THE LABOR SUPPLY EFFECT OF SOCIAL SECURITY EARNINGS TEST REVISITED: NEW EVIDENCE FROM ITS ELIMINATION AND REVIVAL IN JAPAN*

SATOSHI SHIMIZUTANI[+] and TAKASHI OSHIO[++]

This version: November 2008

* This project was undertaken under a research project on the interaction between pension program, retirement, and well-being of the elderly, supported by Health and Labor Sciences Research Grants (FY2007–FY2008). We are grateful to the Ministry of Health, Labor and Welfare for providing us a micro level dataset from the *Survey on Employment of the Elderly*. We also thank Naoki Mitani and Izumi Yokoyama for their useful comments. The views expressed in this paper are personal.

[+] Corresponding author; Senior Research Fellow, Institute for International Policy Studies; Toranomon 30 Mori Building 6th floor, Toranomon 3-2-2, Minato-ku, Tokyo Japan 105-0001; Tel: +81-3-5404-6651, FAX: +81-3-5405-6650; E-mail:sshimizutani@iips.org.

[++]Professor, Graduate School of Economics, Kobe University; E-mail:oshio@econ.kobe-u.ac.jp.

Abstract

Evidence on the effect of the social security earnings test on the labor supply of the elderly continues to be mixed. We utilize micro-level data compiled by the Japanese government in order to examine the labor supply effect for those aged 65-69 before and after two major reforms of the social security earnings test in Japan: its elimination in 1985 and its revival in 2002. We provide little evidence that the changes in the earnings test affected the wage distribution of the elderly after controlling for changes in the attributes of workers and firms. At the same time, the direct survey responses to the effect of the revival in 2002 reveals a large effect on the labor supply of the elderly. These empirical findings indicate the risk that a traditional bunch analysis underestimates the labor supply effect when it is obscured by measurement errors or labor market rigidities.

Keywords: social security earnings test, labor supply of the elderly, Japan, wage distribution, DiNardo-Fortin-Lemieux decomposition.

JEL Classification Codes: H55, J26.

1. Introduction

Downsizing the labor force is one of the major challenges for a country experiencing rapid aging and historically low birth rate, which will lead to a tremendous increase in social security benefits for retired persons and burden for the younger generation. As is often debated in the policy arena, a natural solution for mitigating the negative impacts of a decline in the labor force is to encourage the elderly to remain in the labor force for a longer period of time.

Japan shares the common concern of the effect of the rapid speed of aging on social security programs with other developed countries. Undoubtedly, the latest retirement age in the world is found in Japan, which is often considered as one of the most distinctive features of its labor market. Indeed, the effective retirement age for Japanese workers in the 2002–2007 period was 69.5 years for males and 66.5 years for females (OECD (2008)), both of which are higher than that in all European countries and the United States.[1] However, the current speed of aging in Japan is far more rapid as compared with that in other developed countries, which calls for a drastic reform in the labor supply of the elderly. The proportion of the elderly aged 65 and over exceeded 20 percent in 2005 and the speed of aging is expected to accelerate further. The share of the elderly is projected to exceed 30 percent in 2025 and reach 40 percent in 2050 (National Institute of Population and Social Security Research

[1] The effective retirement age is defined as a weighted average of net withdrawals from the labor market at different ages over a five-year period for workers initially aged 40 and over (OECD 2008).

2007). The extremely rapid pace of aging in Japan is likely to offset the positive effect of late retirement on overall labor force participation.

One popular view prevailing in the policy debates on the labor force participation of the elderly in Japan is that the social security earnings test (*Zaishoku Rorei Nenkin* scheme) is an important disincentive to paid work for the elderly through a high effective tax rate on work. This scheme is a part of the Employee's Pension Insurance (EPI; *Kosei Nenkin*) program, which is the core of the public pension scheme and covers approximately half the pensioners in Japan. The earnings test reduces immediate payments to social security benefits for EPI pensioners whose labor income exceeds a certain exemption threshold. Although benefits are subsequently increased in order to compensate for any such reduction, it is commonly viewed that the earnings test "punishes" the labor supply of the elderly, as in the United States (Gruber and Orszag 2003).

This study provides new evidence on the labor supply effect of the social security earnings test for workers aged between 65 and 69, focusing on two major episodes since the 1980s—the elimination of the earnings test in 1985 and its revival in 2002. We focus on workers aged 65–69 for two reasons. First, the target age group for enhancing labor force participation in Japan is now shifting to those aged 65–69, whose labor force participation is lower than that for those aged 60–64.[2] The mandatory retirement age, which is effectively

[2] The labor force participation rate for males in 2007 was 93.1 percent for workers aged 55–59, 74.4 percent for those aged 60–64, and 48.5 percent for those aged 65–69; the labor force participation rate for females in the same year was 60.8 percent for those aged 55–59, 42.2 percent for those aged between

equivalent to the eligibility age for public pension benefits, is now in transition from 60 to 65, and the labor force participation is expected to increase for workers aged 60–64[3].

Second and more importantly, the rule of the earnings test is simple for workers aged 65–69 with a single threshold and rate of benefit reduction. Moreover, the history of the revisions is clear: it was abolished in 1985 and revived in 2002. In contrast, for workers aged 60–64 years, the earnings test has multiple thresholds and the reduction rates differ across earnings brackets. In addition, their revisions have been gradual, thereby making it difficult to precisely examine the labor supply effect of the revisions of the earnings test.

We view these reforms as clean natural experiments for examining the effect of the revisions of earning test rules with regard to the labor supply of the elderly. This study is empowered by a large micro-level dataset from the *Survey on Employment of the Elderly* (Konenreisha Shugyo Jittai Chosa, henceforth SEE) compiled by the Japanese government with information on both employment status and social security eligibility. This study provides new evidence on the labor supply effect of the social security earnings tests in Japan, thereby enabling us to contribute to the vast literature in certain new aspects. First and most importantly, we adopt two different approaches for examining the labor supply effect of the revision of the earnings test rule. One is to examine the changes in wage distributions before and after the reforms, an idea similar to a bunch analysis, which has been frequently used in

60–64, and 25.8 percent for those aged 65–69.

[3] Shimizutani and Yokoyama (2008) showed that the average tenure for Japanese workers has increased since the 1990s due to the extension of mandatory retirement from 55 to 60 years.

the literature (e.g., Friedberg 2000, Haider and Loughran 2008). The other is to utilize direct survey responses to the revival of the earnings test in 2002, which is unique in the literature. In the 2004 survey, the SEE explicitly asked the respondents how the earnings test discouraged them from working. This feature enables us to examine the labor supply effect of the rule revisions in terms of both the direct survey responses and change in wage distributions, which reinforce our empirical results. While some may argue that survey responses are subjectively biased, the advantage is that we are able to directly measure the discouraging effect for those who quit working altogether and the degree of circumventing a higher effective tax on labor income for incumbent workers under the earnings test rule. This approach complements the traditional bunch analysis and prevents us from underestimating the labor supply effect of the earnings test when measurement errors and labor market rigidities are serious, an aspect emphasized in a recent work by Haider and Loughran (2008).

Second, we employ an alternative econometric technique for examining the change in the wage distributions before and after the reforms. Contrary to preceding empirical studies, the interval of the timing of our dataset is every four or five years, and we are not able to employ a straightforward application of the bunch analysis. Instead, we examine the change in whole wage distributions using the methodology of DiNardo Fortin and Lemieux (1996), which permits us to decompose the change in the wage (wage plus second-tier benefits for the 2002 revival) distributions into two parts: the change in the distributions of the attributes

of workers and firms and the change in the effect of the attributes on wage (wage and second-tier benefits for the 2002 revival). We compare the actual distributions before the reforms with the counter-factual distribution after the reforms, controlling for the changes in the attributes of workers and firms.[4]

Third, we endeavor to extract lessons from the Japanese experience, which are useful for other countries in two aspects. One is to examine two major and clean episodes of the elimination and revival of the earnings test, whose effect on work incentives for the elderly may be asymmetric. To our knowledge, there have been few studies that deal with both repeal and recurrence episodes outside Japan. The other is to relate the sensitivity of work incentive to pension benefits with late retirement in Japan. If we find a large labor supply effect of the rule revisions on work incentives, the highest elderly labor force participation in the world may become even higher on account of lowering disincentives. If this is not the case, the decision to work among the elderly in Japan is less likely to be sensitive to benefits and other factors are more responsible for the late retirement age.

Here, we present a preview of our empirical results. By examining the changes in the wage distributions for those aged 65–69 before and after the elimination of the earnings test in 1985 and its revival in 2002, we provide little evidence that the revisions in the earnings test affected the wage distribution of the elderly. In other words, the elimination and revival

[4] Lemiuex (2002) applied this methodology for examining the effect of change in minimum wage on wage distribution.

of the earnings test did not alter the labor supply decisions of the elderly. However, this is not the end of the story. The survey response to the revival in 2002 reveals a large effect on the labor supply of the elderly. The earnings test discouraged a larger share of people aged 65–69 from working, and the labor income earned by incumbent workers was not close to the threshold. Our empirical findings reveal that a traditional bunch analysis has a pitfall of underestimating the labor supply effect when it is obscured by measurement errors or labor market rigidities.

This paper proceeds as follows. The next section overviews previous research on the labor supply effect of the social security earnings test. Section 3 briefly describes the revisions in the earning rule in Japan. Section 4 explains the dataset used in this study. Section 5 presents the results of the bunch analysis using histograms and decomposition analysis of the wage distribution before and after the elimination and revival of the earnings test. Section 6 uses the survey responses to the revisions of the earnings test and compares the results with those in the previous section. The last section summarizes our findings.

2. Previous studies

Despite the popular view that emphasizes the disincentive effect of social security earnings tests on elderly workers, it is fair to state that a large volume of empirical

investigation has not reached a consensus on this issue. While a majority of the vast existing literature found a small effect of the earnings test on the labor supply of the elderly[5], the conclusion are rather mixed.

The effect of the social security earnings test is often analyzed in a standard textbook labor supply framework with a kinked budget constraint, which corresponds to the threshold (Friedberg 2000). Contrary to the prevailing view, the theory suggests that the elimination of the social security earnings test is unlikely to substantially increase labor supply among retirees since a change in the earnings test shifts the budget constraint, which in turn invites both income and substitution effects and makes the net effect ambiguous (Borjas 2005). In other words, *a priori* it is not possible to predict the sign and magnitude of an earnings test rule on labor supply, and the net effect depends on whether the substitution or income effect dominates, which is an issue that must be empirically examined.

As Gruber and Orszag (2003) summarized, there have been two branches of studies that examine labor supply effects. One is to employ a "bunch analysis" for examining earnings concentration at the threshold of the earnings tests. The other type employs sophisticated econometric analyses of the aggregate impacts of the earnings test on the conditional hours worked by elderly workers of the kinked budget constraint, which requires a variety of structural assumptions.

[5] For example, Burtless and Moffitt (1985) and Gustman and Steinmeier (1985) performed simulations and suggested that the elimination of the earning tests would only have minor effects on labor supply. The revision of the earnings test rule in the United States is summarized in Table 1 of Gruber and Orszag (2003).

Since the number of works in the literature is tremendous, we will focus on several important works published recently after 2000. Along with a majority of the existing large literature that found a small labor supply effect of the earnings test, Gruber and Orszag (2003) utilized the changes in the earnings test over the past three decades in order to identify the effect of exogenous changes in the test rule on labor supply and benefits receipt. By performing both graphical analysis of breaks in labor supply trends and reduced-form regression estimates based on a difference-in-difference approach, they found no robust influence on the labor supply decisions of men and some more suggestive evidence affecting the labor supply decisions of women.

In contrast, two recent studies found a sizable labor supply effect of the earnings test. Friedberg (2000) performed both bunch analysis and structural model estimation of the aggregate impacts of the earnings test on elderly workers of the kinked budget constraint and uncovered a substantial response by workers to three past changes in the earnings test rules. She found that the earnings distribution is clearly concentrated just below the threshold and confirmed that the bunching has shifted in response to the revision of the earnings test. Moreover, the structural estimation yielded sizable impacts of the elimination of the earnings test on number of hours worked by workers aged 65 and over. Haider and Loughran (2008) also found a consistent and substantial response to the earnings test, particularly for younger men, and insisted that the response to the earnings test in survey data is obfuscated by

measurement errors and labor market rigidities.

In addition, there are numerous other studies conducted outside the United States. For example, Disney and Tanner (2002) examined the effect of the elimination of the earnings test in the United Kingdom and Baker and Benjamin (1999) in Canada; both studies found that the elimination of the earnings test increased the employment and earnings of affected male workers. In Japan as well, there were numerous studies on the labor supply effect of the social security earnings test up until around 2000; however, there has been little research on the topic since then. While a majority of the studies found a significant labor supply effect of the earnings test, Ogawa (1998) and Iwamoto (2000) built econometric models in order to simulate the effect of the earnings test and found a sizable effect on labor supply. In contrast, Abe (1998) employed a difference-in-difference estimation in order to estimate the labor supply effect for workers aged 60–64 in the 1989 reform and found little effect of a change in the earnings test rule on the labor supply of these workers. It is fair to conclude that evidence on the labor supply effect of the social security earnings test is still mixed and has not reached a consensus.

3. The social security earnings test in Japan

This section overviews the revisions of the social security earnings test rule, focusing

on that for workers aged 65–69 since the 1980s. The earnings test, which is known as the *Zaishoku Rorei Nenkin* program, was introduced in Japan for the first time in 1965 and has been revised every 4–6 years. Among the three types of public pension programs in Japan, the earnings test rule is applied to the beneficiaries of the Employees' Pension Insurance (EPI; *Kosei Nenkin*) program, which is applicable to employees in the private sector and includes 48 percent of all pensioners in Japan.[6] The pensioners under the other two types of public pension are exempted from earnings tests—the National Pension Insurance (NPI; *Kokumin Nenkin*) for self-employed persons (45.5 percent) and Mutual Aid Insurance (*Kyosai Nenkin*) for employees in the public sector and private schools (6.5 percent).

In this paper, we confine our interest to EPI pensioners aged 65–69. The revisions of the earnings test rule for workers in this age group are rather evident. The earnings test for workers aged 65–69 was eliminated in 1985 and revived in 2002. Under the earnings test between 1980 and 1984, 1 yen of the social security benefit was withheld for every 5 yen earned above 156,000 yen for workers aged 65–69, thereby indicating a marginal tax rate of 20 percent for those who earned labor income amounting to over 156,000 yen.[7]

In 1985, the earnings test rule for workers aged 65–69 was eliminated; however, reduced rates continued to be applied to workers aged 60–64. While the earnings test rule for workers aged 60–64 was revised again in 1989, 1992, and 1996, there was no earnings test for

[6] The figures for the share of EPI, NPI, and MAI pensioners have been obtained from Komamura (2007).
[7] During the same period, the effective rate was higher for workers aged 60–64, whose rate was 20 percent below 95,000 yen, 50 percent for 95,000–130,000 yen, 80 percent for 130,000–155,000 yen and 100 percent for 155,000 yen and over.

workers aged 65–69 between 1985 and 2002. In 2002, the earnings test for workers aged 65–69 was revived. According to the Ministry of Labor, Health and Welfare (MLHW 2002), the revival aimed to improve the fiscal balances of the social security programs and the labor supply effect was largely ignored. After 2002, 1 yen of social security benefits was withheld for every 2 yen if the sum of earned income and second-tier social security benefits is above 370,000 yen. Hence, workers who earned over 370,000 yen from labor income and pension benefits faced a marginal tax rate of 50 percent. It must be noted that the sum of labor income and second-tier benefits, not just labor income, has been tested since 2002.[8] The first-tier pension benefits are not earnings tested for workers aged 65–69, although it is tested for workers aged 60–64. The rule was slightly revised in 2004 to include bonuses as labor income and change the reduction rate correspondingly; however, there has been virtually no change in the rule since 2002.

4. Data description

This study uses micro-level data from the SEE compiled by the MLHW. The Survey has been performed in 1983, 1988, 1992, 1996, 2000, and 2004. We utilize microlevel data from the SEE in 1983, 1988, 2000 and 2004 to examine the effect of the 1985 abolishment

[8] In 2002, the effective rule was revised for workers aged 60–64 too. In this reform, the threshold for 60 percent rate increased from 340,000 yen to 370,000 yen. In 2007, the earnings test began to be applied to workers aged 70 and over.

and the 2002 revival of the earnings test for those aged 65 and over. The individuals in the sample were aged between 55 and 69 randomly chosen from all regions in Japan. The workers aged 65-69 occupied one third of the total number of the individuals in the samples; 7,186 in 1983, 6,702 in 1988, 6,060 in 2000 and 5,260 in 2004.[9]

The SEE is only large-scale dataset compiled by the Japanese government providing detailed information on both employment status (including labor income) and social security eligibility and benefits.[10] The SEE contains a variety of variables that are indispensable for examining the labor supply effect of the earnings test including demographics and employment status of the individual (age, sex health status, type of job, monthly wage, working days per week, and hours worked per day, etc.) and characteristics of firms (firm size, prefecture, etc). Another distinguished feature of the SEE is that it asks the respondents whether the social security earnings test discourages them from working in the 2004 survey. We use this unique information obtained through the survey response to complement what we observed in the change in wage distributions.

First, we must identify those eligible for the EPI benefit because the earnings test is applicable only to them. The 1983, 1988, and 2000 surveys asked whether a respondent

[9] The total sample size was 26,954 in the 1983 survey, 26,290 in the 1988 survey, 19,595 in the 2000 survey, and 17,853 in the 2004 survey.

[10] There are other large-scale datasets on employment collected by the government, represented by *Labor Force Survey* (Rodo Ryoku Chosa) or *Basic Survey on Employment Structure* (Shugyo Kozo Kihon Chosa) with a large number of observations randomly selected from all regions in Japan. Even if we confine the sample to the elderly, their sample sizes are quite large; however, there is no information on pension eligibility, which is indispensable to this study.

actually receives the EPI benefit presently, while the 2004 survey asked whether a respondent actually receives benefits regardless of type of pension program except private pension programs. In addition, the 1983 and 2004 surveys asked the respondent whether he/she is eligible for the EPI benefit, which permits us to identify EPI pensioners directly. However, it is difficult to identify precisely who is an EPI pensioner in 1988 and 2000 based on the information in the dataset, so we assume that all the individuals who are actually receiving benefits from EPI are eligible for EPI benefits. While it is possible that this definition excludes those who are eligible for EPI benefits but do not currently receive any EPI benefit, we believe that this definition is reasonable because the number of such individuals is limited.[11] When we compare the wage distributions between 1983 and 1988, we employ the same definition of EPI eligibility as that in 1983. In the case of the 2004 survey, we use the eligibility information in the dataset since the 2004 survey does not permit us to identify EPI or other benefits.

Further, we exclude the individuals whose monthly wages or non-wages in the month prior to the surveys exceeded a million yen. Moreover, we confine our sample to individuals who responded that they were healthy because those with an adverse subjective health status are less likely to work at full ability and be subjected to the earnings test. We do not employ a difference-in-difference approach by using workers aged 60–64 as a control group because all

[11] In the 1983 survey, we compare the respondents who were eligible and those who actually received pension benefits and found that the correlation coefficient is 0.93.

the changes in the earnings test rule for workers aged 65–69 have been accompanied with those for workers aged 60–64. While it is a good idea to use individuals aged 70 and over as a control group, the individuals in the sample of our dataset are aged between 55 and 69 only.

The sample size after these adjustments is 651 in 1983, 697 in 1988, 986 in 2000, and 890 in 2004. The summary statistics of the variables used in the decomposition analysis will be provided in the Appendix Table[12].

5. Bunch and decomposition analysis of the wage distribution

In this section, we confine our sample to the individuals who actually worked in the month prior to the survey months and perform two types of analysis. First, we perform a bunch analysis by comparing the histogram of monthly labor income (including second-tier benefits in 2000 and 2004) before and after the elimination (1983 and 1988) and the revival (2000 and 2004) of the earnings test. Second, we perform a decomposition analysis in order to explore what accounted for the change in the income distributions before and after the reforms in the earnings test.

Figure 1 compares the distributions of the monthly wage between 1983 and 1988. In each year, the SEE asked respondents to report their monthly labor income (in an integer

[12] The classifications of type of job vary in each survey and we adjusted to make them comparable between 1983 and 1998 and between 2000 and 2004. Since the data on firm size is not available in the 2004 survey, we exclude it as an independent variable.

number with a unit of 10,000 yen) that they earned in May, merely one month prior to June, when the survey was conducted.[13] The range in the X-axis is from zero to 1,000,000 yen (100 ten thousand yen) and each cell stands for a 10,000 yen bracket. If the social security earnings tests restrain labor supply, we would observe a bunch below the threshold (156,000 yen) in 1983 but not in 1988 after the earnings test was eliminated.

First, we observe that the most frequent cell in 1983 was 100,000 yen, accounting for 10.8 percent. The cumulative frequency exceeds 50 percent (the median) in the cell of 120,000 yen. What is interesting is the concentration in the cell of 150,000 yen just below 156,000 yen, which was the threshold prior to 1985. The frequency for the cell of 150,000 yen in 1983 is 7.8 percent and the cumulative frequency is close to 70 percent at the cell. In 1988, it is also evident that the cell with the greatest frequency is the cell of 100,000 yen (9.5 percent), which exceeds the median, followed by 150,000 yen (7.3 percent). Thus, we notice that the share "just below" the threshold prior to the elimination of the earnings test measured in the cell of 150,000 yen declined only slightly (0.5 percentage points). At the same time, we notice the possibility of reporting errors since large frequency is observed in the cells with a multiple of 50,000 yen in the histogram. If we expand the range of "below the threshold" for the broader cells of between 100,000 yen and 150,000 yen, the share of those measured in the broader cells declined by 5.8 percentage points from 31.3 percent. These observations

[13] The time interval is short and the information is reliable. This is also the case for the surveys in 2000 and 2004, which were performed in October and asked the respondents to report their labor income in September.

indicate that the bunch that was observed below the threshold disappeared in response to the elimination of the earnings test in 1985.

Figure 2 represents the histogram of the monthly wage and the full second-tier benefit in 2000 and 2004. If the earnings test creates a bunch, we would not see it below the threshold (370,000 yen) in 2000 but in 2004 after the earnings test was revived in 2002. The full second-tier benefits for each individual in 2004 are calculated in the pension formula.[14] Since the second-tier benefits are not necessarily in an integer value measured in a 10,000 yen bracket, we take a broader range as "below the threshold" for identifying a bunch. The share of individuals whose monthly wage and full second-tier benefit are between 300,000 yen and 370,000 yen was 12.3 percent in 2000 and 10.1 percent in 2004. In both years, the median is located at approximately 210,000 yen; further, the cumulative frequency below 370,000 yen was 83.6 percent in 2000 and 81.0 percent in 2004. These observations indicate that, contrary to the prediction of a bunch analysis, the share of the individuals below the threshold declined after the revival of the earnings test.

[14] The methodology of calculation is summarized as follows. P_F is denoted as the full second-tier pension benefit, P_A as the actual second-tier pension benefit, and W as wage, all of which are on a monthly basis. In the 2004 survey, the data provides the sum of public pensions (first and second tiers) and employer-provided pension programs (third tier). P_A is computed by subtracting the full basic pension benefits (66,208 yen per month in 2004) from the sum of the first- and second-tier benefits. Under the earnings test, a person eligible for the full benefit is entitled to receive P_F when W is zero and $P_F + W$ when the sum of P_F and W is less than 370,000 yen ($P_A = P_F$). When the sum of P_F and W exceeds 370,000 yen, a marginal tax of 50 percent is applied to the additional benefit, i.e., P_A is calculated as $18{,}500 + 0.5P_F - 0.5W$. What we know in the dataset is W and P_A, and P_F is computed as $W + 2 \times (P_A - 18{,}500)$. Finally, when monthly wage exceeds the sum of the full second-tier pension benefits, 370,000, second-tier pension benefits are reduced to zero. Since we assume that all the persons are eligible for full benefits of the first tier, a limited number of individuals in the sample, who were excluded from the estimation, show a negative value for the sum of wage and second-tier benefits. Moreover, we compute the second-tier benefit assuming no third-tier benefit. See footnote 15.

In brief, the share of the individuals whose wage was below the threshold declined between 1983 and 1988, after the elimination of the earnings test in 1985. This change indicates that the elimination of the earnings test eliminated the bunch, although the share of "just below the threshold" remained unchanged. In contrast, it appears that this is not the case for the revival of the earnings test in 2002 because we do not observe any new bunch below the threshold in 2004.

It may be tempting to conclude that the elimination of the earnings test in 1985 affected the labor supply of the elderly and the revival in 2002 did not. However, it is possible that numerous factors other than the revisions of the earnings test contributed to the change in the distributions between 1983 and 1988 or between 2000 and 2004, which may have obscured the effect of the earnings test reforms. In order to address this issue, we employ a DiNardo Fortin and Lemieux decomposition (DiNardo, Fortin and Lemieux (1996), Dinardo (2002), Lemieux (2002)). This is a semi-parametric approach and visually decomposes the change in wage distributions into two parts: the change in the distributions of the attributes and in the effect of attributes on wage distributions.

First, we compare the actual wage distribution in 1983 (before elimination) with the counter-factual distribution defined as what the density of wage would have been in 1988 (after elimination) if the attributes of workers and firms had remained at their 1983 level. Second, we compare the actual distribution of the sum of wages and full second-tier benefit

in 2000 (before revival) with the counter-factual distribution defined as what the density of wage plus full second-tier benefit would have been in 2004 (after revival) if the attributes of workers and firms had remained at their 2000 level.

Our prediction is summarized as follows. Since it is unlikely that the change in the earnings test rule alters the attributes of workers and firms used in the analysis, it is natural to assume that the effect of the earnings test reforms, if any, is observed not in a change in the attributes of workers and firms but in the effect of those attributes on labor income. In other words, if the counter-factual distribution overlaps with the actual distribution prior to the reforms, the change in the distributions is caused by the change in the attributes, not the effect of the attributes on wage, and thus the change in the earnings test is not responsible for the change in wage distributions. The procedure is summarized in the Appendix and the summary statistics of the variables used as the attributes of workers and firms are reported in Appendix Table.

Figure 3 reports the actual distributions in 1983 and 1988 and the counter-factual distribution, assuming that the attributes of workers and firms remained at their 1983 level. First, in comparison with the actual distributions in 1983 and 1988, we find that the peak around 100,000 yen declined and that the distribution in 1988 became flatter than that in 1983. Second, the counter-factual distribution overlaps the actual distribution in 1983, thereby implying that, if the attributes of workers and firms were unchanged between 1983 and 1988,

the entire distribution would have hardly changed. In other words, the difference in the actual distribution in 1983 and 1988 stems from the difference in the attributes of workers and firms, not the effect of the attributes on wages. Since it is unlikely that the elimination of the earnings test altered the attributes of workers and firms, the change in wage distributions between 1983 and 1988 was not caused by the elimination of the earnings test.

Figure 4 presents the actual distribution of wage and full second-tier benefit in 2000 and 2004 as well as the counter-factual distribution in which we assume that the attributes of workers and firms had remained at their 2000 level. First, we notice that the actual distribution of wages and second-tier benefit is flatter in 2004 than in 2000, including the corresponding parts below the threshold (370,000 yen). In this sense, we do not observe any new bunch below the threshold after the revival of the earnings test in 2002. Second, the counter-factual distribution produces a peak but the location is far below the threshold. The density just below the threshold is less than the actual distribution in 2000, thereby implying that even if the attributes of workers and firms had been fixed, we would have not seen any bunch in 2004 after the revival of the earnings test.[15] What we observed in the decomposition analysis of the wage distribution before and after the reforms reveals that the

[15] We acknowledge the limitation of our calculation of the full second-tier benefit from our dataset. One strong assumption is that all the respondents are eligible for the full first-tier pension benefit, which underestimates the second-tier benefit. At the same time, we disregard the employer-provided pension benefit (third-tier) simply because we are not able to compute it from the dataset, which overestimates second-tier benefits. In general, the amount of the employer-provided pension benefit is larger than that of the first-tier benefit, thereby implying that the actual second-tier benefit is smaller than what we compute and that the distribution moves to the left if we use the precise data of the second-tier benefit.

elimination of the earnings test in 1985 did not eliminate the bunch and that the revival in 2002 did not create a new one, though partial effect on labor supply is observed. Based on these observations, we conclude that the labor supply effect of the social security earnings test is negligible and the revisions of the earnings test rule did not affect the labor supply decision of the elderly.

However, a closer look at the frequency in the counterfactual distribution and the actual distribution in 2000 shows that the frequency slightly over 370,000 yen declines and that this change contributes to the larger peaks at around 100,000 yen and 200,000 yen. Since those changes are caused not by the change in the attributes of workers and firms, it is possible that the reintroduction of the earnings test motivated some workers to reduce their labor supply with respect to the earnings test. In order to explore the possibility, we examine the direct survey response in the next section.

6. Direct survey response

A unique advantage of our dataset is that the respondents were explicitly questioned with regard to the effect of the change in the earnings test. We use this survey response to complement what we observed in the change in wage distributions. The 2004 survey asked the respondents who were eligible for EPI benefits: “Do you restrict working hours or days

due to a reduction or no receipt of EPI benefits under the social security earnings test?" Then, each respondent is asked to choose one of the following answers: (1) I did not work at all, (2) I restricted working hours or days, (3) I did not adjust employment even under the earnings test, and (4) others.[16]

This direct response complements what we found in the bunch analysis for three reasons. First, the direct question reveals the proportion of those who did not work at all under the earnings test, while the bunch analysis focused on the individuals who are currently working. However, the most serious effect of the earnings test on labor supply should be found in those who have given up working; thus, the bunch analysis is likely to underestimate the labor supply effect of the earnings test.

Second, a bunch analysis is vulnerable to reporting errors or labor market rigidity (Haider and Loughran 2008). The analysis implicitly assumes that workers can adjust their wages and working hours to the level just below the threshold; however, in reality, this may not be the case. If many workers are not able to adjust their labor supply freely, the effect of the earnings test is not observed "just below" the threshold and we overlook or underestimate the effect. The observation on the comparison between the counterfactual and actual distributions in 2000 suggests that this may be the case. Third, our calculation of the full second-tier benefits is not completely exempt from measurement errors since the exact value of the

[16] The 2000 survey provided a similar question but only the individuals aged between 60 and 64 were asked to respond to it because those aged between 65 and 69 were not included in the earnings test in 2000.

benefits is not available in the 2004 survey.

While we do not insist that the direct survey responses are a perfect measure for evaluating the effect of the earnings test and acknowledge the possibility that the subjective response overestimates the effect, we believe that the direct survey response is an alternative measure that complements the methodology in the literature.

Table 1 summarizes the results of the direct survey responses. First, individuals who have stopped working completely shared approximately 30 percent and those who restricted their working days or hours shared 17 percent, both of which comprised nearly half the respondents.[17] Another 30 percent responded that their labor supply decision was not affected by the earnings test. If we disregard the individuals who chose "others," the earnings test affected labor supply behavior of over half the respondents and close to 40 percent stopped working. This result demonstrates that the labor supply effect of the earnings test is large in contrast to the results of the bunch analysis.

Second, the averages of monthly wage, sum of the monthly wage and (actual) second-tier benefit, and working hours/days were indeed smaller for individuals who reported that they were affected by the earnings test than those who reported that they were not. For example, the average monthly wage was close to zero for those who chose (1), approximately 100,000 yen for those who chose (2), and approximately 200,000 yen for those who chose (3).

[17] Since the monthly wage and pension benefits were received in September 2004, a few people earned some wages occasionally in the month; therefore, the average of monthly wage or working hours/days are not virtually equal to zero.

Hence, the average monthly wage of the individuals who were partially affected by the earnings test is half of that of those who were not affected at all. The hours worked per day and the days worked per week were smaller for those who were affected by the earnings test, although the gap was smaller than in the monthly wage.

Third, the sum of monthly wage and second-tier benefit, which was the objective of the earnings test in 2004, was approximately 120,000 yen for individuals who completely stopped working, 220,000 yen for those who partially stopped working, and 320,000 yen for those who were not affected at all by the earnings test. The most noticeable is the fact that the average for those who adjusted their labor supply due to the earnings test, i.e., 220,000 yen is far below the threshold of 370,000 yen. The individuals who adjusted for their labor supply earned much lower labor income, not just below the threshold, thereby suggesting that a bunch analysis is likely to underestimate the effect of the earnings test on labor supply. This is consistent with the above-mentioned observation in the comparison between the counterfactual and the actual distributions in 2000.

Our discussion based on direct survey responses to a question regarding the labor supply effect of the earnings test echoes with the finding of Haider and Loughran (2008) who argued that the response to the earnings test in survey data is obfuscated by measurement error and labor market rigidities. Figures 1 and 2 indicate that the data contains reporting errors, which is evident from the larger frequency in a multiple of 50,000 yen. Moreover,

labor market rigidity is evident from the fact that working hours are concentrated in certain specific cells in the histogram (see the appendix figure). A bunch analysis disregards individuals who are not incumbent workers as well as earnings and working hours/days of workers who indeed earn much less than the threshold. The labor supply effects are likely to be larger than what is suggested by the observation from the bunch analysis.

7. Concluding remarks

Recent policy reforms in Japan aimed to encourage elderly workers to remain for a longer period in the labor force and retire in subsequent years. The discouraging effects of the social security earnings tests have been debated in both academic and policy arenas; however, a large volume of the literature has not reached a consensus on the labor supply effect of the earnings tests.

We utilize micro-level data from the nationwide survey on employment of the elderly for examining the change in the labor supply effect for those aged 65–69 before and after two major reforms of the social security earnings test in Japan: its elimination in 1985 and its revival in 2002. Our analysis provides two important findings. First, there is little evidence that the revisions in the earnings test affected the wage distribution of the elderly, even after controlling for the changes in the attributes of workers and firms between the surveys. This

finding supports the small effect of the revisions of the earnings test, in keeping with the results of a majority of related papers. Second, the direct responses by the respondents in the SEE to the revival in 2002 revealed a large effect on labor supply of the elderly. The share of the individuals in the sample who responded as having adjusted their labor supply due to the earnings test accounted for half the respondents and indeed their wage and working hours were much smaller than those who answered that they did not adjust their labor supply.

Our empirical findings show that the traditional bunch analysis may overlook and underestimate the labor supply effect when it is obscured by measurement errors or labor market rigidities. We argue that a direct response to the question on the labor supply effect is an alternative to complement the traditional methodology used to examine the labor supply effect of the earnings test. In order to cope with these issues, a further study must examine the effect of the social security earnings test for younger people aged between 60 and 64. Moreover, other important factors that affect the labor supply decision such as health status, family relationship, and labor-leisure choice must be examined together in order to study the relative effect of the earnings test on the labor supply decision of the elderly.

References

Abe, Yukiko. 1998. "1980-1990 Nendai no Dansei Koureisha no Roudou Kyokyu to Zaishoku Rorei Nenkin Seido" (Labor Supply of the Elderly Male and the Earnings-Tested Pension in the 1980s and 1990s) [in Japanese], JCER Economic Journal (Nihon Keizai Kenkyu), Vol. 36, pp. 50-82.

Baker, Michael and Dwayne Benjamin. 1999. "How Do Retirement Tests Affect the Labor Supply of Older Men?" Journal of Public Economic*s*, Vol. 71(1), pp. 27-52.

Borjas, George J. 2005. Labor Economics (Third Edition). Boston: McGraw-Hill Irwin Press.

Burtless, Gary, and Robert A. Moffitt. 1985. "The Joint Choice of Retirement Age and Post-Retirement Hours of Work," Journal of Labor Economics, Vol. 3, pp. 209-36.

DiNardo, John. 2002. "Propensity Score Reweighting and Changes in Wage Distributions," University of Michigan, mimeograph.

DiNardo, John, Nicole Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," Econometrica, Vol. 65, No. 5, pp. 100-1046.

Disney, Richard and Sarah Smith. 2002. "The Labour Supply Effect of the Abolotion of the Earnings Rule for Older Workers in the United Kingdom," Economic Journal, Vol. 112, Issue 478, pp. C136-C152.

Friedberg, Leona. 2000. “The Labor Supply Effects of the Social Security Earnings Test,” Review of Economics and Statistics, Vol. 82, No. 1, pp. 48-63.

Gruber, Jonathan and Peter Orszag. 2003. “Does the Social Security Earnings Test Affect Labor Supply and Benefits Receipt?” National Tax Journal, Vol. 56, No. 4, pp. 755-73.

Gustman, Alan L., and Thomas L. Steinmeier. 1985. “The 1983 Social Security Reforms and Labor Supply Adjustments of Older Individuals in the Long Run,” Journal of Labor Economics, Vol. 3, pp. 237-53.

Haider, Steven and David Loughran. 2008. “The Effect of the Social Security Earnings Test on Male Labor Supply: New Evidence from Survey and Administrative Data,” Journal of Human Resources, Vol. 43, No.1, pp. 57-87.

Iwamoto, Yasushi. 2000. “Zaishoku Rorei Nenkin Seido to Koureisha no Shugyo Kodo” (The Social Security Earnings Test and Labor Supply of the Elderly) [In Japanese], Quarterly of Social Security Research (Kikan Shakai Hosho Kenkyu), Vol. 35, No. 4, pp.364-76.

Komamura, Kohei. 2007. “The 2004 Pension Reform and the Impact Aging in Japan,” The Japanese Journal of Social Security Policy, Vol. 6, No. 1, pp. 144-56.

Lemiuex, Thomas. 2002. “Decomposing Changes in Wage Distributions: A Unified Approach,” Canadian Journal of Economics, Vol. 35, No. 4, pp. 646-88.

Ministry of Labour, Health and Welfare. 2002. *Nenkin Hakusho* (*White Paper on Pensions*).

National Institute of Population and Social Security Research (2007). Population Projection of Japan (as of December 2006).

OECD. 2008. Ageing and Employment Policies-Statistics on Average Effective Age of Retirement.

Ogawa, Hiroshi. 1998. “Nenkin ga Koureisha no Shugyo Kodo ni Ataeru Eikyo nit suite” (On the Impact of the Public Pension on Labor Supply of the Elderly) [in Japanese], Keizai Kenkyu (Economic Review), Vol. 49, No. 3, pp. 245-58.

Shimizutani, Satoshi and Izumi Yokoyama. 2008. “Has Japan’s Long-Term Employment Practice Survived? New Evidence Emerging Since the 1990s,” Industrial and Labor Relations Review, forthcoming.

Appendix

Using a comparison between the 1983 and 1988 distributions as an example, we will briefly describe the procedure of a DiNardo, Fortin, and Lemieux decomposition. The wage distributions in 1983 and in 1988 are written as

$$f^{1983}(Y) = \int f^{1983}(Y \mid X) h(X \mid t = 1983) dX,$$

$$f^{2003}(Y) = \int f^{2003}(Y \mid X) h(X \mid t = 2003) dX,$$

respectively, where $f^{1983}(Y \mid X)$ is the mechanism of wage determination in 1983 that maps the attributes of workers and firms X to the wage distribution Y and $f^{1988}(Y \mid X)$ is the wage determination in 1988. Moreover, what the wage distribution would be in 1988 if the distribution of X is unchanged from that in 1983 is written as

$$f_{1983}^{1988}(Y) = \int f^{1988}(Y \mid X) h(X \mid t = 1983) dX.$$

The DiNardo, Fortin, and Lemieux approach employs a “re-weighting” method for estimating the counterfactual distribution. The counterfactual distribution can be rewritten as

$$f_{1983}^{1988}(Y) = \int f^{1988}(Y \mid X) h(X \mid t = 1983) dX = \int \omega f^{1988}(Y \mid X) h(X \mid t = 1988) dX,$$

where $\omega \equiv \frac{h(X \mid t = 1983)}{h(X \mid t = 1988)}$. The Bayesian rule produces $\omega = \frac{P(t = 1983 \mid X)}{P(t = 1988 \mid X)} \frac{P(t = 1988)}{P(t = 1983)}$, where the conditional probabilities $P(t = 1983 \mid X)$ and $P(t = 1988 \mid X)$ are propensity scores for the specific observations in 1983 and 1988, respectively, conditioned on *X*, which are calculated by the logit model in this analysis (the estimation results of the logit model is available on request). The terms $P(t = 1983)$ and $P(t = 1988)$ are calculated based on the proportion of the observations pertaining to 1983 and 1988 in the pooled data, respectively. The counterfactual distribution is calculated using the kernel density estimation, using calculated weight ω. The kernel density is also useful to adjust for reporting errors in this study. In order to make the results comparable with those from the histogram analysis, we take the level of wages as the dependent variable; however, the results are unchanged when we take the logarithm of wages as the dependent variable.

Figure 1 Histogram of the wage distribution in 1983 and 1988

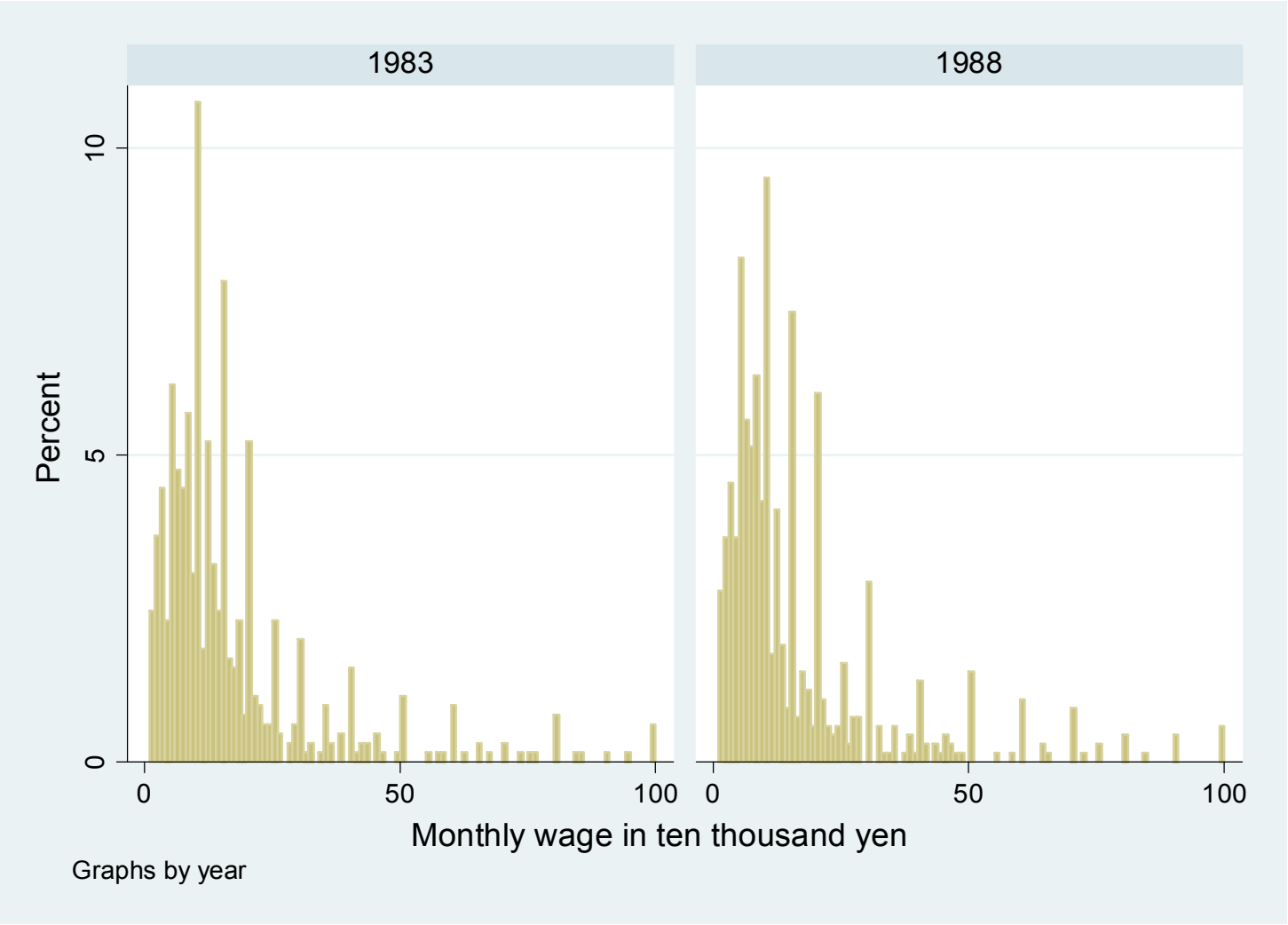


Figure 2 Histogram of the distribution of wage plus full second-tier benefit in 2000 and 2004

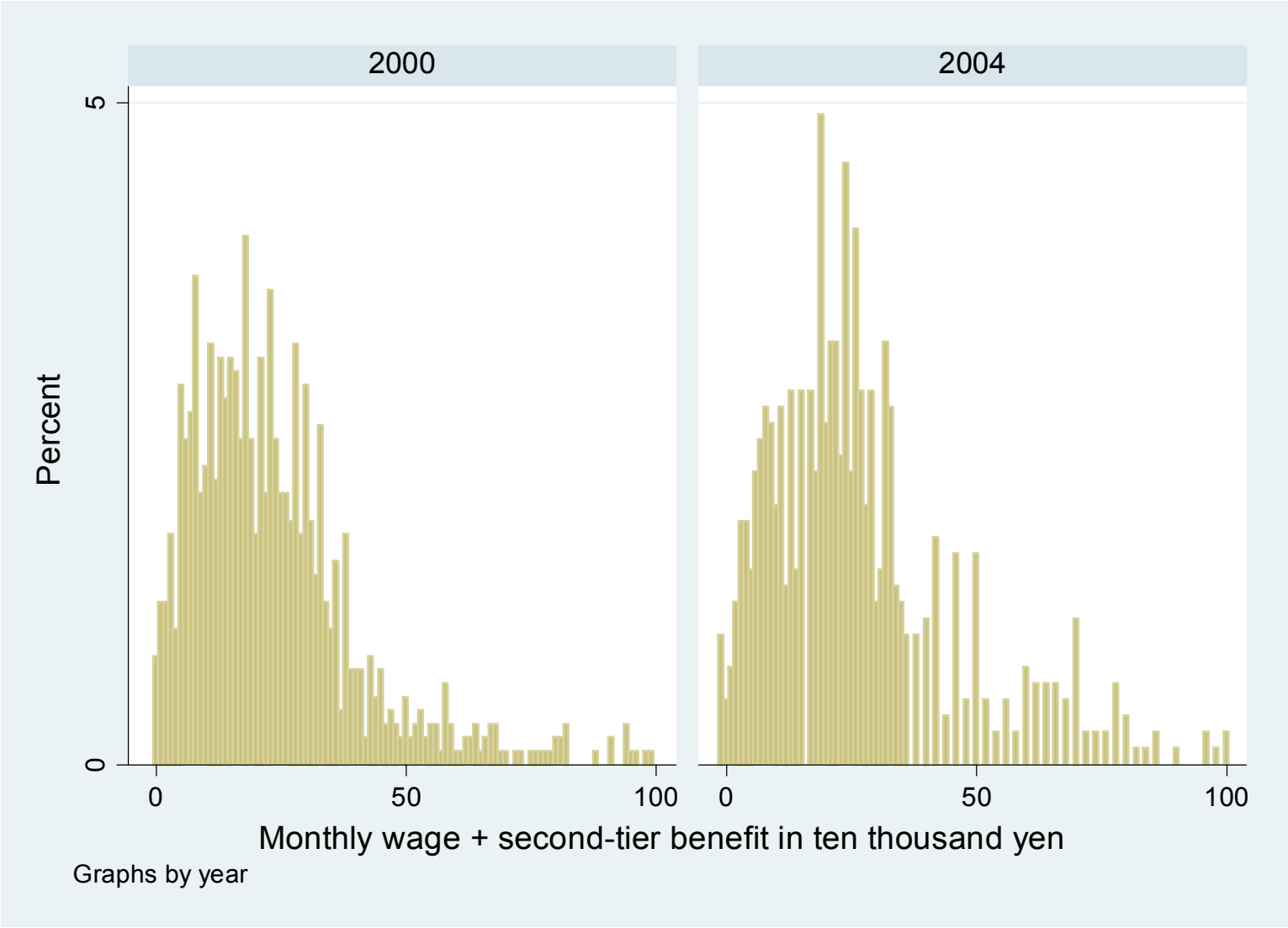


Figure 3 Wage distribution for the age group 65-69 in 1983 and 1988

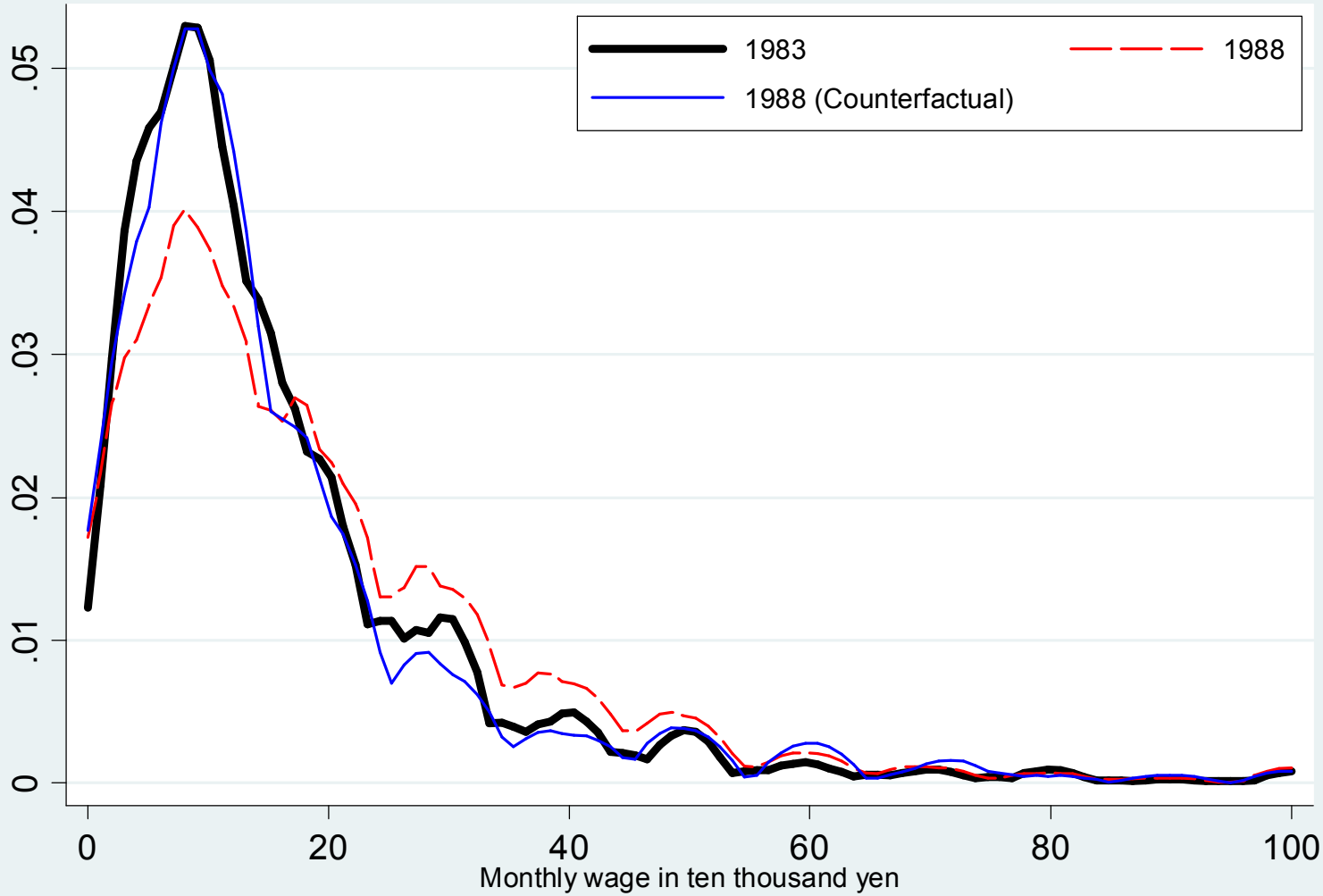


Figure 4 Wage and second-tier benefit distribution for the age group 65-69 in 2000 and 2004

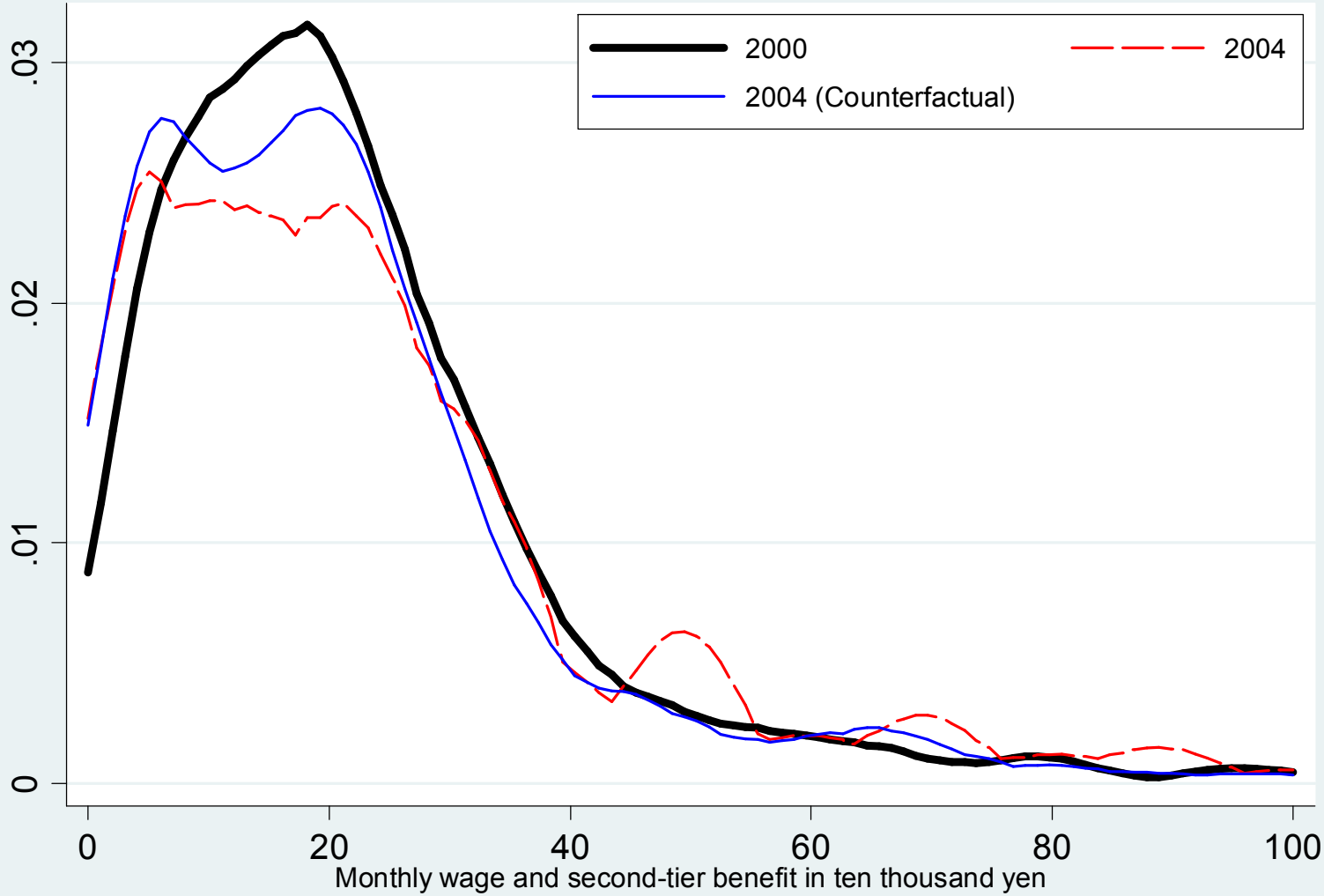


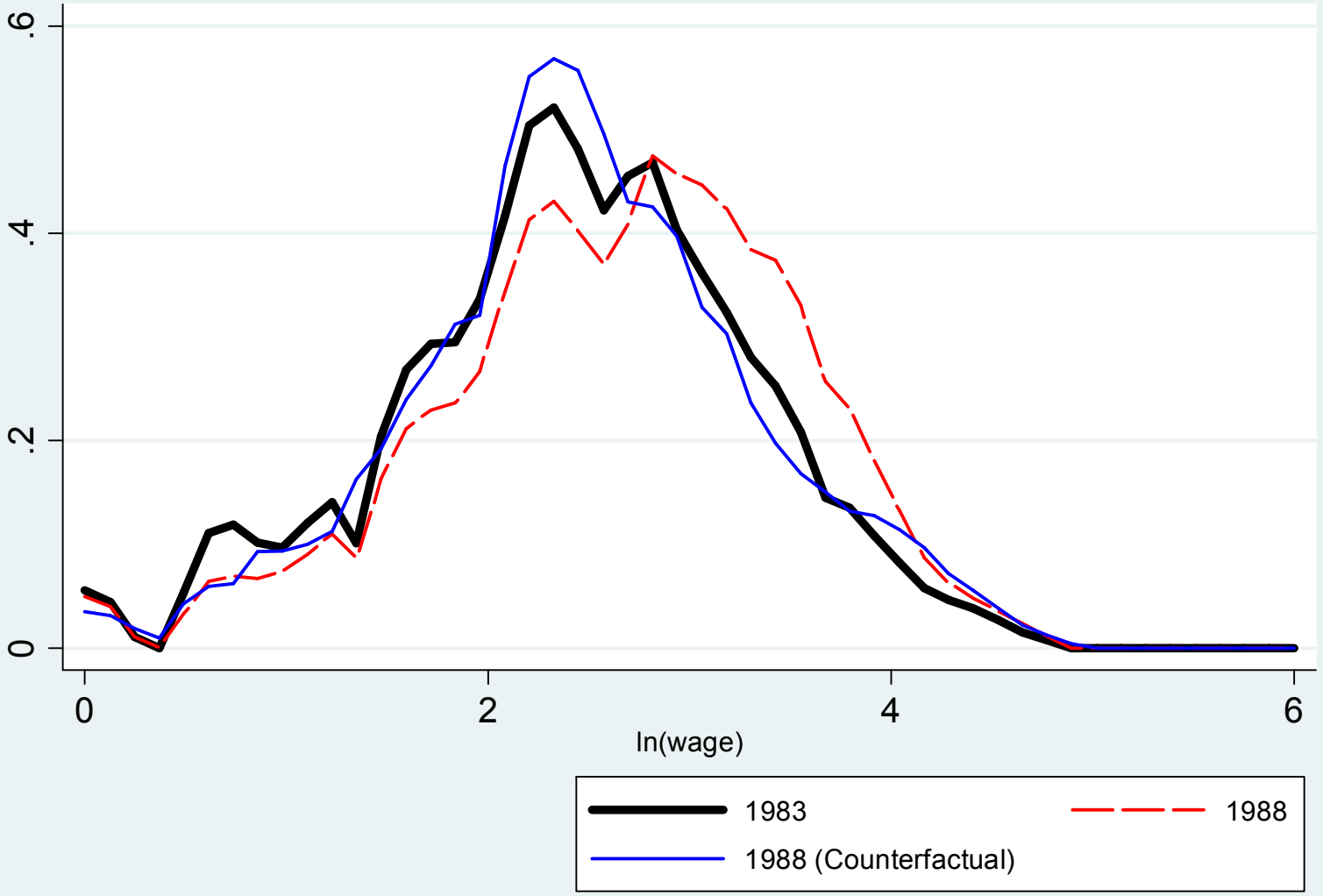
Table 1 Direct Response to the Effect of the Earnings Test: The 2004 survey

	Number of respondents	Monthly wage	Monthly wage + second-tier benefits	Working Hours per day	Working days per week
upper number	average	average	average	average	average
lower number	(share)	(S.D.)	(S.D.)	(S.D.)	(S.D.)
Not working at all	469	0.26	12.03	0.16	0.17
	29.5%	(2.61)	(8.99)	(0.90)	(0.93)
Restrict working days or hours	268	10.29	21.92	5.58	3.50
	16.9%	(13.31)	(23.54)	(2.73)	(1.82)
Not affected by the earnings test	497	19.32	31.79	6.81	4.63
	31.3%	(18.23)	(30.23)	(2.52)	(1.69)
Others	354	5.42	17.19	2.27	1.49
	22.3%	(13.48)	(20.47)	(3.34)	(2.33)
Total	1,588	9.07	21.03	3.63	2.42
		(15.34)	(23.65)	(3.70)	(2.51)

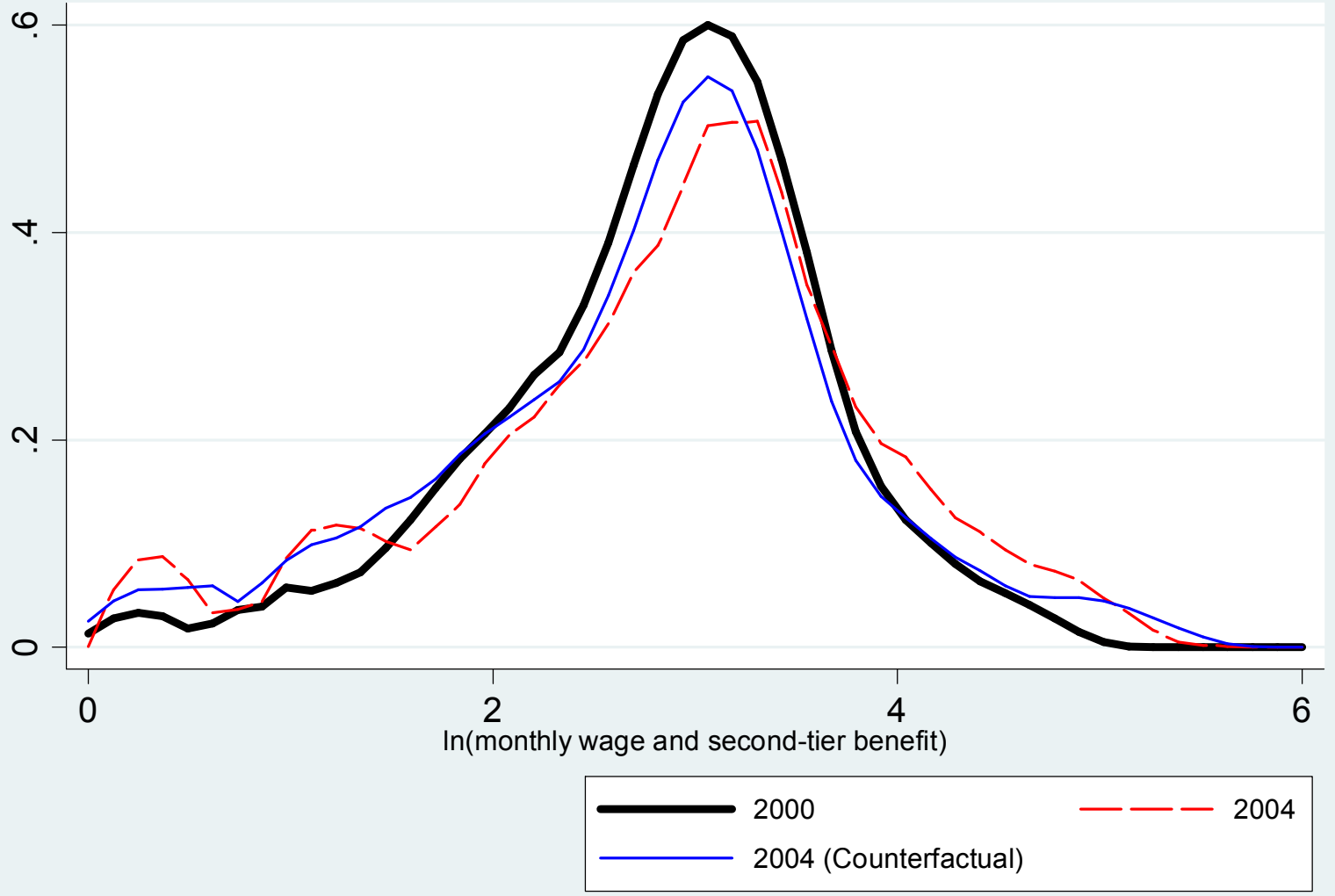
(Note) Author's calculation.

Appendix Figure Working hours per month

Appendix Figure 2 Log wage distribution for the age group 65-69 in 1983 and 1988



Appendix Figure 3 Log (wage and second-tier benefit) distribution for the age group 65-69 in 2000 and 2004



Appendix Table Summary statistics of the main variables in the decomposition analysis

	1983	1988	2000	2004
	share			
Male	0.768	0.696	0.718	0.737
Age				
age65	0.214	0.240	0.224	0.274
age66	0.243	0.247	0.226	0.213
age67	0.194	0.199	0.189	0.200
age68	0.183	0.174	0.193	0.162
age69	0.166	0.141	0.168	0.151
Type of job				
Expert or technical	0.101	0.095	0.109	0.172
Management	0.151	0.148	0.105	0.111
Administration	0.101	0.086	0.086	0.092
Sales	0.118	0.103	0.086	0.067
Services	0.052	0.083	0.096	0.116
Security guard	0.032	0.039	0.044	0.046
Agriculture, Forestry and Fishery	0.083	0.113	0.042	0.056
Transportation	0.014	0.023	-	-
Communication	0.000	0.003	-	-
Mining	0.002	0.000	-	-
Transportation and communication			0.287	0.211
Production workers (all)			0.125	0.122
Production workers (ceramic products)	0.011	0.001	-	-
Production workers (metal materials)	0.006	0.007	-	-
Production workers (chemical products)	0.005	0.003	-	-
Production workers (metal products)	0.029	0.027	-	-
Production workers (machinery products)	0.014	0.020	-	-
Production workers (transportation products)	0.008	0.009	-	-
Production workers (precision products)	0.012	0.003	-	-
Production workers (food products)	0.028	0.020	-	-
Production workers (spining products)	0.012	0.009	-	-
Production workers (clothing products)	0.032	0.029	-	-
Production workers (wood products)	0.012	0.016	-	-
Production workers (pulp/paper products)	0.005	0.010	-	-
Production workers (printing)	0.014	0.001	-	-
Production workers (plastic products)	0.009	0.007	-	-
Production workers (pelt products)	0.003	0.001	-	-
Production workers (others)	0.015	0.011	-	-
Production workers (operation)	0.002	0.003	-	-
Production workers (electronics)	0	0.003	-	-
Production workers (construction)	0.038	0.032	-	-
Production workers (transportation)	0.017	0.014	-	-
Production workers (miscellaneous)	0.072	0.080	-	-
Firm size				
1-4 persons	0.022	0.026		
5-29 persons	0.121	0.129		
30-99 persons	0.083	0.082		
100-299 persons	0.048	0.047		
300- persons	0.074	0.044		
governments	0.018	0.014		
	mean			
Working days per week	5.406	4.709	4.646	4.431
(S.E.)	(1.238)	(1.746)	(1.671)	(1.639)
Working hours per day	7.230	6.686	6.757	6.616
(S.E.)	(2.257)	(2.214)	(2.350)	(2.275)
Number of observations	651	697	986	890

(Note) In addition to the variables in this table, we include indicator variables for each prefecture (47 prefectures) in the estimation.