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 Economics, 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

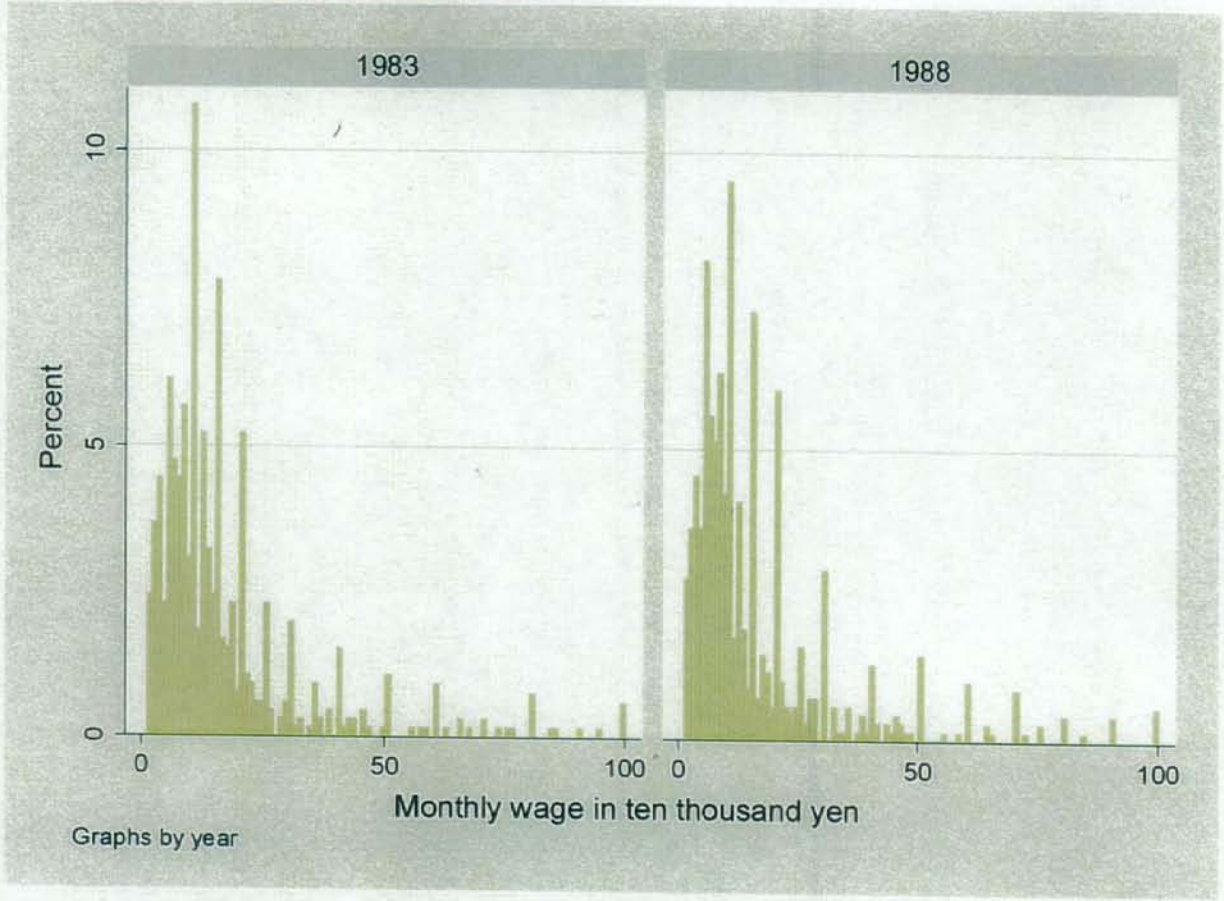


Figure 2 Histogram of the distribution of wage plus full second-tier benefit in 2000 and 2004

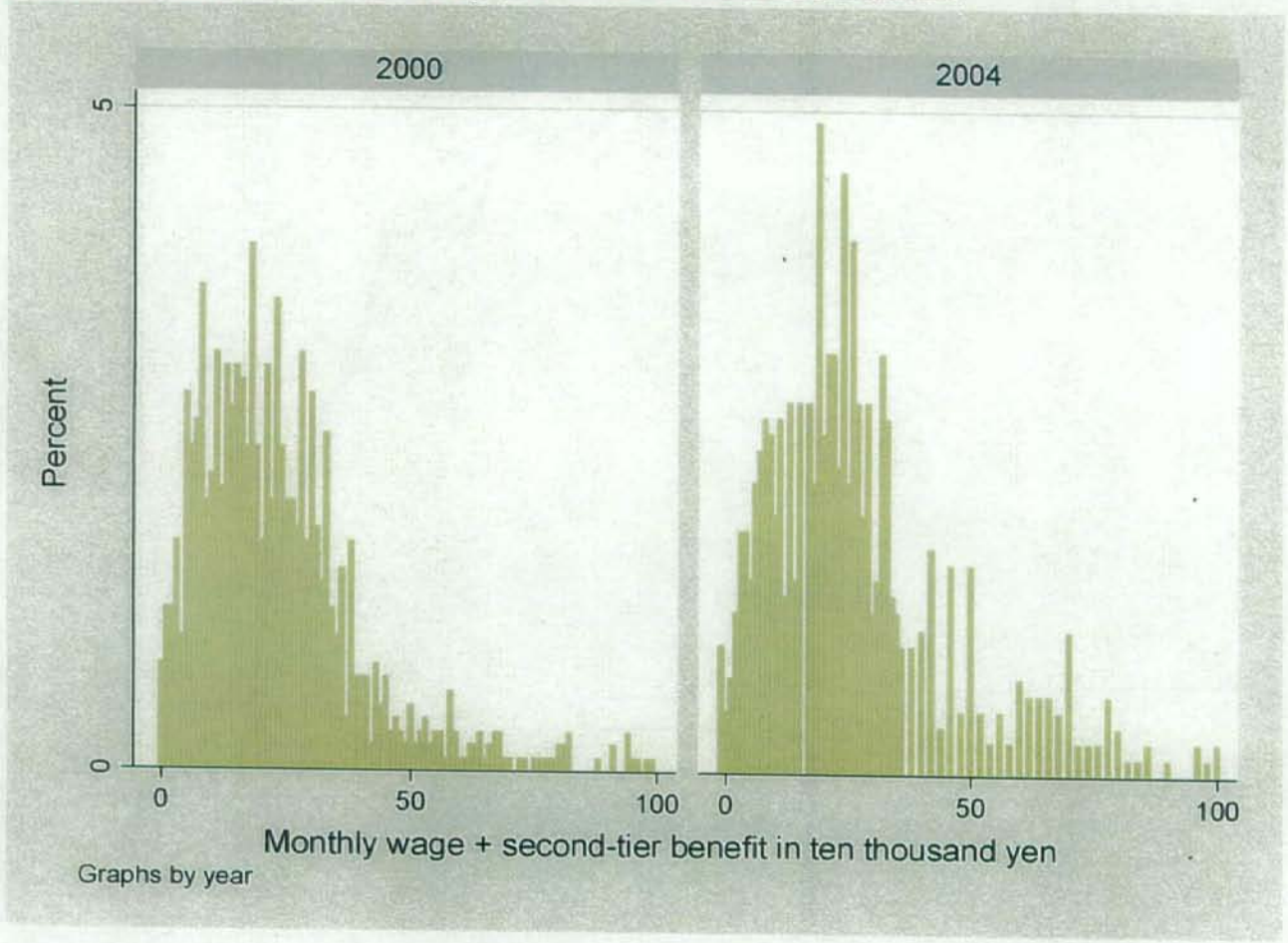


Figure 3 Wage distribution for the age group 65-69 in 1983 and 1988

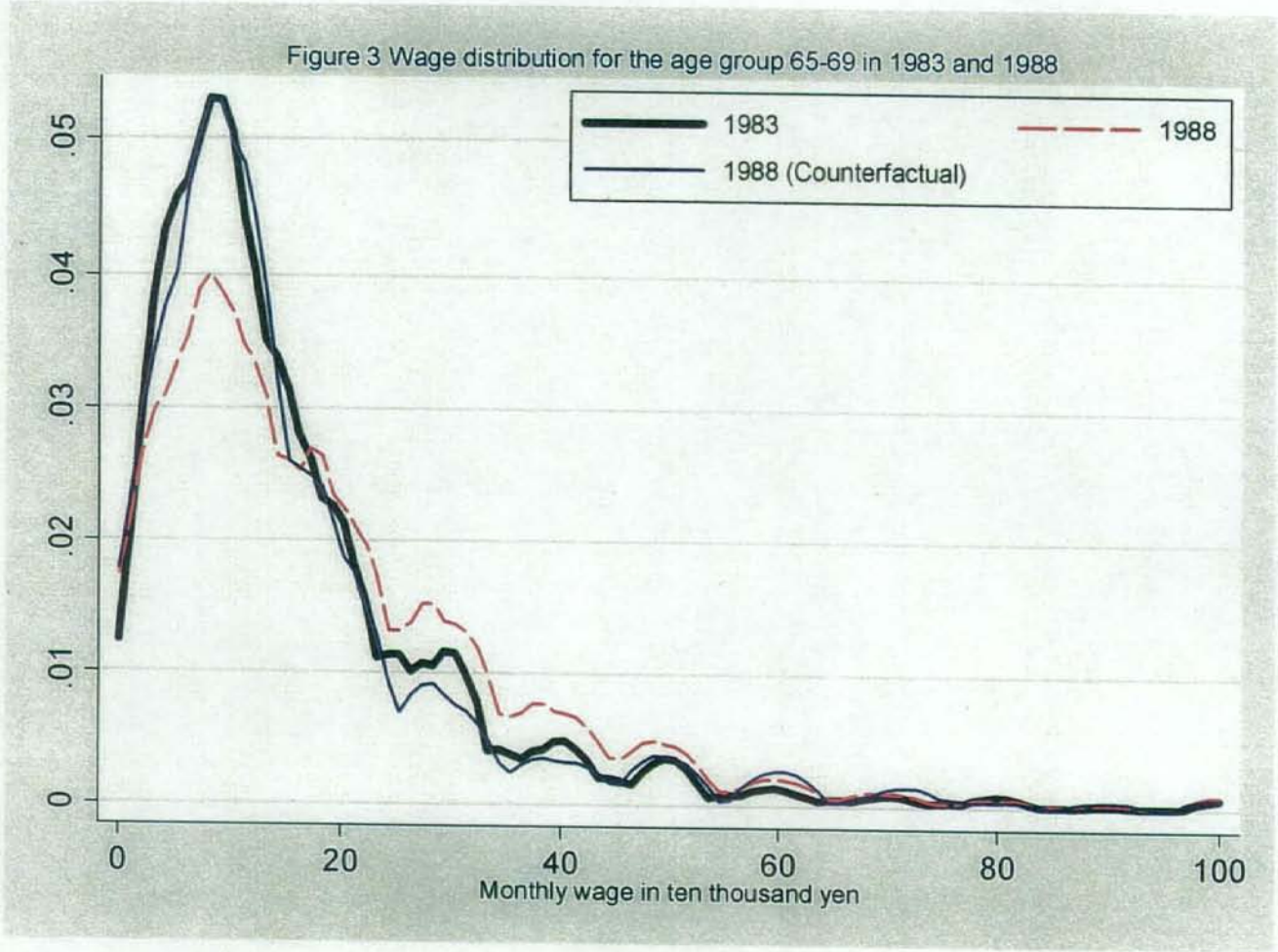


Figure 4 Wage and second-tier benefit distribution for the age group 65-69 in 2000 and 2004

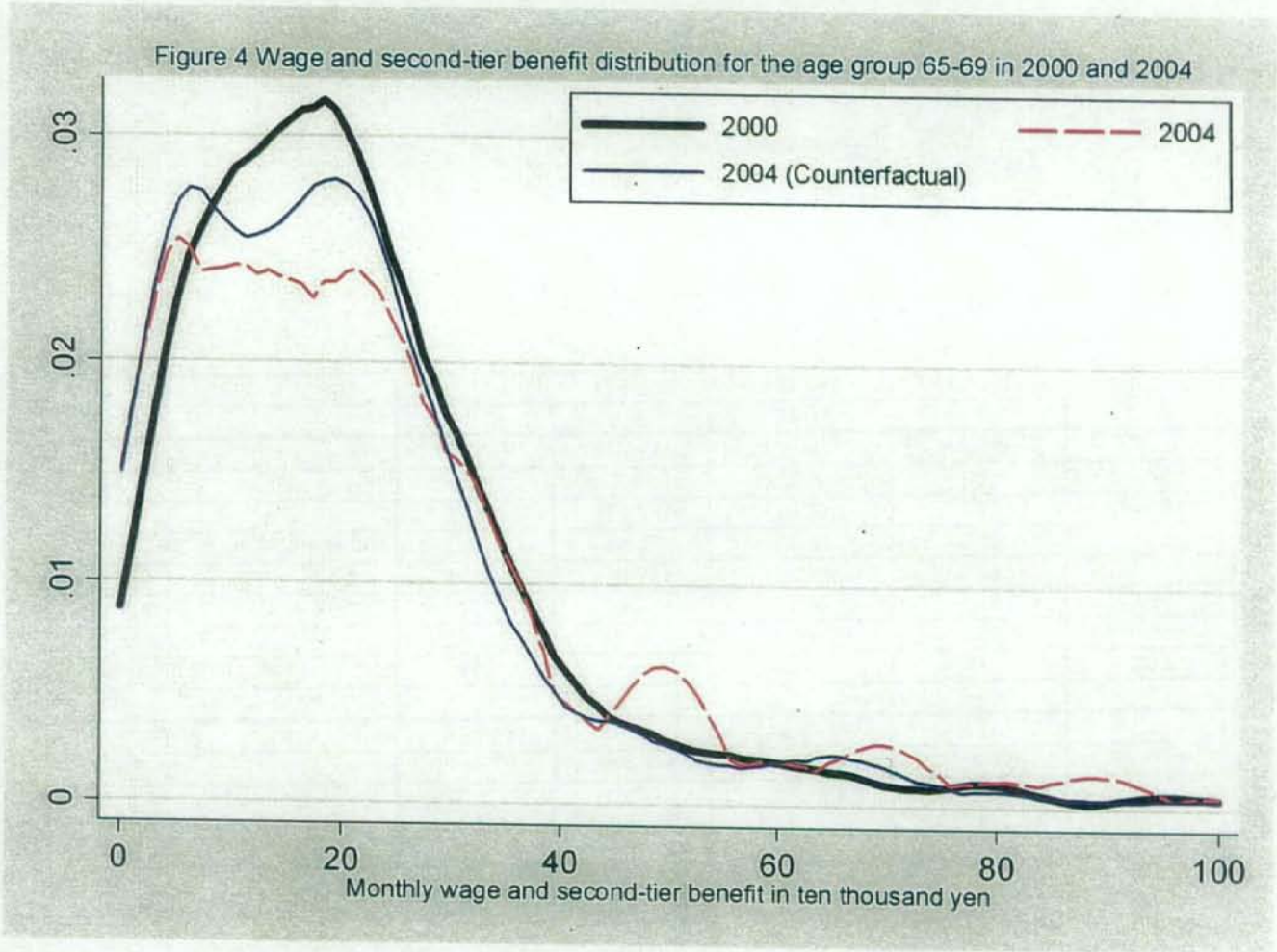
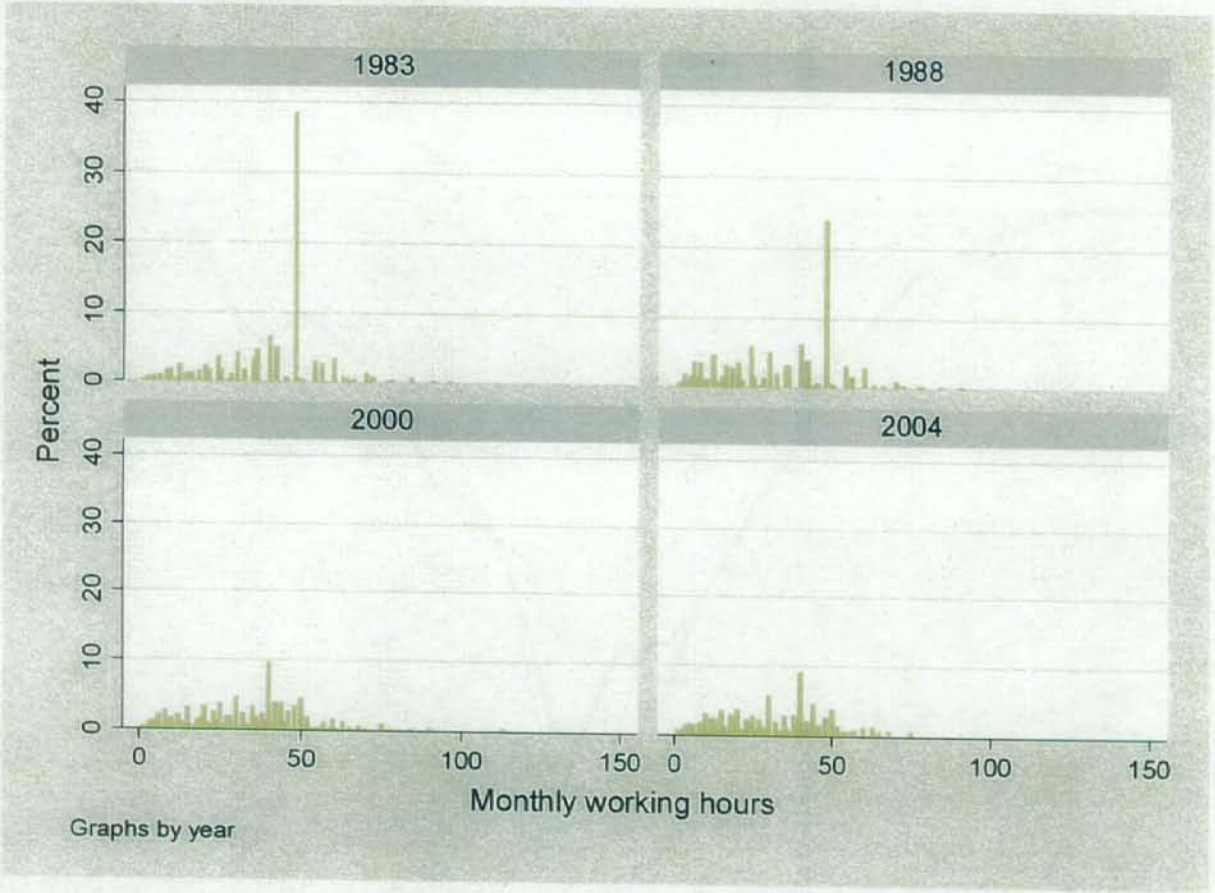


Table 1 Direct Response to the Effect of the Earnings Test: The 2004 survey

	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