# Changes in the Generative Process of Wage Gaps

## An Analysis of Employer-Employee Matched Data in Japan

July 10 (Thu), 2025, RC 28 Poster Session "Cutting-Edge Research in Social Stratification,"
5th ISA Forum of Sociology, Rabat, Morocco

Koji TAKAHASHI (Japan Institute for Labour Policy and Training)

## Introduction

- This poster aims to analyze matched employer–employee wage data and clarify the changes in the generative process of wage gaps in Japan in the 21st century.
- In daily life, two types of generative processes vis-à-vis wage gaps are recognized:
  - **X** (male individuals, college graduates, etc.) tend to earn more because they are paid more within each firm (employer).
  - **X** tend to earn more because they are more likely to work for high-wage firms (employers).
- Although most studies focus on the former process, the latter is also important for understanding the wage gap situation in a country where wage levels differ by firm (employer)—that is, the wage setting is decentralized—such as in Japan.
- Herein, we apply Takahashi's method (Takahashi 2016) to separate these two mechanisms to datasets obtained in different years and identify the changes in the generative process of wage gaps in Japan.

## Framework and Method

- By controlling for the wage level of each firm, we can calculate the wage gaps induced by individual-level factors within each firm (intrafirm wage gap) (**Figure 1**).

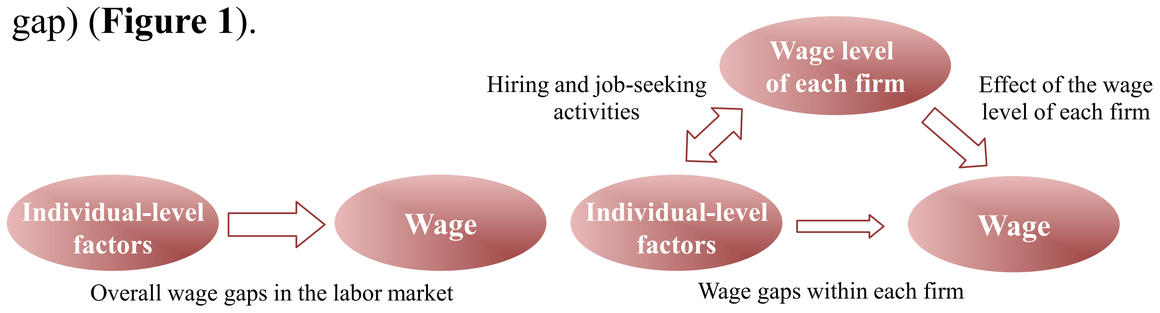



**Figure 1. Framework of the Analysis**

- To achieve this, matched employer–employee data, wherein individual employees' information is nested within their employers', are required. Wage gaps within each firm are estimated using the fixed effects model (Equation (2), $\boldsymbol{\beta_{FE}}$).
- However, the results of the pooled ordinary least squares (OLS) model are considered to correspond to the overall wage gaps in the labor market (Equation (1), $\boldsymbol{\beta_{Pooled}}$).
- Subsequently, wage gaps generated through hiring and job-seeking activities are estimated by subtracting the coefficients of Equation (2) ($\boldsymbol{\beta_{FE}}$) from those of Equation (1) ($\boldsymbol{\beta_{Pooled}}$). The differences between $\boldsymbol{\beta_{Pooled}}$ and $\boldsymbol{\beta_{FE}}$ expand when individual attributes correlate with the wage level of their firms.

Equation (1) $\quad \mathbf{y_{ij}} = \boldsymbol{\alpha} + \mathbf{X_{ij}}\boldsymbol{\beta_{Pooled}} + \boldsymbol{\varepsilon_{ij}}$ [Pooled OLS model]

Equation (2) $\quad \mathbf{y_{ij}} = \boldsymbol{\alpha} + \mathbf{X_{ij}}\boldsymbol{\beta_{FE}} + \boldsymbol{\delta_j} + \boldsymbol{\varepsilon_{ij}}$ [Fixed effects model]

Notation:
- Subscript j is the ID of the firm
- Subscript i is the ID of an individual employee working for firm j
- $\mathbf{y_{ij}}$ is the wage of individual employee i working for firm j
- $\mathbf{X_{ij}}$ denotes the vector of individual-level variables of individual employee i working for firm j
- $\boldsymbol{\alpha}$ is the constant (intercept) for all individual employees
- $\boldsymbol{\beta}$ is the vector of the slope for individual-level variables $\mathbf{X_{ij}}$
- $\boldsymbol{\delta_j}$ is a unique constant (intercept) for individual employees working for firm j
- $\boldsymbol{\varepsilon_{ij}}$ is the error factor for all individual employees

## Data

- Japan's nationwide matched employer–employee data obtained from the "General Survey on Diversified Types of Employment" conducted by the Ministry of Health, Labour and Welfare in 2003 and 2019 were used (**Table 1**).

**Table 1. Data Overview (Unweighted)**

| | | | 2003 | 2019 |
|---|---|---|---|---|
| Overview of the survey | Employers | Distribution | 16,232 | 17,278 |
| | | Response | 11,624 (71.6%) | 7,499 (43.3%) |
| | Employees | Distribution | 35,094 | 36,527 |
| | | Response | 24,930 (71.0%) | 23,521 (64.4%) |
| Target of the analysis | Employers | | 4,849 | 4,244 |
| | Employees | | 14,182 | 13,130 |
| | Employees per employer (mean) | | 2.92 | 3.09 |

Note: Employees who were 60 years or older, those dispatched from other firms, students, those working less than 20 hours per week, and those earning less than 500 or more than 10,000 JPY per hour were excluded from the analysis.

- Dependent variable (**y**): logarithm of the hourly wage (JPY) of individual employees.
- Independent variables (**X**): gender (male dummy), age (age and the square of age), education (college graduate dummy), and employment type (regular employee dummy). Although other variables affect wages in Japan, only basic variables were included in the model for simplification.

## Data (cont'd)

- We weighted these back to all individual employees working for all firms in the population. **Table 2** presents the descriptive statistics.

**Table 2. Descriptive Statistics (Weighted)**

| | 2003 | | | | | 2019 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | S.D. | Min. | Max. | N | Mean | S.D. | Min. | Max. |
| Ln (wage) | 34,644,826 | 7.21 | 0.43 | 6.21 | 8.53 | 36,423,054 | 7.32 | 0.41 | 6.21 | 8.67 |
| Male | 34,644,826 | 0.51 | 0.50 | 0 | 1 | 36,423,054 | 0.55 | 0.50 | 0 | 1 |
| Age | 34,644,826 | 38.41 | 10.38 | 15.00 | 59.00 | 36,423,054 | 42.39 | 10.09 | 17.50 | 57.50 |
| Square of age | 34,644,826 | 1583.37 | 823.94 | 225.00 | 3481.00 | 36,423,054 | 1898.84 | 834.64 | 306.25 | 3306.25 |
| College graduate | 34,644,826 | 0.28 | 0.45 | 0 | 1 | 36,423,054 | 0.41 | 0.49 | 0 | 1 |
| Regular employee | 34,644,826 | 0.73 | 0.44 | 0 | 1 | 36,423,054 | 0.70 | 0.46 | 0 | 1 |

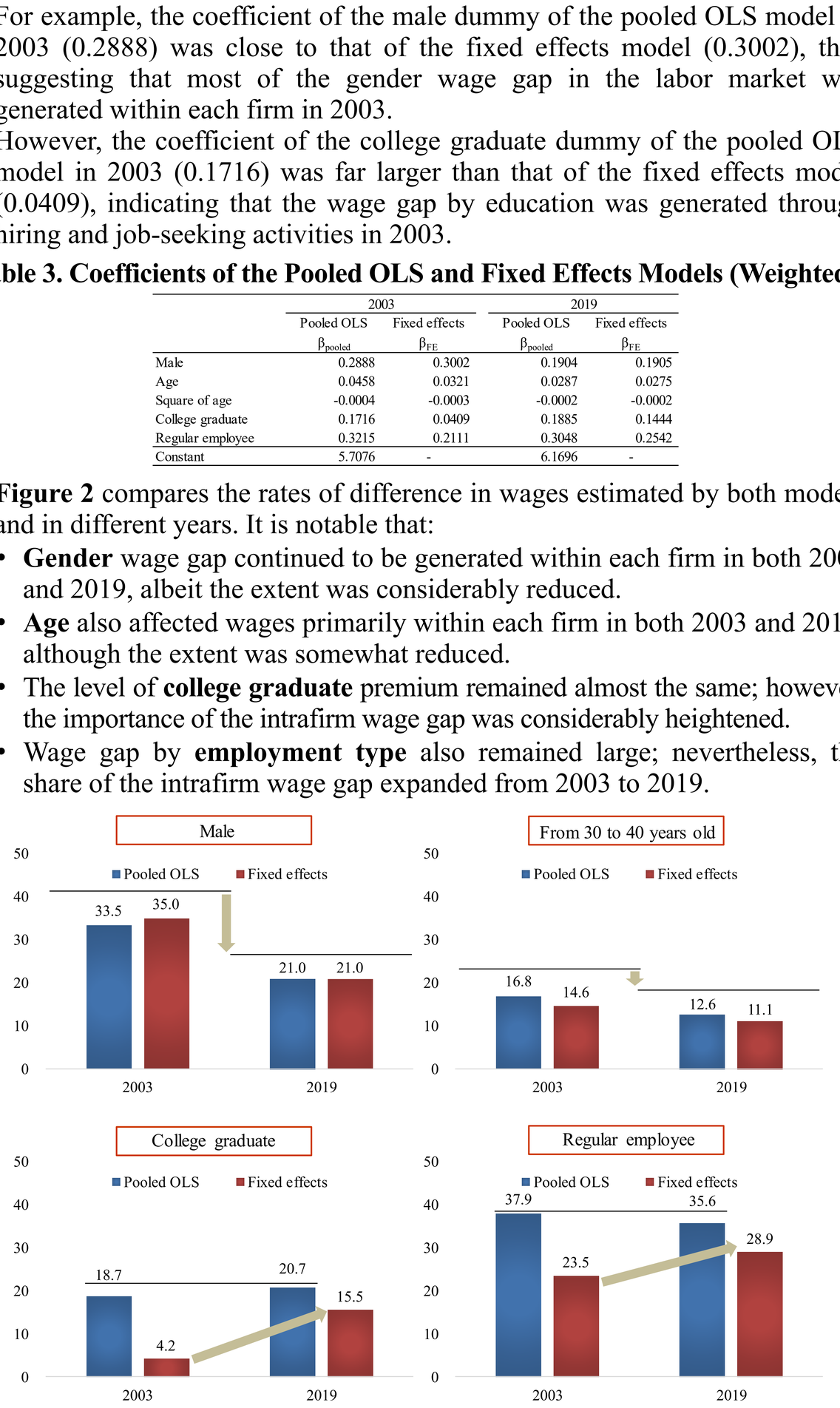
## Results

- **Table 3** presents the results of the pooled OLS and fixed effects models for each dataset.
- For example, the coefficient of the male dummy of the pooled OLS model in 2003 (0.2888) was close to that of the fixed effects model (0.3002), thus suggesting that most of the gender wage gap in the labor market was generated within each firm in 2003.
- However, the coefficient of the college graduate dummy of the pooled OLS model in 2003 (0.1716) was far larger than that of the fixed effects model (0.0409), indicating that the wage gap by education was generated through hiring and job-seeking activities in 2003.

**Table 3. Coefficients of the Pooled OLS and Fixed Effects Models (Weighted)**

| | 2003 | | 2019 | |
|---|---|---|---|---|
| | Pooled OLS | Fixed effects | Pooled OLS | Fixed effects |
| | $\beta_{pooled}$ | $\beta_{FE}$ | $\beta_{pooled}$ | $\beta_{FE}$ |
| Male | 0.2888 | 0.3002 | 0.1904 | 0.1905 |
| Age | 0.0458 | 0.0321 | 0.0287 | 0.0275 |
| Square of age | -0.0004 | -0.0003 | -0.0002 | -0.0002 |
| College graduate | 0.1716 | 0.0409 | 0.1885 | 0.1444 |
| Regular employee | 0.3215 | 0.2111 | 0.3048 | 0.2542 |
| Constant | 5.7076 | - | 6.1696 | - |

- **Figure 2** compares the rates of difference in wages estimated by both models and in different years. It is notable that:
  - **Gender** wage gap continued to be generated within each firm in both 2003 and 2019, albeit the extent was considerably reduced.
  - **Age** also affected wages primarily within each firm in both 2003 and 2019, although the extent was somewhat reduced.
  - The level of **college graduate** premium remained almost the same; however, the importance of the intrafirm wage gap was considerably heightened.
  - Wage gap by **employment type** also remained large; nevertheless, the share of the intrafirm wage gap expanded from 2003 to 2019.




Notes:1) Rates of difference in wages (%) were obtained by calculating 100*(exp (β)-1) based on Table 3.
2) Horizontal lines indicate that the two rates were almost the same.

**Figure 2. Rates of Difference in Wages by Gender, Age, Education, and Employment Type (%)**

- In total, the share of intrafirm wage gaps increased in terms of the levels of college graduate premium and wage disparity by employment type, although wage gaps based on gender and age shrank.

## Conclusion and Further Agenda

- The expansion of intrafirm wage gaps suggests that the wage systems and human resource management practices of each firm are becoming more important in shaping the wage gaps by education and employment type.
- The underlying factors driving the expansion of intrafirm wage gaps, including a narrowing of wage differences by firm size, an increase in the number of college graduates employed by lower-wage firms, and a broader adoption of non-regular employment contracts by high-wage firms, should be investigated.

**Reference:** Takahashi, Koji (2016) "Two Components of Wage Gaps Induced by Individual-level Variables: Intra-firm or Inter-firm?" *International Journal of Japanese Sociology*, Vol. 25, pp. 117-130.