DIPARTIMENTO DI ECONOMIA POLITICA E STATISTICA — DEPS

unisi

QUADERNI DEPS 942/2026

THE IMPACT OF OVERTIME LIMITS ON FIRMS AND WORKERS: EVIDENCE FROM JAPAN'S WORK STYLE REFORM

Gabriel Burdin
Ryo Kambayashi
Takao Kato

April 2026

UNIVERSITÀ DI SIENA 1240

Working Paper Series of the Department of Economics and Statistics of the University of Siena
Piazza San Francesco 7-8, 53100, Siena | Italy
deps.unisi.it

ISSN: 17209668

The Impact of Overtime Limits on Firms and Workers: Evidence from Japan's Work Style Reform*

Gabriel Burdin
University of Siena

Ryo Kambayashi
Musashi University

Takao Kato
Colgate University

March 30, 2026

Abstract

How do limits on working hours affect firms, workers, and households? This paper answers this question by analyzing Japan's 2018 Work Style Reform (WSR), which introduced the first binding cap on overtime hours. Using establishment payroll data and worker surveys in a difference-in-differences design, we show that the reform reduced monthly overtime by 5 hours (25%) and compressed the distribution of overtime within firms. Total earnings fell by 1.4% due to the effect of lower overtime pay. The reform also narrowed overtime gaps between standard and nonstandard jobs and reduced gender differences in long hours. Consistent with a reduction in the importance of extreme overtime as a screening device, women gained increased access to standard, career-track positions. We further document improvements in life and leisure satisfaction among female workers, but not among men. These gender differences are not explained by changes in perceived work intensification or time use. Instead, men partially substituted unpaid for paid overtime, consistent with the absence of well-being gains among male workers. Finally, exploiting information on spouses' working hours, we find suggestive evidence of cross-spousal spillovers on women's well-being, consistent with household-level complementarities.

Keywords: Working Time Regulations, Overtime, Wages, Employment, Subjective Well-being, Gender, Japan, Work Style Reform

*We thank conference and seminar participants at CEP-LSE, LUBS (University of Leeds), Rome Tre University, Trans-Pacific Labor Seminar (UCSB), IAFEP 2024 (Naples), AIEL 2024 (Naples), EALE 2024 (Bergen) and Workshop on Well-Being and The Future of Industrial Relations (Teiyko University) for providing valuable comments. The authors gratefully acknowledge funding from the Joint Usage and Research Center Program at the Institute of Economic Research, Hitotsubashi University. Gabriel Burdin also expresses his gratitude to the Japanese Society for the Promotion of Science (JSPS) for its generous travel funding in support of this research. This paper utilizes the micro data from the Preference Parameters Study of Osaka University's 21st Century COE Program 'Behavioral Macro-Dynamics Based on Surveys and Experiments', its Global COE project 'Human Behavior and Socioeconomic Dynamics' and JSPS KAKENHI 15H05728 'Behavioral-Economic Analysis of Long-Run Stagnation'.

1 Introduction

Many occupations reward long and inflexible work schedules with disproportionate career returns, creating strong incentives—especially within internal labor markets—for workers to supply hours far above their preferred level. Under conditions of asymmetric information, such incentives can generate a "rat race" in which workers engage in inefficiently long hours to signal their ability and commitment to employers (Akerlof, 1976; Landers et al., 1996; Kato et al., 2016). The incentives associated with "greedy jobs" interact with the unequal division of household labor, which raises the cost of supplying long hours for women and makes overwork an even stronger signal of promotion potential, thereby generating sharp gender asymmetries in career advancement (Goldin, 2014; Kato et al., 2016; Cortes and Pan, 2017; Frederiksen et al., 2024).

Overwork and persistent gender disparities pose major policy challenges for advanced economies facing population aging, slowing productivity, and the underutilization of highly educated women, while also striving to sustain workforce well-being and extend healthy working lives.[1] In response, many countries have introduced policies to promote work–life balance, including mandatory reductions in working hours and stricter regulations on overtime.[2] Understanding whether interventions that limit overwork can alleviate these pressures is therefore a central economic question, with implications for productivity, well-being, and gender equality.

This paper contributes to this debate by analyzing Japan's 2018 Work Style Reform (WSR), which introduced a binding cap on overtime hours. The reform offers a rare opportunity to identify the causal effects of limiting long hours on firms, workers, and households. The new regulation came into force in April 2019 and introduced a maximum overtime cap of 360 hours per year, equivalent to 30 hours per month.

Japan offers a compelling case study due to its entrenched long-hours work culture and the close link between overwork and persistent gender inequality. In 2018, roughly 16% of standard workers supplied overtime above the 30-hour-per-month

[1] For recent evidence highlighting the adverse economic impact of work-related stress and burnout, see Nekoei et al. (2024).

[2] A separate rationale for working-time regulations is to reduce unemployment through "work sharing." However, the evidence generally indicates either negligible or adverse employment effects (Crépon and Kramarz, 2002; Batut et al., 2023).

threshold.[3] Compared to other developed nations, Japan ranks in the bottom quintile in work-life balance and struggles with persistent issues such as long working hours and underutilized paid leave (Jones and Seitani, 2019). *Karoshi*—death from overwork—remains a significant social issue (Nishiyama and Johnson, 1997). Gender gaps in the labor market have also been a major concern in the Japanese context over the last decades. Despite substantial gains in female educational attainment, large gender gaps in pay and managerial representation persist. To address these challenges, Japan has introduced various policies over the decades, including a gradual reduction of the standard workweek from 48 to 40 hours (1980s–1990s) and increased overtime premiums for employers (Lee et al., 2012; Hamermesh et al., 2017). The 2018 WSR, described by late Prime Minister Shinzo Abe, as 'the first major reform [to labor laws] in 70 years', marked a critical milestone by implementing, for the first time, a mandatory cap on overtime hours.

To study the causal impact of the reform, we utilize a unique array of establishment, individual-, and household-level data, combining payroll and self-reported information on working hours—including paid and unpaid overtime—wages, employment, and subjective well-being. Our identification strategy leverages exogenous variation in overtime hours across firms and workers resulting from differential exposure to the reform. Using a difference-in-differences (DiD) framework, we compare establishments with varying initial levels of affected jobs before and after the reform. We then analyze the individual-level impacts of the reform by comparing full-time workers whose overtime hours exceeded the new cap before the reform with those whose hours did not. To guide our empirical analysis, we develop a simple conceptual framework in which long working hours serve as a signal of worker commitment and career prospects. The framework yields testable predictions regarding the effects of an overtime cap on paid and unpaid effort, job allocation within firms, and workers' well-being.

We summarize our main findings as follows. First, using establishment-level payroll data, we find that the reform had the intended first-stage effect: monthly overtime per worker in affected establishments decreased by 5 hours (25%) compared to the control group. The implied reduction in total working hours is much more modest

[3] In Japan, a standard employee is defined as a worker termed *"seiki no jyuugyouin"* in the place of his/her employment. For Japan's labor market segmentation into standard, career-track, and non-standard jobs, see for instance Kambayashi and Kato (2016).

(about 2%), reflecting the fact that paid overtime accounts for only a limited share of total monthly hours. Moreover, the distribution of overtime hours within these establishments also became less dispersed. Overtime is reallocated away from extreme levels toward lower intensities: the share of workers in the right tail of the distribution—those working more than 30 hours of overtime per month—declines sharply after the reform, while the shares working zero or moderate overtime increase. Total monthly wages fell by 1.4%, primarily due to a reduction in overtime pay, which was only partially offset by an increase in base wages and bonuses, and total employment remained unchanged. We also show that the reduction in extreme overtime compresses working-time differences between standard and nonstandard jobs, and we find evidence consistent with a reallocation of female employment away from nonstandard (precarious) positions and toward career-track jobs in affected establishments.

Second, in the individual-level analysis using survey data, we document a reduction in self-reported long workweeks and overtime. These reductions translate into significant improvements in both life and leisure satisfaction, but only among female workers. The gains correspond to increases of 7–11% relative to pre-reform sample means (about 0.3-0.4 standard deviations)—a magnitude equivalent to roughly 34% (47%) of the cross-sectional difference in life satisfaction (leisure satisfaction) between women reporting excellent versus poor health in our data. These gender differences are not driven by differential changes in perceived work intensification (*jitan-harasumento*) or time use (e.g., commuting, housework). Among male workers, and consistent with muted well-being effects for this group, we find that the decline in paid overtime was partly offset by an increase in unpaid overtime. Finally, we exploit detailed household-level information to explore cross-spousal effects of the reform. We find suggestive evidence that well-being gains are concentrated among single women and married women in dual-treated couples. For single women, these gains arise from reductions in their own working time, while for married women they reflect spillovers associated with their husbands' reduced work strain. These findings suggest that the impacts of the reform also operate through household-level channels.

The paper makes four distinct contributions. First, we contribute to the literature on working-time regulations by providing the first comprehensive causal evaluation of Japan's mandatory overtime cap—a policy instrument that differs markedly from the standard-workweek reductions studied in Europe. Existing research has mostly focused on European reforms that require firms to hold monthly wages constant (Hunt, 1999; Crépon and Kramarz, 2002; Rocheteau, 2002; Batut et al., 2023; Asai et al., 2024). In contrast, Japan's reform targeted excessive overtime—far more prevalent than in other advanced economies—and imposed no obligation to preserve monthly earnings. By analyzing the effects of this distinct form of regulation, our study offers new evidence on the effectiveness of policies aimed specifically at curbing extreme overwork.

Second, we contribute to the literature on gender inequality and the high returns to long hours (Bertrand et al., 2010; Kato et al., 2013; Cha and Weeden, 2014; Goldin, 2014; Cortes and Pan, 2017; Frederiksen et al., 2024). These studies point to long and inflexible hours as a culprit for the gender gap in wages and promotions. Our findings speak directly to this channel, showing that limits on extreme overtime can alter the relative allocation of women across standard (career-track) and nonstandard jobs within firms, consistent with a reduction in the importance of long hours as a screening device. Interestingly, the effects of a mandatory overtime ceiling documented in our paper contrast with the adverse employment effects of working-time floors on female workers recently reported in the literature (Carry, 2023).

Third, we expand the body of research on overtime regulation and unpaid work. Prior studies examine overtime taxation and regulations (Trejo, 2003; Oaxaca, 2014; Cahuc and Carcillo, 2014; Martins, 2017), as well as Japanese management systems and voluntary work–life balance initiatives (Tanaka et al., 2022; Takahashi et al., 2024). Others analyze the determinants of unpaid overtime (Bell and Hart, 1999; Pannenberg, 2005). We build on this literature by offering the first causal evidence on how a mandatory cap on paid overtime affects unpaid overtime and the distribution of hours within establishments—documenting a reduction in overtime dispersion and showing that men partially substitute toward unpaid overtime while women do not.

Finally, our paper relates to the literature on working hours and subjective well-being (SWB). Prior work documents SWB responses to shorter standard workweeks in Korea (Rudolf, 2014; Hamermesh et al., 2017), Portugal and France (Sánchez, 2017; Lepinteur, 2019; Berniell and Bietenbeck, 2020), and Germany (Cygan-Rehm and Wunder, 2018). Hamermesh et al. (2017) further show positive SWB effects of shorter legal working hours in Japan during the 1990s. A recent study by Carcillo et al. (2023) examines Korea's overtime limits but does not analyze labor-market or SWB outcomes.

We contribute to this literature in two distinct ways. First, we study the causal impact of Japan's Work Style Reform on both working hours and workers' subjective well-being within a unified empirical framework. Importantly, our estimated SWB gains among female workers are larger than those documented in studies of standard-workweek reductions in European countries (Lepinteur, 2019). This pattern is consistent with Japan's WSR targeting the extreme right tail of the overtime-hours distribution—a margin where the marginal disutility of hours is considerably higher, and thus where reductions in working time yield disproportionately larger well-being improvements. Second, our analysis goes beyond individual-level effects by examining cross-spousal spillovers of the working-time reform (Goux et al., 2014; Ma and Shi, 2020; Sun et al., 2026). Most prior studies cannot study household spillovers because they lack information on partners' working hours or treatment exposure. Exploiting matched respondent–spouse data, we show that average effects mask substantial heterogeneity across household types and that spousal treatment status shapes individual outcomes: men reduce overtime when their wives are treated, while women experience substantial SWB gains only when both spouses are treated. Using a four-cell difference-in-differences design, we isolate direct effects, cross-spousal spillovers, and complementarities in joint exposure, thereby uncovering household mechanisms through which overtime caps differentially affect men and women.

The paper is organized as follows. Section 2 outlines the institutional context and details of Japan's Work Style Reform, and sketches a conceptual framework that guides the empirical analysis. Section 3 describes the data. Section 4 presents the identification strategy and main results regarding establishment-level responses to the reform, along with a set of robustness checks. Section 5 delves into the individual-level impacts, while section 6 examines cross-spousal effects and household-level channels. Finally, Section 7 concludes.

2 Institutional context and Japan's Work Style Reform

Long Working Hours, Gender Gaps and Institutions. Japan has long been known for its extensive working hours. According to the Databook of International Labour Statistics (JILPT 2024), Japan has the highest share of employees working 49+ hours per week among G7 countries (15.3% in 2022), nearly twice the G7 average. This overwork culture has been linked to negative health outcomes for workers, including the phenomenon of death from overwork, known as *karoshi* (Nishiyama and Johnson, 1997; Bassanini and Caroli, 2015).

Persistent gender gaps in the labor market also represent a critical challenge for the Japanese economy (e.g., late Prime Minister Shinzo Abe's "target of increasing the share of women in leadership positions to at least 30% by 2020 in all fields in society"). According to the OECD, the gender gap in median earnings for full-time employees in Japan was approximately 21% in 2022 (the third largest in the OECD), almost twice as high as the OECD average. This is particularly troublesome, given that the gender gap in educational attainment has narrowed considerably. In 1980, only 12.3% of women advanced to the university-level. By 2023, it rose to 54.5% (the School Basic Survey, 2022). As the proportion of college-educated women has increased, the worker composition of full-time workers has changed dramatically. In particular, there has been a significant increase in the proportion of female universit graduates among standard employees from 2.3% in 1982 to 13.2% in 2022 (Employment Status Survey). Despite these improvements in average female educational attainment, however, a significant gender wage gap persists in Japan. According to the most recent Labor Force Survey, the proportion of female college graduates in managerial positions was still only 13% in 2022 (far short of late Prime Minister Abe's 30% target by 2020).

Following a series of institutional changes between the mid-1980s and mid-1990s, Japan's standard workweek was set at 40 hours. Employers have also been required to pay overtime premiums, with rates depending on the firm's size. However, prior to the reform, there were no attempts to set upper limits on overtime hours.

The 2018 Work Style Reform. The Work Style Reform Bill, enacted in June 2018, introduced mandatory limits on overtime hours for the first time in Japan.[4] The bill introduced key amendments to two central pieces of labor law: the Labor Standards Act (LSA) and the Industrial Safety and Health Act (ISHA). [5] Under the new legislation, which came into force in April 2019, overtime work is capped at 360 hours per year (30 hours per month, with a maximum of 45 hours in any single month). Non-compliant

[4]The bill was reviewed by the Labor Policy Council (an advisory panel to the Minister of Health, Labour and Welfare) in September 2017 and approved by the cabinet in April 2018. Deliberations in the House of Representatives and House of Councillors (the lower and upper houses of the National Diet of Japan, respectively) took place between April and June 2018. The timing of the legislative process is relevant, as firms and workers may have anticipated the reform's effects. However, event studies reported below do not show signals of anticipatory responses neither at the establishment nor individual level.

[5]For further information, see here.

firms face fines of up to JPY 300,000 (approximately USD 2,000) per worker. In special circumstances, firms can negotiate with employees to require up to 100 hours of overtime per month, provided that the annual limit of 720 hours is not exceeded. Certain highly specialized professionals and occupations—such as managers, drivers, doctors, and research and development (R&D) professionals—are exempt from the overtime limits, and there was a one-year grace period for small and medium-sized enterprises (SMEs).[6] While the overtime cap is the primary focus of our analysis, the legislation introduced additional measures aimed at improving worker health and well-being. For example, starting in April 2023, SMEs no longer benefit from a reduced overtime premium and are required to pay the same 50% additional wage rate for overtime hours as large firms. The new law also mandates that employers allow workers to take at least five days of annual paid leave. Finally, the bill introduces a work-interval system to ensure workers have enough rest between working days. [7] [8]

Conceptual Framework and Testable Predictions. Motivated by the institutional features of the overtime regulation described above, in Appendix A we develop a simple theoretical framework that rationalizes the impact of the reform and yields a set of testable predictions, serving as a guide for the empirical analysis. The framework highlights four key mechanisms: (i) compression of paid overtime under a binding cap, (ii) substitution into unpaid effort, (iii) household spillovers in well-being, and (iv) reallocation across job types within establishments.

First, a binding cap on paid overtime reduces reported overtime hours, with larger effects in establishments and among workers with higher pre-reform exposure.

Second, because effort generates both wage income and career returns, workers with stronger career concerns may partially substitute unpaid for paid overtime when the cap binds. As a result, total effort is expected to decline by less than paid overtime hours, and substitution into unpaid effort may be more pronounced among workers

[6] Historically, the connection between hours worked and compensation has been relaxed for specific occupations in Japan. For example, under the discretionary working system (DWS) introduced in 1987, companies are exempt from complying with working time limits in the case of professionals and managerial employees engaging in corporate planning activities. Only about 1% of Japanese employees are affected by this special regime. Exempt workers under DWS work longer hours but do not report worse health or lower job satisfaction than nonexempt workers (Izumi et al., 2025).

[7] We use data covering 2014-2022, and hence the elimination of the overtime pay exception for SMEs should not affect our difference-in-differences analysis. As shown below, our results are not influenced by variations in the number of working days.

[8] For more details, see here.

facing stronger career incentives.

Third, by compressing long-hours requirements in standard jobs, the reform can alter job assignment within establishments. A reduction in the hours requirement of standard jobs lowers the effort-cost gap between job types and can reduce the assignment threshold, thereby expanding access to standard positions ("good jobs") for workers with higher disutility of long hours. If women face higher marginal costs of long working hours, for example due to a greater burden of family responsibilities, this mechanism predicts a relative increase in women's representation in standard, career-track jobs. Finally, reductions in working time affect well-being both directly and through household interactions. Lower effort reduces fatigue and increases non-work time. When non-work time is complementary within households, well-being gains may be larger when both partners experience reductions in effort, leading to potentially non-additive effects of the reform within couples.

3 Data

To analyze the impact of the WSR on firm and worker outcomes, we utilize three main datasets detailed in this section.

Establishment payroll data. The Basic Survey on Wage Structure (BSWS) is an employer-employee matched survey conducted annually on June 30 by the Japanese Ministry of Labor, Health, and Welfare (MLHW). It primarily replicates company payroll records, providing accurate data on *actual* working hours and wages for June, including overtime hours and pay. Additionally, the survey gathers worker characteristics such as education, gender, tenure, age, and contract type (e.g., fixed-term, nonstandard).[9] Establishment-level attributes include employment size, industry, and location (prefecture). The sample size is large, covering approximately one million workers across fifty thousand establishments each year—representing roughly 5% of all establishments in Japan. Sampling is conducted in two stages: first, establishments with five or more employees are selected using a stratified sampling method; second, employees are sampled within each selected establishment. While the dataset has a panel structure at the establishment level, it does not track individual employees over time. Consequently, we aggregate employee-level information to the establishment

[9]For contract types used for employment in Japan, see Kambayashi and Kato (2016).

level for analysis.

Individual-level survey data. While establishment payroll data allows for accurate measurement of paid working hours, wages, and employment, examining other potentially important outcomes and channels requires the use of self-reported information from workers. To better understand the impact of the Work Style Reform (WSR) at the individual level, we use panel data from the PreferenceParameters Study conducted by the University of Osaka (OPPS). This survey has been conducted annually since 2003, with the exceptions of 2014–2015 and 2019–2020, and is representative of the Japanese population aged 20 to 69 years. Several key features of this dataset are particularly relevant to our study. First, the survey collects self-reported data on wages and *usual* weekly working hours. Specifically, respondents are asked: "How many hours per week do you and your spouse usually work, including overtime?" They then report their weekly overtime hours, distinguishing between paid and unpaid overtime. Second, the survey gathers data on various aspects of subjective well-being (SWB), including cognitive measures (e.g., life and job satisfaction), affective measures (e.g., happiness, stress, anxiety), and eudaimonic measures (e.g., "My daily life is fulfilling").[10] Finally, the survey collects a broad range demographic controls, such as gender, age, tenure, occupation, industry, and employer size. Importantly, having information on spouses' demographics, working time, and employment is crucial for investigating household-level adjustment channels and cross-spousal effects of the reform, as examined in Section 6.

Industrial Relations Survey. We also use the Survey on Labor-Management Communication (SLMC), conducted by the Ministry of Health, Labor, and Welfare in Japan every five years. SLMC examines various aspects of labor-management communication at the workplace level. Using establishment-level data from the SLMC, we calculate the prevalence of worker voice institutions—such as joint labor-management committees, shop-floor committees, and unions—across industries.[11] This aggregated industry-level data is then applied in Section 4.6 to assess whether establishments' responses to the Work Style Reform vary according to their sectoral worker voice regime.

[10] SWB measures have been extensively validated and correlate with neural activity and a range of behaviors (Urry et al., 2004; Clark, 2016; Liberini et al., 2017; Borga et al., 2022). However, there is no consensus on the underlying utility concept behind these measures (Benjamin et al., 2023).

[11] For worker voice institutions in Japan, see for instance Kato and Morishima (2002).

4 Establishment-Level Responses

In this section, we analyze the impact of the overtime cap introduced by the 2018 Work Style Reform at the establishment level, using the BSWS payroll data described earlier.

4.1 Research Design: Establishment-Level Difference-in-Differences

We construct an establishment-level panel spanning 2014–2022 by aggregating employee-level data.[12] As outlined in Section 2, the Work Style Reform, implemented in April 2019, introduced a cap on overtime hours: 360 hours annually, equivalent to 30 hours per month, with a maximum of 45 hours in any single month.

Our strategy exploits differences in establishments' structural dependence on overtime work.[13] To quantify an establishment's exposure to the regulation, we calculate the pre-reform share of non-managerial full-time employees working more than 30 overtime hours per month before 2019, i.e. the share of *high overtime workers*. We exclude managers, as highly specialized professions were exempt from the new regulations and may be covered by discretionary working systems (DWS).[14] To reduce the impact of temporary shocks, we average this exposure variable across the entire pre-reform period (2014–2018). The treatment indicator, denoted as $HighShare30hrs_j$, equals 1 if establishment j's pre-treatment average share of high overtime workers exceeds the median value of all establishments, and 0 otherwise.[15] This research design enables a comparison of establishments with varying initial levels of affected jobs before and after the reform. Importantly, while we rely on a binary treatment indicator in our baseline specification to facilitate interpretation, Section 4.6 shows that our main results are robust to alternative measures of exposure, including continuous pre-reform shares of exposed workers and a measure of the overtime wedge—defined as excess overtime hours relative to the new cap—at the establishment level.

Figures 1–3 provide a descriptive evidence of overtime work, total working hours,

[12] In Appendix TableA1, we report the cross-sectional correlates of overtime hours at the individual level. The provision of overtime hours increases with age, tenure, and is more common among standard workers and individuals employed in large establishments. Holding constant other factors, women supply roughly 6 overtime hours less than men. The likelihood of supplying more than 30 overtime hours per month is 10 percentage points lower among female workers.

[13] Our approach is similar to Carry (2023) and Asai et al. (2024).

[14] To be precise, we exclude the following corporate titles: general or department managers (*buchō*), section managers (*kachō*), supervisors (*kakarichō*) and shop-floor leaders (*shokuchō*).

[15] For the average (median) establishment in the sample, the pre-reform share of workers supplying more than 30 overtime hours is 13% (1%).

and exposure to the overtime cap before and after the Work Style Reform. Panel A of Figure 1 shows a clear compression of the overtime distribution after 2019, with a sharp decline in the right tail above 30 hours and a corresponding increase in the mass at zero and moderate overtime levels. Panel B documents a parallel, though more muted, shift in total monthly hours, indicating that the reform primarily affected overtime rather than base hours. Figure 2 illustrates substantial heterogeneity in pre-reform exposure across establishments, considering both the share of treated workers (Panel A) and excess overtime hours (Panel B), which underpins our difference-in-differences identification strategy. Finally, in Figure 3, we report the differential evolution of overtime hours per worker in treatment and control establishments.[16] Together, these figures suggest that the reform triggered a redistribution of hours away from extreme overtime.

To obtain average post-reform treatment effects, we estimate the following difference-in-differences (DiD) specification:

$$\mathrm{y}_{jt} = \alpha + \beta Post_t + \eta(HighShare30hrs_j \times Post_t) + \mu_j + \psi X_{jt} + \epsilon_{jt} \tag{1}$$

where y_{jt} are the outcomes for establishment j in year t (average overtime per worker, wages, employment), $Post_t$ is a post-reform dummy equals one after 2018 and zero otherwise, $HighShare30hrs_j$ is the above-defined treatment group dummy, while μ_j are establishment fixed effects accounting for time-invariant unobserved attributes. We also control for time-variant establishment-level characteristics, including workforce composition (gender, age, tenure, education) and establishment size. In certain specifications, we additionally account for industry- and region-specific time trends. Coefficient η captures the impact of the reform.

In our analysis, identification relies on the assumption that, absent the reform, establishments with different structural needs for overtime work—despite potentially differing along multiple dimensions—would have followed similar trajectories. Specifically, if the observed effects stem from the new overtime limits introduced by the Work Style Reform, highly exposed establishments should not display differen-

[16]Panel A of Table A2 in the Appendix reports descriptive statistics for the estimation sample. While the two groups are broadly similar in terms of observable characteristics, the treatment group has a higher share of large and manufacturing establishments, reflecting their more intensive use of extreme overtime hours. However, as shown below, the two groups follow broadly similar pre-reform trends. In Section 4.6, we further assess the robustness of our estimates by considering alternative, more homogeneous comparison groups of establishments based on overtime utilization (e.g., medium-high versus high exposure establishments).

tial trends relative to the control group in the pre-reform period. To test for potential pre-trends, we estimate pre-reform effects by plotting year-specific DiD estimates for all outcome variables. These estimates correspond to the interaction between $HighShare30hrs_j$ and a full set of year dummies, with the 2018 coefficient normalized to zero.

4.2 Effects on Working Hours

Average Overtime and Total Working Hours. The primary objective of the reform was to reduce long working hours by capping overtime. A natural starting point is to assess whether the reform achieved its intended first-stage effect: did highly exposed establishments adjust overtime hours as expected in comparison to the control group? To evaluate the reform's impact over time and verify that highly exposed establishments were not on a differential pre-reform trend, Panel (a) of Figure 4 presents year-specific DiD estimates for average monthly overtime hours. Pre-reform interaction coefficients are economically small relative to post-reform estimates, despite some statistical significance.[17] Beginning in 2019, the first year after the reform's implementation, overtime hours declined significantly, with this downward trend continuing into 2020. A slight uptick in overtime hours was observed in subsequent years.[18] Importantly, the timing of the effect, which emerges in 2019, suggests that the change is driven by the reform rather than by differential exposure across establishments to the recessionary effects of the unforeseen COVID-19 pandemic shock.[19]

Panel (B) and (C) of Figure 4 provide estimates for base hours and total monthly working hours, respectively. These event-study graphs reveal that the reduction in overtime hours resulted in a decrease in total working hours, which was only partially offset by an increase in base hours. Table 1 reports the pooled DiD estimates, along with clustered standard errors at the establishment level in parentheses, for the entire post-reform period using Equation (1). Column (1) reports an average reduction of 5 monthly overtime hours per worker in affected establishments relative to the control group—a decrease of approximately 25% compared to the pre-reform mean.

[17] To further assess robustness to potential violations of parallel trends, in Section 4.6, we implement the procedure of Rambachan and Roth (2023), which confirms that our main results remain statistically significant under bounded deviations from parallel pre-trends.

[18] In Section 5.2, we also document a reduction in self-reported paid overtime using individual-level survey data.

[19] The first confirmed case of COVID-19 in Japan was reported on January 15, 2020.

This reduction was partially offset by an increase of 1 base hour (column 2), resulting in a net decrease of 4 total working hours per worker per month (column 3). Relative to the pre-reform average, the decline in total hours is small in magnitude (2%).

Overtime Dispersion. We also investigate whether the reform led to a redistribution of overtime hours within establishments. To capture this, we use the intra-establishment standard deviation of overtime hours as our outcome variable. Panel (D) of Figure 4 presents the event-study graph for this measure of overtime dispersion. The results indicate that overtime hours became less dispersed following the reform. Importantly, pre-trends are minimal in comparison to the post-reform effects, underscoring the impact of the policy. The differential compression of overtime hours in affected establishments relative to the control group is further corroborated by the pooled DiD estimates reported in column (4) of Table 1. Additional analysis in Appendix Table A3 delves deeper into the distributional changes in overtime hours within establishments. We find that the share of workers supplying zero or only a few overtime hours per month increased, while the proportion of high-overtime jobs declined after the reform.

4.3 Effects on Wages

Overtime Pay and Total Monthly Earnings. We now examine how wages evolved in affected establishments compared to the control group. Notably, the Work Style Reform did not mandate firms to compensate high-overtime workers for lost income following the introduction of the overtime cap. Figure 5 presents event-study graphs illustrating changes in various components of workers' compensation packages, including overtime pay, base wages, bonuses, total monthly wages, and hourly wages. These variables are averaged at the establishment level and measured in logs.

Consistent with the reduction in overtime hours noted earlier, Panel (A) of Figure 5 depicts a sharp decline in average overtime pay among affected establishments relative to the control group. As Panel (D) demonstrates, this decline was not offset by increases in other compensation components (base wages, bonuses), resulting in a reduction in total monthly wages.

Table 2 reports the corresponding pooled DiD estimates for the entire post-reform period. Column (2) indicates that overtime pay fell by 0.34 log points (approximately 29%) in affected establishments relative to the control group. Column (3) shows a compensating increase in base wages (1.5%) and bonuses (2.5%), suggesting that firms partially offset workers' lost overtime income, likely reflecting Japan's tight labor market conditions (Kawaguchi, 2019). Column (1) reports a net decline of 1.4% in total monthly wages. Because total working hours fell only slightly more (about 2%), we detect a modest increase in hourly wages (column 5). Unlike the European approach to working-time reductions, which required firms to maintain monthly wages and thereby mechanically induced large increases in hourly wages (Crépon and Kramarz, 2002; Asai et al., 2024), the implementation of an overtime cap under Japan's Work Style Reform did not result in a substantial rise in firms' labor costs.

Within-Establishment Pay Dispersion. We also examine the impact of the reform on within-establishment pay dispersion. Columns (6) and (7) of Table 2 report the DiD estimates for the standard deviation of total monthly wages and overtime pay. Consistent with the observed compression of overtime hours, we find a significant decline in intra-establishment dispersion of overtime pay. By contrast, we detect no differential change in the dispersion of total monthly wages.

4.4 Effects on Employment.

Employment Level. Having documented a reduction in overtime hours and monthly wages, we now analyze the effect on employment. The distinction between standard and nonstandard employment is particularly relevant in the Japanese context. Several studies have highlighted labor market dualization—defined as the coexistence of precarious nonstandard jobs and standard, career-track positions—as a major challenge to Japan's traditional long-term employment system.[20] In our analysis, we adopt the definition of nonstandard jobs used in the BSWS, which is based on workplace titles (*seiki no jyuugyouin*). This definition encompasses the growing prevalence of nonstandard workers with open-ended contracts, offering a more comprehensive measure of the primary "good job" segment and the secondary "bad job" segment of the Japanese labour market (Kambayashi and Kato, 2016; Hijzen et al., 2015).

Figure 6 displays event-study evidence for the log of total, standard, and nonstandard employment. The absence of significant pre-trends differences between the treatment and the control suggests no systematic differences prior to the reform. Ad-

[20] See, for instance, Kawaguchi and Ueno (2013).

ditionally, there is no indication of substantial employment adjustments during the post-reform period. These findings are corroborated by the DiD estimates presented in Table 3. The lack of negative employment effects is not surprising, as employers adjusted monthly salaries downward via overtime pay cuts. As shown in the rest of the table and discussed below, however, there is a notable gender difference in the employment effects.

Workforce Composition. Firms may have responded to the reduction in overtime hours by altering the composition of their workforce. To examine this channel, Table 4 presents DiD estimates for various measures of workforce composition at the establishment level (share of fixed-term, nonstandard workers and college graduates). Relative to the control group, the share of fixed-term and nonstandard jobs decreased by approximately 1 percentage point, representing a 4-5% reduction compared to pre-reform levels (see columns (1) and (2) of Table 4). Column (3) reveals a statistically significant increase in the share of college graduates, suggesting that employers may have responded to the new regulation by enhancing worker quality. However, the magnitude of this effect is relatively small.

4.5 Gender Heterogeneity

We examine establishment-level effects of the overtime cap separately for men and women by replicating the baseline analysis using gender-specific outcomes. The treatment indicator remains the establishment-level exposure defined above.

Figure 7 presents event-study estimates by gender. Panels (A) and (B) show that overtime hours and overtime pay decline for both male and female workers. Turning to employment, Panel (E) indicates no significant changes in male employment, while Panel (F) shows some evidence of an increase in female employment. However, female employment in affected establishments was already on an upward trend prior to the reform, particularly in standard jobs. While both male and female standard employment exhibited pre-reform increases, these trends diverge after the reform: the upward trend continues for women but flattens for men. In contrast, Panel (G) shows a significant decline in nonstandard ("bad") female jobs, with no comparable change for men. These employment effects are further confirmed by the pooled DiD estimates reported in Table 3. Taken together, these patterns suggest a gender-specific reallocation of employment. The decline in nonstandard female employment is not preceded by comparable pre-trends and is therefore more plausibly linked to the reform, while the increase in standard female employment partly reflects a continuation of pre-existing trends.

We next examine how the reform affected the composition of employment within establishments. We find that the share of women among standard jobs increases in more exposed establishments. As shown in Panel A of Figure 8, pre-reform coefficients are slightly negative but stable over time, with no clear evidence of differential pre-trends. The post-reform shift toward positive coefficients therefore reflects a break coinciding with the reform. Pooled DiD estimates reported in column (1) of Table 5 indicates that the share of female workers among standard jobs raises by 1 p.p over the post-reform period. This is suggestive that the reform is associated with improved relative access of women to standard jobs.

To shed light on the mechanism underlying these compositional changes, we examine how the reform affects the relative intensity of working time across job types and genders. We construct establishment-level overtime gaps defined as the difference between average overtime hours in standard and nonstandard jobs, and compute analogous measures separately by gender as well as gender differences within job types. Event studies and pooled DiD estimates are reported in Panels B-D of Figure 8 and columns (2)-(6) of Table 5, respectively. We find a significant compression in these gaps following the reform. The overall standard–nonstandard overtime gap declines, driven primarily by a larger reduction among men (-1.68 hours) than among women (-0.77 hours). Consistent with this pattern, the gender gap in overtime within standard jobs narrows substantially (-1.98 hours), while the corresponding gap in nonstandard jobs declines more modestly (-0.87 hours). Event-study estimates show no evidence of differential pre-trends in these gap measures, supporting a causal interpretation of this compression. These results indicate that the reform disproportionately reduced the long-hours intensity of standard jobs for male workers.

Within the conceptual framework outlined in Section 2 and fully developed in Appendix A, this compression implies a reduction in the hours requirement of standard jobs, thereby narrowing the effort–cost gap between standard and nonstandard positions and lowering the assignment threshold θ, which represents workers' tolerance for long hours, for access to standard jobs. By reducing male workers' relative advantage in supplying long hours, the reform may have contributed to the observed reallocation of women toward standard, career-track jobs.

4.6 Robustness and Additional Analysis

Parallel trends. Our strategy assumes that, in the absence of the WSR, establishments differently exposed to the new regulation would have evolved similarly. For all our outcomes, we report year-specific DiD estimates for the pre-reform period (2014-2018). Pre-trends are not significant or relatively small in magnitude compared to the post-reform estimates. To further assess the validity of our approach, we apply a recent procedure proposed by Rambachan and Roth (2023) and find that the significant reduction in overtime hours and overtime pay remain robust to potential violations of the parallel trends assumption prior to the policy (see Appendix Figure A2).

Staggered treatment timing. We examine the robustness of our main results to staggered treatment timing resulting from the presence of late-treated units. In our setting, the new regulation came into force in April 2019 for large establishments employing 300+ employees, while April 2020 was the starting application date for establishments employing not more than 300 employees (not more than 50 employees for retail businesses, not more than 100 for wholesale retail).[21] Following Callaway and Sant'Anna (2021), we first estimate the individual cohort-time-specific treatment effects, allowing for treatment effect heterogeneity, and then aggregate these individual treatment effects to obtain overall treatment effects. The event study for overtime hours, presented in Appendix Figure A3, yields conclusions consistent with our baseline estimates.

Placebo Outcome: Managers' Overtime. Japanese managers are exempt from the overtime cap and have historically been excluded from working-time limits under discretionary work systems. We conduct a placebo test using mean managerial overtime hours as an outcome (see Figure A4). Consistent with the reform's institutional design, we find only a modest reduction in managerial overtime when considering either top managerial titles (buchō, kachō) or a more comprehensive measure that also includes other mid-level managers (kakarichō, shokuchō) (approximately -0.4 and -1.5 hours

[21] Indeed, previous attempts to study the initial impact of the Work Style Reform find a significant discontinuous reduction in extreme overtime hours around the firm-size threshold in 2019, consistent with the one-year grace period granted to small and medium-sized companies (Toda, 2023; Jiang and Sasaki, 2025).

per month, respectively), which is substantially smaller than the effect observed for rank-and-file workers (-5 hours).[22] The placebo test suggests that the reductions documented among non-managerial employees are not driven by unobserved aggregate shocks or simultaneous trends affecting all workers equally.

Alternative treatment indicators. We evaluate the robustness of our findings using four alternative definitions of the treatment group. First, we use a stricter threshold, considering establishments with a positive share of workers exceeding 45 overtime hours per month. Second, we compare firms in the top quartile of the pre-reform distribution of the share of high-overtime employees (*high exposure*) to firms whose pre-reform share of high-overtime employees is between the 50th and the 75th percentile (*medium-high exposure*). Third, we use a continuous-intensity DiD model, where pre-reform exposure is directly measured as the share of workers supplying overtime hours above the 30-hour cap. Finally, we use an alternative continuous indicator, where pre-reform exposure is measured as the fraction of firm-level overtime hours exceeding the 30-hour cap (for a similar indicator see e.g. Asai et al. (2024)). As shown in Table A4, our main findings remain robust across these alternative specifications.

Reclassification of Overtime Hours. Does the documented reduction in overtime hours merely reflect reporting changes? A potential concern is that firms respond to the overtime cap by relabeling working time rather than by reducing actual hours. Three pieces of evidence suggest that this is not the primary channel in our setting. First, although paid overtime declines and base hours increase modestly, total monthly working hours fall significantly at the establishment level, which rules out pure relabeling of hours. Second, we find no post-reform increase in the share of employees classified as managers, indicating that firms do not systematically avoid the cap through job-title reclassification into exempt categories.[23] Third, as shown below, while treated men display some substitution from paid to unpaid off-the-book overtime, this increase offsets only part of the decline in paid overtime, so that total

[22] This small spillover on mid-level managers (usually supervisors and shop-floor leaders) may reflect establishment-level adjustments due to on-site working-time complementarities—such as rescheduled meetings or tasks requiring the simultaneous presence of subordinates and managers—rather than direct legal exposure.

[23] If anything, as shown in Figure A5, there is a slight reduction in the share of managers at the establishment level.

overtime still falls substantially. Taken together, these results imply that the reform led to genuine reductions in working time, with only limited scope for reclassification at the margin.

Balanced panel. Finally, we compare our baseline estimates using the unbalanced panel of establishments with a restricted sample of establishments that remain in the BSWS for the entire 2014–2022 period. This is a demanding check, as the estimates rely on a small subsample of 1,356 establishments—roughly 3% of the original sample. Despite the substantial drop in sample size, our main results qualitatively hold (see Fig. A6). We continue to observe a significant, though smaller, reduction in overtime per worker and intra-establishment overtime dispersion, as well as no meaningful changes in hourly wages or total employment. However, we do find evidence of a differential expansion in female employment within incumbent establishments as well. These patterns suggest that our findings are not driven by establishment turnover.

Worker Voice Institutions. A well-established set of institutions facilitates labor-management communication in Japanese firms, including Joint Labor-Management Committees (JLMCs), shop-floor committees (SFCs), and unions. Notably, the WSR allows firms to exceed the new overtime limits under special circumstances, provided they negotiate a collective agreement with their workers. Unfortunately, payroll data do not contain information on the extent of labor-management communication. Instead, we use the Survey on Labor Management Communication (SLMC), outlined in Section 3, to calculate the prevalence of worker voice institutions at the industry level and merge this data with our establishment-level payroll information. We then conduct DiD estimates to explore heterogeneous effects based on whether an establishment operates in an industry with a high or low incidence of worker voice institutions. Figure A1 presents event-study graphs of our main outcome variables, distinguishing establishments in industries with above- and below-median prevalence of worker voice institutions. This pattern is consistent with worker-voice channels helping to protect employees who prefer high pay and long hours, potentially preventing a shift from paid to unpaid overtime.[24]

[24]This pattern is also consistent with the rat-race/adverse-selection model of long working hours in Frederiksen et al. (2024). In their framework, promotions depend on sustained long hours, but because only workers observe their true disutility of effort, some high-disutility workers may overwork to mimic low-disutility types. Firms therefore set inefficiently high hour thresholds to deter such behav-

5 Individual-Level Impacts

In this section, we extend the establishment-level analysis by exploring the impact of the 2018 Work Style Reform at the individual level. To do so, we use longitudinal OPPS survey data, as described in Section 3. Specifically, we leverage detailed information on self-reported working hours, including paid and unpaid overtime, perceived effort intensity, time-use indicators, and subjective well-being (SWB).

Our analysis focuses on full-time, non-managerial employees aged 20–65, in line with the establishment-level analysis. We include data from five pre-reform waves (2012, 2013, 2016, 2017, and 2018) and two post-reform waves (2021 and 2022). Panel (B) of Table A2 in Appendix summarizes descriptive statistics for the estimation sample.[25]

5.1 Individual-Level Difference-in-Differences

To examine the impact of the new overtime cap at the worker level, we classify individuals into a treatment group based on whether their paid overtime hours exceeded the new overtime cap (30 overtime hours per month) in at least one year during the pre-reform period.[26] The control group consists of full-time workers who never worked more overtime hours than the new limit in the pre-reform period.

We estimate the following difference-in-differences specification:

$$y_{it} = \alpha + \beta Post_t + \eta(HighOvertime_i \times Post_t) + \psi X_{it} + \mu_i + \epsilon_{it} \tag{2}$$

where y_{it} are the outcomes for individual i in year t (e.g. long working hours, overtime hours, SWB), $Post_t$ is a post-reform dummy equals one from 2019 onward and zero otherwise, $HighOvertime_i$ is the above-defined treatment group dummy, while μ_i

ior. Worker voice institutions improve information flows, reduce adverse selection, and thereby weaken the need for inflated hour thresholds. As a result, the WSR is likely to be less binding in establishments with stronger worker voice, leading to smaller effects on overtime hours and pay.

[25] As noted in Section 2, the reform's enforcement was staggered, impacting large firms from April 2019 and small and medium-sized firms from April 2020. However, by 2021 (the first year with post-reform data available), the legislation applied to firms of all sizes. Our results remain robust if we restrict the analysis to individuals employed in large firms before the reform, as shown in Appendix Panel (E) of Table A6.

[26] One concern is that this definition may capture occasional overtime spikes rather than persistent exposure to high overtime. As a robustness check, we therefore define treatment as average paid overtime during the pre-reform period exceeding the new cap of 30 monthly overtime hours. As shown in Table A5, the results for paid and unpaid overtime and subjective well-being (SWB), by gender, remain unchanged.

are individual fixed effects. We also control for time-variant personal and firm-level characteristics (age, tenure, occupation and employer size). Coefficient η captures the impact of the reform. We estimate equation (2) by OLS, clustering standard errors at the individual level in order to account for serial correlation.

5.2 Self-Reported Working Time: Paid and Unpaid Overtime

First, we verify whether the reduction in overtime hours observed in the establishment-level payroll data also holds for self-reported working hours from individual survey data.[27] The results are presented in Table 6. Panel (A) reports unconditional DiD estimates, while Panel (B) includes controls for individual- and firm-level attributes, along with region-specific time trends to account for time-varying shocks. In Panel (C), we include individual fixed effects to control for unobserved time-invariant characteristics.

Our individual-level DiD estimates align with the results reported in Section 4, showing a reduction in self-reported long working hours and total overtime hours. The preferred estimates in Panel (C), columns (2) and (3), indicate a 7 percentage point reduction in the incidence of long hours (i.e., individuals working more than 60 hours per week) and a reduction of 2.8 total overtime hours per week among treated individuals compared to the control group. For completeness, column (4) of Table 6 (Panel C) reports additional DiD estimates using the log of self-reported monthly wages as the dependent variable. There is a reduction in monthly wages, albeit the effect is imprecisely estimated.[28]

In Table 7, we examine changes in self-reported overtime hours in greater detail by distinguishing between paid and unpaid overtime. Columns (1)–(3) report estimates for the full sample of full-time workers, as in Table 6. Columns (4)–(6) restrict the

[27] Working hours are self-reported and may contain measurement error, although previous research suggests that self-reported hours in Japan are highly reliable (Imai et al., 2016). We further compare our data with national statistics and find similar levels of unpaid overtime (4.1 hours in official data vs. 3.8 hours in our sample). We also compare the sum of standard weekly hours (40 hours) and reported overtime with individuals' reported total usual weekly hours in OPPS. Figure A7 in the Appendix shows the distribution of the differences for treated and control groups. The modal value of zero suggests generally accurate reporting, with symmetric inconsistencies around zero.

[28] As mentioned in Section 2, the WSR mandates that employers provide workers with at least five days of annual paid leave. It is important to note that changes in the number of days worked may not necessarily occur within the segment of the overtime hours distribution we are analyzing. In fact, column (5) of Table 6 presents no evidence of differential changes in annual days worked. Furthermore, we do not observe significant changes in the average number of days worked in affected establishments relative to the control group (results available upon request).

analysis to individuals reporting positive total overtime hours, and column (7) reports estimates for the unpaid-to-total overtime ratio, which by construction is defined only for this latter group.

Interestingly, as shown in column (1)-(3) of Panel C, Table 7, the decrease in total overtime is driven by a reduction in paid overtime (3.6 hours per week), partly offset by an increase in unpaid overtime (approximately 0.6 hours per week), although the latter effect is imprecisely estimated. Paid overtime decreased by 45% relative to the pre-reform mean.[29] The reduction in paid overtime holds for both men and women, although a statistically significant increase in unpaid overtime of about 1 hour per week is observed only among male workers.[30] As shown in columns (6)–(7) of Table 7, the rise in unpaid overtime is larger in magnitude (1.6 hours) and highly significant among male workers reporting positive overtime hours, translating into a 10 percentage-point increase in the ratio of unpaid to total overtime relative to the control group.

Reconciling establishment- and survey-based overtime effects. While the results from our establishment-level and individual-level survey-based DiD estimates are qualitatively similar, the magnitudes differ at face value. Survey respondents report usual working hours in a typical week, whereas establishments report payroll-based measures of actual working hours in a specific month. Moreover, the survey estimates capture treatment effects among directly exposed individuals, while the establishment estimates represent intention-to-treat effects that are mechanically diluted by the large share of workers with low or zero overtime.

Intuitively, a given reduction in average overtime at the establishment level can correspond to a substantially larger reduction among the subset of workers whose pre-reform overtime placed them above the cap. This scaling provides a benchmark for comparing establishment-level reductions with survey estimates that focus on individuals with high baseline overtime exposure. In the survey data, treated work-

[29]The reduction in self-reported paid overtime is robust to several checks. First, while our preferred estimates include individual fixed effects, we address concerns about worker sorting around the reform by estimating DiD models on a subsample of job stayers, defined as individuals with at least five years of tenure with their current employer as of 2021 (the first observed post-reform year). Second, we exclude individuals with inconsistent working-time reports and trim the top 2% of hours. Finally, we re-estimate the DiD models using a balanced panel of individuals. See Table A6 in the Appendix.

[30]Figure 11 shows event-study plots comparing paid and unpaid overtime for male and female workers.

ers—those with more than 30 hours of paid overtime before the reform—reduce paid overtime by about 3.6 hours per week (see Panel C, column 4 of Table 6), or roughly 14 hours per month. In the establishment-level payroll data, average monthly overtime per worker in high-exposure firms falls by about 5 hours.

Formally, let h_{ijt} denote monthly paid overtime hours of worker i in establishment j at time t. Partition workers into two groups based on *pre-reform* exposure to the cap: treated (T), with $h_{ij,pre} > \bar{h}$ (30 overtime hours per month), and untreated (U). Let s_T and s_U denote their pre-reform employment shares within high-exposure establishments, with $s_T + s_U = 1$, and let Δh_T and Δh_U denote reform-induced changes in mean overtime for the two groups. The establishment-level DiD effect on average overtime can be written as:

$$\eta^{EST} \;=\; s_T \Delta h_T + s_U \Delta h_U. \tag{3}$$

Our payroll-based estimate implies $\eta^{EST} \approx -5$ hours per month, and in high-exposure establishments the pre-reform share of treated workers is $s_T \approx 0.29$.

As a benchmark, if the reform primarily reduces overtime among treated workers and has little effect on untreated workers (i.e., $\Delta h_U \approx 0$), equation (3) implies a treated-group effect of about -17.2 hours per month (approximately -4.0 hours per week).[31] This is close to the survey estimate of a -3.6 hour/week reduction among treated individuals. While this approximation abstracts from spillovers toward moderate overtime among untreated workers—which would mechanically attenuate the establishment-level effect—as well as from differences in measurement and timing across data sources, both datasets consistently point to a reduction of roughly 3–4 hours of paid overtime per week for workers most exposed to the cap.

5.3 Effects on Workers' Subjective Well-Being

After documenting a consistently negative first-stage effect on both payroll and self-reported working hours, we now turn to the impact of the reform on workers' subjective well-being (SWB). Specifically, we focus on measures of cognitive well-being, including life satisfaction and satisfaction with key life domains (work, spouse, leisure, and family). Due to the lack of survey data for 2019-2020, our analysis primarily captures medium-term responses to the reform. It is possible that SWB improved in the

[31] Under $\Delta h_U \approx 0$, equation (3) implies $\Delta h_T \approx \eta^{EST}/s_T = -5/0.29 \approx -17.2$ hours per month. Converting to weekly units using 4.33 weeks per month yields $-17.2/4.33 \approx -4.0$ hours per week.

short run following the Work Style Reform, only to return to baseline levels over time due to hedonic adaptation.[32] Moreover, while anticipatory responses could be a concern in our context, such responses would likely make it more difficult to detect any genuine impact of the reform. Importantly, we find no evidence of differences in SWB during the pre-reform years (see Figure 9).

Results are presented in Table 8. The dependent variables are measured on Likert scales ranging from 1 to 5, with higher values indicating greater satisfaction. In Panel (C), we report results from our preferred DiD specifications, which include individual fixed effects. Overall, we find that the reduction in overtime hours did not lead to significant changes in subjective well-being (SWB).[33]

However, these findings mask some heterogeneity by gender. In panel (D) of Table 8), we observe a positive and statistically significant effect on both life satisfaction and leisure satisfaction for female workers. Specifically, the reform increases life satisfaction by 0.25 points (mean: 3.59) and leisure satisfaction by 0.38 points (mean: 3.35) on 1–5 scales. These gains correspond to about 7% and 11% of the respective sample means, or roughly 0.32 and 0.43 standard deviations, respectively. To provide an additional sense of scale, the increase in life satisfaction (leisure satisfaction) induced by the reform is about 34% (47%) of the cross-sectional difference in life satisfaction (leisure satisfaction) between women who report excellent versus poor health in our data.

5.4 Robustness Checks and Additional Results

Parallel Trends. Similar to our analysis using establishment data, we report dynamic DiD estimates to examine whether the outcomes for treated and control individuals were on parallel trends before the reform. Figure 9 presents event-study graphs for our main individual-level outcomes—self-reported overtime hours, and measures of subjective well-being—across the years surrounding the reform. Each estimated coefficient corresponds to the interaction between T_i and a full set of year dummies, with the coefficient for 2018 normalized to zero. We find no evidence of differential

[32]There is no clear evidence of individual adaptation to reductions in working hours in other contexts (Lepinteur, 2019). In light of the missing data for 2019–2020, however, we address the possible confounding effects of COVID-19 in Section 6.4.

[33]Consistent with the lack of significant effects on self-reported job satisfaction, we observe no differences in the likelihood of seeking a new job, which is typically interpreted as a revealed-preference measure of job (dis)satisfaction (see Appendix Figure A8).

pre-reform trends for any of our main variables. The reduction in overtime hours for treated individuals becomes statistically significant in 2021 and 2022. Importantly, this result holds for both unmatched and matched DiD estimates using non-parametric coarsened exact matching (Iacus et al., 2012).[34] In Figure 10, we display the results by gender and find no evidence of pre-reform trends for either male or female workers. The figures also confirm the differential increase in life and leisure satisfaction among women, as discussed earlier.

Covid-19: Individuals' Ability to Work from Home. Our establishment-level DiD estimates show that the reduction in overtime hours is already evident in 2019, suggesting that the Covid-19 pandemic is unlikely to confound our results. However, the first post-reform survey wave used in our individual-level analysis is from 2021. While the pandemic itself should not affect our estimates unless it impacted treated and control individuals differently, we recognize that differences in the ability to work from home could influence both subjective well-being and the reporting of working hours. To address this concern, we conduct additional DiD estimates excluding individuals who were teleworking at least one day per week just prior to the pandemic (January 2020). As reported in Table A7, our main results remain robust to this sample restriction.

Other Well-Being Facets. We also examine potential impacts on other dimensions of well-being, beyond cognitive well-being. In Appendix Table A8, we report DiD estimates analyzing differential effects on a proxy of eudaimonic well-being (e.g., "My life is fulfilling") and affective well-being (measured by happiness on a 1-10 scale). We also examine other subjective indicators, such as health anxiety, feelings of stress, depression, and sleep problems. Consistent with our findings for cognitive well-being, we observe positive effects only among treated women. Specifically, for this subsample, we document a significant increase of 0.44 points in happiness following the reform, corresponding to roughly 7% of the sample mean.

Perceived Effort Intensity. Employers may have responded to the new overtime cap by implementing strategies to extract more effort from workers within the reduced

[34] Specifically, we first match individuals using pre-reform characteristics measured in 2018 (such as gender, age, and firm size), and the matching weights are then used to estimate the DiD model.

overtime hours. Work intensification could dampen the reform's impact on workforce well-being, while allowing firms to offset the reduction in overtime. Specifically, gender differences in work intensification may explain why the reform had heterogeneous impacts on subjective well-being between female and male workers. Anecdotal evidence suggests that firms may have placed additional pressure on workers to maintain output despite shorter overtime. In fact, the term *jitan-harasumento* (reduced-hours harassment) was nominated for the 2018 Buzzword Award in Japan (Japan Times, 2018), becoming popular around the time of the Work Style Reform. The survey provides data on individuals' perceived work intensity, asking them how hard they work each day. Using our DiD framework, we examine whether these perceptions changed differentially around the time of the reform. Our outcome variable is a dummy that takes the value of one if the individual responds with either "Work hard and continuously" or "Could not work any harder than currently." We also allow for heterogeneous treatment effects by gender. As shown in column (1) of Table A9, there is no evidence of gender differences in perceived work intensity among treated individuals relative to the control group.

Use of Freed Time. By capping overtime hours, the reform may have led to changes in time use and a reorganization of the workweek, potentially freeing up time for other activities. Changes in time use may be different for men and women, potentially explaining why the reform has gender-specific effects on SWB, as documented above. The survey provides information on commuting time, housework time, and physical activity frequency. While weekly commuting time is similar for treated men and women (roughly 3 hours), weekly time devoted to home production is three times higher for women than for men in our sample.[35] In columns (2)-(4) of Table A9, we report DiD estimates with gender-specific effects, finding no significant changes in time use as a result of the overtime cap. [36]

[35]Weekly housework is constructed using reported hours of housework on a typical weekday (HW_{wd}) and weekend day (HW_{we}). After trimming values above the 99th percentile, weekly totals are computed as $HW_{\text{week}} = 5 \times HW_{wd} + 2 \times HW_{we}$. Weekly commuting is based on hours of commuting on a typical workday (C_{wd}). Following the same trimming procedure, weekly commuting time is calculated as $C_{\text{week}} = 5 \times C_{wd}$.

[36]Dynamic DiD estimates presented in Appendix Figure A8 further confirm the absence of differences in both commuting and housework time over the entire study period.

6 Cross-Spousal Effects and Household Complementarities

In the previous sections, we document four key findings. First, the WSR substantially reduced paid overtime for both men and women. Second, despite these broad reductions in working hours, improvements in subjective well-being (SWB) are concentrated among women. Third, men partially offset reductions in paid overtime by increasing unpaid overtime, which helps explain the absence of well-being gains for male workers. Finally, the reform induced a shift in the composition of female employment toward relatively fewer nonstandard poor-quality jobs. Until now, however, we have focused on the impact of the reform on respondents' own working hours and subjective well-being, abstracting from household-level mechanisms. In the absence of household data—a common limitation in studies of working-time regulation—the estimated effects necessarily aggregate across heterogeneous household types, implicitly assuming that utility is separable across household members.

In this section, we examine whether the gendered well-being effects documented above are partly shaped by household-level spillovers, rather than solely by individuals' own labor-supply adjustments. In particular, we focus on married respondents, for whom changes in one spouse's working conditions may affect the other spouse's well-being even if their own working hours remain unchanged. To do this, we use information from the OPPS survey distinguishing between single and married (or cohabiting) respondents and providing detailed information on spouses' working hours.

Single vs. Married Respondents. We begin by estimating difference-in-differences models separately for single and married respondents. Estimates are reported in Table A10. Among single workers, the reform reduces paid overtime for both men and women. While we do not detect statistically significant changes in overall life satisfaction for singles, we find a significant increase in leisure satisfaction for single women. Among married workers, reductions in paid overtime are concentrated among men, whereas women experience significant improvements in life satisfaction despite no change in their paid overtime. This contrast suggests that, for married women, well-being gains may operate through channels other than own working-time reductions.[37]

[37] It is worth noting that married women who supply very long hours prior to the reform may constitute a highly selected group with strong career orientation. Long hours play a major signaling role

This possibility motivates a detailed examination of heterogeneity across household exposure types, depending on whether the spouse was also treated.

Household Exposure Types and Empirical Design. To investigate whether household-level exposure helps explain these patterns, we exploit information on spouses' pre-reform overtime, as reported by the main respondent, and classify couples according to whether neither, one, or both partners were exposed to the reform. This allows us to distinguish between direct effects of an individual's own exposure and potential spillovers arising from a partner's exposure. As mentioned, estimates for married individuals reported in Table A10 pool all of these categories, which may average out heterogeneous responses. Table A11 documents the empirical relevance of this heterogeneity: a large share of couples are discordant in treatment exposure (e.g., 72% of female respondents with treated husbands are not treated themselves).

To isolate spousal spillovers and joint-exposure effects among dual-earner households, we estimate the following difference-in-differences model for cohabiting respondents:

$$\begin{aligned} Y_{it} = \alpha + \beta\, Post_t + \lambda_{10}(H_{10,h} \times Post_t) + \lambda_{01}(H_{01,h} \times Post_t) \\ + \lambda_{11}(H_{11,h} \times Post_t) + \psi X_{it} + \gamma_i + \varepsilon_{it}, \end{aligned} \tag{4}$$

where Y_{it} denotes the outcome (paid overtime, SWB) of main respondent i at time t and γ_i are individual fixed effects. $H_{10,h}$ is an indicator for (discordant) couples where only the respondent was treated before the reform, $H_{01,h}$ indicates (discordant) couples where only the partner was treated, and $H_{11,h}$ indicates (concordant) couples where both were treated. Households where neither partner was treated ($H_{00,h}$) serve as the omitted reference group.[38] We test the additivity hypothesis, $H_0 : \lambda_{11} = \lambda_{10} + \lambda_{01}$, which states that the joint effect of both partners being treated is equal to the sum of the individual treatment effects. Rejecting H_0 implies non-linearities: complementarities if $\lambda_{11} > \lambda_{10} + \lambda_{01}$, or substitutabilities if $\lambda_{11} < \lambda_{10} + \lambda_{01}$. Importantly, because household treatment status generates small cells we interpret these results cautiously.

Results. Estimates from our household-level DiD model are reported in Table A12. For male respondents, the reform reduces their paid overtime when they alone are

for this group (Kato et al., 2016). For these workers, reducing overtime—even when a legal cap is introduced—may not be a viable individual margin, as doing so could risk negative career consequences.

[38] The group of married respondents with non-working spouses is excluded from the cross-spousal analysis, since spouses who do not work have no defined overtime exposure.

treated (λ_{10}), with a reduction of roughly 4 hours. There is also a significant spillover effect when only the spouse is treated (λ_{01}), with a reduction of about 2.5 hours. When both spouses are treated, the estimated reduction in men's overtime ($\lambda_{11} \approx -5.7$) is statistically indistinguishable from the sum of the individual and spousal effects. Therefore, we do not reject additivity ($p = 0.76$). Consistent with earlier results, no treatment cell yields significant effects on men's SWB. For female respondents, the pattern differs markedly. When only the respondent is treated, overtime falls by about 1.5 hours (λ_{10}). In contrast, we find no evidence of overtime spillovers from husbands' treatment status (λ_{01}).[39] Importantly, we do not detect statistically significant changes in women's own overtime hours in dual-treated couples. However, the estimates are imprecise and cannot rule out moderate reductions in own hours.

Crucially, when both partners are treated, we observe consistent and statistically significant improvements in women's life satisfaction and leisure satisfaction ($\lambda_{11} > 0$). We reject additivity for leisure satisfaction, suggesting complementarities within dual-earner households.[40] [41] These patterns motivate an exploration of potential household-level mechanisms.[42]

Household-level mechanisms. Do SWB gains among married women in both-treated couples (H_{11}) relate to household-level channels? As a first step, we examine whether women report a reduction in their husbands' overtime hours. As shown in Column 1 of Table A14, a decline in spousal overtime is observed only among married women with treated husbands (a statistically significant reduction of approximatively

[39]The asymmetric cross-spousal labor-supply responses we document—namely, the absence of overtime spillovers for women with treated husbands and the presence of positive spillovers for men with treated wives—are consistent with previous evidence on household-level adjustments to the mandatory workweek reduction in France (Goux et al., 2014).

[40]Formally, $D = \lambda_{11} - (\lambda_{10} + \lambda_{01})$ is positive (0.76) and significant in a one-sided test ($p = 0.026$).

[41]Figure A9 plots event-study estimates for married women across the four household treatment cells, normalized to the last pre-reform OPPS wave (2018). The coefficients show no evidence of differential pre-trends across household treatment groups prior to the reform. In the post-reform period, positive and statistically significant effects on both life and leisure satisfaction emerge only for women in dual-treated couples, with the impact concentrated in the 2021 wave, the first post-reform survey. Women in households with only one treated spouse show no comparable improvements.

[42]One concern is that the heterogeneous responses documented above may reflect marriage itself rather than spouses' exposure to the overtime cap. As a robustness check, we also compare the reform's own effect for single respondents with that for married respondents with never-treated partners, i.e., non-working spouses and working but permanently below-cap spouses. For this restricted sample, we estimate a triple-difference specification allowing the post–treatment interaction to differ between singles and married individuals. The interaction term is insignificant. This suggests that the main household interactions uncovered in the four-cell design arise primarily through spousal exposure rather than through the mere presence of a partner (see Table A13).

5 hours in dual treated couples and close to 2 hours in couples with only her husband treated). Then, we examine potential changes in time availability at the individual level (Panel A), including weekly commuting and housework time. Results are reported in columns (2)–(5) of Table A14. As shown in column (2), female respondents experience modest reductions in commuting time, but these changes are not specific to jointly treated couples. In column (3), we show that partners' (husbands) commuting time does not change. Finally, in columns (4) and (5), we document that neither respondents nor spouses adjust their housework time differently across treatment cells. This rules out the possibility that SWB gains for married women in both-treated couples result from straightforward increases in discretionary time.

Another potential mechanism underlying the SWB gains experienced by this group operates through improvements in household-level functioning associated with a reduction in partners' work-related strain. To investigate this channel, we examine respondents' perceptions of their partners' work intensity. In column (6) of Table A14, we document a differential reduction in perceived husband's work intensity in both-treated households. This pattern appears consistent with the interpretation that improvements in partner's working conditions resulting from reduced overtime may improve household functioning and, in turn, women's well-being. Consistent with these findings, column (7) of Table A14 shows that, among married women in dual-treated households, spouse satisfaction (i.e., satisfaction with their husbands) increases significantly.

Taken together, the results suggest a role for household spillovers. In particular, women in dual-treated couples experience substantial reductions in their partners' overtime hours, alongside improvements in well-being. Although the estimates are too noisy to fully rule out own labor-supply responses among married women, the magnitude of spousal adjustments and the pattern of results across household types point to a prominent role for cross-partner effects. Given the relatively small number of observations in some household treatment categories, we view the household-level analysis as complementary evidence that helps reconcile the gendered well-being effects documented earlier, rather than as a definitive test of household spillover mechanisms.

7 Conclusions

This study provides the first causal estimates of the effects of Japan's Work Style Reform (WSR) on firms and workers. We show that overtime hours in more exposed establishments fell by roughly 5 hours per month (25%), and that the distribution of hours within firms became markedly less dispersed. Monthly earnings declined because the reduction in overtime pay was not fully offset by increases in base wages and bonuses. At the same time, the reform appears to have shifted the composition of female employment toward career-track (standard) positions and away from nonstandard jobs, consistent with the reform narrowing the overtime hours requirements between standard and nonstandard positions.

Using individual worker-level survey data, we corroborate these findings with direct evidence of reduced self-reported long workweeks and overtime. We also find that the reform improved life and leisure satisfaction, but only for women. These gender differences are not explained by perceived work intensification or observable time-use changes. Rather, among men, reductions in paid overtime were partly offset by an increase in unpaid overtime, consistent with the absence of well-being gains. Finally, our household-level analysis reveals that women's well-being improves only in couples where both spouses were treated, highlighting the importance of intra-household channels in shaping welfare responses to working-time regulation.

While Japan's Work Style Reform provides a unique setting to study the consequences of limiting long hours, several limitations point to promising directions for future research. First, our data do not capture firms' internal promotion processes or workers' long-run career trajectories, which are important to understanding how reforms affect incentives embedded in "greedy jobs." In particular, assessing whether the observed shift of women into standard employment translates into greater representation in managerial roles will require longer follow-up data. Second, future work could examine objective health outcomes—such as sickness-related absences or work injuries—to complement our subjective well-being measures. Third, although our household analysis uncovers important mechanisms, we lack detailed information on household bargaining and task allocation, which richer couple-level time-use data could illuminate. Finally, the long-run effects of sustained limits on overwork—particularly whether firms adjust production technologies or career systems—remain an important open question.

Overall, our results align with a growing literature emphasizing the high returns to long and inflexible hours as a key structural barrier to gender equality in labor markets. They suggest that well-designed working-time regulations can relax these constraints and support women's access to high-quality jobs by weakening the link between standard jobs and long working hours and, hence, reducing the comparative advantage of men in supplying long hours.

References

Akerlof, G. (1976). The economics of caste and of the rat race and other woeful tales. *The Quarterly Journal of Economics*, 90(4):599–617.

Asai, K., Lopes, M. C., and Tondini, A. (2024). Firm-Level Effects of Reductions in Working Hours. CEPREMAP Working Papers (Docweb) 2405, CEPREMAP.

Bassanini, A. and Caroli, E. (2015). Is work bad for health? the role of constraint versus choice. *Annals of Economics and Statistics*, (119/120):13–37.

Batut, C., Garnero, A., and Tondini, A. (2023). The employment effects of working time reductions: Sector-level evidence from european reforms. *Industrial Relations: A Journal of Economy and Society*, 62(3):217–232.

Bell, D. N. F. and Hart, R. A. (1999). Unpaid work. *Economica*, 66(262):271–290.

Benjamin, D., Debnam Guzman, J., Fleurbaey, M., Heffetz, O., and Kimball, M. (2023). What do happiness data mean? theory and survey evidence. *Journal of the European Economic Association*, 21.

Benjamin, D. J., Heffetz, O., Kimball, M. S., and Szembrot, N. (2014). Beyond happiness and satisfaction: Toward well-being indices based on stated preference. *American Economic Review*, 104(9):2698–2735.

Berniell, I. and Bietenbeck, J. (2020). The effect of working hours on health. *Economics Human Biology*, 39:100901.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–55.

Borga, L., D'Ambrosio, C., and Lepinteur, A. (2022). *Economic Perspectives on Individual Well-being*, pages 45–61. Springer Fachmedien Wiesbaden, Wiesbaden.

Cahuc, P. and Carcillo, S. (2014). The detaxation of overtime hours: Lessons from the french experiment. *Journal of Labor Economics*, 32(2):361–400.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Carcillo, S., Hijzen, A., and Thewissen, S. (2023). The limitations of overtime limits to reduce long working hours: Evidence from the 2018 to 2021 working time reform in korea. *British Journal of Industrial Relations*, n/a(n/a).

Carry, P. (2023). The effects of the legal minimum working time on workers, firms and the labor market. Technical report.

Cha, Y. and Weeden, K. A. (2014). Overwork and the slow convergence in the gender gap in wages. *American Sociological Review*, 79(3):457–484.

Clark, A. E. (2016). Swb as a measure of individual well-being. In Adler, M. and Fleurbaey, M., editors, *The Oxford Handbook of Well-Being and Public Policy*, pages 518–552. Oxford University Press, USA.

Cortes, P. and Pan, J. (2017). Cross-Country Evidence on the Relationship between Overwork and Skilled Women's Job Choices. *American Economic Review*, 107(5):105–109.

Crépon, B. and Kramarz, F. (2002). Employed 40 hours or not employed 39: Lessons from the 1982 mandatory reduction of the workweek. *Journal of Political Economy*, 110(6):1355–1389.

Cygan-Rehm, K. and Wunder, C. (2018). Do working hours affect health? evidence from statutory workweek regulations in germany. *Labour Economics*, 53:162–171.

Frederiksen, A., Kato, T., and Smith, N. (2024). Working hours, top management appointments, and gender: Evidence from linked employer-employee data. *Journal of Labor Economics*, 0(ja):null.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *The American Economic Review*, 104(4):1091–1119.

Goux, D., Maurin, E., and Petrongolo, B. (2014). Worktime regulations and spousal labor supply. *American Economic Review*, 104(1):252–76.

Hamermesh, D. S., Kawaguchi, D., and Lee, J. (2017). Does labor legislation benefit workers? well-being after an hours reduction. *Journal of the Japanese and International Economies*, 44:1–12.

Hijzen, A., Kambayashi, R., Teruyama, H., and Genda, Y. (2015). The japanese labour market during the global financial crisis and the role of non-standard work: A micro perspective. *Journal of the Japanese and International Economies*, 38:260–281.

Hunt, J. (1999). Has work-sharing worked in germany? *The Quarterly Journal of Economics*, 114(1):117–148.

Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20:1–24.

Imai, T., Kuwahara, K., Miyamoto, T., Okazaki, H., Nishihara, A., Kabe, I., Mizoue, T., Dohi, S., and on Occupational Health Study Group, J. E. C. (2016). Validity and reproducibility of self-reported working hours among japanese male employees. *Journal of Occupational Health*, 58(4):340–346.

Izumi, Y., Kawaguchi, D., Kuroda, S., and Tsubota, T. (2025). Exemption and work environment. *Industrial Relations: A Journal of Economy and Society*, 64(4):478–519.

Jiang, M. and Sasaki, S. (2025). Do overtime regulations reduce overtime work in japan? *SSRN Electronic Journal*.

Jones, R. S. and Seitani, H. (2019). Labour market reform in japan to cope with a shrinking and ageing population. OECD Working Papers 1568.

Kahneman, D. and Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives*, 20(1):3–24.

Kahneman, D., Wakker, P. P., and Sarin, R. (1997). Back to bentham? explorations of experienced utility. *Quarterly Journal of Economics*, 112(2):375–405.

Kambayashi, R. and Kato, T. (2016). Good Jobs and Bad Jobs in Japan: 1982-2007. Working Paper Series 348, Center on Japanese Economy and Business, Columbia University.

Kato, T., Kawaguchi, D., and Owan, H. (2013). Dynamics of the gender gap in the workplace: An econometric case study of a large japanese firm. RIETI Discussion Paper Series 13-E-038.

Kato, T. and Morishima, M. (2002). The productivity effects of participatory employment practices: Evidence from new japanese panel data. *Industrial Relations: A Journal of Economy and Society*, 41(4):487–520.

Kato, T., Ogawa, H., and Owan, H. (2016). Working hours, promotion, and gender gaps in the workplace. *RIETI Discussion Paper Series*, 16(E-060):1–63.

Kawaguchi, D. and Ueno, Y. (2013). Declining long-term employment in japan. *Journal of the Japanese and International Economies*, 28:19–36.

Kawaguchi, D., M. H. (2019). The labor market in japan 2000–2018. Technical report.

Landers, R. M., Rebitzer, J. B., and Taylor, L. J. (1996). Rat race redux: Adverse selection in the determination of work hours in law firms. *The American Economic Review*, 86(3):329–348.

Lee, J., Kawaguchi, D., and Hamermesh, D. S. (2012). Aggregate impacts of a gift of time. *American Economic Review*, 102(3):612–16.

Lepinteur, A. (2019). The shorter workweek and worker wellbeing: Evidence from portugal and france. *Labour Economics*, 58:204–220.

Liberini, F., Redoano, M., and Proto, E. (2017). Happy voters. *Journal of Public Economics*, 146:41–57.

Ma, Y. and Shi, X. (2020). Are spousal labor supplies substitutes? evidence from the workweek reduction policy in china. *Journal of Development Economics*, 145:102472.

Martins, P. S. (2017). Economic effects of overtime premium flexibility: Firm- and worker-level evidence from a law reform. Technical report.

Nekoei, A., Sigurdsson, J., and Wehr, D. (2024). The Economic Burden of Burnout. Technical Report 19091, C.E.P.R. Discussion Papers.

Nishiyama, K. and Johnson, J. (1997). Karoshi–death from overwork: occupational health consequences of japanese production management. *Int J Health Serv*, 27(4):625–641.

Oaxaca, R. (2014). The effect of overtime regulations on employment. Technical report.

Pannenberg, M. (2005). Long-term effects of unpaid overtime. *Scottish Journal of Political Economy*, 52(2):177–193.

Rambachan, A. and Roth, J. (2023). A More Credible Approach to Parallel Trends. *The Review of Economic Studies*, 90(5):2555–2591.

Rocheteau, G. (2002). Working time regulation in a search economy with worker moral hazard. *Journal of Public Economics*, 84(3):387–425.

Rudolf, R. (2014). Work shorter, be happier? longitudinal evidence from the korean five-day working policy. *Journal of Happiness Studies*, 15(5):1139–1163.

Sun, A., Sun, W., Xiang, W., Zhang, H., and Zhang, J. (2026). Narrowing or widening the gender gap in market and domestic work? the impact of workweek reduction reform in china. *Journal of Development Economics*, 179:103672.

Sánchez, R. (2017). Does a mandatory reduction of standard working hours improve employees' health status? *Industrial Relations: A Journal of Economy and Society*, 56(1):3–39.

Takahashi, K., Kodama, N., Arita, K., Kazama, H., Sakai, S., Takeuchi, M., and Owan, H. (2024). Has japan's work style reform had the intended effect? *Applied Economics*, 0(0):1–24.

Tanaka, M., Kameda, T., Kawamoto, T., Sugihara, S., and Kambayashi, R. (2022). Managing long working hours: Evidence from a management practice survey. *Journal of Human Resources*.

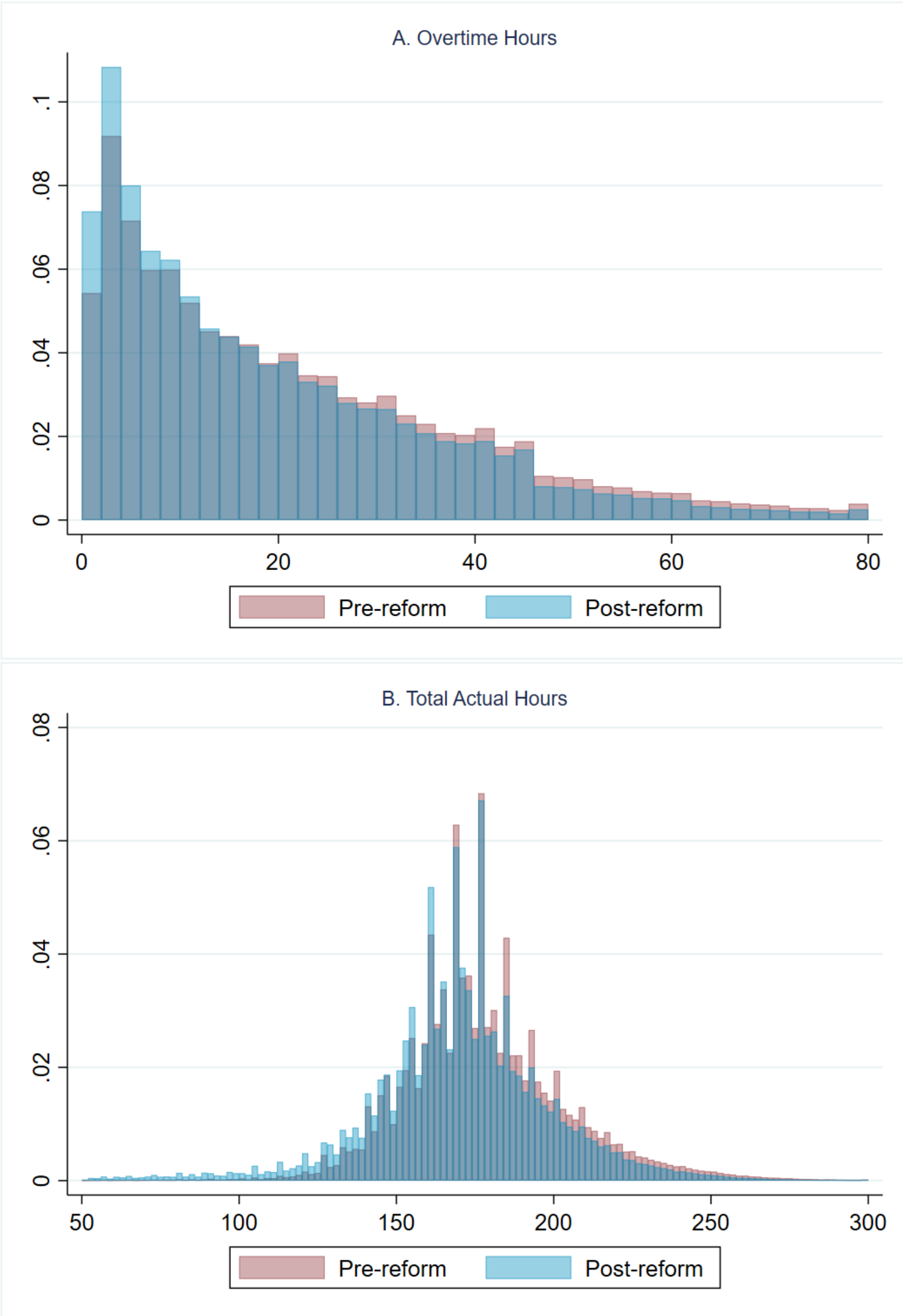
Toda, A. (2023). Assessing the effects of upper limits on overtime work: One example using data to promote evidence-based policy making. *The Japanese Journal of Labour Studies*, 65(752):62–70. Special Issue 2023: The 2022 Conference on Labor Policy Study. Originally published in Japanese.

Trejo, S. J. (2003). Does the statutory overtime premium discourage long workweeks? *ILR Review*, 56(3):530–551.

Urry, H., Nitschke, J., Dolski, I., Jackson, D., Dalton, K., Mueller, C., Rosenkranz, M., Ryff, C., Singer, B., and Davidson, R. (2004). Making a life worth living: neural correlates of well-being. *Psychol Sci*, Jun 15(6):367–72.

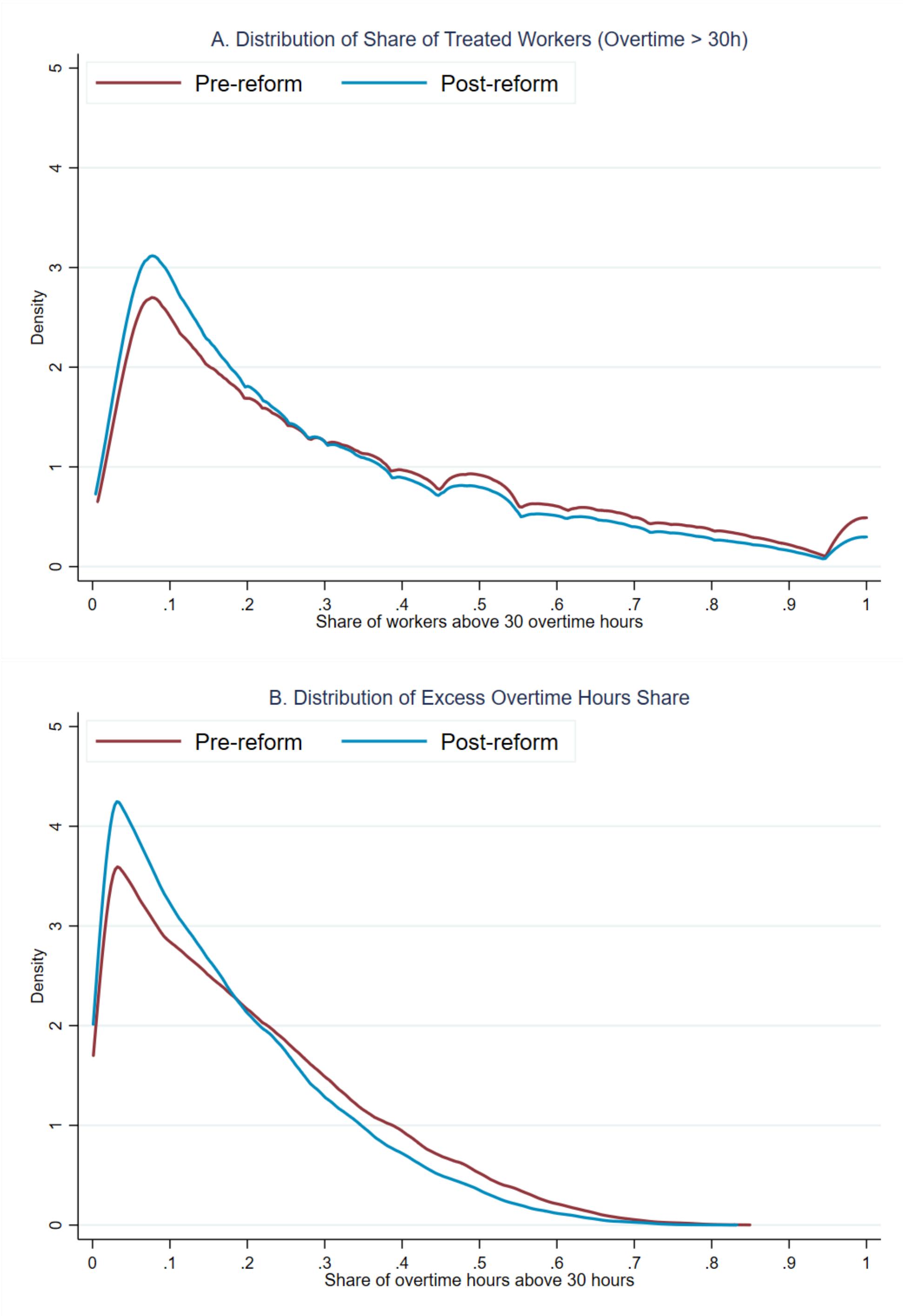
Figures and tables

Figure 1: Distribution of Monthly Hours Worked in BSWS



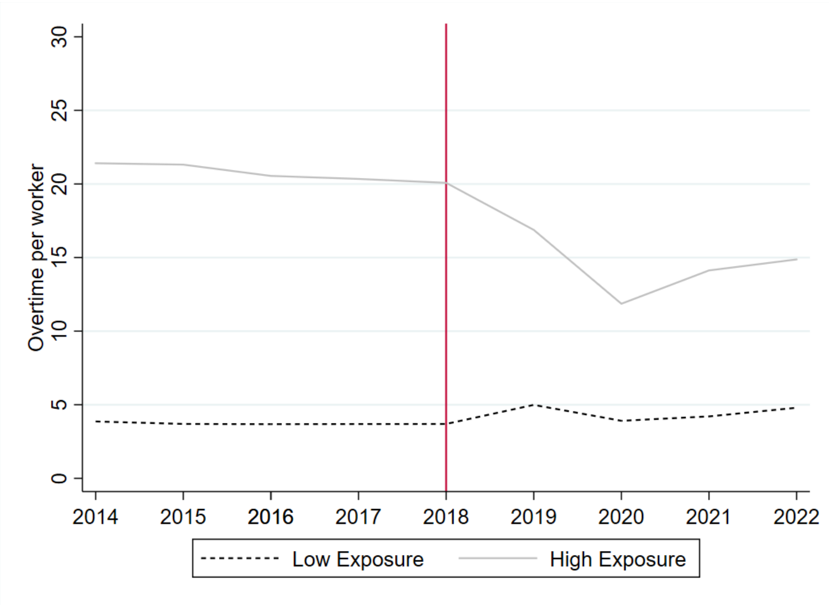
Notes: The figure plots the distribution of monthly overtime hours (Panel A) and total actual hours worked (Panel B) based on worker-level data from BSWS.

Figure 2: Distribution of Share of Treated Workers and Excess Overtime Hours across Firms in BSWS



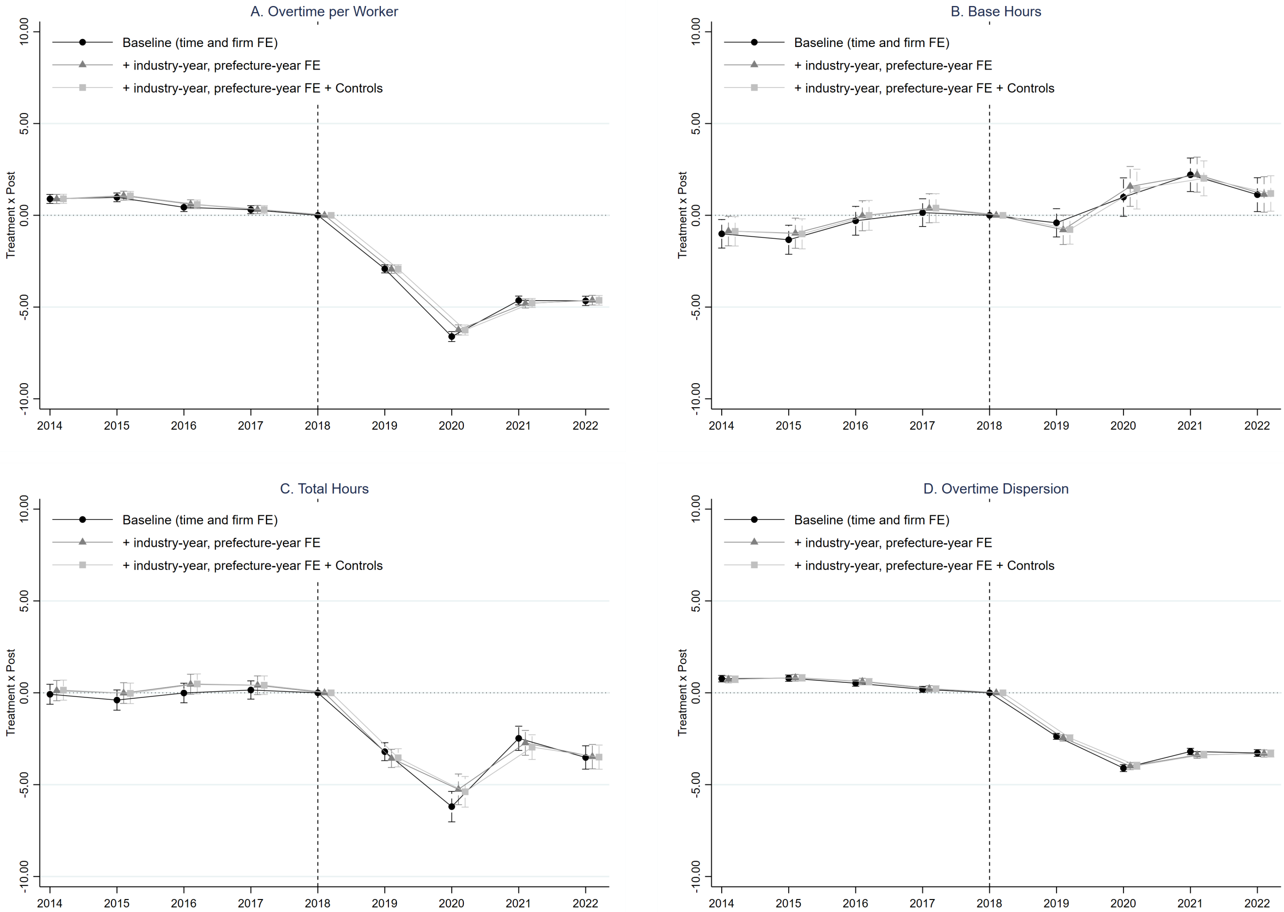
Notes: The figure plots the distribution of share of treated workers (Panel A) and excess overtime hours (Panel B) before (2014-2018) and after (2019-2022). Establishment-level data from BSWS.

Figure 3: Evolution of Overtime Hours per Worker



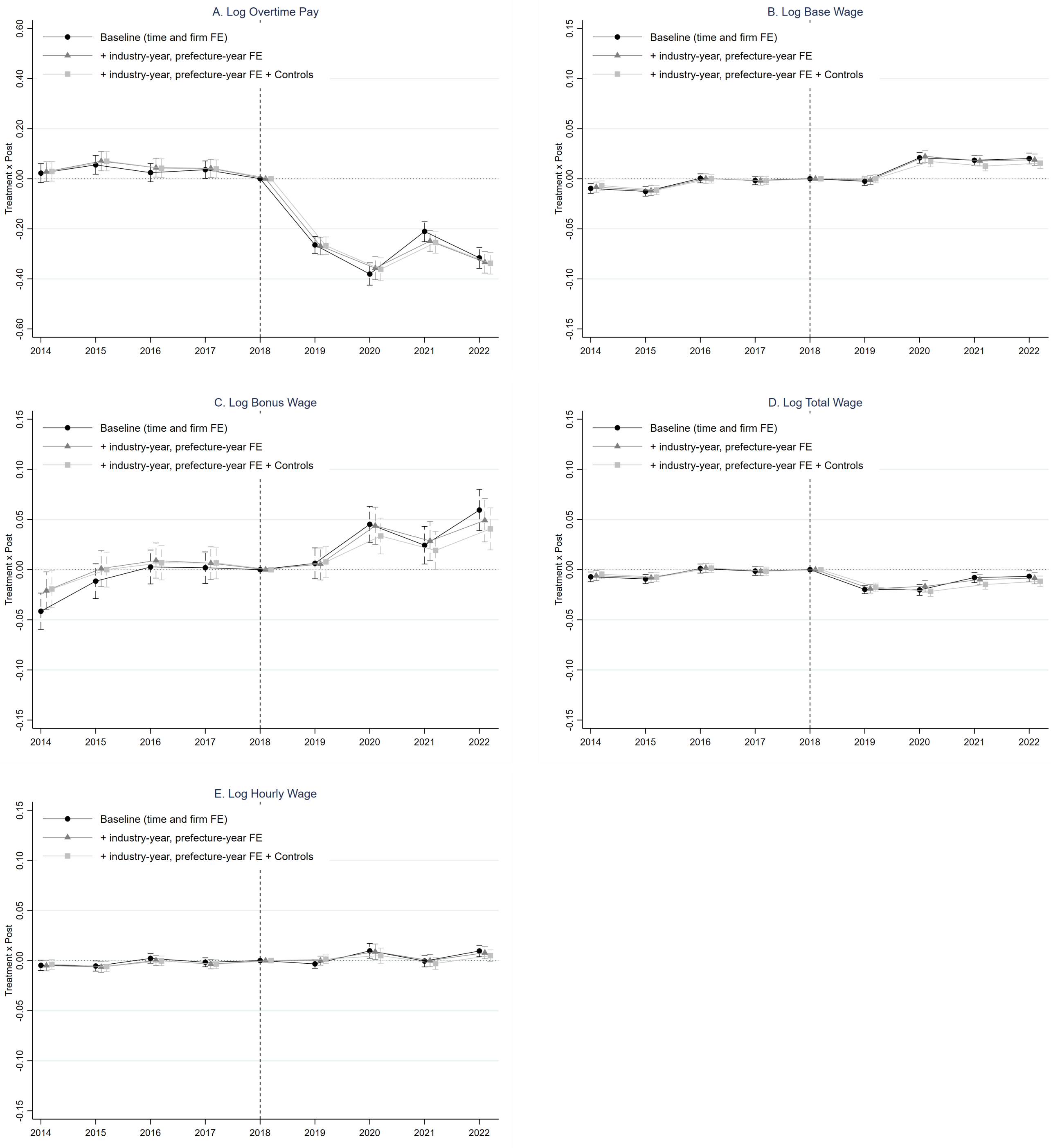
Notes: The figure plots the evolution of overtime hours per worker in "more affected" and "less affected" establishments. Establishments' degree of exposure considers the pre-reform share of workers supplying more than 30h overtime per month, i.e. the new WSR cap.

Figure 4: Event-Study Analysis: Hours



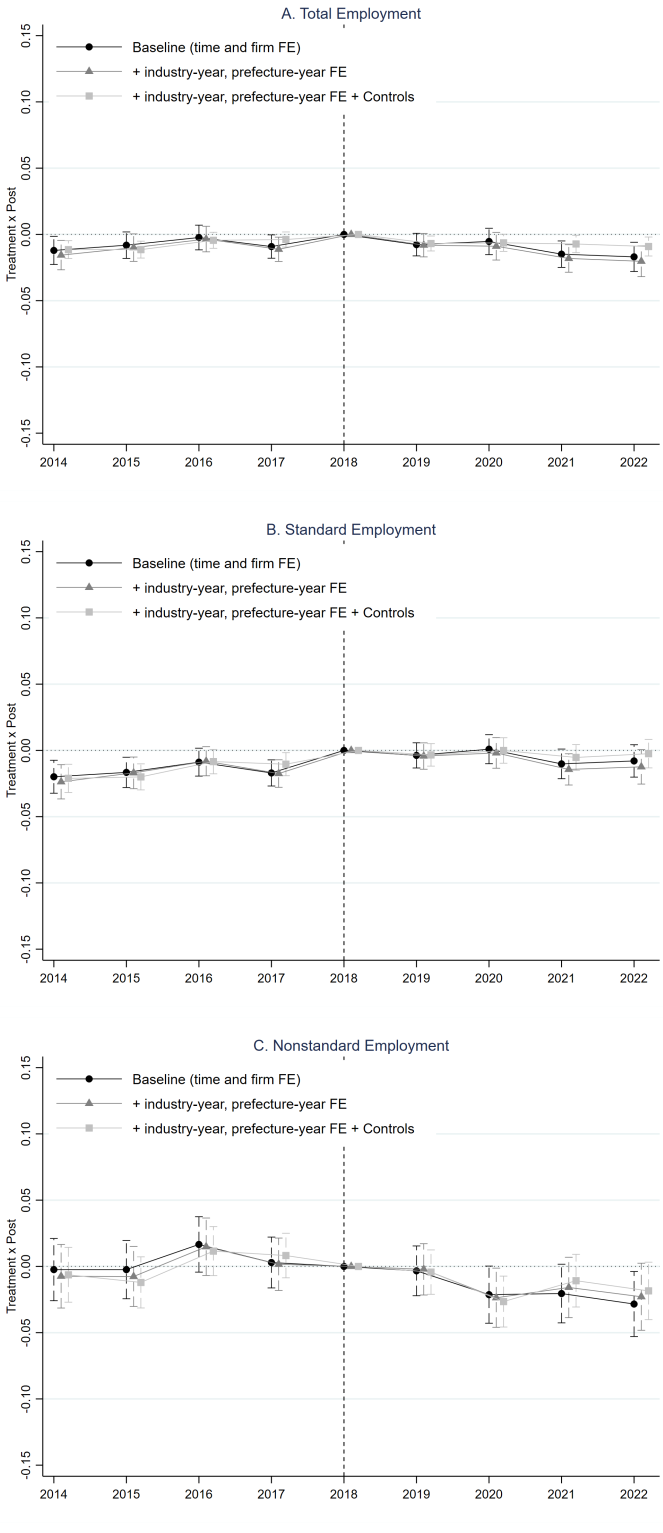
Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Dependent variables: Overtime per Worker (Panel A), Base Hours (Panel B), Total Hours (Panel C) and Intra-Establishment Overtime Dispersion (sd) (Panel D). Working hours variables are measured on a monthly basis. Sample: BSWS collapsed at the establishment level (establishment-level panel 2014-2022). Reported estimates include firm and year-FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies. Standard errors clustered at the establishment level.

Figure 5: Event-Study Analysis: Wages



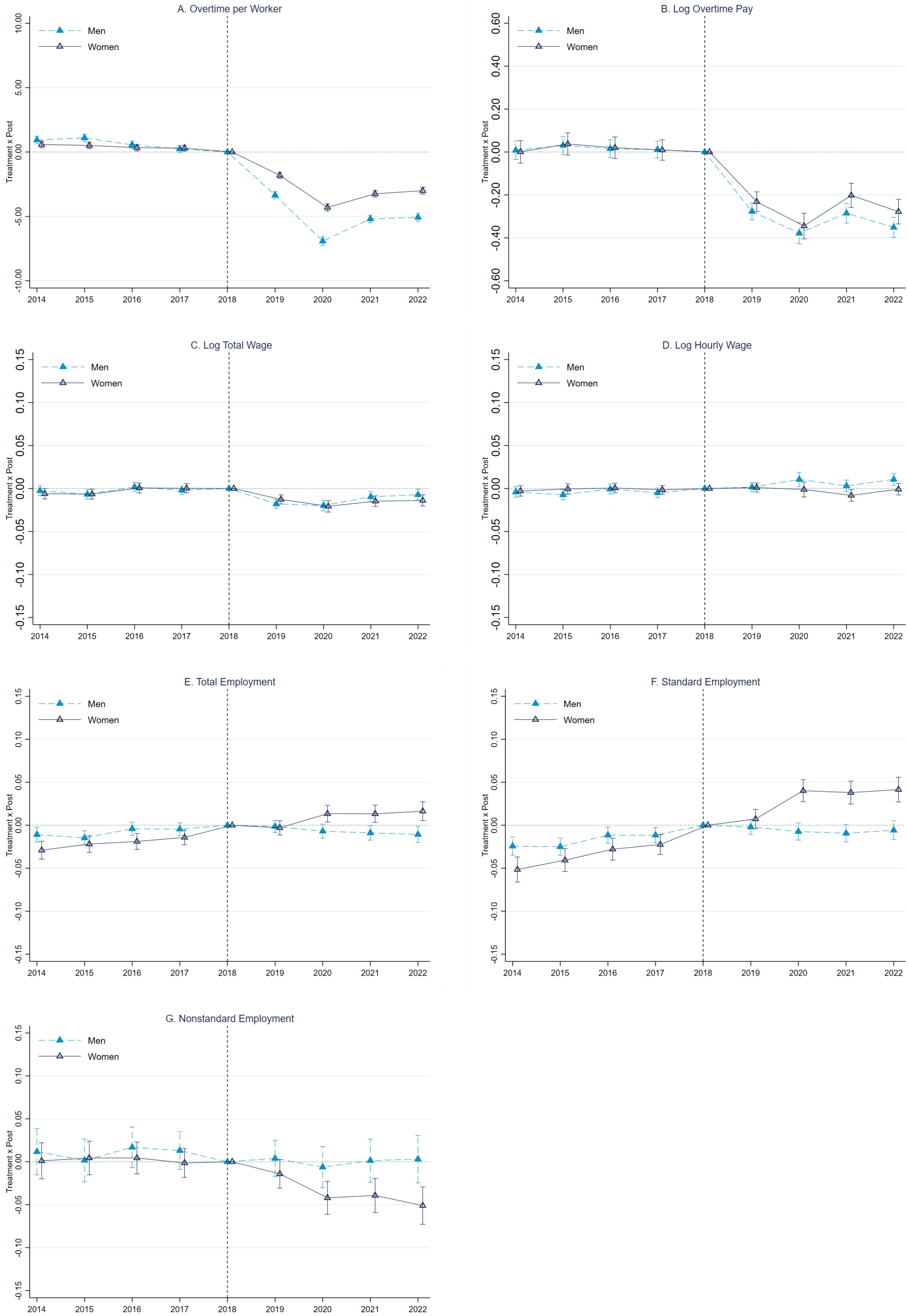
Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Dependent variables (in logs): Overtime Pay (Panel A), Base Wage (Panel B), Bonuses (Panel C), Total Monthly Earnings (Panel D), and Hourly Wage (Panel E). Sample: BSWS collapsed at the establishment level (establishment-level panel 2014-2022). Reported estimates include firm and year-FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies. Standard errors clustered at the establishment level.

Figure 6: Event-study Analysis: Employment



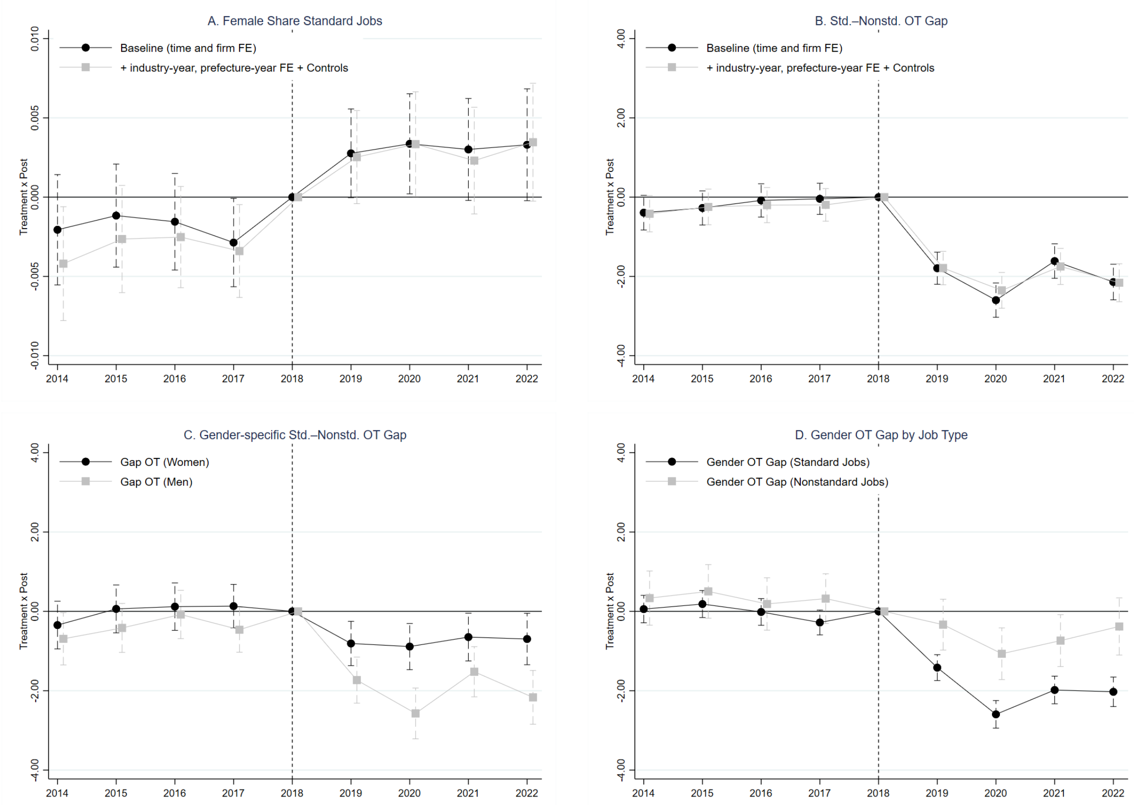
Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Dependent variables (in logs): Total Employment, Standard Employment, and Nonstandard Employment. Sample: BSWS collapsed at the establishment level (establishment-level panel 2014-2022). Reported estimates include firm and year-FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies. Standard errors clustered at the establishment level.

Figure 7: Heterogeneous Effects: Gender



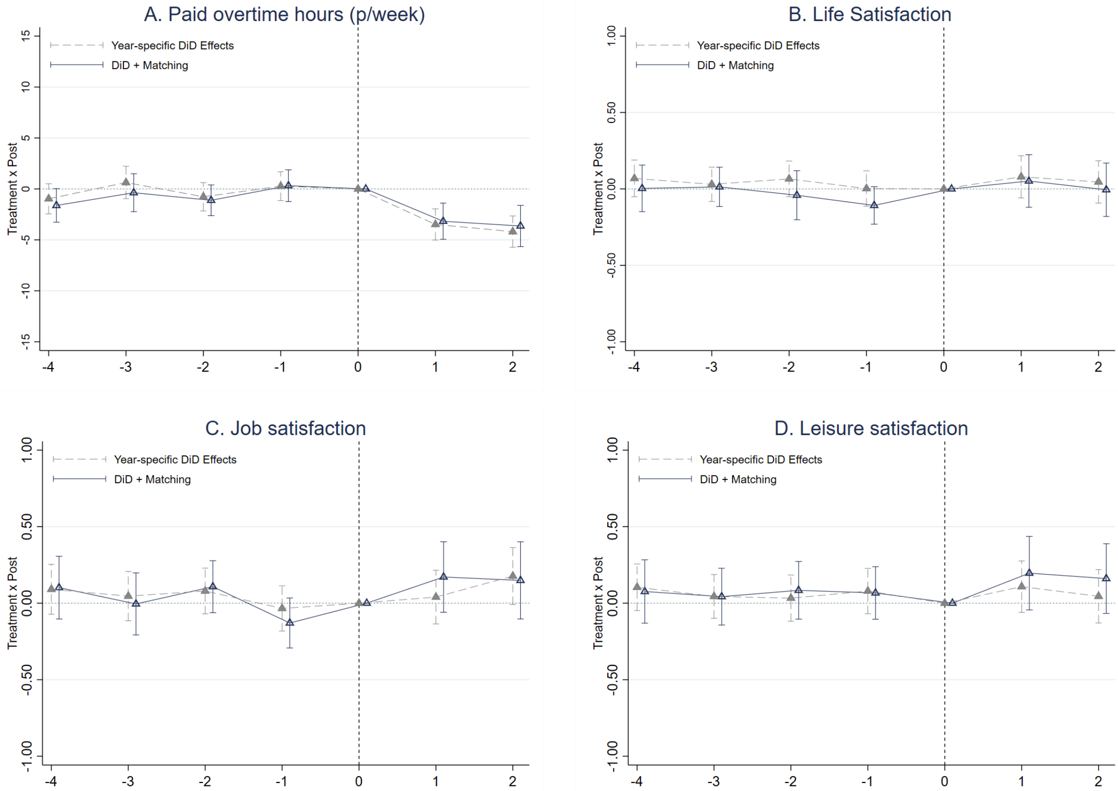
Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Regressions for each gender-specific outcome at the establishment level are estimated separately. Dependent variables: Overtime per Worker (Panel A), Overtime Pay (Panel B), Total Monthly Earnings (Panel C), Hourly Wage (Panel D), Total Employment (Panel E), Standard Employment (Panel F), and Nonstandard Employment (Panel G). Sample: BSWS collapsed at the establishment level (establishment-level panel 2014-2022). Reported estimates include firm and year-FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies. Standard errors clustered at the establishment level.

Figure 8: Female Share of Standard Jobs and Overtime Gaps by Gender and Job Type



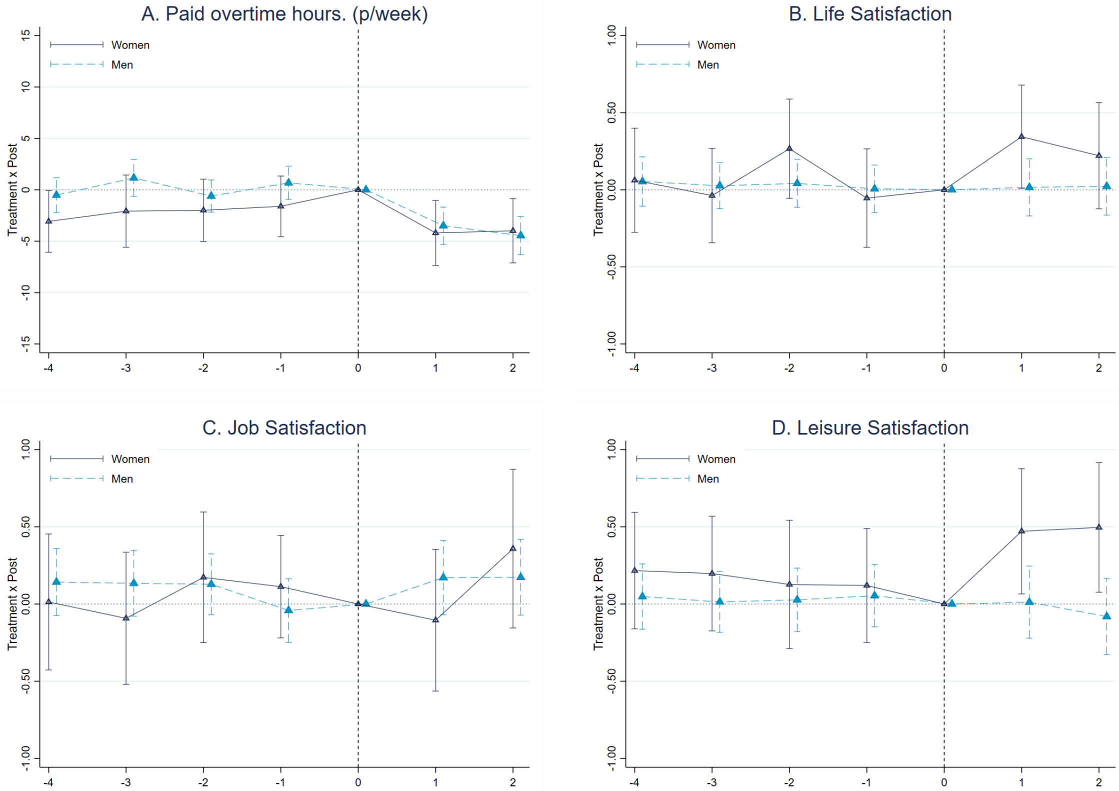
Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Dependent variables: Female Share of Standard Jobs (Panel A), Standard–Nonstandard Overtime Gap (Panel B), Gender-specific Std.–Nonstd. OT Gap (Panel C), Gender OT Gap by Job Type (Standard vs. Nonstandard Jobs) (Panel D). Sample: BSWS collapsed at the establishment level (establishment-level panel 2014-2022). Reported estimates include firm and year-FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, firm-size dummies. Standard errors clustered at the establishment level.

Figure 9: Event-Study Analysis: Self-Reported Overtime and Subjective Well-Being



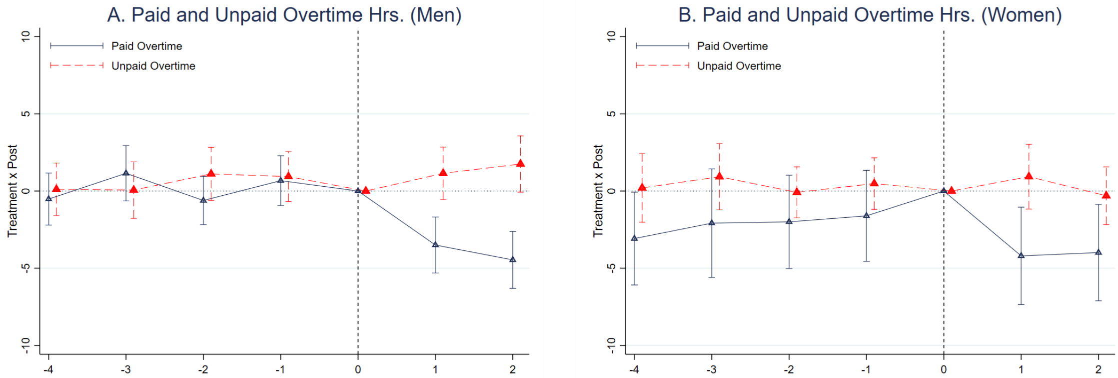
Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Dependent variables: (Weekly) Paid Overtime Hrs. (Panel A), Life Satisfaction (Panel B), Job Satisfaction (Panel C), Leisure Satisfaction (Panel D). SWB variables are 1-5 Likert scales. Sample: Individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). Reported estimates include individual and year-FE. Other controls: tenure group dummies, occupation dummies, firm size class dummies and sector effects. "DiD + Matching" refers to a re-weighted DiD estimates using a coarsened exact-matched sample of treated and control individuals. Standard errors are clustered at the individual level.

Figure 10: Event-Study Analysis: Self-Reported Overtime and Subjective Well-Being by Gender



Notes: The figure displays year-specific DiD estimates by gender and 95% confidence intervals. Dependent variables: (Weekly) Paid Overtime Hrs. (Panel A), Life Satisfaction (Panel B), Job Satisfaction (Panel C), Leisure Satisfaction (Panel D). Sample: Individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). Reported estimates include individual and year-FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies and sector effects. "DiD + Matching" refers to a re-weighted DiD estimates using a coarsened exact-matched sample of treated and control individuals. Standard errors are clustered at the individual level.

Figure 11: Event-Study Analysis: Paid vs. Unpaid Overtime Hours by Gender



Notes: The figure displays year-specific DiD estimates for paid and unpaid overtime by gender and 95% confidence intervals. Sample: Individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). Reported estimates include individual and year-FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies and sector effects. Standard errors are clustered at the individual level.

Table 1: DiD estimates: Working Hours

	(1) Mean Overtime	(2) Base Hours	(3) Total Hours	(4) Overtime Dispersion (sd)
High Exposure × Post	-5.161***	1.231***	-4.024***	-3.708***
	(0.072)	(0.238)	(0.183)	(0.052)
Observations	316,744	316,744	316,744	307,921
Establishment controls	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes
Industry-specific time trends	Yes	Yes	Yes	Yes
Prefecture-specific time trends	Yes	Yes	Yes	Yes
Mean of outcome	20.75	164.9	185.7	15.54

Notes: DiD estimates using establishment-level panel (2014-2022) from Basic Survey on Wage Structure (BSWS). Reported estimates include firm FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies (8). Standard errors clustered at the establishment level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01.

Table 2: DiD estimates: Log Wages

	(1) Monthly Wage	(2) Overtime Pay	(3) Base Wage	(4) Bonus Wage	(5) Hourly Wage	(6) Monthly Wage (sd)	(7) Overtime Pay (sd)
High Exposure × Post	-0.014***	-0.337***	0.015***	0.025***	0.004***	-2.161	-58.219***
	(0.001)	(0.012)	(0.001)	(0.005)	(0.002)	(3.519)	(1.088)
Observations	316,709	259,845	316,709	280,522	316,675	307,921	307,921
Establishment controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-specific time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prefecture-specific time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of outcome	7.953	5.690	7.813	8.560	2.736	822.7	302.2

Notes: DiD estimates using establishment-level panel (2014-2022) from Basic Survey on Wage Structure (BSWS). Reported estimates include firm FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies (8). Standard errors clustered at the establishment level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01.

Table 3: DiD estimates: Employment

	All			Female Employment			Male Employment		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Total	Nonstandard	Standard	Total	Nonstandard	Standard	Total	Nonstandard	Standard
High Exposure × Post	-0.001	-0.015**	0.008***	0.018***	-0.027***	0.049***	-0.003	-0.006	0.003
	(0.002)	(0.006)	(0.003)	(0.003)	(0.006)	(0.004)	(0.003)	(0.008)	(0.003)
Observations	316,744	248,247	313,119	306,662	222,700	277,115	308,905	191,337	301,432
Establishment controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-specific time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prefecture-specific time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of outcome	3.977	2.641	3.505	2.662	2.195	2.074	3.471	2.053	3.200

Notes: DiD estimates using establishment-level panel (2014-2022) from Basic Survey on Wage Structure (BSWS). Reported estimates include firm FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies (8). Standard errors clustered at the establishment level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01.

Table 4: DiD estimates: Workforce Composition

	(1) % Fixed Term	(2) % Nonstandard	(3) % College educ.
High Exposure $\times$ Post	-0.009***	-0.007***	0.006***
	(0.002)	(0.002)	(0.001)
Observations	317,620	317,620	316,744
Establishment controls	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes
Industry-specific time trends	Yes	Yes	Yes
Prefecture-specific time trends	Yes	Yes	Yes
Mean of outcome	0.165	0.177	0.303

Notes: DiD estimates using establishment-level panel (2014-2022) from Basic Survey on Wage Structure (BSWS). Reported estimates include firm FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies (8). Standard errors clustered at the establishment level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01.

Table 5: DiD estimates: Overtime Gaps by Gender

	(1) Female Share Standard Jobs	(2) Std.–Nonstd. OT Gap	(3) Std.–Nonstd. OT Gap (Women)	(4) Std.–Nonstd. OT Gap (Men)	(5) Gender OT Gap Standard Jobs	(6) Gender OT Gap Nonstandard Jobs
High Exposure × Post	0.005***	-1.807***	-0.766***	-1.681***	-1.978***	-0.872***
	(0.001)	(0.129)	(0.167)	(0.181)	(0.096)	(0.184)
Observations	313,915	151,092	99,545	112,388	247,420	80,341
Mean of outcome	0.246	8.788	4.821	8.932	9.044	4.141

Notes: DiD estimates using establishment-level panel (2014-2022) from Basic Survey on Wage Structure (BSWS). Reported estimates include firm FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies (8). Standard errors clustered at the establishment level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01.

Table 6: DiD estimates: Self-Reported (Weekly) Working Hours.

	(1) Total hours	(2) Long hours	(3) OT total	(4) Log wage	(5) Workdays
A. Uncontrolled					
Treat × Post	-2.331***	-0.083***	-3.629***	0.033	-1.590
	(0.772)	(0.025)	(0.644)	(0.027)	(3.340)
Observations	5120	5120	4850	4652	5042
B. Controlled					
Treat × Post	-2.704***	-0.090***	-4.116***	-0.035	-2.586
	(0.764)	(0.025)	(0.641)	(0.022)	(3.219)
Observations	5062	5062	4801	4620	4991
C. Individual FE					
Treat × Post	-1.636**	-0.069**	-2.759***	-0.009	0.710
	(0.792)	(0.027)	(0.619)	(0.017)	(3.404)
Mean outcome	51.849	0.270	12.224	5.693	247.668
Observations	5062	5062	4801	4620	4991
D. Men vs. Women					
Treat × Post	-1.332	-0.041	-2.888***	-0.006	-1.690
	(0.920)	(0.033)	(0.746)	(0.019)	(3.805)
Treat × Post × Female	-1.037	-0.056	0.684	0.006	7.059
	(2.072)	(0.069)	(1.403)	(0.052)	(9.515)
TE women	-2.369	-0.097	-2.204	-0.000	5.370
p-value: TE women	0.199	0.106	0.068	0.997	0.537
Mean outcome men	52.675	0.297	12.625	5.754	250.646
Mean outcome women	47.704	0.134	10.223	5.394	232.380
Observations	5062	5062	4801	4620	4991

Notes: DiD estimates using individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group comprises workers who were supplying more than 30 hrs. of paid overtime in a typical month before the reform. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies, sector fixed effects and region-specific time trends. Standard errors clustered at the individual level and shown in parentheses. Significance levels: * 0.10, ** 0.05, ***0.01

Table 7: DiD estimates: Paid vs. Unpaid Overtime

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All workers			Workers with OT total>0			
	OT total	OT paid	OT unpaid	OT total	OT paid	OT unpaid	Unpaid-to-Total OT Ratio
A. Uncontrolled							
Treat × Post	-3.629***	-3.982***	0.044	-2.584***	-4.201***	1.168*	0.137***
	(0.644)	(0.457)	(0.479)	(0.723)	(0.556)	(0.617)	(0.038)
Observations	4850	4950	4994	3374	3474	3518	3374
B. Controlled							
Treat × Post	-4.116***	-3.862***	-0.476	-3.599***	-3.964***	0.097	0.086**
	(0.641)	(0.443)	(0.468)	(0.710)	(0.527)	(0.602)	(0.035)
Observations	4801	4897	4942	3348	3444	3489	3348
C. Individual FE							
Treat × Post	-2.759***	-3.608***	0.632	-2.547***	-3.727***	0.822	0.081**
	(0.619)	(0.452)	(0.450)	(0.721)	(0.556)	(0.588)	(0.032)
Observations	4801	4897	4942	3348	3444	3489	3348
D. Men vs. Women							
Treat × Post	-2.888***	-4.188***	1.036*	-2.555***	-4.651***	1.629**	0.100***
	(0.746)	(0.555)	(0.578)	(0.899)	(0.722)	(0.777)	(0.038)
Treat × Post × Female	0.684	1.813	-1.017	0.069	2.427*	-2.201*	-0.076*
	(1.403)	(1.120)	(0.966)	(1.617)	(1.412)	(1.219)	(0.077)
TE women	-2.204*	-2.375**	0.019	-2.486*	-2.225*	-0.572	0.023
p-value: TE women	0.068	0.014	0.980	0.074	0.065	0.532	0.723
Mean outcome men	12.625	8.175	4.726	14.573	9.392	5.432	0.311
Mean outcome women	10.223	6.697	3.579	11.042	7.213	3.858	0.317
Observations	4801	4897	4942	3348	3444	3489	3348

Notes: DiD estimates using individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group comprises workers who were supplying more than 30 hrs. of paid overtime in a typical month before the reform. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies, sector fixed effects and region-specific time trends. Standard errors clustered at the individual level and shown in parentheses. Significance levels: * 0.10, ** 0.05, ***0.01

Table 8: DiD estimates: Subjective Well-Being.

	How satisfied are you with each of the following? (1-5 scale)				
	(1)	(2)	(3)	(4)	(5)
	Overall Life	Job	Spouse	Leisure	Family
A. Uncontrolled					
Treat × Post	0.034	0.069	-0.009	0.037	0.009
	(0.062)	(0.079)	(0.085)	(0.072)	(0.069)
Observations	5134	5175	4074	5180	5024
B. Controlled					
Treat × Post	0.018	0.009	0.023	0.058	0.024
	(0.065)	(0.084)	(0.093)	(0.079)	(0.076)
Observations	5069	5112	4030	5116	4961
C. Individual FE					
Treat × Post	0.025	0.036	0.064	0.034	-0.012
	(0.060)	(0.079)	(0.081)	(0.079)	(0.070)
Mean outcome	3.625	3.057	3.777	3.433	3.814
Observations	5069	5112	4030	5116	4961
D. Men vs. Women					
Treat × Post	-0.036	0.068	0.066	-0.086	-0.047
	(0.065)	(0.087)	(0.088)	(0.086)	(0.077)
Treat × Post × Female	0.284**	0.010	0.137	0.463**	0.066
	(0.128)	(0.192)	(0.207)	(0.186)	(0.181)
TE women	0.247**	0.078	0.204	0.377**	0.019
p-value: TE women	0.026	0.651	0.276	0.024	0.909
Mean outcome (men)	3.633	3.041	3.781	3.450	3.825
Mean outcome (women)	3.586	3.134	3.738	3.353	3.762
Observations	5069	5112	4030	5116	4961

Notes: DiD estimates using individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group comprises workers who were supplying more than 30 hrs. of paid overtime in a typical month before the reform. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies, sector fixed effects and region-specific time trends. Standard errors clustered at the individual level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01

Online Appendix

A Conceptual Framework

This section develops a simple framework that clarifies four mechanisms central to our empirical analysis: (i) compression of paid overtime under a binding cap, (ii) substitution into unpaid effort, (iii) gender-differentiated well-being responses, and (iv) reallocation across standard and nonstandard jobs within establishments. The framework is intentionally stylized and serves to organize the empirical findings.

A.1 Overtime Choice and the Paid Overtime Cap

Consider a worker i who chooses paid overtime $h_i \geq 0$ and unpaid overtime $u_i \geq 0$. Total effort is

$$e_i = h_i + u_i.$$

Choice utility is

$$U_i(h_i, u_i) = y(h_i) + \phi_i R(e_i) - C_i(e_i),$$

where $y(h_i)$ captures earnings from paid overtime, $R(e_i)$ reflects career returns as a function of total effort, $C_i(e_i)$ captures fatigue costs, and $\phi_i \geq 0$ measures the strength of career concerns. Fatigue costs are assumed increasing and convex, $C_i'(e) > 0$ and $C_i''(e) > 0$, while career returns are increasing and weakly concave, $R'(e) > 0$ and $R''(e) \leq 0$.

Intuitively, because effort enters utility through total effort $e_i = h_i + u_i$, a worker supplying unpaid effort could always increase utility by reallocating that effort into paid overtime as long as the marginal wage return $y'(h_i)$ is positive. In this unconstrained benchmark, unpaid overtime is absent and total effort coincides with paid overtime[43]:

$$e_i^{UC} = h_i^{UC}.$$

The optimal choice of effort satisfies the first-order condition

$$y'(h_i^{UC}) + \phi_i R'(e_i^{UC}) = C_i'(e_i^{UC}),$$

[43] In practice, however, paid overtime may be contractually or organizationally constrained even before the reform (e.g., due to reporting limits or implicit norms), so that some workers may supply unpaid effort in equilibrium. The key distinction is that, absent a binding cap, workers can typically adjust paid overtime at the margin, making unpaid effort less relevant.

which equates the marginal benefit of additional effort—arising from wage income and career returns—to the marginal fatigue cost.

The reform introduces a binding cap on paid overtime,

$$h_i \leq \bar{h}.$$

For workers with $h_i^{UC} > \bar{h}$, the cap binds and paid overtime is fixed at $\bar{h}$. Workers may still choose unpaid overtime u_i. If $u_i^{cap} > 0$, the post-reform first-order condition becomes

$$\phi_i R'(e_i^{cap}) = C_i'(e_i^{cap}),$$

with

$$e_i^{cap} = \bar{h} + u_i^{cap}.$$

Proposition 1. For workers for whom the cap binds, total effort under the cap is lower than under the unconstrained optimum:

$$e_i^{cap} < e_i^{UC}.$$

Intuitively, when the cap binds, additional effort can only take the form of unpaid overtime and therefore yields no marginal wage return. Workers therefore equate marginal career returns to marginal fatigue costs, whereas before the reform the marginal benefit of effort also included wage income. Because wage incentives previously increased optimal effort, removing them reduces the optimal effort level. Thus the cap reduces total effort, although workers with strong career concerns may still supply some unpaid overtime.[44]

A.2 Household Spillovers, Experienced Utility and Well-Being

To connect effort choices to subjective well-being, we distinguish between the utility workers optimize when choosing effort and the experienced utility captured by SWB survey measures.[45]

We write experienced utility as

$$W_i = \alpha_i - F(e_i) + J(L_i, L_{-i}), \qquad L_i = \bar{T} - e_i,$$

[44]For workers for whom the cap binds, paid overtime is fixed at $\bar{h}$, so total effort satisfies $e_i^{cap} \geq \bar{h} > 0$.

[45]The distinction between decision and experienced utility is standard in the well-being literature. See Kahneman et al. (1997); Kahneman and Krueger (2006); Benjamin et al. (2014).

where $J(L_i, L_{-i})$ captures household interactions in non-work time.

We allow $J(\cdot)$ to be a general function of both partners' non-work time. The cross-partial derivative

$$\frac{\partial^2 J(L_i, L_{-i})}{\partial L_i \partial L_{-i}}$$

determines whether non-work time is complementary or substitutable within the household. If the cross-partial is positive, the marginal value of an increase in one partner's non-work time is higher when the other partner also has more non-work time. If it is zero, preferences are additively separable; if it is negative, non-work time is substitutable across partners.

The household block yields a simple comparative-static implication and provides a direct interpretation of the empirical additivity tests. In the four-cell household design presented in Section 6, additivity corresponds to $\lambda_{11} = \lambda_{10} + \lambda_{01}$, which arises when $J(\cdot)$ is additively separable, while deviations from additivity reflect non-zero cross-partial derivatives. Since $L_i = \bar{T} - e_i$, we have $\partial L_i / \partial e_i = -1$, so

$$\frac{\partial W_i}{\partial e_i} = -F'(e_i) - \frac{\partial J(L_i, L_{-i})}{\partial L_i}.$$

Thus, under the assumptions that $F'(e_i) > 0$ and $\partial J / \partial L_i > 0$, lower effort raises well-being both by reducing fatigue and by increasing non-work time. When non-work time is complementary within couples, i.e.

$$\frac{\partial^2 J(L_i, L_{-i})}{\partial L_i \partial L_{-i}} > 0,$$

the marginal well-being gain from a reduction in own effort is larger when the spouse also has more non-work time. This implies that joint exposure to the reform can generate larger-than-additive improvements in well-being.[46]

A.3 Assignment to Standard Jobs

We now examine how the cap can affect job composition within establishments. Workers differ in their tolerance for long hours. We assume individual fatigue costs take the form $C_i(e) = \frac{1}{\theta_i} c(e)$, where $c(e)$ is a common convex cost function and θ_i captures heterogeneity in tolerance for long hours.

[46] Complementarities in non-work time may arise through several mechanisms, including joint consumption of leisure activities, fixed or coordination costs of shared activities, and improvements in the quality of interactions when both partners experience lower work strain.

Each establishment j offers two job types.

Nonstandard job (N)

$$V_{ij}^N = w_{Nj}h_{Nj} - \frac{1}{\theta_i}c(h_{Nj})$$

Standard job (S)

$$V_{ij}^S = w_{Sj}h_{Sj} - \frac{1}{\theta_i}c(h_{Sj})$$

where $w_{Nj}h_{Nj}$ and $w_{Sj}h_{Sj}$ denote total compensation associated with the two job types and $h_{Sj} > h_{Nj}$ reflects the longer-hours norm typically associated with standard jobs. In this simplified version, the compensation differential should be interpreted broadly, capturing both direct earnings and the career value $R(\cdot)$ associated with standard positions.

We interpret standard jobs as positions that require a minimum tolerance for long hours. For a given establishment j, define θ_j^* as the cutoff worker type that is indifferent between standard and nonstandard jobs:

$$V_{ij}^S = V_{ij}^N.$$

Workers with $\theta_i \geq \theta_j^*$ are more likely to be assigned to standard jobs in establishment j. Solving the indifference condition yields

$$\theta_j^* = \frac{c(h_{Sj}) - c(h_{Nj})}{w_{Sj}h_{Sj} - w_{Nj}h_{Nj}}.$$

The reform affects this assignment threshold through two opposing channels. A reduction in the hours requirement of standard jobs reduces the effort-cost gap between job types and can lower θ_j^*, thereby expanding access to standard jobs for workers with higher disutility of long hours, unless offset by a corresponding compression in earnings differences.

B Formal Derivations

B.1 Proof of Proposition 1

Define

$$f(e) = C_i'(e) - \phi_i R'(e).$$

Under the assumptions that fatigue costs are convex and career returns are concave, we have

$$f'(e) = C_i''(e) - \phi_i R''(e) > 0,$$

so that $f(e)$ is strictly increasing.

Before the reform, the worker chooses paid overtime h_i to satisfy the first-order condition

$$y'(h_i^{UC}) + \phi_i R'(e_i^{UC}) = C_i'(e_i^{UC}).$$

Rearranging yields

$$f(e_i^{UC}) = y'(h_i^{UC}) > 0.$$

After the reform, the cap binds for workers with $h_i^{UC} > \bar{h}$ and the worker chooses unpaid effort u_i such that

$$\phi_i R'(e_i^{cap}) = C_i'(e_i^{cap}),$$

which implies

$$f(e_i^{cap}) = 0.$$

Since $f(e)$ is strictly increasing and

$$f(e_i^{UC}) > f(e_i^{cap}),$$

it follows that

$$e_i^{UC} > e_i^{cap}.$$

B.2 Assignment Threshold and Comparative Statics

The value to worker i of holding a standard or nonstandard job in establishment j is

$$V_{ij}^S = w_{Sj}h_{Sj} - \frac{1}{\theta_i}c(h_{Sj}),$$

$$V_{ij}^N = w_{Nj}h_{Nj} - \frac{1}{\theta_i}c(h_{Nj}),$$

and the cutoff type θ_j^* satisfies

$$\theta_j^* = \frac{c(h_{Sj}) - c(h_{Nj})}{w_{Sj}h_{Sj} - w_{Nj}h_{Nj}}.$$

Differentiating with respect to h_{Sj} yields

$$\frac{\partial \theta_j^*}{\partial h_{Sj}} = \frac{c'(h_{Sj})}{\left(w_{Sj}h_{Sj} - w_{Nj}h_{Nj}\right)^2}\left[\left(w_{Sj}h_{Sj} - w_{Nj}h_{Nj}\right) - \frac{w_{Sj}}{c'(h_{Sj})}\left\{c(h_{Sj}) - c(h_{Nj})\right\}\right].$$

The sign of this derivative is ambiguous. In particular,

$$\frac{\partial \theta_j^*}{\partial h_{Sj}} > 0 \quad \Longleftrightarrow \quad c(h_{Sj}) - c(h_{Nj}) < \frac{c'(h_{Sj})}{w_{Sj}}\left(w_{Sj}h_{Sj} - w_{Nj}h_{Nj}\right).$$

The effect of the reform can be understood by considering a reduction in the hours requirement of standard jobs, h_{Sj}. A decline in h_{Sj} affects the assignment threshold θ_j^* through two opposing channels. On the one hand, lower hours reduce the effort-cost gap $c(h_{Sj}) - c(h_{Nj})$, making standard jobs less demanding and thereby lowering θ_j^*. On the other hand, the same reduction in hours decreases the earnings differential $w_{Sj}h_{Sj} - w_{Nj}h_{Nj}$ by compressing overtime-related pay, which tends to raise θ_j^*.

The relative strength of these two forces depends on how differences in effort costs compare to differences in earnings, once both are expressed in comparable units. The term $\frac{c'(h_{Sj})}{w_{Sj}}$ plays this role: it converts monetary returns into utility-equivalent units and can be interpreted as the the utility cost of earning one additional unit of income through longer hours).

The condition above shows that a reduction in h_{Sj} lowers θ_j^*—thereby expanding access to standard jobs—when the decline in the effort-cost gap is sufficiently large relative to the compression in earnings. Conversely, if the loss in earnings dominates the reduction in effort costs, the assignment threshold increases.

C Supplementary Tables and Figures

Table A1: Who Supplies Overtime Hours?

	(1) Overtime hrs.	(2) Long overtime (30h+)
Female	-6.331***	-0.109***
	(0.018)	(0.000)
Age	0.226***	0.003***
	(0.005)	(0.000)
Age square	-0.005***	-0.000***
	(0.000)	(0.000)
Tenure	0.223***	0.003***
	(0.003)	(0.000)
Tenure square	-0.007***	-0.000***
	(0.000)	(0.000)
Junior High School	0.943***	0.015***
	(0.055)	(0.001)
Junior College	-1.023***	-0.019***
	(0.024)	(0.000)
Over 4-year College	-1.465***	-0.022***
	(0.022)	(0.000)
Fixed term	0.081**	-0.001
	(0.033)	(0.001)
Nonstandard	-0.463***	-0.009***
	(0.033)	(0.001)
Large Establishment	3.179***	0.047***
	(0.026)	(0.001)
Constant	16.984***	0.228***
	(0.104)	(0.002)
Observations	4,161,792	4,161,792
R-squared	0.112	0.074

Notes: Correlates of monthly overtime hours. Individual-level regressions using BSWS. Pre-reform pooled sample (2013-2018). Estimates control for industry and region fixed effects (not reported).

Table A2: Estimation Samples: Descriptive Statistics

	Treated	Controls
A. Establishment-level data (BSWS)		
Mean Age	41.033	42.692
Mean Tenure	10.848	10.499
% Female	0.322	0.435
% Fixed-term	0.163	0.138
% Nonstandard (=1)	0.178	0.156
% College educ.	0.314	0.293
Large Establishment	0.126	0.013
Manufacturing	0.279	0.143
Large City	0.187	0.174
B. Individual-level survey (OPPS)		
Female	0.18	0.37
Age	46.15	48.89
Tenure >20y	0.42	0.48
Clerical worker	0.21	0.25
Service sector	0.17	0.21
Large firm	0.52	0.47
Big city	0.29	0.27

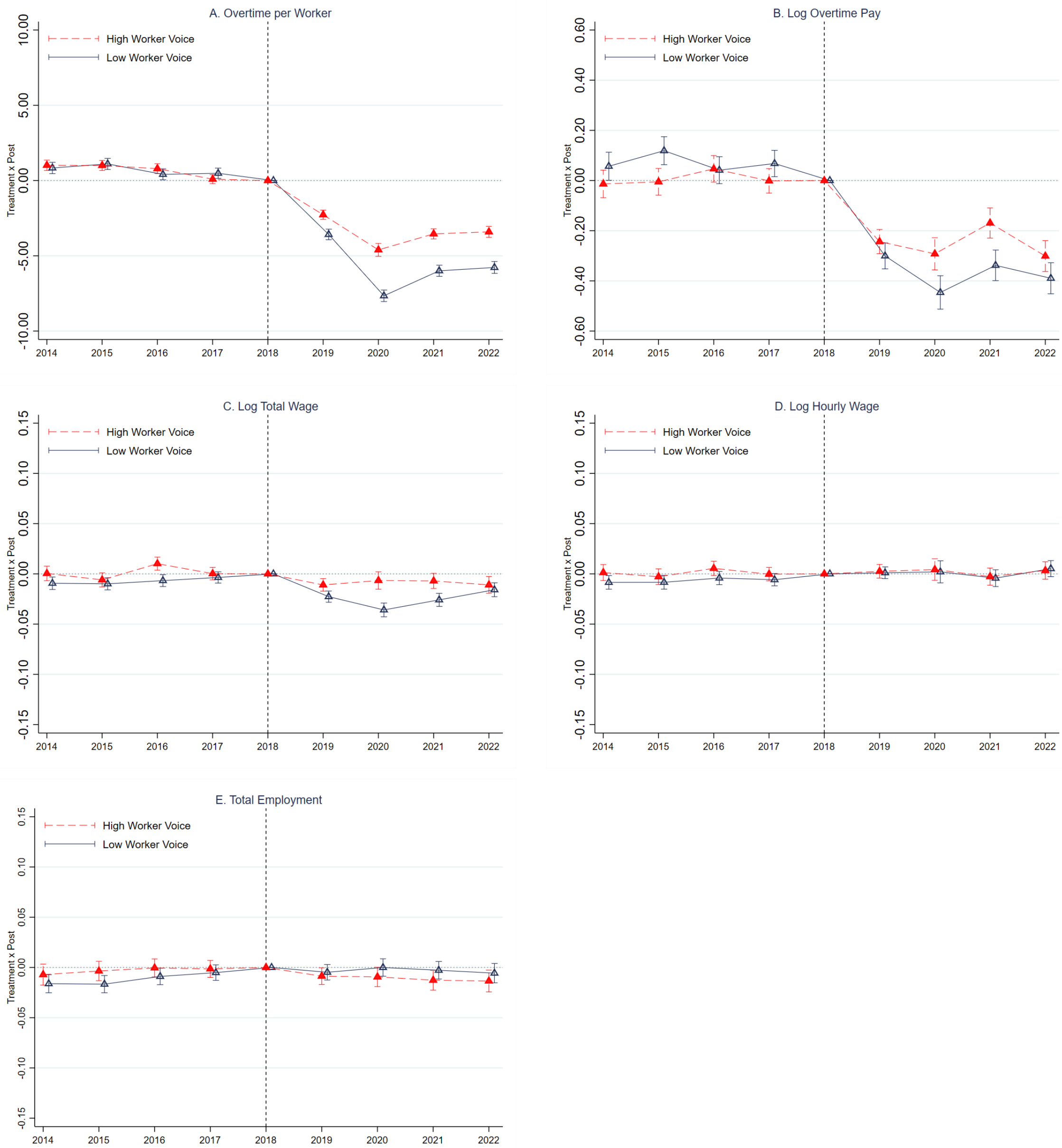
Notes: This table shows summary statistics of the estimation samples: establishments (Panel A) and individuals (Panel B). Sources: Basic Survey Wage Structure (BSWS) and Osaka Preference Parameter Study (OPPS).

Table A3: DiD estimates: Anatomy of Overtime Hours Changes Within Establishments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	% Overt0hrs	% Overt1-10hrs	% Overt11-20hrs	% Overt21-30hrs	% Overt31-40hrs	% Overt41-50hrs	% Overt51-60hrs	% Overt60hrs+
A. All workers								
High Exposure × Post	0.057***	0.050***	0.016***	-0.012***	-0.050***	-0.028***	-0.014***	-0.019***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Observations	316,744	316,744	316,744	316,744	316,744	316,744	316,744	316,744
Mean of outcome	0.249	0.181	0.142	0.131	0.126	0.0779	0.0404	0.0524
B. Male workers								
High Exposure × Post	0.055***	0.051***	0.025***	-0.004**	-0.055***	-0.032***	-0.017***	-0.023***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	297,910	297,910	297,910	297,910	297,910	297,910	297,910	297,910
Mean of outcome	0.239	0.149	0.134	0.137	0.142	0.0906	0.0475	0.0618
C. Female workers								
High Exposure × Post	0.055***	0.038***	-0.005***	-0.022***	-0.036***	-0.016***	-0.007***	-0.007***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)
Observations	285,128	285,128	285,128	285,128	285,128	285,128	285,128	285,128
Mean of outcome	0.321	0.263	0.153	0.110	0.0789	0.0401	0.0174	0.0173

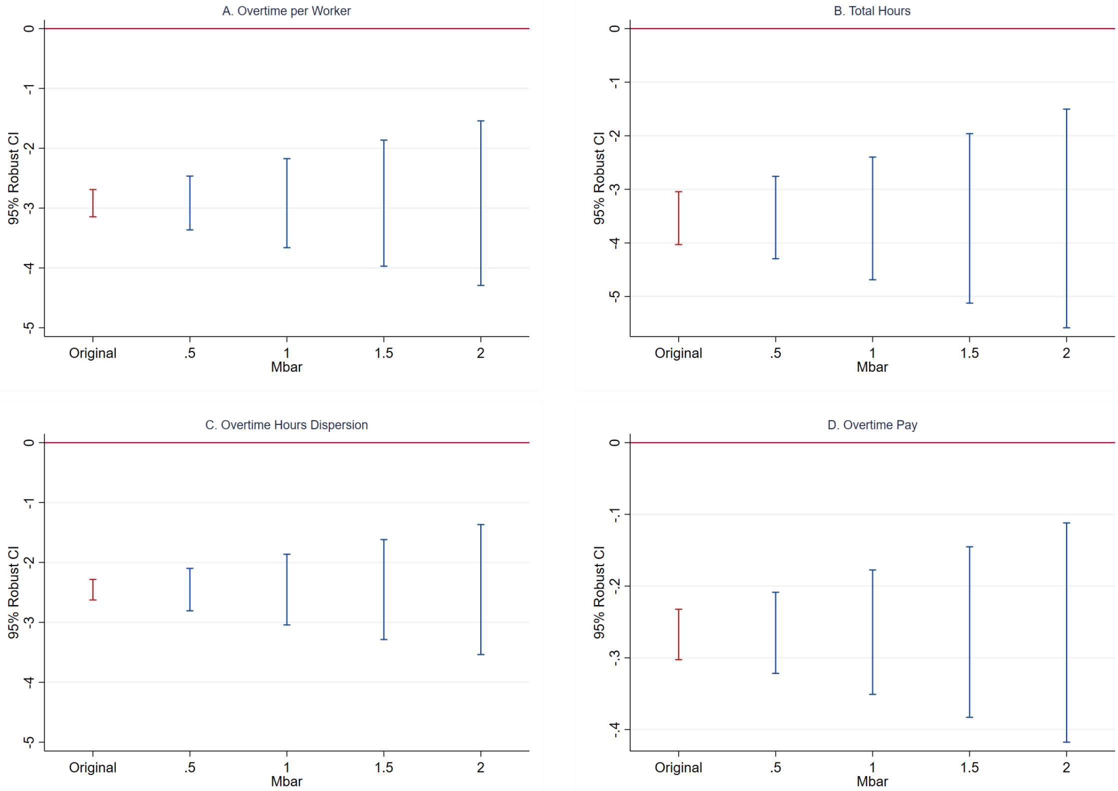
Notes: DiD estimates using establishment-level panel (2014-2022) from Basic Survey on Wage Structure (BSWS). Dependent variables: share of jobs of X overtime Hrs at the establishment level. Reported estimates include firm FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies. Standard errors clustered at the establishment level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01.

Figure A1: Heterogeneous Effects: Worker Voice Institutions



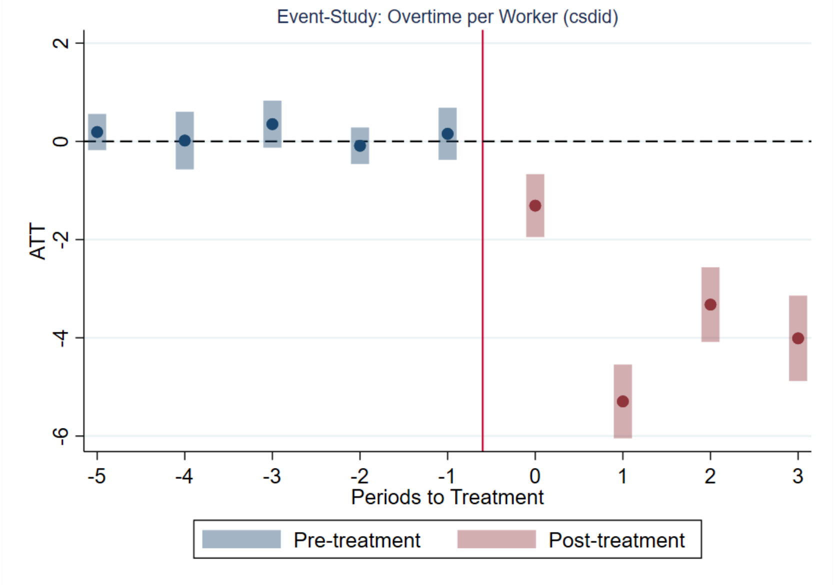
Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Regressions for establishments operating in high/low sectoral worker voice regimes are estimated separately. High (low) worker voice regimes are sectors with above-(below)the-median incidence of worker voice institutions at the workplace level (unions, labor-management committees, shop-floor committees) computed from SLMC. Dependent variables: Overtime per Worker (Panel A), Overtime Pay (Panel B), Total Monthly Earnings (Panel C), Hourly Wage (Panel D), and Total Employment (Panel E). Sample: BSWS collapsed at the establishment level (establishment-level panel 2014-2022). Reported estimates include firm and year-FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies. Standard errors clustered at the establishment level.

Figure A2: Honest Pre-Trends (Rambachan and Roth, 2023)



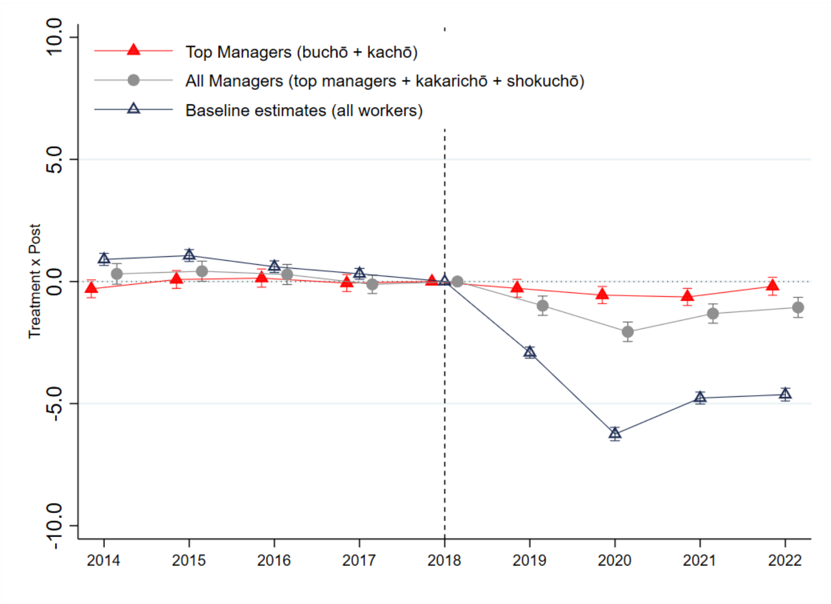
Notes: This Figure plots alternative estimated confidence intervals for η_{2019} (the first post-reform year-specific DiD effect). These 95% confidence intervals allow for deviations from parallel trend in the pre-reform period, following the method proposed by Rambachan and Roth (2023). They are calculated assuming that the post-reform violation of parallel trends is at most Mbar larger than the maximum violation of parallel trends in the pre-reform period. For example, Mbar equals to 2 means that the post-reform deviation is no more than twice that in the pre-reform period.

Figure A3: Staggered Treatment (Callaway and Sant'Anna, 2021): Overtime per Worker



Notes: Event-study analysis accounting for staggered treatment timing. Highly-exposed large firms treated from 2019 onward. Highly-exposed small-medium sized firms treated from 2020 onward. Small establishments employ not more than 300 employees (not more than 50 employees for retail businesses, not more than 100 for wholesale retail). We rely on the approach and Stata routine (*csdid*) proposed by Callaway and Sant'Anna (2021) using never-treated units as controls. using establishment-level panel (2014-2022) from Basic Survey on Wage Structure (BSWS). Reported estimates include controls for workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies, industry- and region-year effects. 95% confidence intervals, wild bootstrap-standard errors.

Figure A4: Placebo Outcome: Overtime among Non-Covered (Managerial) Occupations



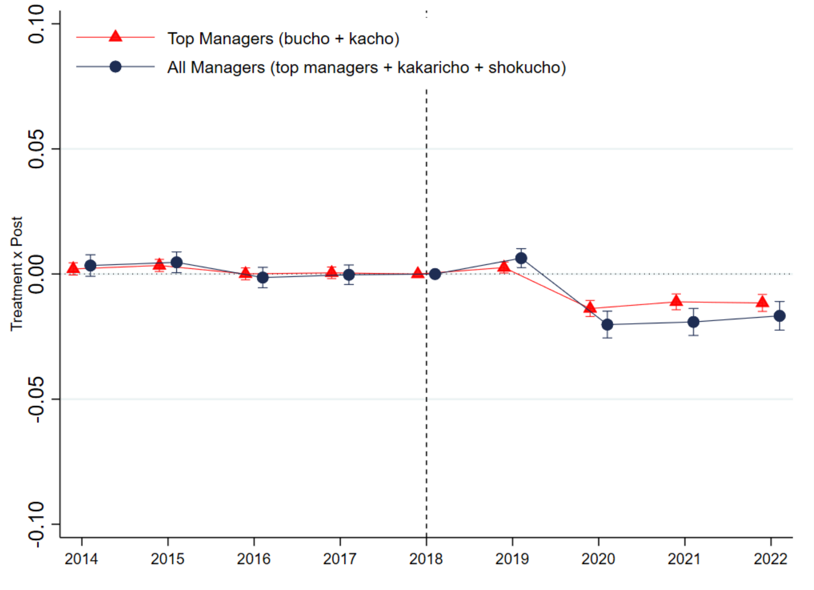
Notes: The figure year-specific DiD estimates and 95% confidence intervals. Dependent variables: Average of overtime of managerial workers at the establishment level. Sample: BSWS collapsed at the establishment level (establishment-level panel 2014-2022). Reported estimates include firm and year-FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, non-standard workers), firm-size dummies. Standard errors clustered at the establishment level.

Table A4: DiD estimates: Alternative Treatment Indicators

	(1) Mean Overtime	(2) Overtime Dispersion (sd)	(3) Overtime Pay	(4) Monthly Wage	(5) Employment
High Exposure^a × Post	-5.639***	-3.943***	-0.228***	-0.015***	0.001
	(0.090)	(0.062)	(0.010)	(0.001)	(0.002)
Observations	316,924	308,101	260,025	316,889	316,924
High Exposure^b × Post	-6.654***	-2.315***	-0.156***	-0.018***	0.013***
	(0.112)	(0.079)	(0.010)	(0.002)	(0.003)
Observations	155,071	153,375	152,361	155,067	155,071
Share Treated × Post	-20.611***	-8.239***	-0.695***	-0.054***	0.022***
	(0.260)	(0.173)	(0.020)	(0.003)	(0.005)
Observations	316,924	308,101	260,025	316,889	316,924
Share Hrs. Treated × Post	-30.080***	-19.728***	-1.110***	-0.093***	0.022***
	(0.479)	(0.310)	(0.036)	(0.005)	(0.008)
Observations	263,593	258,884	253,857	263,585	263,593

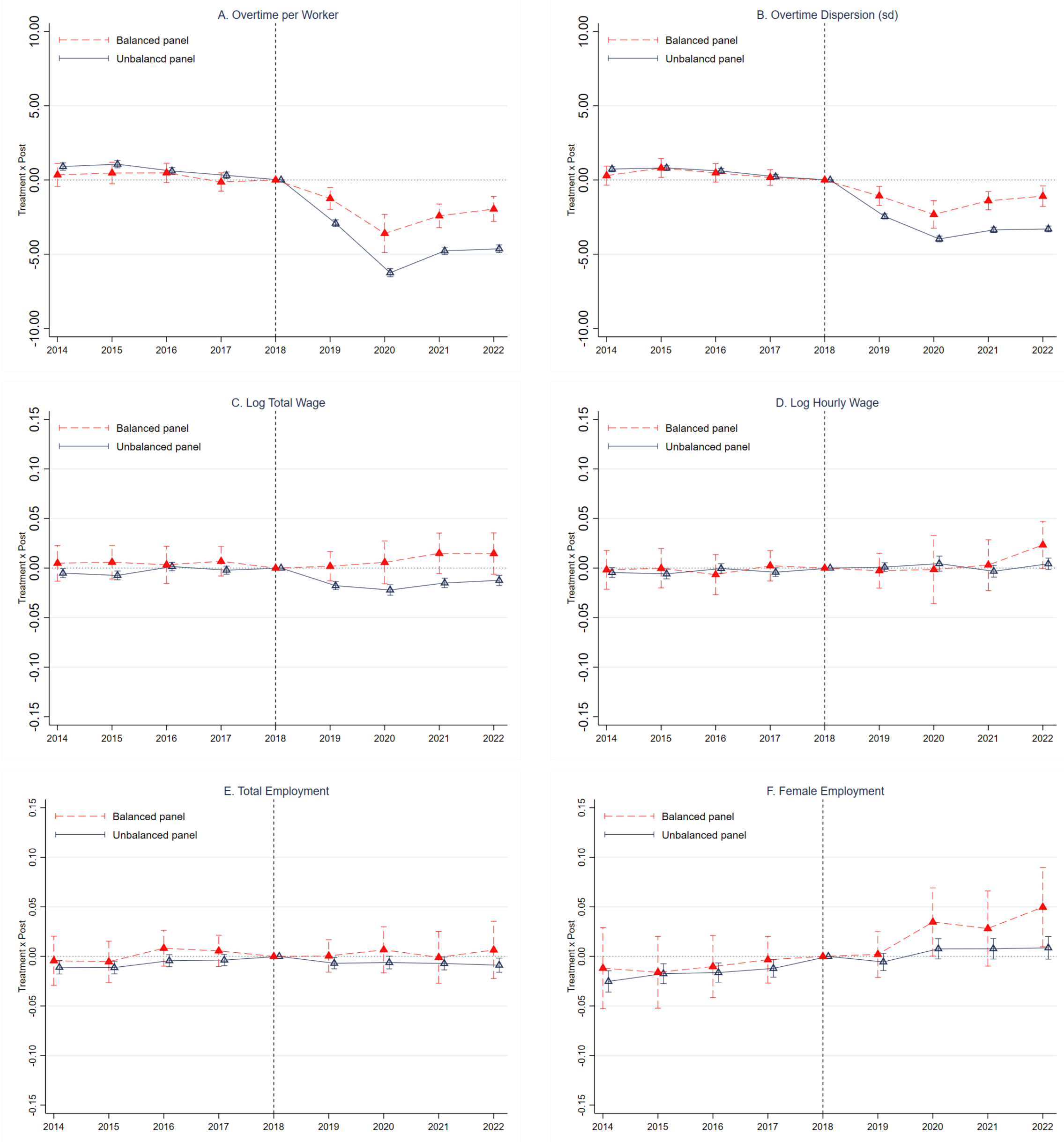
Notes: DiD estimates using establishment-level panel (2014-2022) from Basic Survey on Wage Structure (BSWS). $HighExposure^a$ considers establishments with a positive share of workers exceeding 45 overtime hours per month; $HighExposure^b$ compares establishments in the top quartile of the pre-reform distribution of the share of high-overtime employees to firms whose pre-reform share of high-overtime employees is between the 50th and the 75th percentile (*medium-high exposure*); *Shared Treated* is directly measured as the share of workers supplying overtime hours above the 30-hour cap; *Shared Hrs. Treated* is the fraction of firm-level overtime hours exceeding the 30-hour cap. Reported estimates include firm FE and controls for workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies. Estimates also control for industry and region-specific time trends. Standard errors clustered at the establishment level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01.

Figure A5: Job-title Reclassification: Share of (Managerial) Occupations



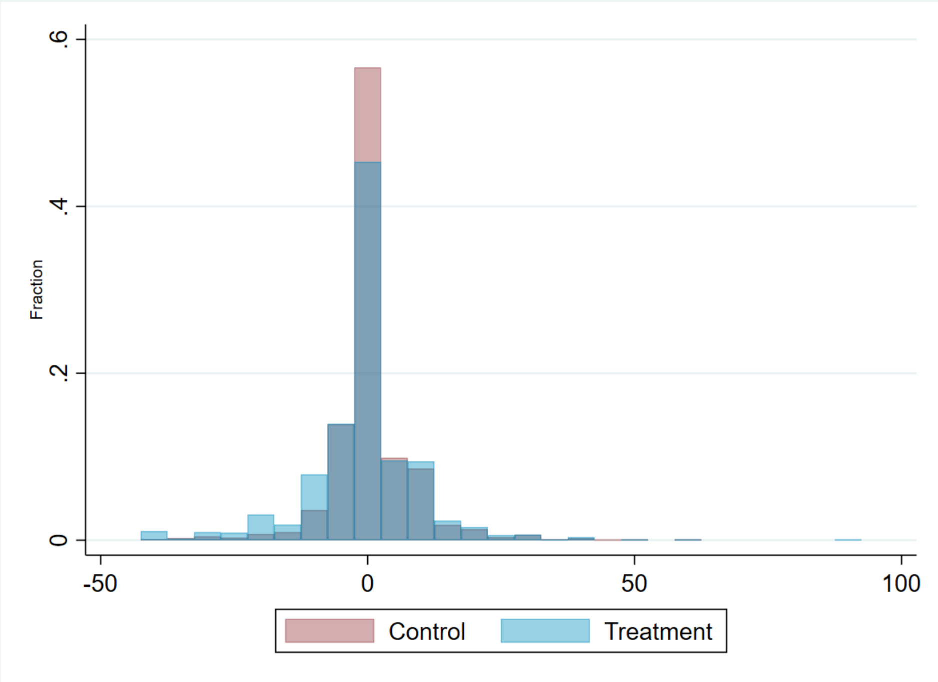
Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Dependent variables: Share of managers at the establishment level. Sample: BSWS collapsed at the establishment level (establishment-level panel 2014-2022). Reported estimates include firm and year-FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies. Standard errors clustered at the establishment level.

Figure A6: Balanced vs. Unbalanced Panel



Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Estimates compare balanced vs. unbalanced panel of BSWS establishments over period 2014-2022. Reported estimates include firm and year-FE, as well as industry- and region-year effects. Other controls: workforce composition (age, tenure, share of female workers, share of workers with college education, share of fixed-term contracts, nonstandard workers), firm-size dummies. Standard errors clustered at the establishment level.

Figure A7: Consistency Check for Self-Reported Working Hours in OPPS



Notes: Distribution of the difference between total working hours as reported by individuals in OPPS and total working hours assuming a standard workweek (40h) plus total overtime hours reported by individuals.

Table A5: Alternative individual-level treatment indicator: average pre-reform overtime

	Paid OT	Unpaid OT	Life sat.	Nonwork sat.
A. Total				
Treatment × Post	-4.572***	1.408***	-0.001	0.079
	(0.671)	(0.538)	(0.076)	(0.089)
Observations	4936	4981	5108	5157
B. Men				
Treatment × Post	-4.473***	1.970***	-0.047	0.016
	(0.718)	(0.607)	(0.080)	(0.095)
Observations	3425	3454	3571	3596
C. Women				
Treatment × Post	-7.648***	-0.125	0.835***	0.881***
	(2.142)	(1.889)	(0.174)	(0.262)
Observations	1511	1527	1537	1561

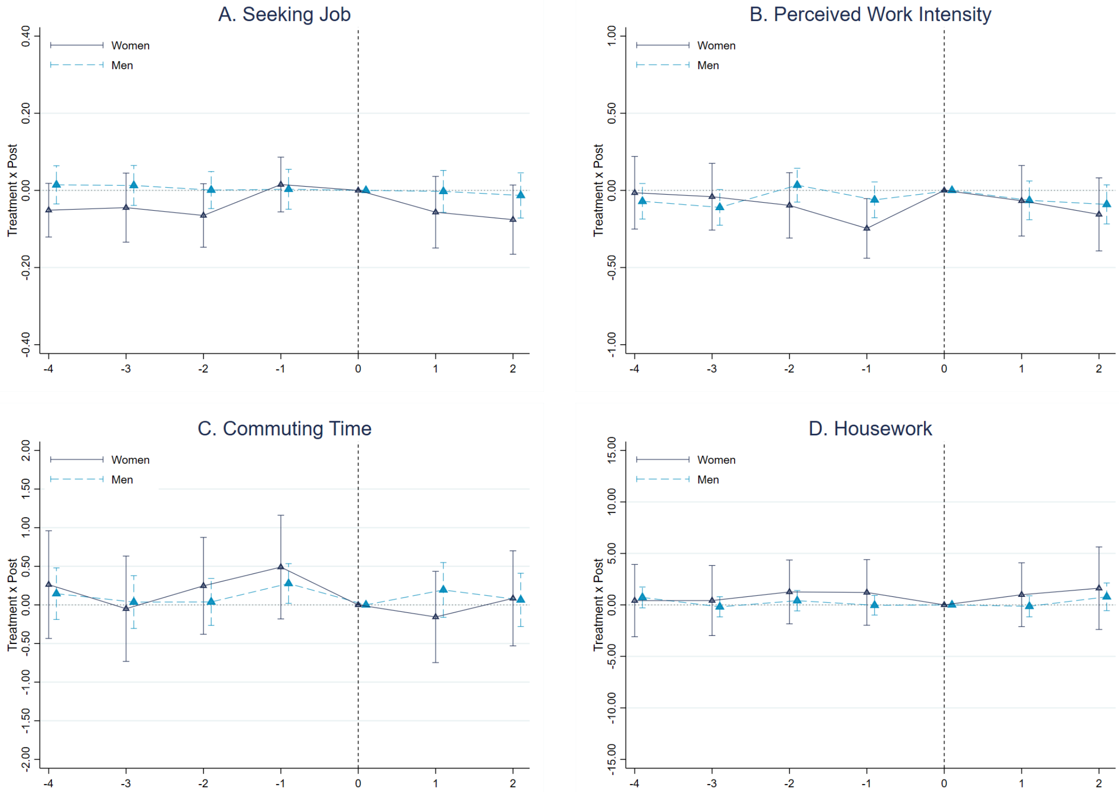
Notes: DiD estimates by gender and 95% confidence intervals. Sample: Individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies and sector fixed effects. Standard errors are clustered at the individual level.

Table A6: Individual-level DiD Estimates: Robustness checks

	(1) Total Hours	(2) Long Hours	(3) Paid Overtime
A. Excluding inconsistent reporting			
Treatment × Post	-1.832**	-0.048	-2.400***
	(0.927)	(0.037)	(0.595)
Observations	1,854	1,854	1,854
B. Trimming top 2% total hours			
Treatment × Post	-2.408***	-0.089***	-3.483***
	(0.672)	(0.023)	(0.427)
Observations	4,957	4,957	4,804
C. Job Stayers			
Treatment × Post	-1.383*	-0.073***	-3.465***
	(0.832)	(0.028)	(0.493)
Observations	3,296	3,296	3,192
D. Balanced panel			
Treatment × Post	-1.651	-0.094***	-2.704***
	(1.027)	(0.034)	(0.583)
Observations	2,207	2,207	2,138
E. Only Individuals Employed in Large firms			
Treatment × Post	-1.666	-0.110***	-4.410***
	(1.079)	(0.038)	(0.689)
Observations	1,969	1,969	1,897

Notes: DiD estimates using individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group comprises workers who were supplying more than 30 hrs. of paid overtime in a typical month before the reform. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies and sector fixed effects. Standard errors clustered at the individual level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01

Figure A8: Event-Study Analysis: Additional Results from OPPS



Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Dependent variables: On-the-Job Search (Panel A), Perceived Work Intensity (Panel B), Commuting Time (Panel C), Housework Time (Panel D). Sample: Individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). Reported estimates include individual and year-FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies and sector effects. Standard errors are clustered at the individual level.

Table A7: DiD estimates excluding Work from Home (WFH) employees

	Paid OT	Unpaid OT	Life sat.	Nonwork sat.
A. Total				
Treatment × Post	-3.613***	0.572	0.031	0.033
	(0.455)	(0.457)	(0.058)	(0.073)
Observations	4835	4877	5003	5050
B. Men				
Treatment × Post	-4.099***	1.024*	-0.018	-0.067
	(0.562)	(0.601)	(0.067)	(0.086)
Observations	3369	3395	3511	3535
C. Women				
Treatment × Post	-2.364**	-0.049	0.249**	0.347**
	(1.009)	(0.719)	(0.115)	(0.163)
Observations	1466	1482	1492	1515

Notes: DiD estimates using individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. WFH employees are excluded. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group comprises workers who were supplying more than 30 hrs. of paid overtime in a typical month before the reform. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies, sector fixed effects and region-specific time trends. Standard errors clustered at the individual level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01

Table A8: DiD estimates: Other Subjective Well-Being Facets

	(1) Fulfilling life	(2) Happiness	(3) Health anxiety	(4) Feeling stressed	(5) Feeling depressed	(6) Sleep problems
Treat $\times$ Post	0.017	-0.074	-0.027	-0.095	-0.040	0.026
	(0.065)	(0.126)	(0.081)	(0.081)	(0.091)	(0.081)
Treat $\times$ Post $\times$ Female	0.149	0.511**	-0.113	-0.014	-0.064	-0.138
	(0.131)	(0.252)	(0.151)	(0.183)	(0.189)	(0.152)
TE women	0.167	0.437	-0.139	-0.109	-0.103	-0.112
p-value: TE women	0.144	0.045	0.269	0.504	0.532	0.382
Mean outcome (men)	3.464	6.516	3.075	3.402	2.790	2.271
Mean outcome (women)	3.581	6.620	3.217	3.480	2.840	2.116
Observations	5135	5096	5133	5135	5136	5137

Notes: DiD estimates using individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group comprises workers who were supplying more than 30 hrs. of paid overtime in a typical month before the reform. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies, sector fixed effects and region-specific time trends. Standard errors clustered at the individual level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01

Table A9: DiD estimates: Perceived Work Intensification and Time Use

	(1) Work intensity	(2) Commuting weekly	(3) Housework weekly	(4) Physical activity
Treat × Post	-0.017	0.038	-0.057	-0.033
	(0.040)	(0.138)	(0.447)	(0.112)
Treat × Post × Female	0.029	-0.280	0.375	0.360
	(0.086)	(0.311)	(1.396)	(0.260)
TE women	0.012	-0.242	0.318	0.327
p-value: TE women	0.877	0.368	0.808	0.165
Mean outcome (men)	0.443	3.317	4.579	1.986
Mean outcome (women)	0.393	3.095	15.340	1.500
Observations	5152	5043	4820	5135

Notes: DiD estimates using individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group comprises workers who were supplying more than 30 hrs. of paid overtime in a typical month before the reform. Commuting and housework time measured as total weekly hours (including weekends). Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies, sector fixed effects and region-specific time trends. Standard errors clustered at the individual level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01

Table A10: DiD estimates - Single vs. Married Respondents

	(1) Paid Overtime	(2) Life Satisfaction	(3) Leisure Satisfaction
A. Single respondents			
Treat × Post	-3.358**	0.047	-0.084
	(1.377)	(0.148)	(0.194)
Treat × Post × Female	-0.940	-0.011	0.513*
	(2.256)	(0.223)	(0.295)
TE women	-4.297***	0.036	0.429*
p-value: TE women	0.007	0.845	0.086
Mean outcome (men)	8.270	3.376	3.451
Mean outcome (women)	7.307	3.627	3.285
Observations	1260	1299	1325
B. Married respondents			
Treat × Post	-4.422***	-0.067	-0.065
	(0.614)	(0.070)	(0.094)
Treat × Post × Female	3.478**	0.417***	0.362
	(1.507)	(0.156)	(0.227)
TE women	-0.944	0.350**	0.296
p-value: TE women	0.485	0.013	0.150
Mean outcome (men)	8.154	3.639	3.415
Mean outcome (women)	6.000	3.590	3.137
Observations	3637	3770	3791

Notes: DiD estimates using individual-level panel from Osaka Preference Parameter Study (OPPS) restricted to full-time, non-managerial employees aged 20-65 years. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group comprises workers who were supplying more than 30 hrs. of paid overtime in a typical month before the reform. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies, sector fixed effects and region-specific time trends. Standard errors clustered at the individual level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01

Table A11: Distribution of Respondents' Treatment, by Spouse's Treatment

A. Men		
	Wife not treated	Wife treated
Non-treated	52.84	30.40
Treated	47.16	69.60
Total	100.00	100.00
B. Women		
	Husband not treated	Husband treated
Non-treated	84.66	72.28
Treated	15.34	27.72
Total	100.00	100.00

Notes: The sample includes full-time employed respondents aged 20-65 years with a cohabiting spouse. The interpretation of figures is as follows: 30.4% of male respondents with treated wives are not treated themselves, while 47% of male respondents with non-treated wives are treated. Treatment group comprises workers who were supplying more than 30h of paid overtime in a typical month before the reform.

Table A12: Difference-in-Differences Household Design: Direct, Spillover, and Joint Effects of the Overtime Cap

	(1) Paid Overtime	(2) Life Satisfaction	(3) Leisure Satisfaction
A. Married Men			
Couple-Only Respondent Treated $(H_{10}) \times$ Post	-3.903***	-0.050	0.085
	(1.503)	(0.135)	(0.184)
Couple-Only Partner Treated $(H_{01}) \times$ Post	-2.530*	0.146	-0.198
	(1.360)	(0.118)	(0.237)
Couple-Both Treated $(H_{11}) \times$ Post	-5.723***	-0.018	-0.270
	(1.667)	(0.155)	(0.201)
p-value: Additivity Test	0.757	0.547	0.600
Observations	1081	1123	1129
B. Married Women			
Couple-Only Respondent Treated $(H_{10}) \times$ Post	-1.495*	0.122	-0.152
	(0.861)	(0.180)	(0.307)
Couple-Only Partner Treated $(H_{01}) \times$ Post	-0.304	0.158	0.051
	(0.595)	(0.129)	(0.177)
Couple-Both Treated $(H_{11}) \times$ Post	0.107	0.471**	0.658***
	(2.192)	(0.221)	(0.221)
p-value: Additivity Test	0.445	0.495	0.052
Observations	784	805	810

Notes: Estimates from a DiD household design as specified in Equation (4). Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Only respondent treated refers to couples where only the main respondent was treated before the reform; only partner treated refers to couples where only the partner was treated before the reform; both treated indicates (concordant) couples where both the main respondent and partner are treated. Households where neither partner was treated are the omitted reference group. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies, sector fixed effects. Standard errors clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

Table A13: Separability Test: Single vs. Married Respondents with Never-Treated Partners

	(1) Paid Overtime	(2) Life Satisfaction	(3) Leisure Satisfaction
A. Men			
Treatment × Post	-4.406***	-0.025	-0.037
	(0.628)	(0.077)	(0.099)
Treatment × Post × Single	1.750	0.114	-0.131
	(1.478)	(0.180)	(0.225)
Observations	3,102	3,241	3,264
B. Women			
Treatment × Post	-1.596*	0.196	-0.148
	(0.885)	(0.179)	(0.312)
Treatment × Post × Single	-2.464	-0.135	0.455
	(1.813)	(0.248)	(0.392)
Observations	1,201	1,214	1,236

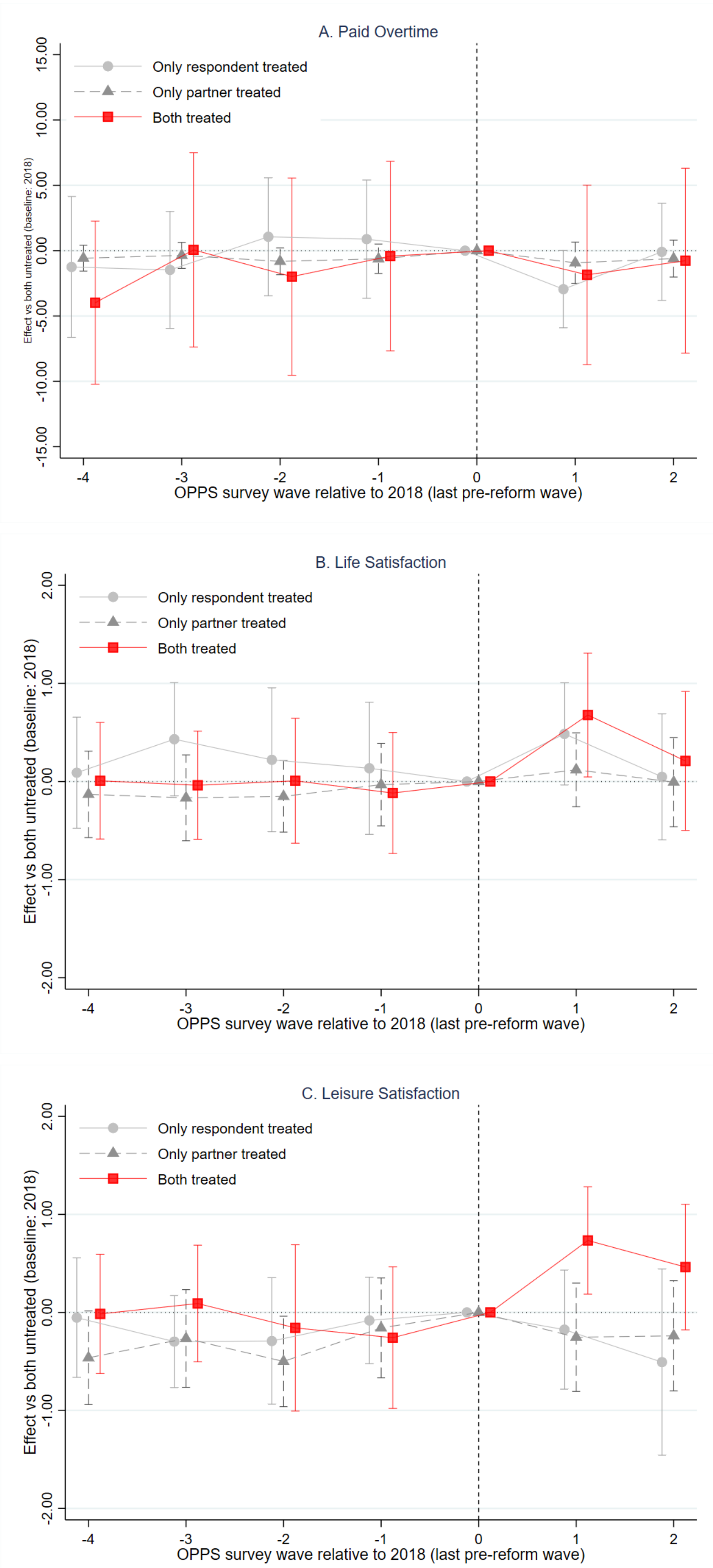
Notes: DiD estimates using individual-level panel from Osaka Preference Parameter Study (OPPS). Estimates from a specification explained in footnote 44. Sample restricted to single and married respondents with untreated partners (main respondents are full-time, non-managerial employees aged 20-65 years). Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group comprises workers who were supplying more than 30 hrs. of paid overtime in a typical month before the reform. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies and sector fixed effects. Standard errors clustered at the individual level and shown in parentheses. Significance levels: * 0.10, ** 0.05, *** 0.01

Table A14: Difference-in-Differences Household Design: Individual Time Use and Household Functioning (Married Women)

	A. Individual Time-Use Channels					B. Household-Level Mechanisms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Husband's OT	Wife's commuting hrs.	Husband's commuting hrs.	Wife's housework hrs.	Husband's housework hrs.	Perceived husband's work intensity	Spouse satisfaction
Couple-Only Respondent (wife) Treated $(H_{10}) \times$ Post	0.388	-0.478**	-0.180	-1.286	-1.052	-0.047	-0.022
	(1.345)	(0.204)	(0.532)	(1.819)	(2.403)	(0.117)	(0.281)
Couple-Only Partner (husband) Treated $(H_{01}) \times$ Post	-1.847**	-0.423**	-0.534	0.946	0.693	-0.014	0.229
	(0.822)	(0.189)	(0.365)	(1.876)	(1.120)	(0.113)	(0.204)
Couple-Both Treated $(H_{11}) \times$ Post	-5.251***	-0.388	0.450	0.969	1.410	-0.213*	0.443**
	(1.379)	(0.294)	(0.355)	(2.229)	(1.765)	(0.114)	(0.213)
Observations	720	807	784	778	761	818	809
p-value: Additivity Test	0.0633	0.179	0.0877	0.658	0.569	0.413	0.520

Notes: Estimates from a DiD household design as specified in Equation (4). Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group made of workers who worked more than 30h of paid overtime in a typical month. Only respondent treated refers to couples where only the main respondent was treated before the reform; only partner treated refers to couples where only the partner was treated before the reform; both treated indicates (concordant) couples where both the main respondent and partner are treated. Households where neither partner was treated are the omitted reference group. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies, sector fixed effects. Standard errors clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

Figure A9: Event-study Analysis: Married Women by Household Type



Notes: The figure displays year-specific DiD estimates and 95% confidence intervals. Sample restricted to married women who are full-time salary workers using a DiD household design as specified in Equation 4. Waves: Pre-reform (2012, 2013, 2016, 2017, 2018), Post-reform (2021, 2022). The post-reform variable equals 1 for years 2021-2022 (policy-on period), and 0 otherwise. Treatment group made of workers who worked more than 30h of paid overtime in a typical month (i.e. >360h per year). Only respondent treated refers to couples where only the main respondent was treated before the reform; only partner treated refers to couples where only the partner was treated before the reform; both treated indicates (concordant) couples where both the main respondent and partner are treated. Households where neither partner was treated are the omitted reference group. Reported estimates include individual FE. Other controls: age, age squared, tenure group dummies, occupation dummies, firm size class dummies, sector fixed effects. Standard errors clustered at the individual.