HOKKAIDO UNIVERSITY

| | |
|---|---|
| Title | Multidimensional inequality for current status of Japanese private companies' employees |
| Author(s) | Hasegawa, Hikaru; Ueda, Kazuhiro |
| Citation | METRON, 74(3), 357-373<br>https://doi.org/10.1007/s40300-016-0094-7 |
| Issue Date | 2016-12 |
| Doc URL | https://hdl.handle.net/2115/67785 |
| Rights | This is a post-peer-review, pre-copyedit version of an article published in METRON. The final authenticated version is available online at: http://dx.doi.org/10.1007/s40300-016-0094-7. |
| Type | journal article |
| File Information | Metron16.pdf |

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

**Noname manuscript No.**
(will be inserted by the editor)

# Multidimensional Inequality for Current Status of Japanese Private Companies' Employees

**Hikaru Hasegawa · Kazuhiro Ueda**

Received: date / Accepted: date

**Abstract** In this paper, we show a method for measuring multidimensional inequality. We use a Bayesian method for a multivariate ordered probit model to obtain a single index, which we refer to as a probability associated with an individual's overall status, from a multivariate distribution representing a number of attributes. We evaluate multidimensional inequality by the Gini coefficient and the generalized entropy measure of inequality using this index. This method is used for the employee data of Japanese private companies, which include income and attributes related to subjective perception such as job satisfaction.

**Keywords** generalized entropy measure · Gini coefficient · Markov chain Monte Carlo (MCMC) · multivariate ordered probit model · polychoric and polyserial correlation coefficients

JEL Classification: C35, D63.

Hikaru Hasegawa
Department of Economics, Hokkaido University,
Kita 9, Nishi 7, Kita-ku, Sapporo 060-0809, Japan.
E-mail: hasegawa@econ.hokudai.ac.jp

Kazuhiro Ueda (corresponding author)
Faculty of Economics, Nihon Fukushi University,
229, Kawaminami-shinden, Ota-machi, Tokai 477-0031, Japan.
Phone: +81-562-393811 fax: +81-562-393281
E-mail: ueda@n-fukushi.ac.jp

## 1 Introduction

Past studies on inequality and poverty, such as Maasoumi (1986), Dardanoni (1995), Seth (2009), and Decancq and Lugo (2012), have yielded that poverty and inequality should not always be measured using a single attribute, especially such pecuniary attributes as income, but should be measured by evaluating other non-pecuniary attributes such as health and education . For this purpose, it is necessary to generate a method to derive an index from data with multiple attributes. After the seminal studies by Kolm (1977) and Atkinson and Bourguignon (1982), there have been many works on the development of multidimensional indices.

According to Tsui (1999, p.146), for the studies on the multidimensional index of inequality, "the first problem confronting researchers in the measurement of inequality is the meaning of multidimensional inequality." One needs to judge whether a multidimensional distribution representing a number of attributes is more unequal than another that measures multidimensional inequality.

When we measure inequality of a population by a single scale, for example, income, it is only necessary to study how to evaluate the distribution of income. However, if we were to measure the overall inequality of a population by not only income but also other variables, a problem arises: how do we determine which of the two populations – one with better income distribution and the other with a more equitable distribution for some other variable – is better. Accordingly, many works, such as Tsui (1995, 1999), and Gajdos and Weymark (2005), have focused on the axioms that the aggregation functions deriving a single index from such multidimensional distributions must satisfy.

In this paper, we attempt to show another method of measuring multidimensional inequality where a single index is derived from multiple variables using a multivariate probability distribution. As an application of our method, we calculate an inequality index using four attributes related to one's well-being: job satisfaction, life satisfaction, health condition, and income.

It is considered that job satisfaction, life satisfaction, health condition, and income are closely related with our daily life. Accordingly, we need to empirically investigate the extent of the correlations among these. We estimated the simple and partial correlation coefficients of those four attributes using polychoric and polyserial correlation coefficients. It is revealed that the correlation coefficients between the four attributes differ: job satisfaction, life satisfaction, and health condition are positively correlated but the correlations between income and the other three attributes are not so strong. Moreover, we consider a way to comprehensively measure people's well-being using a multidimensional index.

We follow the row-first step in Pattanaik et al. (2012) in the sense that we first aggregate across attributes and then across individuals. This two-step method fulfills the axioms discussed in previous studies. The first step aggregating across attributes to calculate an index of individuals' current status fulfils monotonicity, symmetry, separability, rank-dependent separability, weak-scale ratio invariance, strong-scale ratio invariance, and weak translation invariance, mentioned in Decancq and Lugo (2012). The second step entails aggregating the inequality index of current status across individuals using a generalized entropy measure of inequality (GE) and Gini coefficient; GE fulfills weak principle of transfers, decomposability, scale independence, and the population principle, but the Gini coefficient does not

always fulfill population decomposability, according to Amiel and Cowell (1992, p.10).

We use the Bayesian multivariate ordered probit model for the derivation of the index. Since the seminal work of Albert and Chib (1993), the Bayesian method has been popular for the estimation of the ordered probit model. The outline of our estimation strategy is as follows. Using the Markov chain Monte Carlo (MCMC) method to estimate the econometric model proposed in the next section, the posterior distributions of the four attributes are estimated. Using the multivariate normal distribution, the probability associated with an individual's present status regarding the four attributes is calculated. Then, we use this probability to calculate the Gini coefficients and GE values as multidimensional inequality indices.

These inequality indices can also be obtained using individual characteristics such as gender and employment status (the employment status is classified into regular employment and non-regular employment in this paper). Furthermore, we can easily calculate the probability with which one inequality index is higher than another one in each employees' characteristic, for example, the probability that the female Gini coefficient is higher than the male Gini coefficient.

Studies on inequality using Bayesian methods have been growing, for example, Chotikapanich and Griffiths (2000, 2005), Griffiths et al. (2005), Gigliarano and Muliere (2013), Hasegawa and Kozumi (2003), and Nishino and Kakamu (2015). Bayesian method is more appropriate for estimating a multivariate ordered probit model than a non-Bayesian one for the following reasons. While the Bayesian method allows the estimation of the multivariate ordered probit model that has correlation between explained variables, the non-Bayesian approach is difficult when the number of explained variables is more than two.[1] It is also possible to estimate the polychoric and polyserial correlation coefficients in a multivariate ordered probit model that includes explanatory variables using the Bayesian method, although it is difficult using the maximum likelihood method. Furthermore, we need to investigate a partial correlation coefficient in addition to a simple correlation coefficient in order not to misunderstand the relationships between the explained variables. Using our method, it is easy to calculate these statistics.

This paper is organized as follows. In Section 2, we present the Bayesian multivariate ordered probit model. In Section 3, the posterior results regarding individuals' status using microlevel data are shown. In Section 4, the multidimensional inequality indices of individuals' status are calculated. Finally, in Section 5, a brief summary and some extensions of our approach are given.

## 2 Bayesian multivariate ordered probit model

Based on Hasegawa (2013, Section 5), we consider the following model. For $i = 1, \cdots, n$, let $y_{1ij}$ $(j = 1, \cdots, m)$ and $y_{2ij}$ $(j = 1, \cdots, q)$ denote the ordinal discrete and continuous variables, respectively. The discrete variable $y_{1ij}$ makes ordinal choices, that is, $y_{1ij} = c_j$ for $c_j \in \{1, \cdots, C_j\}$. We assume the following model for

[1] For example, if we use STATA for such an estimation, it is possible to estimate a bivariate ordered probit model using `BIOPROBIT`, which computes maximum likelihood estimates, but there is no function to estimate ordered probit models with three or more variables.

the ordinal discrete variable $y_{1ij}$:

$$y_{1ij} = c_j, \text{ if } \ z_{ij} \in (\gamma_{j(c_j-1)}, \gamma_{jc_j}], \ c_j \in \{1, \cdots, C_j\}; \ j = 1, \cdots, m; \ i = 1, \cdots, n, \tag{1}$$

where $z_{ij}$ denotes the latent variable and $\gamma_{jc_j}$ is a cutoff point for the $j$th ordinal response. Following Chen and Dey (2000), we specify that

$$-\infty = \gamma_{j0} < \gamma_{j1} = 0 < \gamma_{j2} < \cdots < \gamma_{j(C_j-1)} = 1 < \gamma_{jC_j} = \infty, \ j = 1, \cdots, m, \tag{2}$$

where conditions $\gamma_{j1} = 0$ and $\gamma_{j(C_j-1)} = 1$ are required to establish the identifiability of the cutoff parameters, and we define $\boldsymbol{\gamma}_j = (\gamma_{j2}, \cdots, \gamma_{j(C_j-2)})'$ $(j = 1, \cdots, m)$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \cdots, \boldsymbol{\gamma}_m')'$. Latent variable $z_{ij}$ is determined by the following model:

$$z_{ij} = \boldsymbol{x}_{1ij}'\boldsymbol{\beta}_{1j} + u_{1ij}, \ j = 1, \cdots, m; \ i = 1, \cdots, n, \tag{3}$$

where

$$\boldsymbol{x}_{1ij} = \begin{pmatrix} x_{1ij1} \\ x_{1ij2} \\ \vdots \\ x_{1ijk_{1j}} \end{pmatrix}, \ \boldsymbol{\beta}_{1j} = \begin{pmatrix} \beta_{1j1} \\ \beta_{1j2} \\ \vdots \\ \beta_{1jk_{1j}} \end{pmatrix}, \ j = 1, \cdots, m.$$

For the continuous variable $y_{2ij}$, we consider the following model:

$$y_{2ij} = \boldsymbol{x}_{2ij}'\boldsymbol{\beta}_{2j} + u_{2ij}, \ j = 1, \cdots, q; \ i = 1, \cdots, n, \tag{4}$$

where

$$\boldsymbol{x}_{2ij} = \begin{pmatrix} x_{2ij1} \\ x_{2ij2} \\ \vdots \\ x_{2ijk_{2j}} \end{pmatrix}, \ \boldsymbol{\beta}_{2j} = \begin{pmatrix} \beta_{2j1} \\ \beta_{2j2} \\ \vdots \\ \beta_{2jk_{2j}} \end{pmatrix}, \ j = 1, \cdots, q.$$

Defining $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}_{11}', \boldsymbol{\beta}_{12}', \cdots, \boldsymbol{\beta}_{1m}')'$, $\boldsymbol{\beta}_2 = (\boldsymbol{\beta}_{21}', \boldsymbol{\beta}_{22}', \cdots, \boldsymbol{\beta}_{2q}')'$, and

$$\boldsymbol{z}_i = \begin{pmatrix} z_{i1} \\ z_{i2} \\ \vdots \\ z_{im} \end{pmatrix}, \ \boldsymbol{X}_{1i} = \begin{pmatrix} \boldsymbol{x}_{1i1}' & \boldsymbol{0}' & \cdots & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{x}_{1i2}' & \cdots & \boldsymbol{0}' \\ \vdots & \vdots & & \vdots \\ \boldsymbol{0}' & \boldsymbol{0}' & \cdots & \boldsymbol{x}_{1im}' \end{pmatrix}, \ \boldsymbol{u}_{1i} = \begin{pmatrix} u_{1i1} \\ u_{1i2} \\ \vdots \\ u_{1im} \end{pmatrix}$$

$$\boldsymbol{y}_{2i} = \begin{pmatrix} y_{2i1} \\ y_{2i2} \\ \vdots \\ y_{2iq} \end{pmatrix}, \ \boldsymbol{X}_{2i} = \begin{pmatrix} \boldsymbol{x}_{2i1}' & \boldsymbol{0}' & \cdots & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{x}_{2i2}' & \cdots & \boldsymbol{0}' \\ \vdots & \vdots & & \vdots \\ \boldsymbol{0}' & \boldsymbol{0}' & \cdots & \boldsymbol{x}_{2iq}' \end{pmatrix}, \ \boldsymbol{u}_{2i} = \begin{pmatrix} u_{2i1} \\ u_{2i2} \\ \vdots \\ u_{2iq} \end{pmatrix},$$

(3) and (4) can be rewritten as

$$\boldsymbol{z}_i = \boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \boldsymbol{u}_{1i}, \ i = 1, \cdots, n \tag{5}$$

$$\boldsymbol{y}_{2i} = \boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + \boldsymbol{u}_{2i}, \ i = 1, \cdots, n. \tag{6}$$

Further, defining $\boldsymbol{w}_i = (\boldsymbol{z}_i', \boldsymbol{y}_{2i}')'$, (5) and (6) can be combined as

$$\boldsymbol{w}_i = \begin{pmatrix} \boldsymbol{z}_i \\ \boldsymbol{y}_{2i} \end{pmatrix} = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{u}_i,$$

where

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}_{1i} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{X}_{2i} \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \ \boldsymbol{u}_i = \begin{pmatrix} \boldsymbol{u}_{1i} \\ \boldsymbol{u}_{2i} \end{pmatrix}.$$

Assuming that $\boldsymbol{u}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, we have

$$\boldsymbol{w}_i | \boldsymbol{\beta}, \boldsymbol{\Sigma} \sim \mathrm{N}(\boldsymbol{X}_i \boldsymbol{\beta}, \boldsymbol{\Sigma}), \ i = 1, \cdots, n, \tag{7}$$

where $\boldsymbol{\Sigma}$ is an $(m+q) \times (m+q)$ positive definite covariance matrix.[2] Further, we collect up $\boldsymbol{y}_{1i} = (y_{1i1}, \cdots, y_{1im})', \boldsymbol{y}_{2i}, \boldsymbol{z}_i, \boldsymbol{w}_i$ for $i = 1, \cdots, n$ as follows:

$$\boldsymbol{y}_1 = \begin{pmatrix} \boldsymbol{y}_{11} \\ \vdots \\ \boldsymbol{y}_{1n} \end{pmatrix}, \ \boldsymbol{y}_2 = \begin{pmatrix} \boldsymbol{y}_{21} \\ \vdots \\ \boldsymbol{y}_{2n} \end{pmatrix}, \ \boldsymbol{z} = \begin{pmatrix} \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_n \end{pmatrix}, \ \boldsymbol{w} = \begin{pmatrix} \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{w}_n \end{pmatrix}.$$

To complete the Bayesian model, we introduce the prior distributions of the parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})$. On the basis of Bayes' theorem, the joint posterior distribution can be written as

$$\begin{aligned} p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{z} | \boldsymbol{y}_1, \boldsymbol{y}_2) &\propto p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{z}) p(\boldsymbol{y}_1, \boldsymbol{y}_2 | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{z}) \\ &= p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) \left[ \prod_{i=1}^{n} p(\boldsymbol{w}_i | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) p(\boldsymbol{y}_{1i} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{z}_i) \right]. \end{aligned} \tag{8}$$

Further, defining

$$\begin{gathered} \mathcal{G}_{ij} = \left( \gamma_{j(c_j - 1)}, \gamma_{j c_j} \right] \ \text{if} \ y_{1ij} = c_j, \ c_j \in \{1, \cdots, C_j\}, \ j = 1, \cdots, m \\ \mathcal{G}_i = \mathcal{G}_{i1} \times \cdots \times \mathcal{G}_{im}, \ i = 1, \cdots, n, \end{gathered}$$

we have

$$p(\boldsymbol{y}_{1i} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{z}_i) = 1(\boldsymbol{z}_i \in \mathcal{G}_i), \ i = 1, \cdots, n,$$

where $1(\cdot)$ is an indicator function.[3] Now, we specify the prior distributions as follows:

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) = p(\boldsymbol{\beta}) p(\boldsymbol{\gamma}) p(\boldsymbol{\Sigma}) = p(\boldsymbol{\beta}) \left\{ \prod_{j=1}^{m} p(\boldsymbol{\gamma}_j) \right\} p(\boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0), \ \boldsymbol{\Sigma}^{-1} \sim \mathrm{W}(\kappa_0, \boldsymbol{Q}_0^{-1}),$$

[2] This multivariate normal model can be extended to a mixture of a countable number of multivariate normal distributions, which would be more flexible than a multivariate normal distribution.

[3] See Chib and Greenberg (1998, p.349).

and $\mathrm{W}(\kappa_0, \boldsymbol{Q}_0^{-1})$ denotes a Wishart distribution with degrees of freedom $\kappa_0$ and scale matrix $\boldsymbol{Q}_0^{-1}$. Further, we introduce the prior distribution of $\boldsymbol{\gamma}_j$, $p(\boldsymbol{\gamma}_j) = p(\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j))$, based on the following transformation for the cutoff points (Chen and Dey, 2000, p.140):

$$\delta_{jc_j} = \log\left(\frac{\gamma_{jc_j} - \gamma_{j(c_j-1)}}{1-\gamma_{jc_j}}\right),\ c_j = 2, \cdots, C_j - 2,$$

where $\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j) = (\delta_{j2}, \cdots, \delta_{j(C_j-2)})'$ $(j = 1, \cdots, m)$.[4] We specify $p(\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j))$ as follows:

$$\boldsymbol{\delta}_j(\boldsymbol{\gamma}_j) \sim \mathrm{N}(\boldsymbol{\delta}_{j0}, \boldsymbol{D}_{j0}),\ j = 1, \cdots, m.$$

Using the sampling scheme of MCMC, we can sample the parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{z})$ from the joint posterior distribution (8).[5]

## 3 Empirical analysis of individuals' status

### 3.1 Data

We use the data of the Questionnaire Survey on Work and Life of Workers by RENGO-RIALS. RENGO stands for the Japanese Trade Union Confederation, Japan's national-level trade union. RENGO-RIALS stands for RENGO's Research Institute for Advancement of Living Standards.

This survey is conducted biannually, in April and October, and investigates private companies' employees on their perceptions of Japanese national economic trends, work, and life using a self-completed questionnaire. The districts of residence of the respondents are prefectures around Tokyo (Tokyo, Saitama, Chiba, and Kanagawa) and Osaka (Osaka, Shiga, Kyoto, Hyogo, Nara, and Wakayama). The age of the respondents ranges from the 20s to early 60s. In our study, we use responses to the following four questions in the 16th wave of the survey, which was conducted in October 2008: job satisfaction (`sat1`), life satisfaction (`sat2`), current health condition (`health`), and annual wage (`income`).

Table 1 shows the survey questions and answer choices, and Table 2a shows the relative frequencies of the responses regarding `sat1`, `sat2`, and `health`, and the summary statistics of `income`. Further, we obtain respondents' socio-demographic data, such as gender, age, marital status, and employment status, from the survey. Table 2b shows these summary statistics.[6]

[4] Since we use the identification restriction on cutoff, $\gamma_{\gamma_{j(C_j-1)}} = 1$, we have $\gamma_{jc_j} \in (0, 1)$. This transformation makes the interval $(0, 1)$ to the real line. For more detailed discussion on this matter, see Jeliazkov et al. (2008, p.124).

[5] Supplementary material provides the details of the sampling algorithms for the Bayesian estimation.

[6] We use wage data as income data. The data provided in the survey are those of the wage band. The number of income bands are 15, and these are 0 yen, less than 1 million yen, 1–2 million yen, $\cdots$, 15–20 million yen, and more than 20 million yen, but data for individuals with zero income are not used in our estimation in practice. We construct the numerical values of income on the basis of income band in accordance with Layard et al. (2008). Layard et al. (2008, p.1850) construct the income data as follows: "In the cross-section surveys, only income bands are available and these we converted into numerical values using the mid-point of each

### 3.2 Posterior results of the estimated equations

The estimated equations are as follows:

$$\begin{aligned} w_j = {} & \beta_{j1} + \beta_{j2}\,\texttt{female} + \beta_{j3}\,\texttt{emp} + \beta_{j4}\,\texttt{emp} \times \texttt{female} \\ & + \beta_{j5}\,\texttt{married} + \beta_{j6}\,\texttt{age} + \beta_{j7}\,\texttt{age}^2 + \beta_{j8}\,\texttt{educ} \\ & + \beta_{j9}\,\texttt{tenure} + \beta_{j10}\,\texttt{tenure}^2 + u_j,\ j = 1, \cdots, 4, \end{aligned} \tag{9}$$

where $j = 1$ to 3 correspond to `sat1`, `sat2`, and `health`, and $j = 4$ corresponds to `log(income)`, respectively.[7] Variables `female`, `emp`, and `married` are dummy variables. When the respondents are women, regular employees, and married, $\texttt{female} = 1$, $\texttt{emp} = 1$, and $\texttt{married} = 1$, respectively. Variables `age`, `educ`, and `tenure` denote the respondents' age, length of education, and years at the current employer, respectively.[8] In the survey, the educational level is classified as junior high school graduate, high school graduate, graduate from vocational schools/schools in the miscellaneous category, junior/technical college graduate, university graduate, or completed graduate school; the values 9, 12, 15, 14, 16, and 18 are given, respectively, as the data for the length of education.

The MCMC simulation was run for 30,000 iterations with the thinning interval being 5; the first 10,000 samples were discarded as the burn-in period.[9] The posterior results obtained thereafter were generated using Ox version 7.0 (Doornik, 2013). We set the prior distributions as follows:

$$\begin{aligned} \boldsymbol{\beta}_j &\sim \mathrm{N}(\mathbf{0}, 100\boldsymbol{I}_{10}),\ j = 1, 2, 3, 4 \\ \boldsymbol{\delta}_j(\boldsymbol{\gamma}_j) &\sim \mathrm{N}(0, 100),\ j = 1, 2, 3 \\ \boldsymbol{\Sigma}^{-1} &\sim \mathrm{W}(6, 50\boldsymbol{I}_4). \end{aligned}$$

Table 3 shows the posterior results of the multivariate ordered probit model. The effects of the differences in gender and employment status on each explained variable are also shown in this table. Coefficients $\beta_{j,2}$ and $\beta_{j,2} + \beta_{j,4}$ represent the effects of being female employees on the explained variables when employees are non-regular and regular, respectively. Coefficients $\beta_{j,3}$ and $\beta_{j,3} + \beta_{j,4}$ represent the effects of being regular employees on the explained variables when employees are male and female, respectively.

According to this table, the 95% credible intervals (95% CIs) of `married` are positive and do not include zero. This means that being married has positive effects on job satisfaction (`sat1`), life satisfaction (`sat2`), and health condition (`health`). For job satisfaction, life satisfaction, and health equations, coefficients of `age` are negative and those of $\texttt{age}^2$ are positive, and their 95% CIs do not include zero. Thus, relationships between these explained variables and age are U-shaped. In

band. For respondents in the lowest income band, we assumed an income of two thirds of the upper limit of the band, and for respondents in the highest income band, we assumed an income 1.5 times the lower income limit of the band."

[7] Variable `log(income)` is the logarithm of annual wage.

[8] Regular employees include the following employees in the survey. Regular employees are full-time employees, and include executive-level employees. Non-regular employees include part-timers, contract workers, dispatched workers, and other non-full-time workers.

[9] We obtained 30,000 samples by running 150,000 iterations and sampling one observation at every five iterations. After obtaining 30,000 samples, we discarded the first 10,000 samples.

addition, being regular employees (`emp`) has a positive effect on life satisfaction (`sat2`) in the case of male employees. Length of education (`educ`) too has a positive effect on life satisfaction.

For the logarithms of income, the 95% CIs of most explanatory variables do not include zero. Thus, these explanatory variables can have either positive or negative noteworthy effects on income. The signs of the effects of these variables are the same as those of their coefficients.

The effects of being female employees on income are represented by $\beta_{4,2}$ in the case of non-regular employment (`emp` $= 0$) and by $\beta_{4,2} + \beta_{4,4}$ in the case of regular employment (`emp` $= 1$). Being female employees has a negative effect on income in any of these cases, and this effect is bigger for non-regular employees.[10]

The effects of being regular employees (`emp`) on income are represented by $\beta_{4,3}$ in the case of males (`female` $= 0$) and by $\beta_{4,3} + \beta_{4,4}$ in the case of females (`female` $= 1$). Being regular employees has a positive effect on income in any of these cases, and this effect is bigger for females.[11]

Furthermore, length of education (`educ`) has a positive effect on income. It is conceivable that age (`age`) and years at the current employer (`tenure`) have inverted U-shaped relationships with income, because of the signs of the coefficients. Age has such an inverted U-shaped relationship with income, because the 95% CIs of `age` and `age`$^2$ do not include zero. However, years at the current employer do not always have such a relationship, because the 95% CI of `tenure`$^2$ includes zero.

The signs of the coefficients whose 95% CIs do not include zero in Table 3 are valid in general. For example, being married may yield a fulfilling home life, motivation to work, and good health. These factors raise their job satisfaction, life satisfaction, and improve their health condition. The relationships between job and life satisfaction and one's age have been confirmed in many previous studies like Clark et al. (1996) and Blanchflower and Oswald (2004). The U-shaped relationship is also observed in Hasegawa and Ueda (2013) who study job satisfaction of employees in Japanese private companies using the same data. The inverted U-shaped relationship between income and age (or labor force experience) is also known in the estimation of wage function (Heckman et al., 2006, pp.320–326). Considering that the average wages of female employees and non-regular employees are lower than that of male employees and regular employees, respectively, it is also reasonable that being female employees has a negative and being regular employees has a positive effect on income.

Regarding the effects of differences in gender and employment status on job satisfaction, life satisfaction, and health condition, for example, it may be considered that being female and being a non-regular employee have negative effects, because the incomes of female employees and non-regular employees are lower than those of male employees and regular employees. On the other hand, it may be considered that female employees and non-regular employees enjoy higher satisfaction from non-pecuniary factors such as convenient work timings and a healthy work–life

[10] The income of females is lower than that of males by 55.1% ($-0.551 = \exp(\beta_{4,2}) - 1$) for non-regular employees, and lower by 33.2% ($-0.332 = \exp(\beta_{4,2} + \beta_{4,4}) - 1$) for regular employees.

[11] The income of regular employees is higher than that of non-regular employees by 64.2% ($0.642 = \exp(\beta_{4,3}) - 1$) for males, and higher by 155.5% ($1.555 = \exp(\beta_{4,2} + \beta_{4,4}) - 1$) for females.

balance. Thus, the results for the effects of differences in gender and employment status may not be clear.

### 3.3 Simple and Partial Correlations

Table 4 shows the medians of polychoric and polyserial simple correlation coefficients and those of polychoric and polyserial partial correlation coefficients between the explained variables in the estimated model (9).[12]

This table shows that all simple correlation coefficients are positive and their 95% CIs do not include zero, except for the correlation coefficient between health condition and income.[13] In particular, the simple correlation coefficient between job satisfaction and life satisfaction is 0.5766 and higher than other correlation coefficients. The simple correlation coefficient between life satisfaction and health condition is nearly equal to that between job satisfaction and health condition. These coefficients are the second and third highest, at 0.3271 and 0.2803, respectively. There exist positive correlations between job satisfaction, life satisfaction, and health condition as described, but the correlations between income and other explained variables are weak.

The simple correlation coefficient measures a correlation between two variables. Since it includes indirect correlations between the two variables via the other variables, we also need to see the partial correlation. As for the partial correlation coefficients, those between any two variables among job satisfaction, life satisfaction, and health condition are positive and their 95% CIs do not include zero. This implies a positive correlation. In particular, the partial correlation between job satisfaction and life satisfaction is the highest at 0.5291. The 95% CIs for the other partial correlation coefficients include zero. Therefore, job satisfaction has a clearly positive correlation with life satisfaction. On the other hand, it is not ascertained that income has positive correlations with the other explained variable. This does not always mean that income is an important factor for them.

Many studies, including Judge and Watanabe (1993) and Diener et al. (1999), on the relationship between job satisfaction and life satisfaction have shown that the two are positively correlated, but the causality between the two is yet to be clarified. Our result also indicates a high correlation. Regarding the relationships between job satisfaction and health condition and between life satisfaction and health condition, a positive correlation is shown as in previous studies such as Faragher et al. (2005). At the same time, it is suggested that a pecuniary factor such as income does not play an important role for job satisfaction, life satisfaction, and health condition.

## 4 Multidimensional inequality indices

In this section, we consider a multidimensional inequality indices using the posterior results of the Bayesian model. An advantage of Bayesian analysis is that the value of latent variable $\boldsymbol{z}_i$ can be directly obtained from the posterior results in

[12] Appendix A1 provides the derivation of partial correlation coefficients.

[13] However, the 90% CI of the simple correlation coefficient between health condition and income does not include zero.

addition to the observed data $\boldsymbol{y}_{i2}$. However, the value of $\boldsymbol{z}_i$ may change with the identification restrictions on the parameters. Therefore, instead of latent variable $\boldsymbol{w}_i = (\boldsymbol{z}_i', \boldsymbol{y}_{i2}')'$, we consider a probability associated with individual $i$'s present status regarding job satisfaction, life satisfaction, health condition, and income.

Define the following set:

$$(-\infty\mathbf{1}, \boldsymbol{w}_i^*] = (-\infty, z_{i1}^*] \times \cdots (-\infty, z_{im}^*] \times (-\infty, y_{2i1}^*] \times \cdots \times (-\infty, y_{2iq}^*],$$

where $\boldsymbol{w}_i^* = (\boldsymbol{z}_i^{*\prime}, \boldsymbol{y}_{i2}^{*\prime})'$, $\boldsymbol{z}_i^*$ is the value of $\boldsymbol{z}_i$ obtained from the posterior results, $\boldsymbol{y}_{i2}^*$ is the observed data of $\boldsymbol{y}_{i2}$, and $\mathbf{1} = (\underbrace{1, \cdots, 1}_{m+q})'$. Using the set $(-\infty\mathbf{1}, \boldsymbol{w}_i^*]$, we define the probability associated with individual $i$'s status including his/her subjective perception as follows:[14]

$$p_i = \Pr\left(\boldsymbol{w}_i \in (-\infty\mathbf{1}, \boldsymbol{w}_i^*] | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{X}\right), \ i = 1, \cdots, n. \tag{10}$$

Probability $p_i$ derived from (10) represents the probability that an individual attains his/her present status regarding job satisfaction, life satisfaction, health condition, and income. A difference in that probability means that there exists a possibility that the individual's present status can be better. In other words, this probability represents the individual's overall status including his/her subjective perception. Thus, it is conceivable that we can show the inequality of individuals' overall status by investigating the distribution of that probability.

To measure a multidimensional inequality, we calculate the individual's index aggregating the four attributes, $p_i$, from (10) in the first step and the overall inequality using the Gini coefficients and GE in the second step. This two-step aggregation to measure the multidimensional inequality fulfills the axioms discussed in the previous studies. In the first step, $p_i$ fulfills monotonicity, symmetry, separability, weak-scale ratio invariance, strong-scale ratio invariance, and weak translation invariance, but does not fulfill normalization.[15] In the second step, GE fulfills the weak principle of transfers, decomposability, scale independence, and the population principle, but Gini does not always fulfill population decomposability, according to Amiel and Cowell (1992, p.10).

GE is defined as follows:

$$\mathrm{GE}(\alpha) = \frac{1}{\alpha(\alpha-1)} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{p_i}{\bar{p}} \right)^{\alpha} - 1 \right\}, \ \alpha \neq 0, 1.$$

In particular, GE(0) represents a mean logarithmic deviation:

$$\mathrm{GE}(0) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{\bar{p}}{p_i},$$

and GE(1) is called Theil's measure:

$$\mathrm{GE}(1) = \frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{\bar{p}} \log \frac{p_i}{\bar{p}}.$$

[14] Appendix A2 provides details of this probability.

[15] Proofs are provided in appendix A3.

Table 5 shows the posterior means of individuals' status inequality, including their subjective perception using the Gini coefficient and GE. They are shown by gender, employment status, and their combinations. The employment status is classified as regular employment and non-regular employment in this table. All inequality indices in Table 5 show that the inequality of female employees is greater than that of male ones and that the inequality of non-regular employees is greater than that of regular ones. Furthermore, when we compare the inequality indices of non-regular employees and those of regular ones by gender, non-regular employees' indices are greater than regular employees' indices in both cases.

To evaluate the appropriateness of these results, we use Table 6 showing the probability by which one inequality index is higher than the other in one attribute.[16] For example, the probability that the Gini coefficient of female employees is greater than that of male ones is 0.9061 and the probability that the Gini coefficient of non-regular employees is greater than that of regular employees is 0.9986. According to this table, the inequality of non-regular employees is greater than that of regular ones with high probability for all employees. Further, similar results are also observed by gender. The probability that the inequality indices of female employees are greater than those of male ones is high.

However, when compared by employment status, the probability that the inequality indices of female employees are greater than those of male ones is not always high. In particular, for regular employees, the probability that the inequality indices of female employees are greater than those of male ones is low except for in the case of GE(0). These results show that the difference in employment status has a larger effect on the inequality of employees' status than the difference in gender.

## 5 Concluding remarks

In this paper, we proposed a method to aggregate multidimensional variates into a single index using a multivariate probability distribution to measure the inequality in a population by evaluating many attributes.

As an application to an empirical analysis, we use the data for four attributes of Japanese private companies' employees: income, and subjective perceptions of job, life, and health condition. Using the multivariate ordered probit model, we estimated the relationships between these four attributes and socio-demographic factors such as gender and tenure, and calculated simple and partial correlation coefficients.

We obtain a single index from the four attributes using a multivariate normal distribution. This index is denoted as a probability $p_i$ that employees attain their present status regarding the four attributes. Then, we measure the inequality of

[16] The probabilities are approximated as follows: for inequality indices of attributes $A$ and $B$,

$$\Pr(A > B|\text{data}) \approx \frac{1}{S}\sum_{s=1}^{S} 1(A^{(s)} > B^{(s)}),$$

where $S$ is the number of simulations.

the employees' present status including their subjective perceptions by calculating their Gini coefficients and GEs of inequality.

In the results, the inequality of female employees is larger than that of male ones for all indices. Even if the data are divided into gender, the inequality of non-regular employees is larger than that of regular ones for both males and females. When the difference in employment status is not taken into account, the probability that the inequality of female employees is larger than that of male ones ranges from 0.64 to 0.97. However, when the difference between regular and non-regular employees is considered, this probability decreases. Thus, it is suggested that the difference in employment status has meaningful effects on the inequality of employees' status including their subjective perceptions.

Further, the Bayesian model used in this paper can be extended to a Bayesian model with more flexible families of distributions for $\boldsymbol{w}_i$. For example, Chen and Dey (2000) propose a Bayesian multivariate ordinal data model with a class of scale mixtures of multivariate normal link functions. Further, using a Dirichlet process (DP) prior, Kottas et al. (2005) provide a Bayesian nonparametric model for multivariate ordinal data. In particular, the Bayesian nonparametric models with a DP prior and a Polya tree have been used in the literature on inequality; see Hasegawa and Kozumi (2003) and Gigliarano and Muliere (2013).

## Appendix

### A1 Correlation matrices

Following Hasegawa (2013), we calculate the *simple* version of the polychoric and polyserial correlation matrix at each MCMC iteration based on the sample of $\boldsymbol{\Sigma}$:

$$\boldsymbol{R}_s = \{r_{ij}^s\} = \boldsymbol{D}_s \boldsymbol{\Sigma} \boldsymbol{D}_s, \tag{11}$$

where $\boldsymbol{D}_s = \mathrm{diag}(1/\sqrt{\sigma_{11}}, \cdots, 1/\sqrt{\sigma_{m+q,m+q}})$ and $\sigma_{jj}$ is the $j$th diagonal element of $\boldsymbol{\Sigma}$. Further, we can calculate the *partial* version of the polychoric and polyserial correlation coefficients defined as the lower triangular part without the diagonal elements of the following matrix:

$$\boldsymbol{R}_p = \{r_{ij}^p\} = -\boldsymbol{D}_r \boldsymbol{R}_s^{-1} \boldsymbol{D}_r, \tag{12}$$

where $\boldsymbol{D}_r = \mathrm{diag}(1/\sqrt{r^{11}}, \cdots, 1/\sqrt{r^{m+q,m+q}})$ and $r^{jj}$ is the $j$th diagonal element of $\boldsymbol{R}_s^{-1}$.

### A2 Calculation of probability associated with employee's current status

Deleting the subscript $i$ in (7) for simplicity, we have $\boldsymbol{w} \sim \mathrm{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$. Since $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix, we can apply the Cholesky factorization to $\boldsymbol{\Sigma}$, that is $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{U}$, where $\boldsymbol{L}$ is a lower triangular matrix and $\boldsymbol{U} = \boldsymbol{L}'$ is a upper triangular matrix. Defining $\boldsymbol{v} = \boldsymbol{L}^{-1}(\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta})$, we have

$$\begin{aligned}
\mathrm{E}(\boldsymbol{v}) &= \boldsymbol{L}^{-1}[\mathrm{E}(\boldsymbol{w}) - \boldsymbol{X}\boldsymbol{\beta}] = \boldsymbol{0} \\
\mathrm{var}(\boldsymbol{v}) = \mathrm{E}(\boldsymbol{v}\boldsymbol{v}') &= \boldsymbol{L}^{-1}\,\mathrm{E}[(\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta})(\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta})'](\boldsymbol{L}^{-1})' \\
&= \boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}')^{-1} = \boldsymbol{L}^{-1}(\boldsymbol{L}\boldsymbol{U})\boldsymbol{U}^{-1} = \boldsymbol{I}.
\end{aligned}$$

Thus, we have $\boldsymbol{v} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{I})$. The distribution function of $\boldsymbol{w}$ for $(-\infty\mathbf{1}, \boldsymbol{w}^*]$ can be calculated as follows:

$$\begin{aligned}
&\Pr\left(\boldsymbol{w} \in (-\infty\mathbf{1}, \boldsymbol{w}^*] | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{X}\right) = \int_{\boldsymbol{w}\in(-\infty\mathbf{1}, \boldsymbol{w}^*]} p(\boldsymbol{w}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{X}) d\boldsymbol{w} \\
&= \int_{\boldsymbol{w}\in(-\infty\mathbf{1}, \boldsymbol{w}^*]} (2\pi)^{-\frac{1}{2}(m+q)} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta})\right] d\boldsymbol{w} \\
&= \int_{\boldsymbol{v}\in(-\infty\mathbf{1}, \boldsymbol{L}^{-1}(\boldsymbol{w}^* - \boldsymbol{X}\boldsymbol{\beta})]} (2\pi)^{-\frac{1}{2}(m+q)} \exp\left(-\frac{\boldsymbol{v}'\boldsymbol{v}}{2}\right) d\boldsymbol{v} \\
&= \prod_{j=1}^{m+q} \int_{-\infty}^{v_j^*} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{v_j^2}{2}\right) dv_j,
\end{aligned}$$

where $\boldsymbol{v}^* = (v_1^*, \cdots, v_{m+q}^*)' = \boldsymbol{L}^{-1}(\boldsymbol{w}^* - \boldsymbol{X}\boldsymbol{\beta})$.

A3 Axioms the first aggregation step fulfills

Let $F$ be a distribution function of a multivariate normal distribution, and omit subscript $i$ in (10):

$$p = \Pr\left(\boldsymbol{w} \in (-\infty\mathbf{1}, \boldsymbol{w}^*] | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{X}\right) = F(\boldsymbol{w}^*) = F(w_1^*, \cdots, w_{m+q}^*),\ \boldsymbol{w}^* \in \mathbb{R}^{m+q}. \tag{13}$$

Monotonicity: If $\boldsymbol{w}^* > \tilde{\boldsymbol{w}}^*$ for any vector $\boldsymbol{w}^*$ and $\tilde{\boldsymbol{w}}^*$, then

$$p = F(\boldsymbol{w}^*) > F(\tilde{\boldsymbol{w}}^*) = \tilde{p}.$$

(Proof) It holds because a distribution function of a multivariate normal distribution is monotonic. □

Symmetry: For any permutation matrix $\boldsymbol{P}$,

$$F(\boldsymbol{P}\boldsymbol{w}^*) = F(\boldsymbol{w}^*).$$

(Proof) Permutation of the attributes does not change the probability calculated from a multivariate normal distribution. □

Separability: Let us define $\boldsymbol{w}^* = \binom{\boldsymbol{w}_1^*}{\boldsymbol{w}_2^*}$, $\tilde{\boldsymbol{w}}^* = \binom{\tilde{\boldsymbol{w}}_1^*}{\tilde{\boldsymbol{w}}_2^*}$, $\boldsymbol{v}^* = \binom{\boldsymbol{v}_1^*}{\boldsymbol{v}_2^*}$, $\tilde{\boldsymbol{v}}^* = \binom{\tilde{\boldsymbol{v}}_1^*}{\tilde{\boldsymbol{v}}_2^*}$. If $\boldsymbol{w}_1^* = \boldsymbol{v}_1^*$, $\tilde{\boldsymbol{w}}_1^* = \tilde{\boldsymbol{v}}_1^*$, $\boldsymbol{w}_2^* = \tilde{\boldsymbol{w}}_2^*$, and $\boldsymbol{v}_2^* = \tilde{\boldsymbol{v}}_2^*$, then

$$F(\boldsymbol{w}^*) > F(\boldsymbol{v}^*) \ \Leftrightarrow\ F(\tilde{\boldsymbol{w}}^*) > F(\tilde{\boldsymbol{v}}^*).$$

(Proof) ($\Rightarrow$) Since a distribution function of a multivariate normal distribution is monotonic, $\boldsymbol{w}^* > \boldsymbol{v}^*$ because $F(\boldsymbol{w}^*) > F(\boldsymbol{v}^*)$. As $\boldsymbol{w}_1^* = \boldsymbol{v}_1^*$, $\boldsymbol{w}_2^* > \boldsymbol{v}_2^*$. In addition, $\tilde{\boldsymbol{w}}^* > \tilde{\boldsymbol{v}}^*$, because $\tilde{\boldsymbol{w}}_1^* = \tilde{\boldsymbol{v}}_1^*$ and $\tilde{\boldsymbol{w}}_2^* = \boldsymbol{w}_2^* > \boldsymbol{v}_2^* = \tilde{\boldsymbol{v}}_2^*$. Therefore, $F(\tilde{\boldsymbol{w}}^*) > F(\tilde{\boldsymbol{v}}^*)$. The converse can be similarly shown. □

Weak-ratio scale invariance: For any vector $\boldsymbol{w}^*$, $\tilde{\boldsymbol{w}}^*$, and any $\lambda > 0$,

$$F(\boldsymbol{w}^*) > F(\tilde{\boldsymbol{w}}^*) \ \Leftrightarrow\ F(\lambda\boldsymbol{w}^*) > F(\lambda\tilde{\boldsymbol{w}}^*).$$

(Proof) Since a distribution function of a multivariate normal distribution is monotonic, we have

$$F(\boldsymbol{w}^*) > F(\tilde{\boldsymbol{w}}^*) \;\Leftrightarrow\; \boldsymbol{w}^* > \tilde{\boldsymbol{w}}^* \;\Leftrightarrow\; \lambda\boldsymbol{w}^* > \lambda\tilde{\boldsymbol{w}}^* \;\Leftrightarrow\; F(\lambda\boldsymbol{w}^*) > F(\lambda\tilde{\boldsymbol{w}}^*). \;\square$$

Strong-ratio scale invariance: For any vector $\boldsymbol{w}^*, \tilde{\boldsymbol{w}}^* \in \mathbb{R}^{m+q}$, and any diagonal matrix with positive diagonal elements $\boldsymbol{\Lambda}$,

$$F(\boldsymbol{w}^*) > F(\tilde{\boldsymbol{w}}^*) \;\Leftrightarrow\; F(\boldsymbol{\Lambda}\boldsymbol{w}^*) > F(\boldsymbol{\Lambda}\tilde{\boldsymbol{w}}^*).$$

(Proof) Since a distribution function of a multivariate normal distribution is monotonic, we have

$$F(\boldsymbol{w}^*) > F(\tilde{\boldsymbol{w}}^*) \;\Leftrightarrow\; \boldsymbol{w}^* > \tilde{\boldsymbol{w}}^* \;\Leftrightarrow\; \boldsymbol{\Lambda}\boldsymbol{w}^* > \boldsymbol{\Lambda}\tilde{\boldsymbol{w}}^* \;\Leftrightarrow\; F(\boldsymbol{\Lambda}\boldsymbol{w}^*) > F(\boldsymbol{\Lambda}\tilde{\boldsymbol{w}}^*). \;\square$$

Weak translation invariance: For any vectors $\boldsymbol{w}^*, \tilde{\boldsymbol{w}}^* \in \mathbb{R}^{m+q}$, any $\kappa \in \mathbb{R}$, and a vector $\mathbf{1}$ with $m+q$ entries equal to one,

$$F(\boldsymbol{w}^*) > F(\tilde{\boldsymbol{w}}^*) \;\Leftrightarrow\; F(\boldsymbol{w}^* + \kappa\mathbf{1}) > F(\boldsymbol{\Lambda}\tilde{\boldsymbol{w}}^* + \kappa\mathbf{1}).$$

(Proof) Since a distribution of a multivariate normal distribution is monotonic, we have

$$\begin{aligned} F(\boldsymbol{w}^*) > F(\tilde{\boldsymbol{w}}^*) \;\Leftrightarrow\; \boldsymbol{w}^* > \tilde{\boldsymbol{w}}^* \;\Leftrightarrow\; \boldsymbol{w}^* + \kappa\mathbf{1} > \boldsymbol{\Lambda}\tilde{\boldsymbol{w}}^* + \kappa\mathbf{1} \\ \Leftrightarrow\; F(\boldsymbol{w}^* + \kappa\mathbf{1}) > F(\boldsymbol{\Lambda}\tilde{\boldsymbol{w}}^* + \kappa\mathbf{1}). \quad \square \end{aligned}$$

**Acknowledgements** The authors appreciate the comments of an anonymous referee, which improve the article greatly. The data for this secondary analysis, "The Questionnaire Survey on Work, Life of Workers, RENGO-RIALS (Research Institute for Advancement of Living Standards)," was provided by the Social Science Japan Data Archive, Center for Social Research and Data Archives, Institute of Social Science, The University of Tokyo. Further, the work of the first author was supported in part by a Grant-in-Aid for Scientific Research (No.25380261) from the Japan Society for the Promotion of Science (JSPS), while that of the first and second authors was supported in part by a Grant-in-Aid for Scientific Research (No.24530236) from the JSPS.

## References

Albert, J.H., Chib, S.: Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88(422), 669–679 (1993)

Amiel, Y. and Cowell, F.A.: Measurement of income inequality: Experimental test by questionnaire. Journal of Public Economics 47(1), 3–26 (1992)

Atkinson, A.B., Bourguignon, F.: The comparison of multi-dimensioned distributions of economic status. Review of Economic Studies 49(2), 183–201 (1982)

Blanchflower, D.G., Oswald, A.J.: Well-being over time in Britain and the USA. Journal of Public Economics 88(7–8), 1359–1386 (2004)

Chen, M.H., Dey, D.K.: Bayesian analysis for correlated ordinal data models. In: Dey D.K., Ghosh S.K., Mallick B.K. (eds.) Generalized Linear Models: A Bayesian Perspective, Marcel Dekker, New York, 133–157 (2000)

Chib, S., Greenberg, E.: Analysis of multivariate probit models. Biometrika 85(2), 347–361 (1998)

Chotikapanich, D., Griffiths, W.E.: Posterior distributions for the Gini coefficient using grouped data. Australian and New Zealand Journal of Statistics 42(4), 383–392 (2000)

Chotikapanich, D., Griffiths, W.E.: Averaging Lorenz curves. Journal of Economic Inequality 3(1), 1–19 (2005)

Clark, A.E., Oswald, A.J., Warr P.: Is job satisfaction U-shaped in age? Journal of Occupational and Organizational Psychology 69(1), 57–81 (1996)

Dardanoni, V.: On multidimensional inequality measurement. In: Dagum C., Lemmi, A. (eds.) Research on Economic Inequality 6, JAI Press, Greenwich, 201–207 (1995)

Decancq, K., Lugo, M.A.: Inequality of wellbeing: A multidimensional approach. Economica 79(316), 721–746 (2012)

Diener, E., Suh, E., Lucas, R., Smith, H.L.: Subjective well-being: Three decades of progress. Psychological Bulletin 125(2), 276–302 (1999)

Doornik, J.A.: An Object-oriented Matrix Programming Language Ox$^{TM}$ 7. Timberlake Consultants Ltd, London (2013)

Faragher, E.B., Cass, M., Cooper, C.L.: The relationship between job satisfaction and health: A meta-analysis. Occupational and Environmental Medicine 62(2), 105–112 (2005)

Gajdos, T., Weymark, J.A.: Multidimensional generalized Gini indices. Economic Theory 26(3), 471–496 (2005)

Gigliarano, C., Muliere, P.: Estimating the Lorenz curve and Gini index with right censored data: A Polya tree approach. Metron 71(2), 105–122 (2013)

Griffiths, W.E., Chotikapanich, D., Rao, D.S.P.: Averaging income distributions. Bulletin of Economic Research 57(4), 347–367 (2005)

Hasegawa, H.: On polychoric and polyserial partial correlation coefficients: A Bayesian approach. Metron 71(2), 139–156 (2013)

Hasegawa, H., Kozumi, H.: Estimation of Lorenz curves: A Bayesian nonparametric approach. Journal of Econometrics 115(2), 277–291 (2003)

Hasegawa, H., Ueda, K.: Analysis of job satisfaction: The case of Japanese private companies. Labour, 30(1), 109–134 (2016)

Heckman, J.J., Lochner, L.J., Todd, P.E.: In Hanushek, E.A., Welch, F. (Eds.), Handbook of the Economics of Education, Volume 1, North-Holland, Amsterdam, 307–458 (2006)

Jeliazkov, I., Graves, J., Kutzbach, M.: Fitting and comparison of models for multivariate ordinal outcomes. In: Chib, S., Koop, G., Griffiths, W., Terrell, D. (eds.) Advances in Econometrics: Bayesian Econometrics, Volume 23, Emerald, Bingley, 115–156 (2008)

Judge, T.A., Watanabe, S.: Another look at the job satisfaction-life satisfaction relationship. Journal of Applied Psychology 78(6), 939–948 (1993)

Kolm, S.C.: Multidimensional egalitarianisms. Quarterly Journal of Economics 91(1), 1–13 (1977)

Kottas, A., Müller, P., Quintana, F.: Nonparametric Bayesian modeling for multivariate ordinal data. Journal of Computational and Graphical Statistics 14(3), 610–625 (2005)

Layard, R., Mayraz, G., Nickell, S.: The marginal utility of income. Journal of Public Economics 92(8–9), 1846–1857 (2008)

Maasoumi, E.: The measurement and decomposition of multi-dimensional inequality. Econometrica 54(4), 991–997 (1986)

Nishino, H., Kakamu K.: A random walk stochastic volatility model for income inequality. Japan and World Economy 36, 21–28 (2015)

Pattanaik, P.K., Reddy, S.G., Xu, Y.: On measuring deprivation and living standards of societies in a multi-attribute framework. Oxford Economic Papers 64(1), 43–56 (2012)

Seth, S.: Inequality, interactions, and human development. Journal of Human Development and Capabilities 10(3), 375–396 (2009)

Tsui, K.-y.: Multidimensional generalizations of the relative and absolute inequality indices: The Atkinson-Kolm-Sen approach. Journal of Economic Theory 67(1), 251–265 (1995)

Tsui, K.-y.: Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation. Social Choice and Welfare 16(1), 145–157 (1999)

**Table 1** Questions and choices regarding `sat1`, `sat2`, `health`, and `income`

| |
|---|
| `sat1`: Overall, how satisfied are you with your job? |
| "`1`: Quite unsatisfied," "`2`: Somewhat unsatisfied," |
| "`3`: Somewhat satisfied," "`4`: Quite satisfied." |
| `sat2`: Overall, how satisfied are you with your life? |
| "`1`: Quite unsatisfied," "`2`: Somewhat unsatisfied," |
| "`3`: Somewhat satisfied," "`4`: Quite satisfied." |
| `health`: How would you describe your current health condition. |
| "`1`: Bad," "`2`: Somewhat bad," |
| "`3`: Somewhat good," "`4`: Good" |
| `income`: Please state your annual wage. |
| "`1`: not earning," "`2`: under 1 mil. yen," "`3`: 1-2 mil. yen," "`4`: 2-3 mil. yen," |
| "`5`: 3-4 mil. yen," "`6`: 4-5 mil. yen," "`7`: 5-6 mil. yen," "`8`: 6-7 mil. yen," |
| "`9`: 7-8 mil. yen," "`10`: 8-9 mil. yen," "`11`: 9-10 mil. yen," "`12`: 10-12 mil. yen," |
| "`13`: 12-15 mil. yen," "`14`: 15-20 mil. yen," "`15`: over 20 mil. yen." |

**Table 2a** Summary statistics (explained variables)

| | Satisfaction | | | | | | |
|---|---|---|---|---|---|---|---|
| Satisfaction with | 1 | 2 | 3 | 4 | | | |
| Work | 84 | 278 | 446 | 91 | | | |
| (sat1) | (9.3) | (30.9) | (49.6) | (10.1) | | | |
| Life | 53 | 275 | 497 | 74 | | | |
| (sat2) | (5.9) | (30.6) | (55.3) | (8.2) | | | |
| Health | 29 | 143 | 498 | 229 | | | |
| (health) | (3.2) | (15.9) | (55.4) | (25.5) | | | |
| | Annual wage (ten thousand yen) | | | | | | |
| | Mean | Sd | Min | 25% | 50% | 75% | Max |
| (income) | 438.9 | 313.5 | 66.7 | 150.0 | 350.0 | 650.0 | 3000.0 |

**Table 2b** Summary Statistics (explanatory variables)

| Gender (`female`) | | | | | | |
|---|---|---|---|---|---|---|
| Male | Female | | | | | |
| 547 | 352 | | | | | |
| Number of Employees (`emp`) | | | | | | |
| RE(M) | RE(F) | Non-RE(M) | Non-RE(F) | RE | Non-RE | |
| 447 | 135 | 100 | 217 | 582 | 317 | |
| Marital Status (`married`) | | | | | | |
| Not married | Married | | | | | |
| 252 | 647 | | | | | |
| Age (`age`, year) | | | | | | |
| Mean | Sd | Min | 25% | 50% | 75% | Max |
| 43.4 | 13.0 | 20 | 32 | 43 | 55 | 64 |
| Length of education (`educ`, year) | | | | | | |
| Mean | Sd | Min | 25% | 50% | 75% | Max |
| 14.3 | 2.0 | 9.0 | 12.0 | 14.0 | 16.0 | 18.0 |
| Length of service (`tenure`, year) | | | | | | |
| Mean | Sd | Min | 25% | 50% | 75% | Max |
| 11.0 | 10.6 | 1.0 | 2.0 | 7.0 | 17.0 | 45.0 |

**Table 3** Posterior median of coefficient parameters

| | | sat1 | sat2 | health | log(income) |
|---|---|---|---|---|---|
| intercept | $\beta_{j,1}$ | $0.7341^{**}$ | $0.7685^{**}$ | $1.2679^{**}$ | $3.1163^{**}$ |
| female (emp=0) | $\beta_{j,2}$ | 0.0662 | 0.0557 | 0.0081 | $-0.8013^{**}$ |
| emp (female=0) | $\beta_{j,3}$ | $-0.0545$ | $0.0779^{*}$ | $-0.0242$ | $0.4957^{**}$ |
| emp*female | $\beta_{j,4}$ | 0.0399 | 0.0265 | 0.0677 | $0.4424^{**}$ |
| female (emp=1) | $\beta_{j,2}+\beta_{j,4}$ | 0.1061 | 0.0822 | 0.0758 | $-0.3589^{**}$ |
| emp (female=1) | $\beta_{j,3}+\beta_{j,4}$ | $-0.0146$ | 0.1044 | 0.0435 | $0.9381^{**}$ |
| married | $\beta_{j,5}$ | $0.1437^{**}$ | $0.1340^{**}$ | $0.1097^{**}$ | 0.0071 |
| age | $\beta_{j,6}$ | $-0.0266^{**}$ | $-0.0320^{**}$ | $-0.0330^{**}$ | $0.0667^{**}$ |
| age2 | $\beta_{j,7}$ | $0.0003^{**}$ | $0.0004^{**}$ | $0.0004^{**}$ | $-0.0007^{**}$ |
| educ | $\beta_{j,8}$ | 0.0077 | $0.0118^{*}$ | 0.0063 | $0.0672^{**}$ |
| tenure | $\beta_{j,9}$ | 0.0074 | 0.0011 | 0.0052 | $0.0197^{**}$ |
| tenure2 | $\beta_{j,10}$ | $-0.0002$ | 0.0000 | $-0.0002$ | $-0.0001$ |
| | $\gamma_{j,2}$ | $0.4039^{**}$ | $0.4036^{**}$ | $0.3884^{**}$ | |

$j = 1$ to 3 correspond to sat1, sat2, and health, respectively, and $j = 4$ corresponds to log(income). "$**$" and "$*$" denote that zero is not included in the 95% and 90% CIs, respectively.

**Table 4** Median of simple and partial correlation coefficients between explained and explanatory variables

| | | Simple Cor. | Partial Cor. |
|---|---|---|---|
| sat1 | sat2 | 0.5766** | 0.5291** |
| sat1 | health | 0.2803** | 0.1174** |
| sat1 | log(income) | 0.1300** | 0.0665 |
| sat2 | health | 0.3271** | 0.2089** |
| sat2 | log(income) | 0.1260** | 0.0569 |
| health | log(income) | 0.0734* | 0.0261 |

"Simple Cor." and "Partial Cor." denote the posterior median of simple and partial (polychoric and polyserial) correlation coefficients, respectively. "∗∗" and "∗" denote that zero is not included in the 95% and 90% CIs, respectively.

**Table 5** Median of inequality indices

| | Total | Male | Female | RE | Non-RE |
|---|---|---|---|---|---|
| Gini | 0.6523** | 0.6446** | 0.6630** | 0.6358** | 0.6808** |
| GE(0) | 1.2556** | 1.2022** | 1.3333** | 1.1809** | 1.3819** |
| GE(1) | 0.7677** | 0.7487** | 0.7950** | 0.7224** | 0.8534** |
| GE(2) | 1.0515** | 1.0336** | 1.0697** | 0.9660** | 1.2250** |
| | RE(M) | Non-RE(M) | RE(F) | Non-RE(F) | |
| Gini | 0.6362** | 0.6764** | 0.6297** | 0.6798** | |
| GE(0) | 1.1520** | 1.4176** | 1.2691** | 1.3551** | |
| GE(1) | 0.7263** | 0.8430** | 0.7041** | 0.8493** | |
| GE(2) | 0.9912** | 1.1894** | 0.8719** | 1.2087** | |

"Total" denotes the posterior median of inequality indices for all employees. "Male," "Female," "RE," and "Non-RE" denote male, female, regular, and non-regular employees, respectively. "RE(M)," "Non-RE(M)," "RE (F)," "Non-RE(F)" denote male regular, male non-regular, female regular, and female non-regular employees, respectively. "$**$" denotes that zero is not included in the 95% CI.

**Table 6** Probability related to an inequality index

| | Female > Male[a] | Non-regular > Regular[b] |
|---|---|---|
| Gini | 0.9061 | 0.9986 |
| GE(0) | 0.9722 | 0.9974 |
| GE(1) | 0.8683 | 0.9984 |
| GE(2) | 0.6384 | 0.9893 |
| | Non-regular > Regular[c] | |
| | Male | Female |
| Gini | 0.9652 | 0.9889 |
| GE(0) | 0.9932 | 0.7887 |
| GE(1) | 0.9591 | 0.9875 |
| GE(2) | 0.8652 | 0.9816 |
| | Female > Male[d] | |
| | Regular | Non-regular |
| Gini | 0.3693 | 0.5634 |
| GE(0) | 0.9119 | 0.3130 |
| GE(1) | 0.3378 | 0.5360 |
| GE(2) | 0.1697 | 0.5329 |

This table shows the probability related to each inequality index.
*a*: Probability that a female employee's inequality index is higher than a male employee's.
*b*: Probability that a non-regular employee's inequality index is higher than a regular employee's.
*c*: Probability that a non-regular employee's inequality index is higher than a regular employee's in each gender.
*d*: Probability that a female employee's inequality index is higher than a male employee's in each employment status.