RIETI Discussion Paper Series 10-E-006

A Macrocounterfactual Analysis of Group Differences: An application to an analysis of the gender wage gap in Japan

YAMAGUCHI Kazuo
RIETI

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 10-E-006

A Macrocounterfactual Analysis of Group Differences: An application to an analysis of the gender wage gap in Japan[1]

YAMAGUCHI Kazuo

The University of Chicago & RIETI

Abstract

This paper introduces a new method for a statistical simulation of macrosocietal counterfactual situations. In particular, this method is concerned with decomposing group differences in the mean of a variable into various within-group and between-group components with respect to group categories of intermediary variables. In modeling counterfactual situations, I juxtapose two different mechanisms, the mechanism of realizing the counterfactual state that deviates least from the existing state, and the mechanism of holding other irrelevant-to-counterfactual relations of variables unchanged, and demonstrate that despite the big difference in the mechanisms, the two counterfactual models generally yield highly consistent outcomes. As an illustrative example, the paper analyzes gender inequality in hourly wages in Japan and thereby demonstrates the usefulness of the new method for deriving policy implications.

[1] This research has been supported by the visiting fellowship of the Research Institute of Economy, Trade, and Industry (RIETI), Japan, to the author.

A MACROCOUNTERFACTUAL ANALYSIS OF GROUP DIFFERENCES: AN APPLICATION TO AN ANALYSIS OF THE GENDER WAGE GAP IN JAPAN

1. INTRODUCTION

1.1 The Objective of this Research

This paper introduces a method for the decomposition of differences in the mean of a variable among groups in order to provide a prescriptive tool for a macrosocietal counterfactual analysis for assessing relative importance of the elements of group differences in policies intended to reduce the differences. It is often claimed that while economics is a prescriptive social science, sociology is a descriptive social science. Since policy makers are concerned with finding an answer to the question whether a particular policy is effective in attaining an intended societal outcome, a prescriptive analysis is considered relatively useful in providing guidance for social engineering. A descriptive analysis is indeed relatively useless if a policy to be considered is unprecedented and, therefore, no information about the consequence of the policy is empirically available.

A prescriptive analysis is possible in economics because economic theory primarily relies on mathematical reasoning on causality. However, prescriptions or predictions based on a theory that lacks an empirical ground will not be reliable. An alternative to theoretical reasoning is a simulation of counterfactual situations that can in part reflect existing empirical states. The method introduced in this paper can be regarded as such a method of statistical simulation.

The method is concerned with modeling of macrosocital counterfactual situations. While modeling of counterfactual situations is important in microcausal analysis, that is, to model what the outcome would be if the subject were placed in an alternative treatment, as in Rubin's causal models (RCM) (Rubin 1985; Rubin and Rozenbaum 1986; Robins 2002), it is also important in causal analysis at the macrosocietal level, especially for analyses where we consider what would happen in society if a particular relation, or a set of relations, among variables, were eliminated.

A long time ago, Blalock, Duncan, and others considered "causal analysis" based on the path-analytic model and covariance structures (Blalock 1964, 1971, Duncan 1966). Their causal idea was later criticized because it reflected neither a counterfactual conception of causality nor the elimination of selection bias from the "treatment" states (Holland 1986). These criticisms are certainly justified. However, there was one thing in their "causal" thinking that is worthy of an elaboration – not in a statistical sense of introducing latent variables, correlated measurement errors, simultaneity, and so on, which were introduced later in structural equation models and in confirmatory factor analysis. It was the macrosocial consideration of what would happen if a particular path, or a trajectory of paths, were eliminated from the observed societal state. A distinction between the direct effect and indirect effects or, more generally, a decomposition of the effect of X on Y into trajectory components through intermediary variables, is based on such a consideration. A particular assumption the method implicitly made in such a decomposition analysis was that the elimination of an effect does not change the other effects or the assumption of "other things being equal" for path coefficients other than those that are modified. However, that assumption may not hold in reality.

In this paper, the method introduced considers two very different alternative mechanisms for realizing a given counterfactual situation as a tool to provide a decomposition analysis of differences in the mean of a variable. Partly based on empirical data and partly based on the assumptions for the realization of counterfactual situations, the analysis can numerically assess the relative effectiveness of alternative policies that are intended to change the effects of a particular variable on its consequence. The two mechanisms are juxtaposed in order to make the analysis rely on the consistency in the characteristics of outcome between the two predicted outcomes for counterfactual situations. The consistency is attained mainly because, as shown in this paper, the outcomes do not depend on alternative assumptions under a general condition and differ only when the condition is not met. We may regard the method as a variation of simulation analysis that utilizes information from empirical data. At the same time, it can be considered a method that extends the old path-analytic idea to a different direction from purely statistical elaborations.

The method introduced below may also be considered to be related to decomposition methods introduced early by Kitagawa (1955) and elaborated by others such as Das Guputa (1978) and Liao (1989), to name only a few, for the decomposition of rates. It is also related to methods for a decomposition of a difference in the interval-scale variable such as a method introduced early by Blinder (1973) and Oaxana (1973), and a method introduced later by DiNardo-Fortin-Lemieux (hereafter DFL). (DiNardo et al, 1996). They are all standardization methods.   However, my approach in this paper does not consider the decomposition method to be merely a technique for standardization, but rather a method for modeling the social consequence of a macrocounterfactual situation.

As shown in an illustrative analysis of gender inequality in hourly wages in Japan, an analysis based on the new method can provide some policy implications for reducing gender inequality in the hourly wage by providing numerical measures in assessing the relative importance of various factors that generate the gender inequality.

1.2 A Review of the Blinder-Oaxana Method and the DFL Method

The Blinder-Oaxana method relies on (1) a conventional standardization method of assuming one group as the standard group and (2) linear regression equations. For example, given a set of equations of *y* for two groups such as for men and women that

$$y^M = \mathbf{X}^M \beta^M + \varepsilon \text{ and } y^F = \mathbf{X}^F \beta^F + \varepsilon,$$

we obtain,

$$\overline{y}^M - \overline{y}^F = \left[(\overline{\mathbf{X}}^M - \overline{\mathbf{X}}^F)\beta^F \right] + \left[\overline{\mathbf{X}}^M (\beta^M - \beta^M) \right]$$

where the first component is the difference in *y* that would be explained by the gender difference in **X** because it gives the difference in *y* if women had the same **X** as men's, and the second component is the unexplained difference and reflects both the main gender effect and the interaction effects of gender and **X** on *y*. If we can assume a causal order among covariates **X**, we can further separate the explained part into components by sequentially applying the method for each X.

As can be easily seen, this method employs the multivariate distribution of **X** for one sex as the standard distribution in standardization. In other words, it implicitly assumes that the joint distribution of gender (variable G) and **X** under the counterfactual situation where G is independent of **X** is equal to the product of the conditional distribution of **X** for one group and the marginal probability of the group, such that $f(\mathbf{X}, G) = f(\mathbf{X} \mid G=1)P(G)$. While this assumption is a standard option for the method of standardization, it is an unrealistic assumption for modeling a counterfactual situation. The results of the analysis depend on the choice of the standard group. Another limitation is a strong dependence of the results on the adequacy of the two linear regression equations.

The DFL method tries to improve the Blinder-Oaxana method by relaxing the assumption of linear relationship between y and **X**. The method assumes a functionally unconstrained relation between y and **X** for each sex such that

$$\overline{y}^M = \int_X f(y \mid X, G=1) f(X \mid G=1) dX \text{ and } \overline{y}^F = \int_X f(y \mid X, G=2) f(X \mid G=2) dX.$$

where G = 1 for men and G = 2 for women. The method then considers a counterfactual mean of y for women when they had the same distribution of **X** as men's such that:

$$\overline{y}^F_{D(X)} \equiv \int_X f(y \mid \mathbf{X}, G=F) f(\mathbf{X} \mid G=M) d\mathbf{X} = \int_X \omega_X f(y \mid \mathbf{X}, G=F) f(\mathbf{X} \mid G=F) d\mathbf{X},$$

where

$$\begin{aligned}\omega_X &= \frac{f(\mathbf{X} \mid G=1)}{f(\mathbf{X} \mid G=2)} = \frac{f(\mathbf{X}, G=1)/f(G=1)}{f(\mathbf{X}, G=2)/f(G=2)} = \frac{f(G=1 \mid \mathbf{X})/(f(\mathbf{X})f(G=1))}{f(G=1 \mid \mathbf{X})/(f(\mathbf{X})f(G=2))} \\ &= \left[P(G=2)/P(G=1)\right]\left[(P(G=1 \mid \mathbf{X})/P(G=2 \mid \mathbf{X})\right].\end{aligned}$$

The nicety of this formulation is that weight ω_X can be estimated for each sample by modeling the logistic regression $\log(P(G=1)/P(G=2))$. The method also recommends the use of this logistic regression equation as the propensity-score weighting by limiting the analysis to samples with a common areas of support for $P(G{=}1)$ and $P(G{=}2)$.

Then, equation

$$\bar{y}^M - \bar{y}^F = \left(\bar{y}^M - \bar{y}^F_{D(X)=M}\right) + \left(\bar{y}^F_{D(X)=M} - \bar{y}^F\right)$$

also leads to a decomposition of the mean difference into the explained and the unexplained components, and if we can assume a causal order among covariates **X**, we can further decompose the explained part into components by sequentially applying the decomposition for each *X*.

While the DFL method is certainly more elaborated than the Blinder-Oaxana method, it also assumes the multivariate distribution of **X** for a group as the standard distribution. However, if the distribution of **X** in question characterizes job characteristics such as occupation and employment status, an implicit assumption such that if women had the same distribution as men may not be the best way to model a counterfactual situation because jobs available in the labor market will depend on demand for jobs. It seems more reasonable to assume that the marginal distribution of each variable remains the same under the counterfactual situation, and the counterfactual condition changes only the matching of job characteristics with the gender of the job occupants.

Second, while the DLF method eliminated a strong assumption on the linear relationship between *Y* and **X**, the results heavily depend on the adequacy of modeling propensity-score weighting by a logistic regression model.

In this paper, I introduce an alternative method that (1) does not require any regression equation, (2) does not assume the multivariate distribution of **X** for one sex as the standard distribution and, instead, models the joint distribution of **X** under a given counterfactual situation, and (3) enables further decomposition not only for the explained part but for the “unexplained” part. Regarding (2), we assume two disttinct mechanisms in realizing counter-factual distributions of variables and show theoretically and empirically that the results from the method introduced in this paper depend rather little on the two alternative assumptions. The method has one major limitation, however. It assumes that all covariates **X** are categorical variables. This assumption, however, brings significant benefits by enabling the standardization and modeling of cross-classified frequency data developed in loglinear models to be incorporated into the method.

2. New Decomposition Method

2.1 Notes on Considering Macrosocietal Counterfactual Situations

In order to provide a simplified image of what this paper tries to accomplish, suppose that we have a path model as in Figure 1. In Figure 1, variable *A* is the key categorical variable whose effect on an interval-scale dependent variable *Y* is the quantity we wish to decompose into components, and variables *B* and *C* are intermediary categorical variables. The path diagram is simplified, because there can be many interaction effects among variables on the variables affected by them.

(Figure 1 about Here)

Suppose, as a concrete example, that variable *A* is the distinction between sexes, variable *B* is the distinction between full-time and part-time work, and variable *C* is occupation, with the assumption here that people's choices of hours of work precede their choices of occupation, and variable *Y* is hourly wage. We can assume an opposite order between *B* and *C* if that is more reasonable. Generally, given this model, a gender difference in the hourly wage may exist because (1) women tend to hold part-time jobs more than men do, and hourly wages are lower for part-time jobs, (2) men and women have different occupations, and occupations held disproportionately by women have lower hourly wages, and (3) women may get lower hourly wages than men even when men and women hold the same occupation and have the same proportions of full-time and part-time workers. We can see in the path diagram of Figure 1 that factor 1 is represented by the effect of *A* on *Y* though *B*, factor 2 by the effect of *A* on *Y* through *C*, and factor 3 by the unique, or direct, effect of *A* on *Y*, and that factors 1 and 2 overlap to the extent that *B* has a significant effect on *C*.

If variables *B* and *C* are interval-scale variables and if no interaction effects between any two variables on a third variable exist, a decomposition of the total effect of *A* on *Y* into the three factors described above is a simple task through the path-analytic decomposition of direct and indirect effects. However, if variables *B* and *C* as well as *A* are categorical variables, and there are higher-order interactions, or associations, not only among *A*, *B*, and *C*, but also in their effects on *Y*, such a decomposition of the gender difference into components is not as simple.

There is another important issue concerning the consideration of counterfactual situations not reflected in path-analytic decomposition. Suppose that we consider a counterfactual situation where the direct effect of *A* on *C* becomes absent when *A* affects *C* directly as well as indirectly through *B*. Then we may reason that the total effect of *A* on *C* would be reduced to the extent to which *A*'s direct effect on *C* was eliminated. However, this reasoning is based implicitly on an assumption that the elimination of a direct path from *A* to *C* will not affect the extent of the indirect effect of *A* on *C* through *B*. Empirically, however, this assumption may not hold true. For a concrete example, suppose that *A* is father's occupation, *B* is son's education, and *C* is son's occupation and that a counterfactual state as well as the present state realizes an "equilibrium" that partly reflects the consequence of people's choices in the respective situations. Suppose now that we consider a hypothetical societal situation where self-employment that can be passed from fathers to sons is made negligible and occupational opportunities among people with different family backgrounds are made completely equal at each level of education by, for example, legally prohibiting, and socially sanctioning against, employment decisions based on job applicants' family backgrounds. Then, cognizant of the situation that the investment in son's education is the only means for making son's occupational attainment advantageous, parents' with more resources than others may try to exert a stronger influence than before on son's educational attainment, which will then lead to an increase in the effect of father's occupation (A) on son's education (B). In addition, since the direct influence of family background on son's occupation is eliminated, occupational attainment may become more strongly dependent on educational credentials, which will lead to an increase in the effect of son's education (B) on son's occupation (C). Hence, both the effect of A on B and the effect of B on C will increase, thereby increasing the indirect effect of A on C through B and partially compensating for the loss of the direct effect of A on C. More generally, we may expect that there will be some underlying social forces that make the existing structure resistant to change under a new social condition imposed externally.

This consideration may seem to require a behavioral model, as in economics, to predict the relationship among variables in a counterfactual situation. This paper, however, provides an alternative approach by considering two extreme situations between which the realization of a counterfactual situation is expected to take place, and it demonstrates the usefulness of such an analysis.

2.2 Basic Formulation

Below, we first consider some counterfactual situations with two intermediary variables. We also assume here, for simplicity of description, a temporal order in the realization of categorical variables *A*, *B*, and *C* and an interval-scale variable *Y* as depicted in Figure 1. We are concerned with a decomposition of the effect of *A* on *Y*.

Let $\overline{y}_{ijk}^{ABC}$ and w_{ijk}^{ABC} be the mean value of *Y* and the number of people in the sample (or in the population) with *i*, *j*, *k* as categories of variables *A*, *B*, and *C*, respectively. The difference between *A* = 1 and *A* = 2 in the mean of *Y* is then given as

$$\begin{aligned}\overline{y}_1^A - \overline{y}_2^A &= \frac{\sum_j \sum_k w_{1jk}^{ABC} \overline{y}_{1jk}^{ABC}}{\sum_j \sum_k w_{1jk}^{ABC}} - \frac{\sum_j \sum_k w_{2jk}^{ABC} \overline{y}_{2jk}^{ABC}}{\sum_j \sum_k w_{2jk}^{ABC}} \\ &= \sum_j \sum_k \left(\overline{w}_{jk|1}^{BC|A} \overline{y}_{1jk}^{ABC} - \overline{w}_{jk|2}^{BC|A} \overline{y}_{2jk}^{ABC} \right), \qquad (1)\end{aligned}$$

where $\overline{w}_{jk|i}^{BC|A} \equiv w_{ijk}^{ABC} / w_i^A$ and $w_i^A \equiv \sum_j \sum_k w_{ijk}^{ABC}$. Note that the weight for the mean of *Y* differs between *A* = 1 and *A* = 2, with respective weights being equal to the proportion of the combined *B* and *C* states for each category of *A*.

Generally, given the assumed temporal order among variables, we can express the joint frequency distribution of *A*, *B*, and *C* as the product of the marginal frequency of *A* and two conditional probabilities such that

$$w_{ijk}^{ABC} = w_i^A (w_{ij}^{AB} / w_i^A)(w_{ijk}^{ABC} / w_{ij}^{AB}) = w_i^A \overline{w}_{j|i}^{B|A} \overline{w}_{k|ij}^{C|AB}, \qquad (2)$$

where $w_i^A \equiv \sum_j \sum_k w_{ijk}^{ABC}$ and $w_{ij}^{AB} \equiv \sum_k w_{ijk}^{ABC}$ are one-way and two-way marginal frequencies, respectively, $\overline{w}_{j|i}^{B|A} \equiv w_{ij}^{AB} / w_i^A$ is the conditional probability of *B* for a given *A*, and $\overline{w}_{k|ij}^{C|AB} \equiv w_{ijk}^{ABC} / w_{ij}^{AB}$ is the conditional probability of *C* for a given set of *A* and *B*. We consider below various counterfactual situations that impose some constraints on equation (2).

2.3 When *A* Is Independent of Both *B* and *C*

Suppose now that we consider a counterfactual situation where *A* becomes independent of *B* and *C*, while *B* and *C* are associated. Substantively, this implies using the concrete example of *A*, *B*, *C*, and *Y* described before, a situation where men and women come to have the same composition of hours of work and the same composition of occupation.

The consideration for a situation where *B* and *C* become independent of *A* leads to a modification of equation (2) such that

$$W_{ijk}^{ABC} = w_i^A \overline{W}_j^B \overline{W}_{k|j}^{C|B} = w_i^A \overline{W}_{jk}^{BC}, \tag{3}$$

where a capital letter for frequencies W and (conditional) probabilities $\overline{W}$ indicates a hypothetical value realized in the counterfactual situation, and $\overline{W}_{jk}^{BC} \equiv \overline{W}_j^B \overline{W}_{k|j}^{C|B}$ is the hypothetical joint probability distribution of *B* and *C*. Equation (3) indicates that when *B* and *C* are independent of *A*, the conditional probabilities of *B* for a given category of *A* become unconditional probabilities and the conditional probabilities of *C* for given categories of *A* and *B* depend only on *B*'s category. However, without knowing the mechanism of their realizations, $\overline{W}_j^B$ and $\overline{W}_{k|j}^{C|B}$ remain unknown. For the moment, I defer my discussion on this mechanism.

In this counterfactual situation, the difference in the mean of *Y* between *A* = 1 and *A* = 2, which we denote $\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)$, becomes

$$\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC) = \frac{\sum_j \sum_k \left(w_1^A \overline{W}_{jk}^{BC} \right) \overline{y}_{1jk}^{ABC}}{\sum_j \sum_k \left(w_1^A \overline{W}_{jk}^{BC} \right)} - \frac{\sum_j \sum_k \left(w_2^A \overline{W}_{jk}^{BC} \right) \overline{y}_{2jk}^{ABC}}{\sum_j \sum_k \left(w_2^A \overline{W}_{jk}^{BC} \right)}$$
$$= \sum_j \sum_k \overline{W}_{jk}^{BC} \left(\overline{y}_{1jk}^{ABC} - \overline{y}_{2jk}^{ABC} \right). \qquad (4)$$

Equation (4) shows that when *A* is independent of *B* and *C*, the difference in *Y* between *A* = 1 and *A* = 2 becomes a weighted average of the *within*-*B*-and-*C*-group differences in*Y* between *A* = 1 and *A* = 2 with the probability distribution of the *B*-and-*C* states realized in this counterfactual situation as weights. By calculating the ratio of $\overline{W}_{ij}^{BC} (\overline{y}_{1jk}^{ABC} - \overline{y}_{2jk}^{ABC})$ to $\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)$, we can also determine how much the within-group difference for each combined state of *B* and *C* contributes to the overall within-group difference. On the other hand, the difference between $\overline{y}_1^A - \overline{y}_2^A$ and $\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)$ represents the *between*-*B*-and-*C*-group difference in *Y*. Similarly, we can regard $\dfrac{\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)}{\overline{y}_1^A - \overline{y}_2^A}$ and $\dfrac{\left(\overline{y}_1^A - \overline{y}_2^A\right) - \left(\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)\right)}{\overline{y}_1^A - \overline{y}_2^A}$ as representing the relative proportions of the within-group and the between-group inequality in $\overline{y}_1^A - \overline{y}_2^A$, respectively. In short, by considering the counterfactual situation where *A* becomes independent of *B* and *C*, we can decompose the difference in *Y* into within-group and between-group components.

An important fact here is that if there are neither interaction effects of *A* and *B* nor interaction effects of *A* and *C* on *Y*, the value of $\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)$ actually does not depend on weights $\overline{W}_{jk}^{BC}$.

Suppose that

$$\overline{y}_{ijk}^{ABC} = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{jk}^{BC}, \tag{5}$$

where $\sum_i \beta_i^A = \sum_j \beta_j^B = \sum_k \beta_k^C = \sum_j \beta_{jk}^{BC} = \sum_k \beta_{jk}^{BC} = 0,$ holds, so that there are neither *AB* nor *AC* interaction effects on *Y*. Then, since $\overline{y}_{1jk}^{ABC} - \overline{y}_{2jk}^{ABC} = \beta_1^A - \beta_2^A$, we obtain

$$\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC) = \sum_j \sum_k \overline{W}_{jk}^{BC} \left(\overline{y}_{1jk}^{ABC} - \overline{y}_{2jk}^{ABC} \right) = \beta_1^A - \beta_2^A, \tag{6}$$

and, therefore, the value of $\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)$ does not depend on weights $\overline{W}_{jk}^{BC}$. This finding is a special case of a more general theorem that Little and Pullum (1979) introduced in the generalized linear framework for standardization. Note that we also implicitly assume here that the effects of *A*, *B*, and *C* on *Y* do not change in the counterfactual situation where *A* becomes independent of *B* and *C*, and this assumption also relies on one of the two mechanisms assumed below.[2]

It follows that only when the interaction effects either *A* and *B* or *A* and *C* on *Y* exist do we need to know a mechanism that generates the joint probability distribution of *B* and *C* in order to obtain the value of $\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)$, and this fact relieves us of the

[2] The fact that the effects of *A*, *B*, and *C* on *Y* do not change is evident for the case of the standardization method that assumes that all unmodified effects remain the same. In the case of the maximum likelihood method, the joint likelihood of *A*, *B* and *C*, and $\overline{Y}$ can be expressed as the product of the marginal likelihood of *A*, *B*, and *C* and the conditional likelihood of $\overline{Y}$ for a given set of *A*, *B*, and *C*, and the change in the marginal likelihood under a given counterfactual situation does not affect parameter estimates that maximize the conditional likelihood, and thereby keeps the effects of *A*, *B*, and *C* on *Y* unchanged

burden of making some additional assumptions about the mechanism in determining of the joint distribution of *B* and *C* realized under the counterfactual condition.

2.4 Two Alternative Assumptions for the Realization of Counterfactual Situations

Now we ask what mechanisms we can reasonably assume for obtaining the $\overline{W}_{jk}^{BC}$ in equation (3) in the counterfactual situation where *A* is independent of *B* and *C*. We consider two "extreme" mechanisms between which we expect the real outcome to lie. One mechanism assumes that there will be strong resistance to change in the joint distribution of variables under the externally imposed conditions of the given counterfactual situation so that the new joint distribution of variables under the externally imposed condition will have a minimal deviation from the existing one. The other mechanism assumes that there will be no change in the relationship among variables other than those that are modified by the externally imposed conditions of a given counterfactual situation.

For the first mechanism, we employ, as the criterion of measuring deviation, the statistical significance of residuals, and this criterion leads to the use of the maximum likelihood (ML) estimates of frequencies for a given condition. Although the ML estimation requires a model, we apply the model of equation (3) with elements of probabilities $\overline{W}_{j}^{B}$ and $\overline{W}_{k|j}^{C|B}$ as parameters by assuming the multinomial distribution for each set of response probabilities. We then obtain $\overline{W}_{j}^{B} = \overline{w}_{j}^{B} = w_{j}^{B} / N$ and $\overline{W}_{k|j}^{C|B} = \overline{w}_{k|j}^{C|B} = w_{jk}^{BC} / w_{j}^{B}$ and, therefore,

$$W_{ijk}^{ABC}(ML) = w_{i}^{A} w_{ij}^{BC} / N \ . \tag{7}$$

It is noteworthy that the adjusted *BC* marginal frequencies, W_{jk}^{BC}, are the same as the unadjusted *BC* marginal frequencies, w_{jk}^{BC}. Thus, even though the indirect association between *B* and *C* due to their common association with *A* is eliminated under the independence of *A* from *B* and *C*, the newly realized situation that minimizes the deviation from the existing structure fully recovers the association between *B* and *C*.

The second mechanism that eliminates a particular effect while retaining other effects is related to standardization in demography. We consider adjusted conditional probabilities $\overline{W}_{k|j}^{C|B}$ that retain the unique effects of *B* on *C* in $\overline{w}_{k|ij}^{C|AB}$ while eliminating the unique effects of *A* on *C* from $\overline{w}_{k|ij}^{C|AB}$. By "unique effects" we mean the set of partial odds ratios in the loglinear or logit (multinomial logit) characterization among categorical variables. However, "retaining the unique effects of *B* on *C*" has an unambiguous meaning only when there are no interaction effects of *A* and *B* on *C*.

Suppose that the observed conditional probabilities $\overline{w}_{k|ij}^{C|AB}$ have only additive effects of *A* and *B* on *C* in the multinomial logit equation, such that

$$\log(\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB}) = \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \text{ f.or } k = 1,\ldots, K-1, \tag{8}$$

where $\sum_i \lambda_{ik}^{AC} = \sum_j \lambda_{jk}^{BC} = 0$, and the (*K*-1) parameters of λ_k^C characterize the odds among categories of *C*, the set of (*I*-1)(*K*-1) parameters of λ_{ik}^{AC} characterizes the partial odds ratios between *A* and *C*, and the set of (*J*-1)(*K*-1) parameters of λ_{jk}^{BC} characterizes the partial odds ratios between *B* and *C* , when variables *A*, *B*, and *C* have *I*, *J*, and *K* categories, respectively.

From equation (8), we obtain

$$\log(\overline{w}_{k|ij_1}^{C|AB} / \overline{w}_{1|ij_1}^{C|AB}) - \log(\overline{w}_{k|ij_2}^{C|AB} / \overline{w}_{1|ij_2}^{C|AB}) = \lambda_{j_1k}^{BC} - \lambda_{j_2k}^{BC}\,. \qquad (9)$$

This implies that the log odds ratio between *B* and *C* in $\overline{w}_{k|ij}^{C|AB}$ for each given level of *A* is characterized by $\lambda_{j_1k}^{BC} - \lambda_{j_2k}^{BC}$ and does not depend on the category of *A*.

Hence, if we obtain the adjusted conditional probabilities $\overline{W}_{k|j}^{C|B}$ that satisfy the equation

$$\log(\overline{W}_{k|j_1}^{C|B} / \overline{W}_{1|j_1}^{C|B}) - \log(\overline{W}_{k|j_2}^{C|B} / \overline{W}_{1|j_2}^{C|B}) = \lambda_{j_1k}^{BC} - \lambda_{j_2k}^{BC}\,, \qquad (10)$$

such conditional probabilities satisfy the condition that they retain the unique effects of *B* on *C* in $\overline{w}_{k|ij}^{C|AB}$ while eliminating the unique effects of *A* on *C*.

The method that Xie (1989) introduced as "partial CD purging," where C stands for the compositional variable and D stands for the dependent variable, yields such adjusted conditional probabilities. Generally, the lambda parameters in equation (8) require sets of normalizing constraints to be individually identifiable, and if we employ the deviation contrast for λ_{ik}^{AC} such that $\sum_i \lambda_{ik}^{AC} = 0$, as Xie did, we obtain, by taking the arithmetic mean of equation (8) over the categories of *A*, that

$$\left(\sum_i \log(\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})\right) / I = \log\left(\prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I}\right) = \lambda_k^C + \lambda_{jk}^{BC}\,, \qquad (11)$$

and therefore

$$\overline{W}_{k|j}^{C|B} = \theta_{k|j}^{C|B} \Big/ \sum_k \theta_{k|j}^{C|B}, \text{where } \theta_{k|j}^{C|B} = \prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I}\,, \qquad (12)$$

satisfy equation (10) because $\log(\overline{W}_{k|j_1}^{C|B} / \overline{W}_{1|j_1}^{C|B}) - \log(\overline{W}_{k|j_2}^{C|B} / \overline{W}_{1|j_2}^{C|B}) = \log(\theta_{k|j_1}^{C|B}) - \log(\theta_{k|j_2}^{C|B})$.

These adjusted probabilities are equivalent to the standardized probabilities based on the method introduced by Teachman (1977), satisfy equation (10), and retain the effects of *B* on *C*, λ_{jk}^{BC}, while eliminating the effects of *A* on *C*, λ_{ik}^{AC}.

This method can be extended to cases where interaction effects of *A* and *B* on *C* exist, even though the meaning of the unique effects of *B* on *C* becomes somewhat ambiguous in such cases, because the method "purges" factor λ_{ijk}^{ABC} by assuming the deviation contrast for this factor, $\sum_i \lambda_{ijk}^{ABC} = 0$, or the uniform distribution of *A* as the standard distribution in purging the interaction effects, for the saturated model

$$\log(\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB}) = \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} \quad .$$

While uniform distribution is a very strong assumption, this purging of the *ABC* interaction effects has one very desirable characteristic: it makes the estimates of $\overline{W}_{k|j}^{C|B}$, on which the decomposition in $\overline{y}_1^A - \overline{y}_2^A$ depends, independent of the distribution of *A*.

Hence, this method seems quite promising. There is one important additional consideration, however. While the effects of *B* on *C* characterized by equation (10) are invariant for the normalizing constraints we make among lambda parameters under the absence of interaction effects of *A* and *B* on *C*, because partial odds ratios are independent of the choice on the normalizing constraints, the values of adjusted conditional probabilities depend on the choice of the normalizing constraints, because the values of $\lambda_k^C + \lambda_{jk}^{BC}$ in equation (11) depend on them. Yamaguchi (2009) points out, however, that giving categories *A* with different size the same weights by assuming the uniform distribution of *A* can make the average adjusted probability determined by the CD-purging method very different from the unadjusted average probability, thereby possibly causing a great change in the marginal distribution of *C*. In particular, if categories of *A* with smaller sizes than others have larger conditional probabilities for *C*, the average adjusted conditional probability tends to be greater than the average unadjusted conditional probability, and the opposite holds true if categories of *A* with smaller sizes than others have smaller conditional probabilities for *C*. Conceptually, however, the retention of the effects of *B* on *C*, while eliminating the effects of *A* on *C*, does not imply any specific change in the marginal distribution of *C*, because relations here imply the characteristics of odds ratios among variables.

Then the question is whether we can modify Xie's CD-purging method to readjust the adjusted conditional probabilities of equation (12) to preserve the marginal probability of *C* while retaining the same unique effects of *B* on *C* characterized by equation (10). The answer is affirmative, because this leads to the adjustment of λ_k^C parameters while retaining the same λ_{jk}^{BC} parameters. If we denote by $\lambda_k^C *$ the adjusted parameters that preserve the marginal probability of *C*, and $\gamma_k^C \equiv \exp(\lambda_k^C * - \lambda_k^C)$, then $\theta_{k|j}^{C|B} * \equiv \exp(\lambda_k^C * + \lambda_{jk}^{BC}) = \gamma_\mathrm{k}^\mathrm{C} \exp(\lambda_k^C + \lambda_{jk}^{BC}) = \gamma_\mathrm{k}^\mathrm{C} \theta_{k|j}^{C|B}$, where $\theta_{k|j}^{C|B} \equiv \prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I}$.

Hence, since $\overline{W}_{k|j}^{C|B} * \equiv \theta_{k|j}^{C|B} * \big/ \sum_k \theta_{k|j}^{C|B} *$, where the asterisk indicates the quantity after the readjustment, and the set of *K* gamma parameters satisfies the set of *K* equations $\sum_j \overline{w}_j^B \overline{W *}_{k|j}^{C|B} = \overline{w}_k^C$, we obtain the following set of equations:

$$\sum_j \overline{w}_j^B \gamma_k^C \theta_{k|j}^{C|B} \Big/ \left(\sum_k \gamma_k^C \theta_{k|j}^{C|B} \right) = \overline{w}_k^C \text{, for } k=1,\ldots,K, \qquad (13)$$

where $\overline{W *}_{k|j}^{B|C} = \gamma_k^C \theta_{k|j}^{B|C} \big/ \sum_k \gamma_k^C \theta_{k|j}^{B|C}$.

Here, the joint readjusted probability distribution, $\overline{W*}_{ij}^{BC} = \overline{w}_j^B \gamma_k^C \theta_{k|j}^{C|B} / \left(\sum_k \gamma_k^C \theta_{k|j}^{B|C}\right)$, has odds ratios $(\overline{W*}_{jk}^{BC}\overline{W*}_{mn}^{BC})/(\overline{W*}_{jn}^{BC}\overline{W*}_{mk}^{BC}) = (\theta_{k|j}^{C|B}\theta_{n|m}^{C|B})/(\theta_{n|j}^{C|B}\theta_{k|m}^{C|B})$ that do not depend on the values of γ_k^C. Hence, the solution to the nonlinear equation (13) of γ_k^C can be easily obtained by the Stephen-Deming iterative proportional adjustment by adjusting the marginal distributions of *B* and *C* to become equal to $\overline{w}_j^B$ and $\overline{w}_k^C$, respectively, starting from $\overline{W}_{ij}^{BC} = \overline{w}_j^B \theta_{k|j}^{C|B} / \left(\sum_k \theta_{k|j}^{C|B}\right)$. The adjustment multiples we obtain for categories of *C* are the estimates for γ_k^C. We will refer to this method as the *modified CD-purging method of standardization with rescaling* and employ it as the method that realizes counterfactual situations by the second mechanism we consider.

By applying the same procedure of adjustment and readjustment to obtain $\overline{W*}_j^B$ by eliminating the effects of *A* on *B* from $\overline{w}_{j|i}^{B|A}$, we simply obtain $\overline{W*}_j^B = \overline{w}_j^B$. Below, we denote readjusted probabilities $\overline{W*}$ simply as $\overline{W}$.

It follows that the adjusted three-way frequency that makes *A* independent of both *B* and *C* under this standardization mechanism, which we denote by $W_{ijk}^{ABC}(ST)$, is given as

$$W_{ijk}^{ABC}(ST) = (w_i^A w_j^B / N) \frac{\gamma_k^C \prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I}}{\sum_k \left\{ \gamma_k^C \prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I} \right\}} , \qquad (14)$$

where the set of γ_k^C satisfies the set of equations (13). Note that unlike the ML estimates of adjusted frequencies $W_{ijk}^{ABC}(ML)$ given by equation (7), values of $W_{ijk}^{ABC}(ST)$ depend on the temporal order we assume between *B* and *C*.

2.5 When *A* Becomes Independent of *B*

For another counterfactual situation, let us consider a situation where variable *B* becomes independent of variable *A*, while *A*'s effects on *C* remain. Then we obtain

$$W_{ijk}^{ABC} = w_i^A \overline{W}_j^B \overline{W}_{k|ij}^{C|AB}, \tag{15}$$

The ML estimate for this model made by assuming probability elements of $\overline{W}_j^B$ and $\overline{W}_{k|ij}^{C|AB}$ as parameters, is

$$W_{ijk}^{ABC}(ML) = w_i^A \overline{w}_j^B \overline{w}_{k|ij}^{C|AB} = (w_i^A w_j^B / N)(w_{ijk}^{ABC} / w_{ij}^{AB}). \tag{16}$$

For the standardization-method estimate, we obtain $\overline{W}_{k|ij}^{C|AB} = \overline{w}_{k|ij}^{C|AB}$ from the assumption that this method retains unmodified effects. Since $\overline{W}_j^B = \overline{w}_j^B$ for this method, we also obtain

$$W_{ijk}^{ABC}(ST) = w_i^A \overline{w}_j^B \overline{w}_{k|ij}^{C|AB} = W_{ijk}^{ABC}(ML) \tag{17}$$

The difference in *Y* between *A* = 1 and *A* = 2 under this assumption in the mechanism of realizing independence between *A* and *B* is given as

$$\begin{aligned} \overline{y}_1^A - \overline{y}_2^A \,|\, (A)(B) &= \frac{\sum_j \sum_k (w_1^A \overline{w}_j^B \overline{w}_{k|1j}^{C|AB}) \overline{y}_{1jk}^{ABC}}{\sum_j \sum_k w_1^A \overline{w}_j^B \overline{w}_{k|1j}^{C|AB}} - \frac{\sum_j \sum_k (w_2^A \overline{w}_j^B \overline{w}_{k|2j}^{C|AB}) \overline{y}_{2jk}^{ABC}}{\sum_j \sum_k w_2^A \overline{w}_j^B \overline{w}_{k|2j}^{C|AB}} \\ &= \sum_j \overline{w}_j^B \sum_k \left(\overline{w}_{k|1j}^{C|AB} \overline{y}_{1jk}^{ABC} - \overline{w}_{k|2j}^{C||AB} \overline{y}_{2jk}^{ABC} \right). \end{aligned} \tag{18}$$

Hence, weights for variable *B* are "standardized," that is, common across categories of *A*, and are equal to proportions in *B*'s marginal distribution. On the other hand, weights for categories of *C* differ among categories of *A*. We can regard the difference between $\overline{y}_1^A - \overline{y}_2^A$ of equation (1) and $\overline{y}_1^A - \overline{y}_2^A \,|\, (A)(B)$ of equation (18) as representing the extent of reduction in the inequality by the elimination of the association between *A* and *B*. This is a subcomponent of the between-group component $\left(\overline{y}_1^A - \overline{y}_2^A\right) - \left(\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)\right)$. On the other hand, the difference between $\overline{y}_1^A - \overline{y}_2^A \mid (A)(B)$ of equation (18) and $\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)$ of equation (4) represents the extent of reduction in the inequality by the elimination of the association between *A* and *C, after having eliminated the AB association.* This is the other subcomponent of $\left(\overline{y}_1^A - \overline{y}_2^A\right) - \left(\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)\right)$. We can regard $\dfrac{\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)}{\overline{y}_1^A - \overline{y}_2^A}$,

$\dfrac{\left(\overline{y}_1^A - \overline{y}_2^A\right) - \left(\overline{y}_1^A - \overline{y}_2^A \mid (A)(B)\right)}{\overline{y}_1^A - \overline{y}_2^A}$, and $\dfrac{\left(\overline{y}_1^A - \overline{y}_2^A \mid (A)(B)\right) - \left(\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)\right)}{\overline{y}_1^A - \overline{y}_2^A}$ as representing the relative proportion of the within-group and the two between-group components in $\overline{y}_1^A - \overline{y}_2^A$.

2.6 When *A* Becomes Conditionally Independent of *C*

Suppose now that we consider another counterfactual situation, where *A* becomes conditionally independent of *C* by holding *B* constant. Then we obtain

$$W_{ijk}^{ABC} = w_i^A \overline{w}_{j|i}^{B|A} \overline{W}_{k|j}^{C|B} . \tag{19}$$

The ML estimate made for this model by assuming the probability elements of $\overline{W}_{k|j}^{C|B}$ as parameters is given as

$$W_{ijk}^{ABC}(ML) = w_i^A (w_{ij}^{AB} / w_i^A)(w_{jk}^{BC} / w_j^B) = w_{ij}^{AB} w_{jk}^{BC} / w_j^B \tag{20}$$

On the other hand, the standardization-method estimate is given as

$$W_{ijk}^{ABC}(ST) = w_{ij}^{AB} \frac{\gamma_k^C \prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I}}{\sum_k \left\{\gamma_k^C \prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I}\right\}} \tag{21}$$

where the set of γ_k^C satisfies the set of equations (13).

The difference in the mean of Y between A = 1 and A = 2 in this situation becomes

$$\begin{aligned} \overline{y}_1^A - \overline{y}_2^A \mid (AB)(BC) &= \frac{\sum_j \sum_k w_{1j}^{AB} \overline{W}_{k|j}^{C|B} \overline{y}_{1jk}^{ABC}}{\sum_j \sum_k w_{1j}^{AB} \overline{W}_{k|j}^{C|A}} - \frac{\sum_j \sum_k w_{2j}^{AB} \overline{W}_{k|j}^{C|B} \overline{y}_{2jk}^{ABC}}{\sum_j \sum_k w_{2j}^{AB} \overline{W}_{k|j}^{C|B}} \\ &= \sum_j \sum_k \overline{W}_{k|j}^{C|B} \left(\overline{w}_{j|1}^{B|A} \overline{y}_{1jk}^{ABC} - \overline{w}_{j|2}^{B|A} \overline{y}_{2jk}^{ABC}\right), \end{aligned} \tag{22}$$

where $\overline{W}_{k|j}^{C|B} = w_{jk}^{BC} / w_j^B$ for the ML estimates and

$\overline{W}_{k|j}^{C|B} = \gamma_k^C \prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I} \Big/ \sum_k \left\{\gamma_k^C \prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I}\right\}$ for the standardization-method estimates.

Hence, while the conditional distribution of B depends on A, the conditional distribution of C for each category of B is independent of A and is "standardized" in this respect. We can regard the difference between $\overline{y}_1^A - \overline{y}_2^A$ of equation (1) and $\overline{y}_1^A - \overline{y}_2^A \mid (AB)(BC)$ of equation (22) as representing the extent of reduction in the inequality by the elimination of the unique effect of A on C. This is a subcomponent of the between-group component $\left(\overline{y}_1^A - \overline{y}_2^A\right) - \left(\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)\right)$. On the other hand, the difference between $\overline{y}_1^A - \overline{y}_2^A \mid (AB)(BC)$ of equation (22) and $\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)$ of equation (4) represents the extent of reduction in the inequality by the elimination of the association between A and B, *after having eliminated the unique effects of C on A*. This is the other subcomponent of $\left(\overline{y}_1^A - \overline{y}_2^A\right) - \left(\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)\right)$. We can regard

$\dfrac{\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)}{\overline{y}_1^A - \overline{y}_2^A}$, $\dfrac{\left(\overline{y}_1^A - \overline{y}_2^A\right) - \left(\overline{y}_1^A - \overline{y}_2^A \mid (AB)(BC)\right)}{\overline{y}_1^A - \overline{y}_2^A}$, and

$\dfrac{\left(\overline{y}_1^A - \overline{y}_2^A \mid (AB)(BC)\right) - \left(\overline{y}_1^A - \overline{y}_2^A \mid (A)(BC)\right)}{\overline{y}_1^A - \overline{y}_2^A}$ as representing the relative proportion of the

within-group and the two between-group components in $\overline{y}_1^A - \overline{y}_2^A$.

2.7 Change of Relationship between Intermediary Variables

As another counterfactual situation, we can consider the situation where the effects of *B* on *C*, with *A* held constant, becomes absent. Generally,

$$W_{ijk}^{ABC} = w_i^A \overline{w}_{j|i}^{B|A} \overline{W}_{k|ij}^{C|A} = w_{ij}^{AB} \overline{W}_{k|ij}^{C|A} \tag{23}$$

holds, and its ML estimate and the standardization estimate are given, respectively, as

$$W_{ijk}^{ABC}(ML) = w_{ij}^{AB}(w_{ik}^{AC} / w_i^A) = w_{ij}^{AB} w_{ik}^{AC} / w_i^A \tag{24}$$

and

$$W_{ijk}^{ABC}(ST) = w_{ij}^{AB} \frac{\gamma_k^C \prod_j (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/J}}{\sum_k \left\{ \gamma_k^C \prod_j (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/J} \right\}}, \tag{25}$$

where the set of γ_k^C satisfies

$\sum_i \overline{w}_i^A \gamma_k^C \theta_{k|i}^{C|A} / \left(\sum_k \gamma_k^C \theta_{k|i}^{C|A}\right) = \overline{w}_k^C$, for k=1,…,K, and $\theta_{k|i}^{C|A} \equiv \prod_j (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/J}$.

(26)

Like the solution for equation (13) the estimates of γ_k^C that satisfy the set of equations (26) can be easily obtained by applying the Stephen-Deming proportional adjustment for the joint probability distribution of *A* and *C*.

3. A GENERALIZATION OF THE METHOD

This section generalizes the decomposition analysis. As an illustration, Figure 2 adds one more intermediary categorical variable, *D*. As was Figure 1, this figure is simplified, because there can be many interaction effects. Even though the method is described for the case with three intermediary variables, it is easy to extend it to a general case with an unspecified number of intermediary categorical variables.

(Figure 2 About Here)

3.1 Backward Sequential Decomposition

Backward sequential decomposition considers the following three counterfactual situations in sequence:

(B1) What would happen to the effect of *A* on *Y* if the direct effect of *A* on *D* became absent?

(B2) What would happen to the effect of *A* on *Y* if the direct effects of *A* on *C* and *D* became absent?

(B3) What would happen to the effect of *A* on *Y* if the direct effects of *A* on *B*, *C*, and *D* became absent?

Generally, we have $w_{ijkm}^{ABCD} = w_i^A \overline{w}_{j|i}^{B|A} \overline{w}_{k|ij}^{C|AB} \overline{w}_{m|ijk}^{D|ABC}$, and the counterfactual situations (B1), (B2), and (B3) imply that we have models $W_{ijkm}^{ABCD} = w_i^A \overline{w}_{j|i}^{B|A} \overline{w}_{k|ij}^{C|AB} \overline{W}_{m|jk}^{D|BC}$,

$W_{ijkm}^{ABCD} = w_i^A \overline{w}_{j|i}^{B|A} \overline{W}_{k|j}^{C|B} \overline{W}_{m|jk}^{D|BC}$, and $W_{ijkm}^{ABCD} = w_i^A \overline{W}_j^{B} \overline{W}_{k|j}^{C|B} \overline{W}_{m|jk}^{D|BC}$. We can easily prove that the ML estimate for the models of (B1), (B2), and (B3) becomes, respectively,

$$W_{ijkm}^{ABCD}(ML) = w_i^A \overline{w}_{j|i}^{B|A} \overline{w}_{ijk|ij}^{C|AB} (w_{jkm}^{BCD} / w_{jk}^{BC}) = w_{ijk}^{ABC} w_{jkm}^{BCD} / w_{jk}^{BC}, \tag{27}$$

$$W_{ijkm}^{ABCD}(ML) = w_i^A \overline{w}_{j|i}^{B|A} (w_{jk}^{BC} / w_j^B)(w_{jkm}^{BCD} / w_{jk}^{BC}) = w_{ij}^{AB} w_{jkm}^{BCD} / w_j^B, \tag{28}$$

And

$$W_{ijkm}^{ABCD}(ML) = w_i^A (w_j^B / N)(w_{jk}^{BC} / w_j^B)(w_{jkm}^{BCD} / w_{jk}^{BC}) = w_i^A w_{jkm}^{BCD} / N. \tag{29}$$

On the other hand, by using the standardization method we employed in the previous section, we obtain adjusted frequencies for situations (B1), (B2), and (B3) by applying $\overline{W}_j^B = \overline{w}_j^B, \overline{W}_{k|j}^{C|B} = \dfrac{\gamma_k^C \prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I}}{\sum_k \left\{ \gamma_k^C \prod_i (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/I} \right\}}$ and

$W_{m|jk}^{D|BC} = \dfrac{\gamma_m^D \prod_i (\overline{w}_{m|ijk}^{D|ABC} / \overline{w}_{1|ijk}^{D|ABC})^{1/I}}{\sum_m \left\{ \gamma_m^D \prod_i (\overline{w}_{m|ijk}^{D|ABC} / \overline{w}_{1|ijk}^{D|ABC})^{1/I} \right\}}$. The set of γ_k^C parameters in $\overline{W}_{k|j}^{C|B}$ satisfy equation (13), and the set of γ_m^D parameters in $\overline{W}_{m|jk}^{D|BC}$ satisfy

$$\sum_{j,k} \overline{w}_{jk}^{BC} \gamma_m^D \theta_{m|jk}^{D|BC} / \left(\sum_m \gamma_m^D \theta_{m|jk}^{D|BC} \right) = \overline{w}_m^D \text{ for } \theta_{m|jk}^{D|BC} \equiv \prod_i (\overline{w}_{m|ijk}^{D|ABC} / \overline{w}_{1|ijk}^{D|ABC})^{1/I}. \tag{30}$$

By using one set of these adjusted frequencies as weights, we obtain $y_1^A - y_2^A \mid (ABC)(BCD)$, $y_1^A - y_2^A \mid (AB)(BCD)$, and $y_1^A - y_2^A \mid (A)(BCD)$ as the extent of inequality in *Y* between *A* = 1 and *A* = 2. The outcome $y_1^A - y_2^A \mid (A)(BCD)$ represents the within-*BCD*-group component, and the between-*BCD*-group component can be decomposed into three components, namely, $\left(y_1^A - y_2^A\right) - \left(y_1^A - y_2^A \mid (ABC)(BCD)\right)$, $\left(y_1^A - y_2^A \mid (ABC)(BCD)\right) - \left(y_1^A - y_2^A \mid (AB)(BCD)\right)$, and

$\left(y_1^A - y_2^A \mid (AB)(BCD)\right) - \left(y_1^A - y_2^A \mid (A)(BCD)\right)$, that indicate, respectively, the reduction in between-group inequality when the *AD* association, the *AC* association, and the *AB* association are eliminated in that order.

3.2 Forward Sequential Decomposition

Forward sequential decomposition considers the following three counterfactual situations in sequence:

(F1) What would happen to the effect of *A* on *Y* if the direct effect of *A* on *B* became absent?

(F2) What would happen to the effect of *A* on *Y* if the direct effects of *A* on *B* and *C* became absent?

(F3) What would happen to the effect of *A* and *Y* if the direct effects of *A* on *B*, *C*, and *D* became absent?

The counterfactual situations (F1), (F2), and (F3) imply that we have models $W_{ijkm}^{ABCD} = w_i^A \overline{W}_j^B \overline{w}_{k|ij}^{C|AB} \overline{w}_{m|ijk}^{D|ABC}$, $W_{ijkm}^{ABCD} = w_i^A \overline{W}_j^B \overline{W}_{k|j}^{C|B} \overline{w}_{m|ijk}^{D|ABC}$, and $W_{ijkm}^{ABCD} = w_i^A \overline{W}_j^B \overline{W}_{k|j}^{C|B} \overline{W}_{m|jk}^{D|BC}$. We can easily prove that the ML estimates for the models of (F1) and (F2) become, respectively,

$$W_{ijkm}^{ABCD}(ML) = w_i^A (w_j^B / N)(w_{ijk}^{ABC} / w_{ij}^{AB})(w_{ijkm}^{ABCD} / w_{ijk}^{ABC}) = (w_i^A w_j^B / N)(w_{ijkm}^{ABCD} / w_{ij}^{AB}) \quad (31)$$

and

$$W_{ijkm}^{ABCD}(ML) = w_i^A (w_j^B / N)(w_{jk}^{BC} / w_j^B)(w_{ijkm}^{ABCD} / w_{ijk}^{ABC}) = (w_i^A w_{jk}^{BC} / N)(w_{ijkm}^{ABCD} / w_{ijk}^{ABC}), \quad (32)$$

while while the formula for (F3) is the same as that for (B3), For the standardization-method estimates, the outcome of (F1) is the same as the ML estimate, the outcome for (F2) is obtained by applying the formula for $\overline{W}_{k|j}^{C|B}$ described before, and the outcome for (F3) is similar to that of (B3) except that we have to employ the adjusted joint distribution of *B* and *C*, $\overline{W}_{jk}^{BC}$, in solving equation (30).

By using these adjusted frequencies as weights, we obtain $y_1^A - y_2^A \mid (A)(B)$, $y_1^A - y_2^A \mid (A)(BC)$, and $y_1^A - y_2^A \mid (A)(BCD)$ as the extent of inequality in *Y* between *A* = 1 and *A* = 2. The differences $\left(y_1^A - y_2^A\right) - \left(y_1^A - y_2^A \mid (A)(B)\right)$, $\left(y_1^A - y_2^A \mid (A)(B)\right) - \left(y_1^A - y_2^A \mid (A)(BC)\right)$, and $\left(y_1^A - y_2^A \mid (A)(BC)\right) - \left(y_1^A - y_2^A \mid (A)(BCD)\right)$, respectively, indicate the reduction in inequality when the *AB* association, the *AC* association, and the *AD* association are eliminated in that order.

3.3 Change of Relationship between Intermediary Variables

We can also consider counterfactual situations where conditional independence is attained between two intermediary variables. We can consider three cases.

First, when variables *C* and *D* become conditionally independent by holding *A* and *B* constant, the model can be written as $W_{ijkm}^{ABCD} = w_{ijk}^{ABC} \overline{W}_{m|ij}^{D|AB}$.

Their ML estimates and the standardization-method estimates are, respectively, given as

$$W_{ijkm}^{ABCD}(ML) = w_{ijk}^{ABC} w_{ijm}^{ABD} / w_{ij}^{AB}, \tag{33}$$

and

$$W_{ijkm}^{ABCD}(ST) = w_{ijk}^{ABC} \frac{\gamma_m^D \prod_k (\overline{w}_{m|ijk}^{D|ABC} / \overline{w}_{1|ijk}^{D|ABC})^{1/K}}{\sum_m \left\{ \gamma_m^D \prod_k (\overline{w}_{m|ijk}^{D|ABC} / \overline{w}_{1|ijk}^{D|ABC})^{1/K} \right\}}, \tag{34}$$

where the set of γ_m^D parameters satisfy equation (30).

Second, when variables *B* and *C* become conditionally independent by holding *A* constant, the model can be written as $W_{ijkm}^{ABCD} = w_{ij}^{AB} \overline{W}_{k|i}^{C|A} \overline{W}_{m|ijk}^{D|ABC}$. Then the ML estimates and the standardization-method estimates are, respectively, given as

$$W_{ijkm}^{ABCD}(ML) = (w_{ij}^{AB} w_{ij}^{AC} / w_i^A)(w_{ijkm}^{ABCD} / w_{ijk}^{ABC}) \tag{35}$$

and

$$W_{ijkm}^{ABCD}(ST) = (w_{ij}^{AB} w_{ijkm}^{ABCD} / w_{ijk}^{ABC}) \frac{\gamma_k^C \prod_j (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/J}}{\sum_k \gamma_k^C \left\{ \prod_j (\overline{w}_{k|ij}^{C|AB} / \overline{w}_{1|ij}^{C|AB})^{1/J} \right\}}, \tag{36}$$

where the set of γ_k^C parameters satisfies equation (26).

Finally, when variables *B* and *D* become conditionally independent by holding *A* and *C* constant, or in other words, if *B* affects *D* only through *C* by holding *A* constant, the model can be written as $W_{ijkm}^{ABCD} = w_{ijj}^{ABC} \overline{W}_{m|ik}^{D|AC}$, and its ML estimates and the standardization-method estimates are given, respectively, as

$$W_{ijkm}^{ABCD}(ML) = w_{ijj}^{ABC} w_{ikm}^{ACD} / w_{ik}^{AC} \tag{37}$$

and

$$W_{ijkm}^{ABCD}(ST) = w_{ijk}^{ABC} \frac{\gamma_m^D \prod_j (\overline{w}_{m|ijk}^{D|ABC} / \overline{w}_{1|ijk}^{D|ABC})^{1/J}}{\sum_m \left\{ \gamma_m^D \prod_j (\overline{w}_{m|ijk}^{D|ABC} / \overline{w}_{1|ijk}^{D|ABC})^{1/J} \right\}}, \tag{38}$$

where the set of γ_m^D parameters satisfy $\sum_{i,k} \overline{w}_{ik}^{AC} \gamma_m^D \theta_{m|jk}^{D|AC} / \left(\sum_m \gamma_m^D \theta_{m|jk}^{D|AC} \right) = \overline{w}_m^D$

for $\theta_{m|ik}^{D|AC} \equiv \prod_j (\overline{w}_{m|ijk}^{D|ABC} / \overline{w}_{1|ijk}^{D|ABC})^{1/J}$.

3.4 Notes on the Zero-Frequency Observation

If the cross-classification of covariates reveals a combination of states with zero sampling observation, the standardization method that takes a geometric mean of conditional probabilities generates adjusted conditional probabilities that are inefficient as statistics because an adjusted conditional probability that involves a zero observation becomes zero regardless of other values involved (Yamaguchi 2009). Hence, if a zero observation exists, it is highly desirable to use the set of expected frequencies from a loglinear model of covariates that fits the data rather than observed frequencies. The loglinear model that fits the data normally eliminates zero-frequency observations without distorting the outcome, and the use of the model's expected frequencies that do not involve zero values provides much more efficient statistics.

4. APPLICATION

4.1 Data

The application employs data from 2005 Japan's Basic Census of Wage Structure. This is a government census of Japan and collects wage data from all employees in Japan except for nonresponses. The present analysis is based on the class-classified mean of hourly wages for employees by gender, age, the distinction between full-time and part-time employment, and the distinction between regular employees (with an unspecified term employment contract) and irregular employees (with a term employment contract), and cross-classified frequencies of those four covariates of hourly wage. I assume below that gender affects age composition during the period of labor-force participation first, and both gender and age affect the choice between full-time and part-time jobs, and these three variables affect the choice between regular and irregular employments, and these four affect hourly wages. The causal order between the third and fourth variables may be debatable, however. The ML estimates of decomposition shown below in Table 2 do not depend on this order – while the standardization-method estimates do. It is well known that the average hourly wage of women is lower than that of men in Japan partly because female workers are disproportionately represented in irregular employment and part-time work, and this study identifies the magnitude of the within-group and between-group components of gender differences in the hourly wage for four groups of employment types defined below, and assesses the effectiveness of alternative strategies to reduce the between-group component.

4.2 Descriptive Statistics

Table 1 gives for each sex the proportion of four employment types classified by a combination of regular-versus-irregular employment distinction and full-time-versus-part-time work distinction, and hourly wage for each combination. The last row in Table 1 gives the wage ratio between women and men.

There are several noteworthy facts in Table 1. First, women are underrepresented in regular full-time employment, for which the hourly wage is the highest among the four employment types, and they are overrepresented in irregular part-time employment, for which the hourly wage is the lowest. Second, regular part-time employment is nearly absent and applies to less than 1% of the labor force, regardless of gender. In other words, almost all regular employees are full-time workers. Third, for each of the employment types, the wage ratio indicates that women's hourly wage is lower than men's hourly wage, and the women-to-men wage ratio is lowest among regular full-time workers (0.698) and highest among irregular part-time workers (0.887). Since women are underrepresented in the most advantaged employment type, and overrepresented in the most disadvantaged employment type, the women-to-men wage ratio for the total labor force becomes lower (0.617) than the type-specific wage ratio.

Figure 3 describes how hourly wage changes as a function of age for three major employment types for men and women. As shown in the figure, average hourly wages do not differ greatly between men and women and across employment types among people under age 20, but as age increases, regular workers get higher wage returns for age because many Japanese firms adopt wage system with tenure-based wage premiums, called *nenko*-wage system, among regular employees. And yet there is a big difference between men and women in the slope of increase in wages with age among full-time regular employees. On the other hand, the extent of increase in the hourly wage with age is relatively small among irregular employees, especially among part-time irregular employees. While gender inequality in the hourly wage still tends to become larger as age increases among irregular employees, the extent of gender inequality generated among them is much smaller than that among regular employees.

4.3 Decomposition Analysis

Suppose we refer to gender as variable *A*, age as variable *B*, the distinction of full-time and part-time workers as variable *C*, and the distinction of regular and irregular employees as variable D. Then, using backward decomposition analysis, we consider (a) the situation where gender (variable A) is conditionally independent of D when B and C are held constant, (b) the situation where gender is conditionally independent of C and D when B is held constant, and (c) the situation where gender is independent of B, C, and D. Then we obtain $y_1^A - y_2^A \,|\, (ABC)(BCD)$, $y_1^A - y_2^A \,|\, (AB)(BCD)$, and $y_1^A - y_2^A \,|\, (A)(BCD)$ as the extent of inequality remaining in Y between A = 1 and A = 2 in these three situations for both the ML and the standardization method of decomposition. The outcome $y_1^A - y_2^A \,|\, (A)(BCD)$ represents the within-BCD-group component, and the differences $\left(y_1^A - y_2^A\right) - \left(y_1^A - y_2^A \,|\, (AB)(BCD)\right)$ and $\left(y_1^A - y_2^A \,|\, (AB)(BCD)\right) - \left(y_1^A - y_2^A \,|\, (A)(BCD)\right)$, respectively, indicate the reduction in between-group inequality when both AC and AD associations are eliminated, and when the AB association is eliminated, in that order. Substantively, $\left(y_1^A - y_2^A\right) - \left(y_1^A - y_2^A \,|\, (AB)(BCD)\right)$ represents the gender wage gap due to the compositional difference of men and women in four employment types, and $\left(y_1^A - y_2^A \,|\, (AB)(BCD)\right) - \left(y_1^A - y_2^A \,|\, (A)(BCD)\right)$ represents the gender wage gap due to age differences between male and female employees that remains after the gender wage gap due to gender differences in the composition of employment types is eliminated.

A further decomposition of $\left(y_1^A - y_2^A\right) - \left(y_1^A - y_2^A \,|\, (AB)(BCD)\right)$ into $\left(y_1^A - y_2^A\right) - \left(y_1^A - y_2^A \,|\, (ABC)(BCD)\right)$ and $\left(y_1^A - y_2^A \,|\, (ABC)(BCD)\right) - \left(y_1^A - y_2^A \,|\, (AB)(BCD)\right)$ yields, respectively, the decomposition of gender wage gap due to gender differences in the composition of employment types into a component that will be eliminated when only the gender difference in the composition of regular and irregular employment is eliminated, and the remaining gender wage gap due to gender difference in the composition of full-time and part-time work combined with the difference in the hourly wage between full-time and part-time workers. On the other hand, the within-*BCD* gender gap, $y_1^A - y_2^A \mid (A)(BCD)$, contains the element for each combined state of variables *B*, *C* , and *D*, and therefore, when they are summed with weights across categories of age (*B*), we obtain the contribution of the gender wage gap *within* each of the four employment types.

Table 2 presents a decomposition of the gender wage gap obtained by this procedure. The contribution of each element is expressed in proportion by dividing each element by the overall gender difference in the hourly wage, $y_1^A - y_2^A$.

The table indicates that the largest component of the gender wage gap, about 51%~52%, is explained as the result of the within-group gender wage gap among regular employees, and the second largest component, about 36%~37%, is explained as the result of gender differences in the composition of employment types. Together, they explain about 88% of the gender wage gap. It is noteworthy that despite differences in the method of estimation based on very different underlying mechanisms, the two estimates are very close.

Hence, in order to reduce gender inequality in the hourly wage, it is most important to abolish certain institutions such as the distinction of *sogo shoku* (comprehensive-task jobs), which nearly all men and less than 5% of women are believed to hold, and *ippan shoku* (general-task jobs), which almost no men and more than 95% of women are believed to hold, among white-collar workers, because such institutions classify men and women in many Japanese firms largely into different wage-tracking and *nenko*-wage premiums among regular employees.

Second, it is important to equalize the opportunity for regular employment between men and women. Since the inequality is largely generated by the fact that women can obtain only irregular employment at reentry into the labor market after leaving their jobs for either marriage or child rearing because Japanese firms recruit regular employees mostly from among recent graduates from schools, it is important to make workplaces more flexible in time so that women during child-rearing periods need not leave their jobs. It is also important to expand opportunities for regular employment for reentrants in the labor market.

Table 2 also shows that if the difference in men and women's composition of full-time and part-time work is maintained, and only the gender inequality in the regular versus irregular employment is eliminated, the gender wage gap will be reduced by only about 9%. This reduction in the gender wage gap is small because women have more part-time jobs than men do: a large gender difference in the proportion of full-time regular workers remains if only the gender inequality in the composition of regular and irregular employment is eliminated within full-time workers and within part-time workers. This occurs mainly because since part-time regular workers are scarce, women's choice of part-time jobs necessarily leads to irregular employment. Indeed, during child-rearing periods, many women in Japan change jobs from those in regular employment to those in irregular employment, despite the fact that irregular employment is much more disadvantageous in job security and hourly wage, because part-time regular employment is not available.

This leads to an expectation that in a hypothetical society where regular employment becomes available regardless of the choice in the hours of work, and therefore there is no statistical association between the distinction between regular and irregular employment and the distinction between full-time and part-time work, women's choices of part-time work will not put them at as great a disadvantage as the present situation with a high crystallization between part-time work and irregular employment.

Table 3 gives the decomposition, for the situation where variables *C* and *D* become independent, of the proportion of the between-*B*-and-*C* inequality $\left(y_1^A - y_2^A\right) - \left(y_1^A - y_2^A \mid (AB)(BC)(BD)\right)$ and its two components,

$\left(y_1^A - y_2^A\right) - \left(y_1^A - y_2^A \mid (ABC)(BC)(BD)\right)$ and

$\left(y_1^A - y_2^A \mid (ABC)(BC)(BD)\right) - \left(y_1^A - y_2^A \mid (AB)(BC)(BD)\right)$.

The table indicates that when women's choice of part-time work does not lead to their overrepresentation in irregular employment because no association between the two dimensions of employment status exists, the realization of gender equality in the opportunity for regular employment can reduce the gender wage gap by about 20%~21%, which is more than twice as much as the estimates in Table 2. This indicates the importance of making regular part-time employment available widely so that we eliminate the societal condition where women's preference for part-time employment because of the incompatibility of full-time employment with their family roles, especially during child-rearing periods, does not deprive them of the opportunity for regular employment. Needless to say, it is also important to equalize the household division of labor between husbands and wives to reduce the extent of this incompatibility, so that women who wish to retain full-time employment can do so, because the traditional gender division of household labor that still persists strongly in Japan even among dual-earner families makes it very difficult for women to retain full-time jobs during child-rearing periods.

CONCLUSION

This paper introduces a new decomposition analysis of group differences. The major objective of the paper, however, is not just to introduce the method for its own sake, but also to stimulate discussion on how we can model counterfactual situations where, in the "causal" diagram, a particular path or a particular set of paths is eliminated. The modeling of such counterfactual situations is important because, as illustrated in the application to an analysis of gender inequality in the hourly wage in Japan, the analysis gives numerical measurements about which aspects of inequality are causing more of the overall inequality and about how a particular aspect of inequality may be reduced by changing the societal situation where variables that play intermediary roles may be modified in their relationships to one another. As is well known from the classical study of Blau (1977), structural constraints imposed by the macro association of variables reduce social opportunities. A difficulty of assessing the consequence of a macrosocietal change, however, is that people may not simply accept different social conditions but may rather utilize new situations for their benefit. A juxtaposition of the ML method, by which I represent the principle of resistance to change due to people's change of choices to recover their lost benefit under the new situation imposed externally, and the standardization method, by which I represent the principle of ceteris paribus, as well as the consistency of the two predictions, demonstrate the usefulness of the new method in describing what would happen in macrosocietal counterfactual situations.

Table 1. Average Hourly Wage by Gender and Employment Pattern

		Full-time, regular	Full-time, nonregular	Part-time, regular	Part-time, nonregular	Total/ average
Population proportion	Male	0.840	0.075	0.003	0.082	1.000
	Female	0.474	0.146	0.009	0.371	1.000
Hourly wage (in yen)	Male	2,094	1,324	1,342	1,059	1,949
	Female	1,462	1,041	1,068	939	1,203
Ratio of wage (female vs. male)		0.698	0.786	0.796	0.887	0.617

{

Table 2 Decomposition of Gender Wage Gap

	ML method	Standardization method
Gender difference in the composition of employment types	36.5	35.8
(a) due to regular-irregular composition within full-time work and within part-time work	(9.1)	(9.4)
(a) remaining difference	(27.4)	(26.4)
Within-group gap among regular full-time workers	51.0	51.8
Within-group gap among irregular full-time workers	4.0	4.1
Within-group gap among regular part-time workers	0.2	0.2
Within-group gap among irregular part-time workers	4.6	4.4
Gender differences in age distribution	3.7	3.7

Table 3. Decomposition under Independence of the Regular-Irregular Employment Distinction and the Full-Time-versus-Part-Time Work Distinction

	ML method	Standardization method
Gender difference in the composition of employment types	40.4	39.3
(b) due to regular-irregular composition within full-time work and within part-time work	(21.3)	(20.4)
(c) remaining difference	(19.1)	(18.9)

Figure 1: Path-Analytic Diagram 1

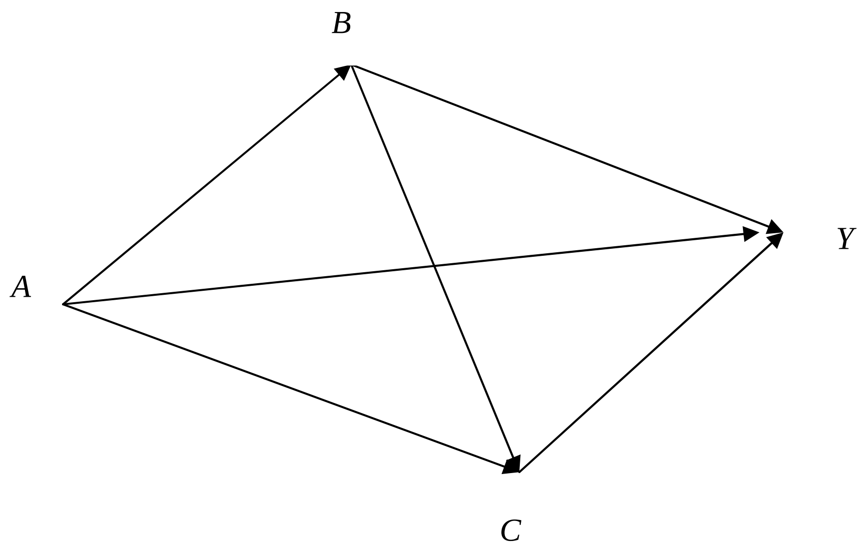
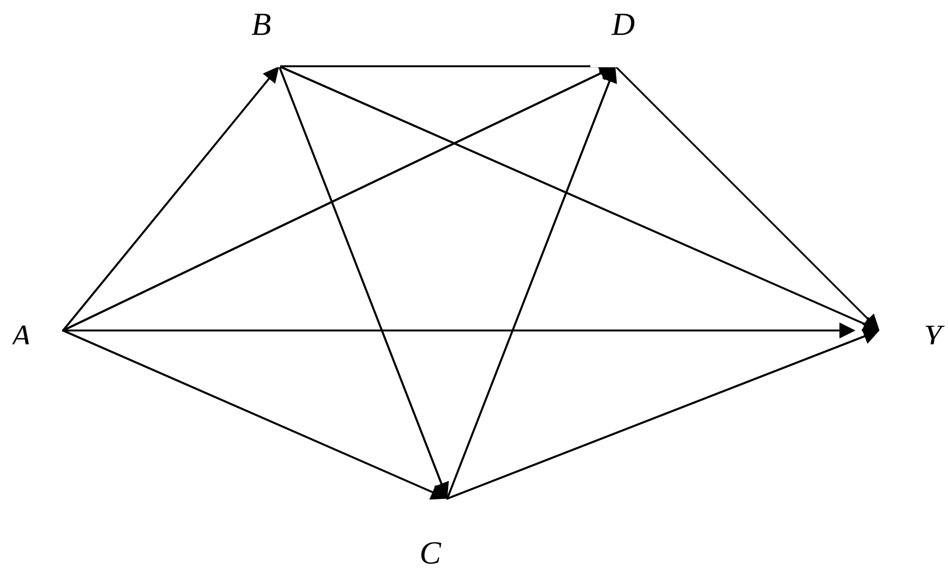


Figure 2. Path-Analytic Diagram 2



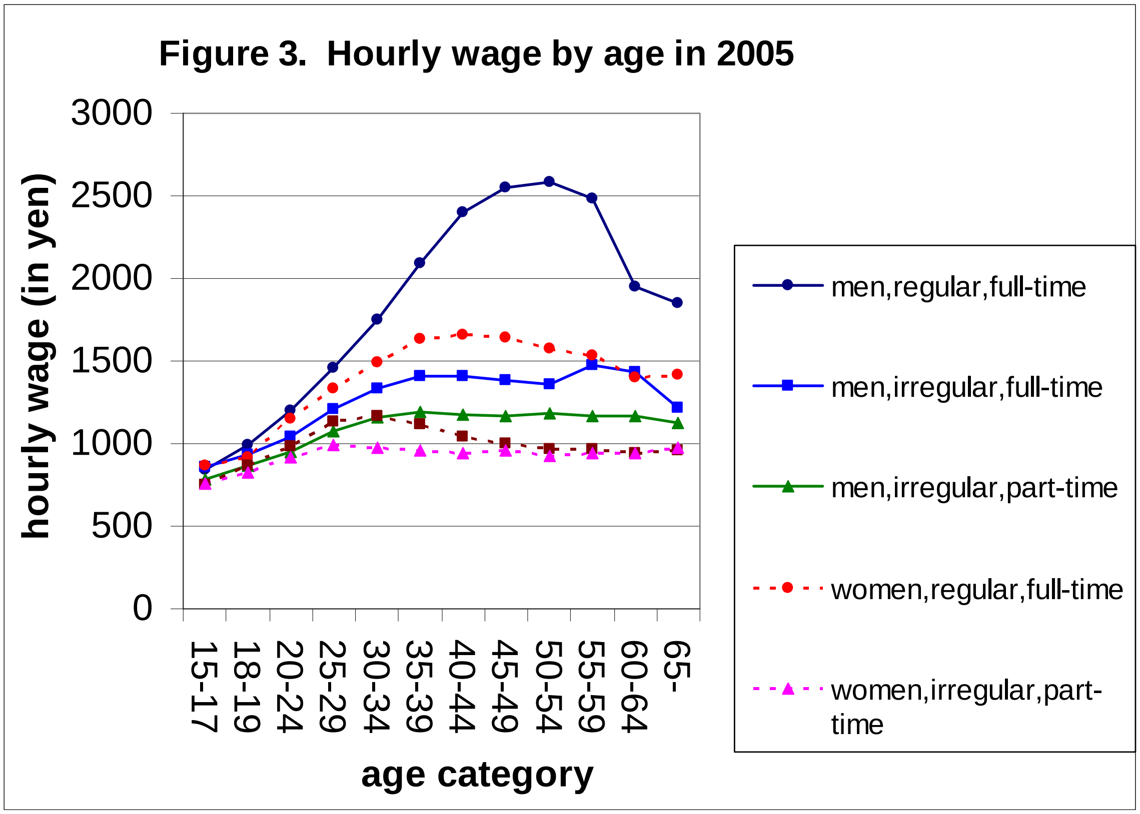


Figure 3. Hourly wage by age in 2005
hourly wage (in yen)
3000
2500
2000
1500
1000
500
0
15-17
18-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-
age category
men,regular,full-time
men,irregular,full-time
men,irregular,part-time
women,regular,full-time
women,irregular,part-time

REFERENCES

Blalock, H.M. 1964. *Causal Inferences in Nonexperimental Research*. Chapel Hill: University of North Carolina Press.

Blalock, H.M.(ed.) 1971. *Causal Models in the Social Sciences.* Chicago: Aldine-Anderton.

Blau,P.M. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: Free Press.

Blinder. A. 1973. "Wage Discrimination: Reduced Form and Structural Variables." Journal of Human Resources 8: 436-455.

De Gupta, E.M. 1978. "A General Method of Decomposing a Difference between Two Rates into Several Components." *Demography* 26: 361-372.

DiNardo, J., N. Fortin, and T. Limlieu 1966. "Labor Market Institution and the Distribution of Wages," *Econometrica* 64:1001-1044.

Duncan, O.D. 1966. "Path Analysis: Sociological Examples." *The American Journal of Sociology* 72: 1-16.

Holland, P.W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81: 945-60.

Kitagawa, E. 1955. "Components of a difference in Two Rates." *Journal of the American Statistical Association* 68: 1168-94.

Liao T.F. 1989. "A flexible Approach to the Decomposition of Rate differences." *Demography* 26: 717-26.

Little, R.J.A. and T. W. Pullum. 1979. The General Linear Model and the Direct Standardization: A Comparison." *Sociological Methods and Research* 7: 475-501.

Oaxana, R. 1973. “Male-Female Wage Differentials in Urban labor Markets.” *International Economic Review* 14: 693-709.

Teachman, J. D. 1977. “The Relationship between Schoen’s ∇ and a Loglinear Measure.” *Demography* 14: 239-241.

Robins, J. A., Miguel A. H. and B. Brumback 2000. “Marginal Structural Models and Causal Inference in Epidemiology.” *Epidemiology* 11: 550-560.

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

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

Xie, Y. 1989. “An Alternative Purging Methods: Controlling the Composition-Dependent Interaction in the Analysis of Rates.” *Demography* 26: 711-16.

Yamaguchi, K. 2009. “A Generalized Linear Approach To the Standardization of Rates and Proportions” Paper presented in the 2009 Annual Meetings of the American Sociological Association.