Cross-sectional Facts in Japan using Keio Household Panel Survey*

Tomoaki Yamada†
Meiji University
September 18, 2013

Abstract

We investigate economic inequalities of Japanese economy from 2004 to 2012 using the Keio household panel survey. We present cross-sectional dispersion earnings, consumption expenditure, and wealth inequalities from time-series and life cycle dimensions. Wage and hours inequalities, which are calculated from the earnings of male and female, full-time and part-time workers and correlations are provided. We also show that the residual inequalities, which are usually interpreted as idiosyncratic income risks that households face, rise over the life cycle.

Keywords: Economic inequality, Wage, Hours worked, Consumption, Wealth
JEL Classification: D12, D31, E21

*This research is financially supported by the Ministry of Education, Science, Sports, and Culture, Grant-in-Aid for Scientific Research (B), 22330090. The author thanks the Keio University 21st Century Center of Excellence program for providing me the Keio Household Panel Survey.

†School of Commerce, Meiji University. 1-1 Kanda-Surugadai, Chiyoda-ku, 101-8301, Tokyo, Japan. E-mail: tyamada@meiji.ac.jp. Phone/Fax: +81-3-3296-2274.
1 Introduction

Recently macroeconomists have taken the heterogeneity of households and/or firm into serious consideration. Thanks to development of analytical tools such as tractable models and numerical methods, these lines of research have produced fruitful results.\(^1\) To consider the heterogeneity, we need empirical backgrounds of cross-sectional dispersions from several respects. The *Review of Economic Dynamics* (RED) provides a special issue for “Cross Sectional Facts for Macroeconomists” in which several authors estimate economic inequalities for international comparison.\(^2\) In line with the special issue, Lise et al. (2013) document the main features of dispersions in wages, earnings, consumption, and wealth in Japan from 1981 to 2008.

We compute the time series and life cycle dimension of cross-sectional inequalities using the Keio Household Panel Survey (KHPS). The RED special issue aims to provide several dimensions of economic inequality in many countries with common definitions. Because definitions of earnings and consumption usually differ between countries and data sets, it is important to develop common measures of economic inequalities for comparability. Lise et al. (2013) is also written in the same manner. Due to limitations of the KHPS, we cannot satisfy some respects of the requirement in the RED special issue. This study provides empirical backgrounds of cross-sectional dispersions in Japan for macroeconomic models. Fortunately, we find that economic inequalities in KHPS are comparable with Lise et al. (2013), which uses a large sample size of cross section data collected by the government.\(^3\)

In this study, we provide a cross-sectional dispersion in wages, hours worked, earnings, consumption and wealth because the KHPS includes information about all these variables. We investigate wage and hours worked inequalities for male and female workers, and for full-time and contingent-job workers by education. We also provide wage, hours worked and consumption inequality from life cycle perspectives.

The paper proceeds as follows. In Section 2, we describe details of our data set and define the economic variables to be estimated. In Section 3, we discuss the estimated results from the time-series dimension. In Section 4, we provide economic inequalities from the life-cycle dimension. Section 5 briefly concludes.

---

\(^1\) For the developments, see Guvenen (2011).


2 Data

2.1 Keio Household Panel Survey

We use the *Keio Household Panel Survey* to investigate cross-sectional inequality in Japan. The KHPS is a panel survey started in 2004 and conducted by Keio University’s 21st Century Center of Excellence program.\(^4\) Survey targets in the KHPS are determined based on the Basic Resident Registration (*Jyumiun Kihon Daicho*). The survey contains singles, couples, and extended families. Question items include annual earnings, typical hours worked, household expenditures, financial/real assets, liabilities, mortgages, and individual characteristics such as sex, age, employment status, education background, and family structure.\(^5\) The survey asks the same questions to the spouse of the survey target if she/he is married. Thus, we have panel data for both the husband’s and wife’s labor earnings and hours worked in addition to household level variables such as consumption, and assets. Hourly wages are also available by dividing labor earnings by hours worked.


2.2 Definition of Variables

This study provides several aspects of economic inequality that are comparable with the RED special issue. To do so, we define the economic variables as consistently as possible with the special issue. In particular, we will compare the results with Lise, et al. (2013), who determine definitions of income and consumption from the special issues of *Review of Economic Dynamics*, “cross sectional facts for macroeconomists”.

We focus on working-age households. Thus, we drop households whose head is younger than 24 or older than 60. We drop 60 year old individuals because they may receive a sizeable amount of retirement payments, which may overestimate earnings inequality. We accept the self-reported household head directly and do not select the household head based on age and sex because the survey specifies the main earner as household head. In married households, 95% of household heads were male. We have checked whether the age of the survey target or the spouse coincides with the

\(^4\)The center of excellence program was completed by the academic year 2012 and Kaken took over to conduct the survey.

\(^5\)Basically employment status and other current statuses are based on January of the survey year.
self-reported household head. If not, we dropped them because we could not construct household earnings from the survey target and the spouse’s earnings.

**Earnings, Wage, and Hours Worked:** The KHPS collects two types of labor earnings. First, it asks for last year’s (before tax) annual earnings for both the survey target and the spouse. We define “household earnings” as the sum of husband and wife, although they may have a third earner. For singles, their own earnings are the household earnings. Our sample contains both full-time and part-time workers. In particular, many female workers, regardless of marital status, work part-time jobs.

Second, survey targets also provide their household’s annual income, including husband and wife after 2005: the figure may include not only labor earnings but other income such as rent and asset income. Before 2008, the survey collected details of January’s income of both husband and wife, in addition to annual income. In 2009, the question was changed and the survey now collects details of annual income. Accordingly, we cannot deconstruct income data provided before 2008 into labor earnings and other income. Although for the second type of income, we have information about asset income, rent, and public transfers after 2009, we mainly focus on the sum of labor earnings as defined above because the latter may contain non-labor income and the third earner’s income.

Following Heathcote et al. (2010) and Lise et al. (2013), we investigate the time-series and life cycle dimensions of the following variables:

- \( \{y^m_{L}, y^f_{L}\} \): From labor earnings of husband and wife or singles alone, we construct earnings of male and female workers. We denote the annual earnings of male and female workers as \( \{y^m_{L}, y^f_{L}\} \). They work in full-time or part-time employment. We define “jyokin” as full-time employment. Contracted jobs, part-time jobs, dispatched workers (haken), and temporary workers (shokutaku) are classified as contingent job workers.\(^6\)

- \( y_{L} \): As stated above, household earning is defined as the sum of the survey target and his/her spouse, i.e., \( y^m_{L} + y^f_{L} \). If the head of household is single, either \( y^m_{L} \) or \( y^f_{L} \) is zero. We also examine equivalized labor earnings using the OECD equivalent scale.

- \( y_{T} \): We use the second type of household income for comparison purpose. Note that this definition may include asset income, transfers, and any third earner’s earnings.\(^6\)

\(^6\)In Japan, co-residence with parents is usual, especially in rural areas. Thus, parents or children of the husband and wife may work as the third (or forth) earner.

\(^7\)For details on the contingent job workers, see Esteban-Pretel et al. (2011).
income.

- \{h^m, h^f\}: The survey asks average hours worked per week. The hours worked includes extended hours (zangyo).

- \{w^m, w^f\}: We obtain hourly wage by dividing annual earnings, \{y^m_L, y^f_L\}, by annualized hours worked.

We could not investigate the relationship between before-tax earnings, household income and disposable income because the tax payment information is restricted in the KHPS.\(^8\)

**Consumption** We mainly focus on nondurable expenditure, but also use food expenditure and total expenditure. While labor earnings are last year’s annual earnings, expenditures are based on monthly expenses in January. Because all expenditures are for the household as a unit, we equivalize expenditures using the OECD equivalent scale. The definitions are as follows:\(^9\)

- \(c_{FD}\): Food expenditure (eating-out is excluded.)

- \(c_{ND}\): Nondurable expenditure consists of (1) food at home, (2) food away from home (eating-out), (5) fuels, utilities, and public services, (8) apparel, (9) medical expenses, (11) communication expenses, (12) expenses for internet, (13) education expenditures, (14) entertainment, (15) social expenses (kozukai), (16) private transfers for children (shiokuri), and (17) others. We exclude transportation because it may include buying a car.

- \(c_T\): Total expenditure includes durable expenditures such as (3) rent and maintenance of houses, (4) common-area charges (kyoeki-hi), (6) furniture, (7) electric devices such as PCs and TVs, in addition to nondurable expenditures.

Note that timing of receiving labor earnings (last year) and consumption expenditure (this January) differs.

**Assets** We use financial and real assets. Financial assets are the sum of bank deposits and market values (self-reported) of securities. Real assets are the sum of land and house prices; both are self-reported market values (when selling land and houses). We exclude mortgages. The labor earnings, wages, and assets are deflated by last year’s CPI and consumption expenditures are deflated by the CPI in January.

\(^8\)For the redistribution effects of tax and social security systems, see Sudo et al. (2012).

\(^9\)We do not exclude low food consumption expenditures because consumption is based on monthly expenditures. It is not surprising that the food expenditures are extremely low if a household has food stocks with some home-grown food.
2.3 Sample Selection

Following guidelines to achieve cross-country data comparability, we dropped individuals with extremely low wage. More concretely, we consider wages below 325 yen as extremely low hourly wages because the legal minimum wage is approximately 650 yen.\textsuperscript{10} We also trimmed the top and bottom 0.25\% of labor earnings. We drop zero earnings and consumption expenditure because we need to compute logarithms.

3 Inequality over Time

This section documents the evolution of cross-sectional inequality from 2004 to 2012.

3.1 Household Earnings

As explained in Section 2, we use two types of earnings; (i) the sum of a husband’s and wife’s earnings, (ii) household income, which is available after 2005. The left panel of Figure 1 plots equivalized annual earnings and household incomes. As may be expected, average household income $y_T$, labeled “Household Income”, is higher than the sum of husband’s and wife’s earnings $y_L$, “Husband+Wife”, because asset income and any third earner’s income may be included in household income. The right panel of Figure 1 shows averaged equivalized monthly expenditures. Notice that both types of expenditures decline after 2009 because of the so-called “Lehman shock”. Since consumption is based on January’s expenditures, the impacts of the Lehman shock that occurred in September 2008 appears after 2009.

Figure 2 shows basic inequalities in household earnings. We plot three types of earnings, $y_L$ represents raw household earnings defined as the sum of husband’s and wife’s earnings or the sole earnings of singles. $eqy_L$ is equivalized labor earnings, and $eqy_T$ is equivalized household income. Earnings inequalities measured by the variance of logarithm and Gini coefficient appear to decline over the sample period. It is, however, difficult to conclude that economic inequality has shrank in this period because of short length of the data set and the attrition problem.

We should mention two findings. First, the level of inequalities as measured by the variance of logarithm is much higher than that estimated in Lise et al. (2013). For example, the variance of logarithm in earnings ranges between 0.5 and 0.6 in the KHPS. On the other hand, the variance of log earnings increases from 0.15 in 1980 to 0.26 in 2008. This reflects the fact that our data set includes single earners and contingent

\textsuperscript{10}The minimum wages by law differs among prefectures. The lowest minimum wage was 653 yen in 2012, which is applied in Iwate, Okinawa, and other prefectures.
job workers. Because Lise et al. (2013) mainly focus on the earnings from households with two-or-more members, inequality is estimated to be lower than that given in the KHPS. Second, the inequality in household income $y_T$ is much lower than in the sum of husband’s and wife’s earnings $y_L$. From P50-P10 ration, it is clear that the high inequality in $y_L$ can be attributed to the fact that there are many poor workers, but not many super rich.

3.2 Consumption Expenditure

In Figure 3, we show basic inequalities in three types of consumption expenditures; food, nondurable, and total expenditures. These inequalities in expenditures are comparable with Lise et al. (2013) and Sudo et al. (2012) because they also use monthly expenditure in the Family Income and Expenditure Survey. The level shown here is slightly higher than their estimates. This finding is not surprising because our data set includes single households. Consumption inequalities also declined from 2004 to 2012.

3.3 Wages and Hours Worked by the Household Head

We deconstruct labor earnings of household heads into wages, hours worked, and their correlation. Figure 4 plots the variance of logarithm in the wages of the household head, the hours worked, equivalized nondurable expenditure, and the correlation between them. We compare the plot of these figures with those found in Figure 4 of Heathcote et al. (2012). Both the variance of log for wages and hours worked in Japan are strictly higher than those in the U.S. perhaps due to the difference in sample selection: Heathcote et al. (2012) used only full-time male workers. The correlation between wages and hours worked is much lower in Japan than in the U.S. We will come back to this point later. In contrast with the high wage and hours inequalities, the correlation between consumption and wages/hours is much lower than those in the U.S.

3.4 Assets

As it has been well documented that in the U.S., wealth inequalities are much higher than earnings and consumption inequalities (Heathcote et al., 2010). The same pattern holds in Japan. As shown in Figure 5, the Gini coefficients are over 0.6 measured in both financial and real assets.
3.5 Wage, Hours Worked and Female Labor

We deconstruct household earnings into male and female earnings, and further dissect those into wages and hours worked. Figure 6 is comparable with Figure 3 in Lise et al. (2013). Average hourly wages for male and female workers in the KHPS are slightly higher than those in Lise et al. (2013). But the differences are very small, only 100-200 yen. The level of hours worked for males slightly declined over the period. Hours worked by female workers are much lower than those by male workers because many females work in contingent employment. From panel (a) of Figure 7, the gender wage/earnings premium in Japan can be seen to be much higher than in the U.S., because female individuals work in contingent employment, which decreases average earnings. In panel (b), we compute the gender premium for full-time workers. These numbers are comparable with Lise et al. (2013) and Heathcote et al. (2010): earnings (wages) of male workers are approximately 80% (50%) higher than those of female workers.

Figure 8 plots the variance of logarithm in wages, hours worked, and earnings by gender and employment status. While earnings inequality for female full-time workers is strictly larger than that for male full-time workers, the difference in wage inequality between males and females appears to be small for full-time workers. It is generated from the inequality found in differences in hours worked: the dispersion of hours worked for female full-time workers is much higher than that of male workers. This may be evidence of women’s diversity of work styles. For part-time employment, where the wage inequality of male workers is higher than that of female workers, the hours inequality for male workers is lower. As a result, there is no difference in earnings inequality for male and female workers engaging in contingent employment. Correlations between wage and hours worked are computed in Figure 9. In all cases, the correlations are negative.

3.6 Education

Lastly, we compute the earnings, wages and hours worked inequality by education. Figure 10 plots the college premium for earnings and wage by gender. We define the college premium as the ratio of “high school” graduates to “college” graduates; “high school” graduates include individuals who graduated from junior high and high school. “College” graduates include two-year college, Master and Doctoral degree holders. Workers with college degrees earn approximately 30% more than workers with only a high school degree.\footnote{Lise et al. (2013) discuss the relationship between the college premium and gender premium using the Basic Survey on Wage Structure.}

From Figure 11, we can see that earnings, wages, and hours worked inequality do
not significantly differ between low- and high-educated workers. We then drop the contingent job workers from this Figure, and Figures 12 plots the resulting correlations between wages and hours worked. Again, the correlations are negative and there is no large difference by the educational level attained.

4 Inequality over Life Cycle

In this section, we investigate the life cycle dimension of economic inequality. Figure 13 show variances of logarithms in wages of household heads, hours worked, and nondurable expenditures. This Figure is comparable with Figure 1 in Heathcote et al. (2012). From the upper left panel, where wage inequality rises by age before 50, the hours inequality remains almost constant over the life cycle. The wage inequality over the life cycle in Japan is much lower than that in the U.S., implying that wage heterogeneity within the same cohort in Japan is lower than in the U.S. The correlation between wages and hours worked also increases by age, a trend similar to that is seen in the U.S.

Figure 14 divides earnings inequality into wages, and hours worked by full-time and part-time workers for both male and female workers. From the last row of Figure 14, it is apparent that female workers’ earnings are more unequal compared with male workers at almost every phase of the life cycle. For full-time workers, the patterns of earning and wage inequality are in line with previous research: inequality rises by age. However, it is not clear whether the inequality of contingent-job workers increases over the life cycle. For male workers, the pattern of wage and hours inequality are unclear partially because of small sample size. The inequalities of female contingent job workers rather seem to decline over the life cycle.

As may be expected, financial and real wealth holdings increase with age (Figure 15). From panel (a), it is clear that average real wealth sharply increased between age 25 to 40 because households tend to purchase houses in this age bracket. Because the difference between the housing have and the have-nots is rather substantial, the Gini coefficients of real assets are extremely high for workers in their 20s and early 30s, but declines sharply before the age of 40.

4.1 Residual Inequality

In the consumption insurance literature, variances of residuals of earnings are interpreted as idiosyncratic income risks. Following Aguiar and Hurst (2013), we compute the

\[ \text{Residual Inequality} = \text{Variance of Residuals} \]

\footnote{See Cunha et al. (2005), Guvenen (2009) and Krueger et al. (2010).}
residuals as follows:

$$\ln Y_{it} = \beta_0 + \beta_{age} D_{it}^{age} + \beta_{\text{cohort}} D_{it}^{\text{cohort}} + \beta_t D_{it}^{\text{time}} + \beta_{edu} D_{it}^{\text{edu}} + \beta_{part} D_{it}^{\text{part}} + \beta_{\text{Xit}} X_{it} + \varepsilon_{it}$$

where $Y_{it}$ is the earnings of household $i$ during year $t$, $D_{it}^{\text{age}}$ is a vector of age dummies (for ages 25-59), $D_{it}^{\text{cohort}}$ is a vector including eleven five-year age of birth cohort dummies, and $D_{it}^{\text{edu}}$ education dummy, $D_{it}^{\text{part}}$ education dummy, and $X_{it}$ is a vector of family structure dummies that include a gender dummy, a marital status dummy, the number of adults (above 19), and the number of children (below 18). We also impose $\sum_{t=2004}^{2012} \beta_t = 0$ and $\sum_{t=2004}^{2012} t\beta_t = 0$ to control for time effects and avoid colinearity.

Figure 16 plots the residual inequality over the entire life cycle. As shown in the Figures, both labor earnings and consumption inequalities rise with age. These rises in inequality are usually interpreted as evidence of the existence of permanent shocks. For labor earnings, residual inequality is high for people in their 20s, declines moderately once they reach their 30s, and rises again after their late 40s, which is consistent with Ohtake and Saito (1998) and Abe and Yamada (2009). On the contrary, consumption expenditure inequality appears flat before the age of 45 although there are many spikes, and it rises sharply after the late 40s. Note that the level of residuals in labor earnings is lower than in consumption inequality. It is evidence that labor earnings risks are partially shared by savings or other channels of insurance.

4.2 Differences

Lastly, we compute variances of the first difference in wage, hours worked and their correlation. In the consumption insurance literature, researchers focus not only on the level of cross-sectional dispersion but also on dispersions in the growth rate of labor earnings and consumption expenditures. We have investigated inequalities in the Japanese economy based on the first view point thus far. To compute the first (or higher order) differences, we need to use panel data. Because the KHPS is panel data, we can compute the first difference.

Heathcote et al. (2012) compute variances of the first (and second in their paper) difference in earnings and consumption. Figure 17 is based on Figure 2 and Figure 5 in Heathcote et al. (2012). Variances of first differences in wage and hours worked in Japan are higher than those in the U.S., as is the same with Figure 4 in this paper. Since Heathcote et al. (2012) estimate structural parameters of a partial insurance model to match these moments, preference parameters and idiosyncratic shocks may be

\[13\] See, for example, Blundell et al. (2008) and Heathcote et al. (2012).
very different between Japan and the U.S.

5 Conclusion

In this study, we showed cross-sectional inequalities from both time-series and life-cycle dimensions. We believe that these estimates are useful as calibration targets for heterogeneity in macroeconomic models. We could go further to estimate the extent of permanent and transitory shocks (Storesletten et al., 2004), the extent of partial insurance (Blundell et al., 2008), or distinguish between heterogeneity and uncertainty (Cunha et al., 2005). These are remaining area for future research. In addition, our estimate reveals that the Keio Household Panel Survey is reliable for economic inequality research because the results are comparable with other empirical researches. All estimated results are available from the web: https://sites.google.com/site/tyamadaeconomics/home/research

References


A Figures

Figure 1: Average labor earnings and consumption expenditures. “Husband+Wife (blue solid line)” in the right panel plots equivalized average earnings which is defined as the sum of annual earnings of husband and wife. “Household Income (red dashed line)” is annual household income, which includes labor earnings, asset income, and transfers of husband and wife.
Figure 2: Basic inequality in household earnings. \( y_L \) is the sum of annual earnings of husband and wife and \( \text{eqy}_L \) is the equivalized household earnings. \( \text{eqy}_T \) is household income.
Figure 3: Basic inequality in equivalized expenditures. \( c_{FD} \) is equivalized food expenditure, \( c_{ND} \) is equivalized nondurable expenditure, and \( c_{T} \) is total expenditure, respectively.
Figure 4: Inequality in wages, hours worked, consumption, and their correlations over time. Both wage and hours worked are measured by those of the household head. As for consumption expenditure, we used nondurable expenditures.

Figure 5: Equivalized financial and real assets.
Figure 6: Average wages and hours worked for male and female workers.
Figure 7: Gender premium with/without part-time workers. LHS: the ratio of average earnings (wages) between male and female workers including full-time and contingent employment. RHS: the ratio of average earnings (wages) between male and female full-time workers.

Figure 8: Variance of logarithm in wages and hours worked for men and women over time. The blue solid line represents male workers. The red dashed line represents female workers.
(a) Correlation between Wages and Hours  
(b) Correlation between Wages and Hours (Full-time)

Figure 9: Correlation between wages and hours worked for men and women.

(a) Earnings  
(b) Wages

Figure 10: College premium. When computing the college premium, we dropped contingent-job workers.
Figure 11: Variance of logarithm in wages and hours worked by education level.
Figure 12: Correlation between wages and hours worked by education level.
Figure 13: Inequalities in wages, hours worked, consumption, and their correlation over the life cycle. Both wages and hours worked are measured by those of the household head. As for consumption expenditure, we used nondurable expenditures.
Figure 14: Variance of logarithm in wages and hours worked for men and women over the life cycle. The blue solid line represents male workers and the red dashed line represents female workers.

Figure 15: Equivalized financial and real assets.
Figure 16: Residual inequality over the life cycle.

Figure 17: Variances of the first difference in wages and hours worked.