent average treatment effect for the treated for treatment l and denote by L-ATT(l), is defined as

$$\text{L-ATT}(l) = E(Y_{lt} - Y_{1t} \mid L = l). \quad (24)$$

If we assume the equivalent of equation (20), that is,

$$E(Y_{0t} - Y_{0s} \mid L = l) = E(Y_{0t} - Y_{0s} \mid L = 1), \quad (25)$$

and indicate that the time effect under no treatment is independent of treatment assignment L, and in addition assume for treatment 1, based on the data condition (3) given above, that its treatment effect is nearly 0 and therefore has no significant variability among subjects of different latent classes, so that

$$E(Y_{1t} - Y_{0t} \mid L = l) = E(Y_{1t} - Y_{0t} \mid L = 1) \quad (26)$$

holds. We then obtain L-DID(l) = L-ATT(l). Note that $E(Y_{1t} - Y_{0t} \mid L = l)$ indicates the effect of treatment 1 compared with no treatment (treatment 0) among people who were assigned to treatment l. Note that unlike the DID, the L-DID assumes that the effects of a particular treatment, treatment 1, on Y are homogenous and do not depend on treatment assignment L, and this new condition is introduced because we do not observe an outcome under the control state at time t, Y_{0t}, for anybody in the sample when the treatment variable is a latent-class variable, A further extension of the L-DID for Abadie's generalized DID, which we will refer to the generalized L-DID, permits a replacement of the assumption of equation (25) by the following weaker assumption:

$$E(Y_{0t} - Y_{0s} \mid L = l, \mathbf{v}) = E(Y_{0t} - Y_{0s} \mid L = 1, \mathbf{v}), \qquad (27)$$

which follows from the assumption of $(Y_{0t} - Y_{0s}) \perp L \mid \mathbf{V}$. Then we can calculate the L-DID estimate for the L-ATT for people assigned to latent class *l* by applying weights

$$\omega_{01}(\mathbf{v}) \equiv \frac{p(\mathbf{v} \mid L = l)}{p(\mathbf{v} \mid L = 1)} = \frac{p(L = 1)\, p(L = l \mid \mathbf{v})}{p(L = l)\, p(L = 1 \mid \mathbf{v})} \qquad (28)$$

for the "observations" in latent classes L = 1 and a constant weight of 1 for $L = l$, respectively, under the weaker assumption. The covariates here are those which we expect to have effects on both $Y_{0t} - Y_{0s}$ and L.

We may be interested in the difference in the average treatment effect between treatment *l* and treatment 1 for the total population rather than ATT(*l*), which is the average treatment effect among people of latent class *l*. Unlike the method for cross-sectional survey data that makes the SITA assumption, however, the generalized L-DID as well as the generalized DID method does not allow us to convert the estimate of the ATT to the estimate of the average treatment effect by changing the covariate distribution by the propensity-score weights. Generally, the SITA assumption of $(Y_1, Y_0) \perp X \mid \mathbf{V}$ consists of (1) the conditional independence of the pretreatment outcome from the treatment assignment, $Y_0 \perp X \mid \mathbf{V}$, and (2) the conditional independence of the treatment effect from treatment assignment, $(Y_1 - Y_0) \perp X \mid \mathbf{V}$. The latter implies that the treatment effect is independent of the assignment, controlling for **V,** and is satisfied if the treatment effect is a function of **V** plus some variability independent of *X*. That is the key condition that enables the conversion of the ATT to the average treatment effect by changing the covariate distribution. However, the generalized L-DID makes a much weaker ignorability assumption of $(Y_{0t} - Y_{0s}) \perp L \,|\, \mathbf{V}$, which does not satisfy this key condition.

3. APPLICATION

3.1. Data and Preliminary Latent-Class Analysis

The application presented here employs data from the Comparative Survey of Work-Life Balance conducted in 2009 by the Research Institute of Economy, Trade and Industry (RIETI) in Japan for the population of employees in four countries. The survey for Japan includes a nationally representative sample of white-collar regular employees in firms with 100 or more employees, with information collected from both employees and their employers. The following analysis uses the sample of employed women of ages 23–59 in Japan, with a sample size of 2,349, from 1,103 employers. Hence, while it is not suitable for a two-level analysis, because the average number of employees per employer is too small (2.3), the data set still has the merit of having information about firms' policies on work-life balance collected directly from employers. The dependent variable is annual earned income measured in 10,000 yen (about 107 dollars in 2009), collected from the employee survey, with a mean of 341.0 and a standard deviation of 158.9, and ranges in the sample from 100 (1 million yen) to 3,300 (33 million yen). Since we are interested in evaluating whether firms' policies on work-life balance affect hourly wages among women, we use average hours of work per week as a control variable.

As the indicators of employers' work-life balance policies, six dichotomous items collected from the employer survey are used. Their average probabilities of positive responses, which imply the presence of a particular policy, are given in Table 1 for the total sample and for samples by firm-size categories.

(Table 1 about here)

As Table 1 shows, one particular policy, namely, a part-time work system for child and family care, was quite prevalent, being available to an average of 73% of employees. In fact, this policy became a statutory entitlement for employees employed by firms with 100 or more employees in the Revised Act on Child and Family Care in 2009, and the law was implemented in 2010, within one year after the survey was conducted. Hence, many employers may have enforced the new legal obligation before its implementation. Table 1 shows that the prevalence of all other policies related to balancing of work and family roles is low, less than 30%, especially for the telecommuting system, which was available to only 4% of employees in the sample. The results by firm size, however, indicate that the prevalence of each policy significantly increases with firm size. Policy I5 may need a description. In Japan it is typical for the personnel department to determine personnel matters, such as recruitments, positional assignments, job transfers, and promotions, with little consideration for employees' intentions or career plans. The presence of an in-house career development system or an in-house open recruitment system, which characterizes policy I5, indicates a nontraditional personnel system where vacant positions to be filled in the firms are based on internal applications by employees and/or a related system that takes into consideration employees' careers plans. As shown in Table 1, only slightly more than 10% of employers with 100 or more employees made such a system available to employees.

Table 2 presents the results from the application of latent-class models with the six indicators without covariates for different number of latent classes. Although none of the latent-class models fit the data according to the likelihood-ratio test, we employ the Bayesian information criterion (BIC) as the criterion of model selection here because we are interested in using the model that adequately characterizes the data with the most parsimonious use of parameters and from which we can therefore expect a robust characterization of data. The BIC here indicates a comparison with the saturated model, with a negative number indicating more parsimony in the use of parameters than the saturated model. A model with the highest absolute value for a negative BIC value is the most parsimoniously fitting model according this criterion. The results thus show the model with three latent classes is the most parsimoniously fitting model.

(Table 2 About Here)

3.2. Main Analysis

3.2.1. Model selection of the latent-class model

In identifying the best-fitting type-I and type-II latent-class models with covariates described in Section 2.2, the following model selection strategy was adopted. Based on the preliminary analysis reported above, I assumed that the three-class model was adequate and imposed this characteristic on all models tested with covariates. Six covariates were considered potential confounders. Three were human-capital variables: (1) educational attainment (with 4 categories), (2) duration of employment for the same employer (with 8 categories), and (3) age (with 7 categories). One variable for labor, namely, (4) hours of work per week, was employed as the control variable because of our analytical interest in hourly wages. Two firm-level variables, (5) firm size (with 3 categories) and (6) industry (with 3 categories—[a] manufacturing, [b] wholesale and retail trades, and [c] other), are also considered. Both duration of employment with the employer and age affect individual income very strongly in Japan, since many firms still have an age- and/or duration-based wage raise system, called the *nenko* (year-based merit) system. The three industries distinguished by variable (6) are the major industries into which firms in the data for women in the sample are divided, with 59.1% in manufacturing, 33.6% in wholesale and retail trade, and 7.3% in "other" industries. Those six variables all affect the individual income of women strongly.

Type-I models assume the presence of covariate effects only on the latent-class variable. For identifying the best-fitting type-I model, the additive effects of the six covariates, all expressed as categorical variables, were initially tested, but three of them (age, duration of employment, and firm's industry) were dropped because of the statistical insignificance of their effects. For the remaining three covariates (educational attainment, firm size, and hours of work), pairwise interaction effects among those covariates were also tested but were not retained in the final model because of their insignificance. The best-fitting type-I model thus includes the main effects of those three covariates.

Type-II models introduce covariate effects on the intercepts of indicator variables as long as they improve the fit with the data. The criterion I employed for model selection for this purpose is the BIC, in order to include only the major covariate effects on the intercepts. In order to identify the best-fitting type-II model, the following procedure was employed: Starting from the best-fitting type-I model (model 1), the effect of each of the three confounding covariates on the intercept of each of the six indicator variables was tested first. Eighteen models were thus compared with the type-I model. Among those that showed a significant improvement over model 1, each factor whose addition showed a significant improvement was added sequentially in the order of importance, measured by the extent of improvement in the BIC, and the addition was stopped when no further improvement was made. Only the effects of firm size on the intercept of four indicators were found significant and retained in the best-fitting type-II model. The statistical program LEM (Vermunt 1997) was employed in order to test various latent-class models with data.

3.2.2. Results for the measurement component of the latent-class model

Table 3 presents the outcome of the *measurement component* of latent-class models, that is, the proportion of each latent class and the conditional probabilities of responses to the six indicator variables for the best-fitting type-I model (model 1) and the best-fitting type-II model (model 2). For comparison, results from the latent-class model without covariates are also presented. The LEM syntax for the application of models 1 and 2 is described in the Appendix.

(Table 3 About Here)

As Table 3 shows, one latent class, L1, has consistently higher probabilities of the presence of the six work-life balance policies. Even the telecommuting system, whose prevalence is about 4%, as shown in Table 3, is available to about 56% of employees in latent class L1 (by model 2). This latent class, which we can regard as the class most supportive of work-life balance policies, is very small (5.4%) in the results from model 2, however.

On the other hand, the largest latent class, class L3, which contains 48% of employees in the model 2, has very low response probabilities for five policies other than policy I2. Even for the relatively prevalent policy I2, a system of part-time work for child and family care, which was made a legal entitlement during the survey year, the probability of its presence in this latent class is less than 60%, and it is much less than that of the other two latent classes. Hence, we may regard latent class L3 as the class least supportive of work-life balance policies.

Latent class L2, which is close in size to latent class L3 and contains 47% of employees in the sample, according to model 2, has a high probability of supporting policy I2, part-time work for child and family care, and middle-level support for two other policies, namely, childcare leave above the statutory minimum (I1) and establishment of a special department or center for the promotion of work-life balance (I6). However, this latent-class model is not very supportive, at least compared with latent class L1, for three other policies, namely, flextime work, telecommuting, and the adoption of an in-house career development system or an open recruitment system. Hence, we can regard this latent class as a class that is moderately supportive of family and childcare but is not supportive of policies to promote workplace flexibility.

Table 3 also shows a variable extent of association between the latent-class variable and the indicators—though the associations are all strong and significant. The results from model 2 suggest that the association is strongest, in terms of the Wald chi-square statistic, for I5, the presence of an in-house career development system or an open recruitment system, and the weakest for I6, the establishment of a department or center to support work-life balance.

3.2.3. Results for the regression component of the latent-class model

Table 4 presents the results of the *regression component* of each of the two models. Model 1 includes only the effects of covariates on the latent-class composition, and model 2 adds significant covariate effects on the intercept of the logit for the indicator variables. It turns out that only firm size affects the indicator variables beyond its effect on the latent-class variable.

Table 4 shows consistently across the two models that compared with latent class L3, which is least supportive of work-life balance policies, latent class L1, which is most supportive of the policies, has three major characteristics. First, it has more regular employees working less than 40 hours a week. In Japan, employers that permit regular employees to work part-time are still very rare, and employees who wish to work part-time typically work as irregular employees. This is one major aspect of the lack of workplace flexibility in Japan. However, Table 4 suggests that the situation differs for employees working for the L1-type firms. Second, there are a higher proportion of college graduates in latent class L1. Third, the relative proportion of the latent class L1 compared with L3 increases with firm size.

On the other hand, selection bias in latent class L2 exists only for the tendency that as firm size increases, the relative proportion of employees in latent class L2, which supports child and family care but not workplace flexibility, increases, compared to employees in latent class L3, which is least supportive of work-life balance policies.

Results from model 2 show further that probabilities of the presence of some work-life balance policies vary with firm size beyond their variability with the latent class variable. Policies I1, I3, I5, and especially I6, the presence of an in-house career development system or an in-house open recruitment system, seems to vary more with firm size over and above the association between firm size and the latent-class variable.

3.2.4. A diagnosis of the independence between the latent-class variable and covariates after the IP weighting

Based on the results from two latent-class models, models 1 and 2, I constructed the IP weights, and Table 5 presents the diagnostic analysis of the attainment of independence of the latent-class variable from each of the three confounding covariates in the IP-weighted sample. The chi-square tests of statistical independence before and after the application of the IP weights are presented for latent-class models 1 and 2.

(Table 5 About Here)

Table 5 shows that propensity scores obtained from models 1 and 2 are both successful in attaining statistical independence of the latent-class variable from each of the three covariates according to this diagnostic test.

3.2.5. Treatment effects

Table 6 presents the main analytical results: the estimates of the treatment effects on individual income with and without a covariance-based control for covariates. For comparison, the estimates of the treatment effects without applying the IP weights are also presented. The results from a model with neither control of covariates by IP weights nor control through covariance reflect the effect including the selection bias of covariates. Alternative standard errors (M1 and M2) of the treatment effect based on the M-estimation are also presented in Table 6, and they are estimated in ways described in Section 2.3 as well as in a footnote to Table 6.

(Table 6 About Here)

Table 6 consistently indicates that female employees employed by firms that is supportive of work-life balance policies (latent class L1) have significantly the largest average income, those employed by firms that support childrearing and family care but introduce little workplace flexibility (latent class L2) have the second-largest average income, and those employed by firms that neither support childrearing and family care nor introduce workplace flexibility (L3) have the lowest average income, controlling for hours of per week, educational attainment, and firm size, which affect both the latent-class composition and income. However, a comparison between the results before the IP weighting and after the IP weighting, without the control of covariate effects on the outcome, shows that more than 60% [0.63 = (89.03-33.04)/89.03] of the difference in annual income between L1 and L3, which amounts to 890,300 yen, according to model 2, is explained by selection bias in those two states, and the estimate of the difference in income between employees in L1-type firms and those in L3-type firms is 330,400 yen after controlling for the selection bias. On the other hand, the difference in annual income between employees in L2-type firms and those in L3-type firms, 249,800 yen, according to model 2, is subject to less selection bias, since less than 30% [0.27 = (24.98-18.26)/24.98] of the difference is explained by selection bias.

Table 6 shows a high consistency in the results from IP-weighted models between the model that does not include confounding covariates in the outcome regression and the model that includes them, and, therefore, the results pass the criterion of the second diagnosis regarding the attainment of statistical independence between the treatment variable and the confounding covariates by the IP weighting. On the other hand, a comparison of treatment effects by the two models with a covariate control for the outcome regression shows a large discrepancy in the effects of L1 and L3 on income between the model with the IP weighting and the model without it. This result indicates that the covariance-based control for confounders is not effective in this case.

The standard errors of the treatment effect vary to some extent with the choice of methods between M1 and M2, indicating its slight dependence on whether we take into account the fact that weights are estimates from a specific latent-class model or not. The results suggest that standard errors are slightly underestimated by the simplified M-estimation method. However, the simplified method may not throw off the test of significance for strong treatment effects, because its bias is small.

Since the failure of covariance-based control for confounding variables may be a result of using a linear regression model on income, where the nonlinearity problem may occur, Table 7 replaces the dependent variable by the logarithm of the income. The results, including the significance of the two treatment effects, and the extent to which each is explained as a result of selection bias, are very similar to the results of Table 6. The covariance-based control for the confounding variable is not successful here either, because of the large discrepancy in the effects of L1 and L3 between the model with the IP weighting and the model without.

Table 7 also indicates that firm type L1 leads to about 10% [1.097 = exp(0.093)] higher income on average than firm type L3, and firm type L2 leads to about 4% [1.044 = exp(0.043)] higher oncome on average than firm type L3 among female regular white-collar employees, controlling for hours of work and selection bias in employment in different types of firms.

(Table 7 About Here)

Table 8 presents the treatment effects on income and on log income by education. The results consistently show that although work-life balance policies of employers, both comprehensive policies represented by L1, and childcare support policies represented by L2, raise income of college graduates and the graduates of advanced vocational schools, they do not make any difference for junior college graduates and for those with high school graduation or below.

(Table 8 about here)

CONCLUDING REMARKS

This paper introduced a method using the inverse-probability weighting of the propensity score for causal analysis where the treatment variable is a latent-class variable with a set of indicators. This method will be useful if the treatment variable of interest cannot be indicated by a single observable variable but is indicated by a set of categorical variables, as in the case of gender-role attitude, political attitude, or other attitudinal or psychological states. In addition, if we have two-level data for individuals and their social contexts, such as firms, schools, and neighborhoods, and a set of categorical variables that characterize the social contexts, we may have a latent-class variable that characterizes a specific aspect of the social contexts, and it can be used as a treatment variable to assess school effects, firm effects, or neighborhood effects regarding that aspect of the social context on the outcome.

Using the method introduced in this paper, an application to an analysis of the effect of employers' policies on work-life balance on female regular workers' income in Japan has shown that the presence of certain policies by employers affects income. Differences in income among those employed by different types of firms arise in part as a result of selection bias, and in particular from the fact that larger firms are more likely to adopt those policies and, at the same time, female employees of those larger firms tend to have higher income. However, the results for the estimate of the treatment effects that controls for such selection bias based on the method introduced in this paper show that among female regular white-collar employees of three types of firms, those employed by firms which are most supportive of work-life balance policies have the highest average income, and those employed by firms that are least supportive to those policies have the lowest average income, controlling for hours of work, educational attainment, and firm size, which affect both the probabilities of being employed in different types of firms and annual individual income. In particular, the results suggest that employers that introduce workplace flexibility in addition to supports for family and childcare provide more opportunities for women to attain higher incomes, especially among college graduates and graduates of advanced vocational schools.

REFERENCES

Abadie, Alberto. 2005. “Semiparametric Difference in Difference Estimators.” *Review of Economic Studies* 72: 1–19.

Bang, Heejung, and James M. Robins. 2005. “Doubly Robust Estimation in Missing Data and Causal Inference Models.” *Biometrics* 61: 962–72.

Bakk, Zsuzsa, Fetene B. Tekle, and Jeroen K. Vermunt 2013. “ Estimating the Association between Latent Class Membership and External Variables Using Bias-adjusted Three-step Approaches.” *Sociological Methodology* 43: 372–311.

Collins, Linda M., and Stephanie T. Lanza. 2009. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences.* New York: Wiley.

Huber, Peter J. 1964. “Robust Estimation of a Location Parameter.” *Annals of Mathematical Statistics* 35: 73–101.

Huber, Peter J. 1967. “The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions.” *Proceedings of the 5th Berkeley Symposium*, 1, 221–33.

Morgan Stephen L., and Christopher Winship, 2007. *Counterfactuals and Causal Inference*. Cambridge: Cambridge University Press.

Pearl, Judea. *Causality* (2nd edition). 2009. Cambridge: Cambridge University Press.

Robins, James. M., A. H. Miguel, and B. Brumback. 2000. “Marginal Structural Models and Causal Inference in Epidemiology.” *Epidemiology* 11: 550–60.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70: 41–55.

Rosenbaum, Paul R., and Donald B. Rubin. 1984. “Reducing Bias in Observational Studies Using Sub-classification on the Propensity Scores.” *Journal of the American Statistical Association* 79: 516–24.

Rubin, Donald B. 1985. “The Use of Propensity Scores in Applied Bayesian Inference”. Pp. 463–72 in J. M. Bernardo, M. H. De Groot, D. V. Lindley, and A. F. M. Smith (eds.), *Bayesian Statistics*, vol. 2. North-Holland: Elsevier.

Vermunt, Jeroen K. 1997. *LEM: A General Program for the Analysis of Categorical Data.* Tilburg, Tilburg University.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA. The MIT Press.

Table 1. Prevalence of Various Work-Life Balance Systems/Policies

	All employees	Employees by firm size		
		100-299	300-499	500-
Sample size	2,349	1,778	212	359
Indicators[1]	% employed by firms with each policy/system			
I1	28.8	23.5	27.8	55.7
I2	72.6	68.8	76.4	80.8
I3	24.3	19.8	30.2	43.2
I4	3.8	3.0	1.4	9.2
I5	12.3	6.8	12.7	39.6
I6	26.1	20.1	28.8	54.3

[1] I1. Parental leave policy over and above statutory minimums.
I2. System of part-time work for child and family care.
I3. Flextime work system.
I4. Telecommuting system.
I5. In-house career development or open recruitment system.
I6. Department/center for the promotion of work-life balance policies.

Table 2. Results from latent-class models without covariates

Model	L^2	*df*	*P*	BIC
Independence	2,017.62	57	0.000	631.19
Two classes	348.39	50	0.000	-39.70
Three classes	178.90	43	0.000	-154.86
Four classes	133.86	36	0.000	-145.57
Five classes	113.20	29	0.000	-111.81

Table 3. Conditional Probabilities of the Presence of Each Work-Life Balance Program by Latent Class and Model

		Model without covariates			Model 1: Type-I			Model 2: Type-II		
		L1	L2	L3	L1	L2	L3	L1	L2	L3
Class Composition		0.063	0.435	0.502	0.104	0.390	0.506	0. 054	0.468	0.478
I1	Response Prob.	0.758	0.477	0.066	0.723	0.446	0.077	0.744	0.467	0.062.
	Wald	75.98***			76.69***			47.78***		
I2	Response Prob.	0.724	0.893	0.583	0.769	0.897	0.586	0.711	0.887	0.571
	Wald	66.75***			64.88***			74.16***		
I3	Response Prob.	0.801	0.299	0.125	0.649	0.294	0.120	0.826	0.301	0.121
	Wald	--------[a]			139.55***			93.73***		
I4	Response Prob.	0.469	0.000	0.017	0.299	0.000	0.017	0.558	0.000	0.016
	Wald	122.20***			82.31***			71.76***		
I5	Response Prob.	0.858	0.138	0.019	0.778	0.079	0.022	0.843	0.145	0.022
	Wald	169.96***			169.96***			113.36***		
I6	Response Prob.	0.748	0.462	0.027	0.722	0.440	0.028	0.717	0.451	0.025
	Wald	34.99***			33.51***			19.27**		

[a]This Wald statistics did not converge.

Table 4: Results from the two latent-class models: Regression Component

	Model 1		Model 2	
I. Regression on the Logit of Latent-Class Proportions				
Contrast	L1 vs L3	L2 vs. L3	L1 vs. L3	L2 vs. L3
1. Hours of work per week (vs 40-44)				
Less than 40 hours	1.218***	0.373	1.393***	0.472*
45-49 hours	0.165	-0.210	0.501	-0.249
50-59 hours	0.257	-0.170	0.505	-0.091
60 hours or more	0.559	0.000	0.657	0.085
Missing	-0.009	0.212	0.245	0.190
2. Educational attainment (vs. high school or less)				
4-year college or more	0.994***	0.061	0.841**	0.217
Junior college	0.368	0.220	-0.043	0.260
Advanced vocational	-0.717	0.103	-0.629	0.038
3. Firm size (vs 100-299 employees)				
300-499 employees	0.847*	0.628*	0.252	0.309
500 employees or more	4.062***	2.062***	2.431***	1.549***
4. Intercept	-3.487***	-0.600	-3.532***	-0.413
I. Regression on the Intercept of Indicator Variables				
The Effects of Firm size (versus 100-299 employees)				
1. On Indicator I1 (presence versus absence)				
300-499 employees	-----		0.052	
500 employees or more	-----		0.750***	
2. On Indicator I3 (presence versus absence)				
300-499 employees	-----		0.505**	
500 employees or more	-----		0.600***	
3. On Indicator I5 (presence vs. absence)				
300-499 employees	-----		0.698*	
500 employees or more	-----		1.796***	
4. On Indicator I6 (presence vs. absence)				
300-499 employees	-----		0.352	
500 employees or more	-----		0.869***	

Table 5. Diagnostic Tests of Statistical Independence of the Latent-Class Variable from Covariates

Covariates	Before the IP weighting			After the IP weighting		
	L2	df	sig.	L2	df	sig
I. Latent-Class Model 1						
(1) Hours of Work	48.67	10	0.000	3.63	10	>0.90
(2) Education	118.51	6	0.000	7.96	6	>0.20
(3) Firm size	526.50	4	0.000	1.53	4	>0.80
II. Latent-Class Model 2						
(1) Hours of Work	59.95	10	0.000	4.61	10	>0.90
(2) Education	73.61	6	0.000	2.59	6	>0.80
(3) Firm size	237.15	4	0.000	0.42	4	>0.90

Table 6. Treatment effects on income[1]

	Before the IP weighting			After the IP weighting		
	Effect	S.E.(M1)	S.E.(M2)	Effect	S.E.(M1)	S.E.(M2)
I. Latent-Class Model 1						
1. The outcome regression without covariates						
L1 vs. L3	94.23***	(11.16)	(12.62)	35.37**	(11.91)	(12.23)
L2 vs. L3	25.76***	(4.30)	(5.50)	18.25***	(4.84)	(5.15)
2. The outcome regression with three confounding covariates						
L1 vs. L3	57.82***	(10.36)	(12.96)	37.18**	(11.28)	(11.26)
L2 vs. L3	17.71***	(4.51)	(4.61)	17.59***	(4.71)	(4.87)
II. Latent-Class Model 2						
1. The outcome regression without covariates						
L1 vs. L3	89.03***	(11.98)	(13.60)	33.04**	(9.70)	(11.51)
L2 vs. L3	24.98***	(4.39)	(5.05)	18.26**	(5.07)	(5.68)
2. The outcome regression with three confounding covariates						
L1 vs. L3	59.63***	(10.70)	(11.74)	35.91***	(9.17)	(10.82)
L2 vs. L3	15.73 **	(4.94)	(5.12)	17.71***	(5.00)	(5.33)

***p<.001,**p<.01,*p<.05 based on the revised M-estimate of the standard error.

[1] S.E.(M1) shows the estimates of the standard error based on the simplified M-estimation that treats both latent-class probabilities and the IP weights as independently observed random variables. S.E.(M2) shows the revised estimates of the standard error based on the M-estimation that treats both latent-class probabilities and the IP weights as estimates from the particular latent-class model employed in the first step.

Table 7. Treatment effects on log(income)

	Before the IP weighting			After the IP weighting		
	Effect	S.E.(M1)	S.E.(M2)	Effect	S.E.(M1)	S.E.(M2)
I. Latent-Class Model 1						
1. The outcome regression without covariates						
L1 vs. L3	0.236***	(0.024)	(0.027)	0.093**	(0.032)	(0.033)
L2 vs. L3	0.069***	(0.010)	(0.010)	0.043***	(0.011)	(0.014)
2. The outcome regression with three confounding covariates						
L1 vs. L3	0.134***	(0.024)	(0.026)	0.101**	(0.031)	(0.032)
L2 vs. L3	0.044***	(0.010)	(0.010)	0.043***	(0.011)	(0.011)
II. Latent-Class Model 2						
1. The outcome regression without covariates						
L1 vs. L3	0.220***	(0.026)	(0.030)	0.090**	(0.027)	(0.033)
L2 vs. L3	0.062 ***	(0.009)	(0.011)	0.038**	(0.009)	(0.012)
2. The outcome regression with three confounding covariates						
L1 vs. L3	0.141***	(0.024)	(0.027)	0.096**	(0.026)	(0.037)
L2 vs. L3	0.034 ***	(0.009)	(0.010)	0.036 ***	(0.009)	(0.010)

***p<.001,**p<.01,*p<.05 based on the revised M-estimate (M2) of the standard error.

Table 8. Treatment effects by education

	Treatment effects of the latent class			
	on income		on log-income	
	L1 vs. L3	L2 vs. L3	L1 vs. L3	L2 vs. L3
total	33.04**	18.26**	0.090**	0.038**
4-year college or more	75.68***	22.90*	0.188***	0.066*
junior college	5.42	12.60	0.019	0.033
advanced vocational	55.19*	54.60***	0.179***	0.080*
high school or less	12.93	9.74	0.039	0.004

***p<.001,**p<.01,*p<.05.

APPENDIX− LEM SYNTAX FOR LATENT-CLASS MODELS

Table A-1 presents the LEM syntax for applying latent-class model 1. The “man” statement specifies the number of manifest (or observed) variables, and the “lat” statement specifies the number of latent variables. The label (“lab”) statement specifies variable labels, and label *X* is employed for the latent-class variable, labels *A*, *B*, *C*, *D*, *E*, and *F* are employed for the six indicators of the latent-class variable, and labels *I*, *J*, and *K* are employed for covariates that indicate education, hours of work, and firm size, respectively. The “dim” statement indicates the number of those 10 variables.

(Table A-1 here)

The model (“mod”) statement specifies a set of simultaneous equations for *X* (the latent variable) and their indicator variables and how the latent variable depends in logit on covariates. The specification X|IJK implies that the latent-class variable depends on the three covariates, and further specification of it by {X, XI, XJ, XK} indicates that the logit model includes the intercept and the main categorical effects of the three covariates on the logit. The specification of A|X, B|X, C|X, D|X, E|X, and F|X as a set indicates that the six indicator variables are conditionally independent, controlling for *X*, and each depends only on the latent variable *X*.

The record (“rec”) statement indicates the number of records to be read from the input data file, labeled “WLF_POLICY.dat” here. The dummy (“dum”) statement indicates the baseline category for each variable. The statement “nfr” indicates the suppression of the cross-classified data for observed and expected frequencies, and residuals in the output file. The last data (“dat”) statement indicates the name of the input data that include the values of the nine observed variables specified in the “man” statement in this order for each sample subject as a free-format text file.

Table A-2 presents the LEM syntax for model 2. This differs from the content of table A-1 only for the model statement. While the specification for the dependence of the latent-class variable on three covariates is the same as that of model 1, model 2 specifies the set of A|XK, B|X, C|XK, D|X, E|XK, and F|XK for the conditional independence among the six indicator variables, specifying that indicators *A*, *C*, *E*, and *F*, but not *B* and *D*, depend on covariate *K*, which indicates firm size. The specification A|XK {X, XA, AK}| indicates that the logit for the presence versus absence of policy I1 depends on two variables, *X* and *K*, and includes the intercept (A), the effect of the latent-class variable (XA), and the effect of firm size (AK) on the logit. The meaning of the specification for C|XK, E|XK, and F|XK is the same.

Table A-1: LEM syntax for model 1

```
man 9
lat 1
dim 3 2 2 2 2 2 2 4 6 3
lab X A B C D E F I J K
mod X|IJK {X, XI, XJ, XK}
    A|X
    B|X
    C|X
    D|X
    E|X
    F|X
rec 2349
dum 3 1 1 1 1 1 1 4 2 1
nfr
dat  WLF_POLICY.dat
```

Table A-2: LEM syntax for model 2

```
man 9
lat 1
dim 3 2 2 2 2 2 2 4 6 3
lab X A B C D E F I J K
mod X|IJK {X, XI, XJ, XK}
    A|XK {A, XA, AK}
    B|X
    C|XK {C, XC, CK}
    D|X
    E|XK {E, XE, EK}
    F|XK {F, XF, FK}
rec 2349
dum 3 1 1 1 1 1 1 4 2 1
nfr
dat  WLF_POLICY.dat
```