UNIVERSITAT D'ALACANT
UNIVERSIDAD DE ALICANTE
Facultat de Ciències Econòmiques i Empresarials
Facultad de Ciencias Económicas y Empresariales

MÁSTER EN ECONOMICS WITH DATA SCIENCE

CURSO ACADÉMICO 2025 - 2026

EVALUATING THE 2019 MANDATORY TIME-TRACKING REFORM IN SPAIN: A DIFFERENCE-IN-DIFFERENCES APPROACH

ALEJANDRO J. GARCÍA PÉREZ

MARIANNA BATTAGLIA
DEPARTAMENTO DE FUNDAMENTOS DEL ANÁLISIS ECONÓMICO

San Vicente del Raspeig, 7 de septiembre de 2026.

Evaluating the 2019 Mandatory Time-Tracking Reform in Spain: A Difference-in-Differences Approach

Alejandro José García Pérez

September 9, 2026

Abstract

This study evaluates the impact of Real Decreto-ley 8/2019, which mandated strict time-tracking for full-time employees in Spain to curb unpaid overtime and excessive working hours. Using a Difference-in-Differences approach on microdata from the Spanish Labor Force Survey (EPA) between 2018 and 2020, this paper identifies the short-run causal effect of the policy. By exploiting pre-reform variations in overtime intensity, the findings reveal that the reform led to a major structural decline in total overtime hours among highly exposed segments, falling by 77.1% relative to the control group's mean (-0.168 hours/week). Crucially, this contraction was primarily driven by a widespread shift of workers reducing their uncompensated hours to exactly zero, effectively eliminating unpaid overtime (-0.100 hours/week), while paid overtime experienced a more moderate decrease of 55.0% (-0.071 hours/week). These adjustments contributed to an overall reduction of 1.1 actual working hours per week (-3.5%). Robustness tests and subgroup analyses confirm that these reductions are structurally consistent across contract types, age brackets, and educational levels.

1 Introduction

In Spain, the prevalence of unpaid overtime work has long been a structural challenge in the labor market. There is a persistent belief among firms that longer working hours signal higher productivity and commitment. Consequently, millions of hours worked remain uncompensated, generating severe negative externalities: workers suffer a deterioration in their work-life balance, and the public administration faces significant tax revenue and social security losses.

The economic literature has long debated whether state interventions actually succeed in reducing working hours and curbing unpaid overtime. Classic empirical studies, such as Hunt (1999), show that while changes in working time policies affect actual hours worked, their success largely depends on law enforcement. In fact, these studies warn that in rigid labor markets, simply mandating shorter hours without proper monitoring can backfire, leading to more evasion and unrecorded overtime.

In the context of the Spanish labor market, recent literature has examined the feasibility of tightening working time regulations. For instance, Conde-Ruiz, Lahera, and Viola (2024) analyze proposed statutory reductions in the workweek, emphasizing a crucial point: without strict time-tracking and enforcement, policies to reduce working hours are likely to fail. Instead of working less, employees might just end up doing more unpaid overtime to absorb the reduction. Furthermore, they point out the deeply rooted culture of "presenteeism" in Spain, where formal laws often struggle to change daily workplace habits.

To directly address this enforcement gap and combat these structural inefficiencies, the Spanish Government introduced the Real Decreto-ley 8/2019 in May 2019. This legislation mandated all companies to strictly record the daily working hours of their full-time employees. Part-time workers were excluded from this specific reform, as their working hours were already subject to mandatory tracking under previous regulatory frameworks. This paper pursues two main objectives: first, to evaluate whether this legislation has affected the actual weekly hours worked per employee; and second, to determine if it has successfully decreased the volume of unpaid overtime hours, particularly within segments of the labor market that historically exhibited the highest incidence of uncompensated extra hours.

To achieve these objectives, this paper employs a Difference-in-Differences framework using microdata from the Spanish Labor Force Survey (EPA) between 2018 and 2020. By exploiting the variation in pre-reform overtime intensity across different labor market segments, I isolate the causal effect of the mandate. The estimates indicate that the introduction of the mandate was followed by a reduction in excessive working hours, particularly unpaid work. The most striking evidence of this success is the collapse of total extra hours, which dropped by 77.1% among highly exposed workers relative to the baseline mean. Crucially, the policy effectively eradicated unpaid overtime relative to the control group mean, a result driven by a substantial increase in workers reporting zero uncompensated hours after the reform, while also reducing paid extra hours by 55.0%. Driven by this notable contraction in overtime, the treated group also experienced a significant overall reduction of approximately 1.1 actual working hours per week. These findings remain strictly robust across multiple alternative specifications, including different treatment definitions and institutional control groups, and demonstrate a remarkably uniform impact across most worker demographics. The notable exception is immigrant workers, highlighting both the statistical and enforcement complexities of applying labor mandates in more vulnerable market segments.

Through these findings, this paper contributes to the ongoing debate on working time regulations. While recent studies like Conde-Ruiz et al. (2024) focus on the theoretical and policy implications of limiting working time, this study measures how firms and workers actually react to a real enforcement shock. By analyzing the impact on actual working hours, alongside both paid and unpaid overtime, this research demonstrates that forcing companies to strictly record employee hours is an effective tool to overcome informal evasion practices.

Importantly, while the empirical setting is the Spanish labor market, the implications of these findings could be highly generalizable. The Spanish reform aligns perfectly with the 2019 European Court of Justice ruling (Case C-55/18), which mandates all EU Member States to implement objective time-measuring systems to prevent unpaid overtime. Recent evidence from other rigid European labor markets underscores the relevance of this mandate; for instance, Nold and Backhaus (2022) examine the German context, highlighting how formalized working time recording fundamentally alters the incidence and compensation of overtime, moving the labor market away from informal, unrecorded hours. Consequently, Spain may serve as a pioneering natural experiment for a policy shift that affects the entire European Union. Furthermore, similar recent enforcement shocks in other advanced economies, such as Japan's 2019 Work Style Reform Act, demonstrate that strict time-tracking mechanisms are increasingly viewed globally as essential tools to curb systemic presenteeism. Thus, this paper provides broader empirical evidence on how institutional enforcement mechanisms influence corporate compliance and working conditions in modern labor markets.

The remainder of the paper is organized as follows. Section 2 describes the data and provides descriptive statistics. Section 3 outlines the Difference-in-Differences methodology and its key identifying assumptions. Section 4 presents the baseline results, followed by robustness checks in Section 5. Section 6 explores heterogeneous effects across different worker characteristics, and Section 7 concludes.

2 Data

The empirical analysis of this study relies on microdata from the Spanish Labor Force Survey (EPA), conducted quarterly by the Spanish National Statistics Institute (INE). The EPA follows a rotating panel design where participating households are surveyed for a maximum of six consecutive quarters. Furthermore, unlike empirical literature that assumes a random sample of strictly balanced panel data, my design is specifically adapted to the unbalanced panel structure of the Spanish Labor Force Survey. To accurately capture the causal effect of the mandatory time-tracking legislation (Real Decreto-ley 8/2019), the analysis is restricted to the period between 2018 and 2020, excluding data from 2021 onwards due to a major methodological change introduced in the Spanish Labor Force Survey (EPA). This timeframe provides an observation window of six quarters before the reform was introduced, May 2019, and six quarters after its implementation. The estimation sample comprises all wage workers from both the public and private sectors and thus, exclude self-employed workers, who are out of the sample because they are exempt from the time-tracking mandate.

Table 1 provides the descriptive statistics of the study data, illustrating the baseline characteristics of the treatment and control groups prior to the reform. For the main specification, the sample is divided based on the median of the pre-reform overtime intensity. This intensity measure is defined as the historical average of overtime hours worked within specific labor market cells, which are constructed by interacting each worker's economic sector with their occupation. The control group comprises individuals in segments with an overtime rate below the median, while the treatment group includes those highly exposed to the regulation, situated above the median (see the Methodology section for the formal setup). Notably, as confirmed by the baseline data, the treated group is effectively more exposed to the regulatory change, displaying a significantly higher intensity of working time. Prior to the legislation, treated individuals worked almost four actual hours more per week than the control group (34.6 versus 30.6 hours). More importantly for the main hypothesis, workers in the treated group performed more than twice the amount of total overtime hours of the control group. This gap is particularly pronounced in unpaid overtime hours (0.218 versus 0.094 hours), confirming their proper assignment to the high-intensity group.

Beyond working hours, the summary statistics reveal profound structural and socio-demographic differences linked to overtime. The control group is highly feminized (61.9% female), whereas the treated group shows a higher proportion of men (64.0%), suggesting a clear gender dimension in the prevalence of unpaid overtime. Regarding age distribution, the sample is divided into three mutually exclusive brackets: young workers (16 to 29 years), middle-aged workers (30 to 44 years), and older workers (45 to 65 years). This specific stratification ensures a sufficiently large sample size across all generations, granting the model the necessary statistical power for subsequent estimations. As can be seen in Table 1, treated individuals are younger than workers in the control group. Additionally, unlike the control group where higher education is predominant (51.2%), treated individuals display a higher concentration of low educational attainment (38.4% compared to 25.4% in the control group). They are also disproportionately concentrated in the industry and construction sectors.

Table 1: Descriptive Statistics and Covariate Balance (Pre-Reform)

Variable	**Control**	**Treated**	**Difference**	**Cohen's d**
Working Hours				
Actual Hours	30.638	34.592	3.953***	0.205
	(13.936)	(13.351)	(0.000)	
Total Overtime Hours	0.218	0.465	0.247***	0.084
	(1.744)	(2.386)	(0.000)	
Paid Overtime Hours	0.129	0.258	0.129***	0.057
	(1.407)	(1.750)	(0.000)	
Unpaid Overtime Hours	0.094	0.218	0.124***	0.062
	(1.070)	(1.689)	(0.000)	
Demographics				
Male	0.381	0.640	0.258***	0.378
	(0.486)	(0.480)	(0.000)	
Female	0.619	0.360	-0.258***	-0.378
	(0.486)	(0.480)	(0.000)	
Immigrant	0.058	0.076	0.018***	0.050
	(0.234)	(0.265)	(0.000)	
16 to 29 years	0.106	0.158	0.052***	0.108
	(0.308)	(0.365)	(0.000)	
30 to 44 years	0.350	0.419	0.069***	0.101
	(0.477)	(0.493)	(0.000)	
45 to 65 years	0.544	0.423	-0.121***	-0.173
	(0.498)	(0.494)	(0.000)	
Labor Contract				
Temporary	0.261	0.250	-0.011***	-0.018
	(0.439)	(0.433)	(0.000)	
Permanent	0.739	0.750	0.011***	0.018
	(0.439)	(0.433)	(0.000)	
Fixed-Discontinuous	0.017	0.024	0.006***	0.032
	(0.131)	(0.152)	(0.000)	
Education Level				
Low Education	0.254	0.384	0.130***	0.200
	(0.435)	(0.486)	(0.000)	
Medium Education	0.234	0.239	0.005***	0.008
	(0.424)	(0.427)	(0.001)	
High Education	0.512	0.377	-0.135***	-0.194
	(0.500)	(0.485)	(0.000)	
Sector				
Agriculture	0.053	0.003	-0.050***	-0.217
	(0.223)	(0.053)	(0.000)	
Industry	0.037	0.291	0.254***	0.515
	(0.190)	(0.454)	(0.000)	
Construction	0.009	0.100	0.091***	0.291
	(0.092)	(0.300)	(0.000)	
Services	0.901	0.606	-0.295***	-0.515
	(0.298)	(0.489)	(0.000)	
Observations	**155,918**	**149,052**	**304,970**	

Notes: This table presents the baseline characteristics for the control and treated groups prior to the reform. Standard deviations are reported in parentheses below the means, and p-values from a two-sample t-test are reported below the differences. The Cohen's d column reports the standardized mean difference to account for the large sample size. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Crucially, assessing covariate balance between these two groups requires contextualizing traditional inferential metrics. With a pre-reform cross-section of $N = 304,970$, classical two-sample t-tests are highly sensitive, yielding p-values below 0.01 for almost all variables (as shown in the third column of this table), even when the economic magnitude of the difference is negligible. To address this mechanical issue, Table 1 reports the standardized mean difference (Cohen's d). While variables such as contractual arrangements and certain age brackets show excellent balance ($d < 0.20$), the structural differences in gender and economic sectors exhibit larger standardized differences. These baseline imbalances do not invalidate the research design; rather, they justify the inclusion of strict socio-demographic controls and population weights in the main regressions to ensure that the estimated treatment effects are not driven by these compositional disparities.

Finally, Table A.2 in the Appendix describes the full temporal evolution of treated and control groups over time. The most salient finding pertains to the evolution of working hours, providing preliminary descriptive evidence of the reform's impact. While actual weekly hours slightly decreased across the entire sample, likely reflecting broader macroeconomic trends during the period, the trajectories of overtime hours diverged significantly between the two groups. For the control group, total and unpaid overtime hours experienced an upward trend. Conversely, the treated group saw a marked reduction in total overtime, driven primarily by a notable drop in unpaid overtime hours (from 0.218 to 0.169). This initial descriptive divergence aligns perfectly with the core objective of Real Decreto-ley 8/2019 to uncover and reduce uncompensated work.

Beyond working hours, the demographic and educational compositions of both groups exhibit parallel structural shifts over time, such as a general aging of the workforce and a systematic increase in the proportion of highly educated workers. As in the previous table, I evaluate the magnitude of these compositional changes using the standardized Cohen's d metric. As reported in the table, the absolute values of Cohen's d for all socio-demographic, educational, and sectoral variables remain strictly below the 0.10 conventional threshold for a negligible effect size. Even the most notable asymmetrical adjustment, where the treated group experienced a reduction in temporary contracts paired with an increase in permanent employment, yields a Cohen's d of just -0.071. This confirms that the overall structural composition of the groups remained remarkably stable over time, reinforcing the validity of the empirical design before moving into the formal econometric analysis.

3 Methodology

The canonical framework for Difference-in-Differences, as defined by Baker et al. (2025), consists of two time periods, one before and one after treatment, and two groups: one that remains untreated in both periods and one that becomes treated in the second period. Using these basic ingredients, I can define a 2×2 DiD design, which is composed of a causal target parameter, a treatment variable, an assumption under which it is identified, and an estimation approach based on the classic difference of two differences.

$$Y_{it} = \beta_0 + \beta_1 \text{Treated}_i + \beta_2 \text{Post}_t + \beta_3 (\text{Treated}_i \times \text{Post}_t) + \theta X_{it} + \delta_t + \varepsilon_{it} \tag{1}$$

First, I define a treatment group dummy variable, Treated_i. To operationalize the treatment assignment, I first classify the labor force into 90 distinct labor market cells by interacting each worker's economic sector with their specific occupation. Next, I calculate the average pre-reform overtime intensity rate for each of these 90 cells. Finally, each worker is assigned the intensity of their specific group. Specifically, I assign treatment status using the median of this intensity variable: the control group comprises individuals in segments with an overtime rate below the median (those minimally exposed to the policy), while the treatment group consists of individuals with an intensity rate above the median (those historically inclined to working overtime). The empirical application also contains a time dummy variable, Post_t, which takes the value of one for all observations in the periods following the implementation of the policy (from the third quarter of 2019 onwards), and zero for the preceding quarters. The effective treatment status, and hence, the parameter which measures the causal effect of the policy, is captured by the interaction term, which equals one exclusively for treated individuals during the post-reform quarters.

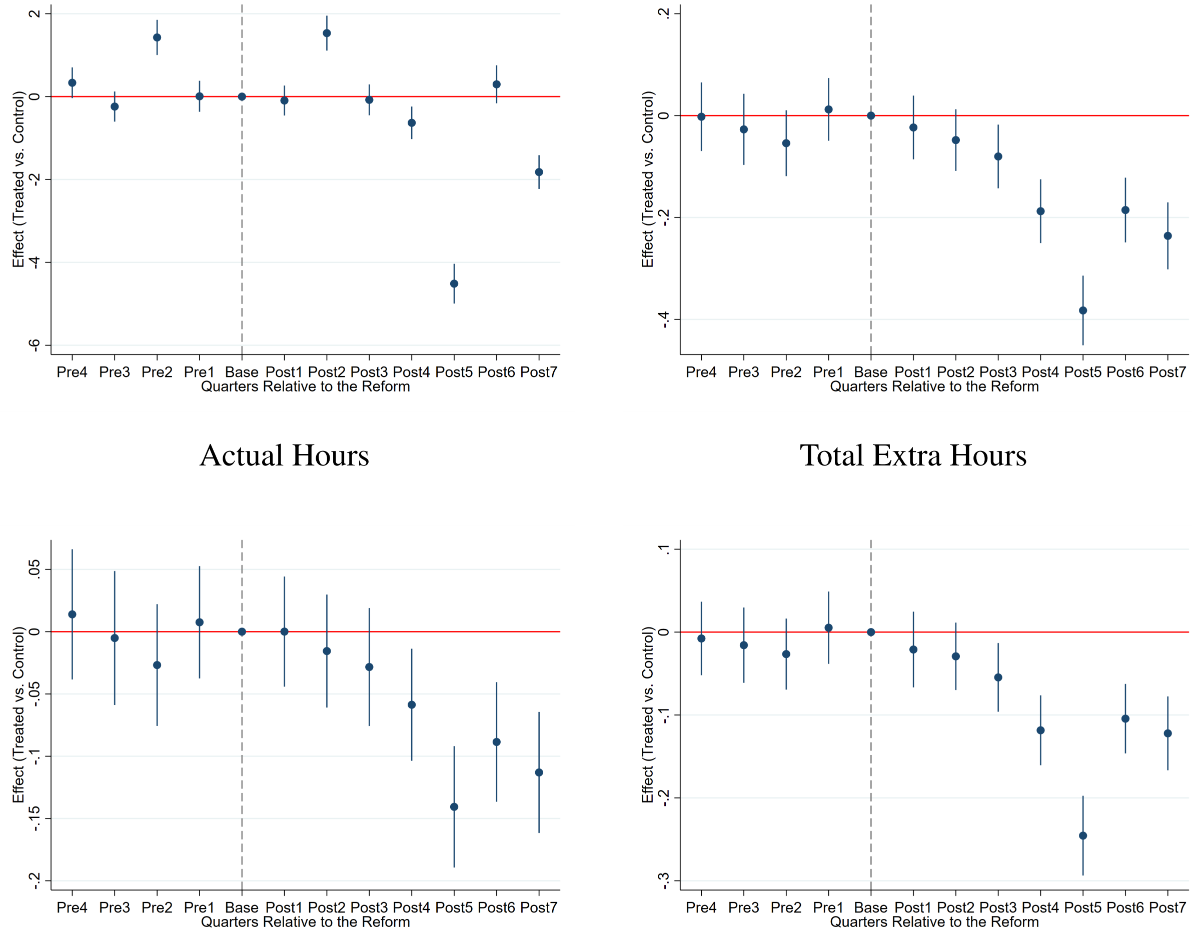
Additionally, X_{it} represents a vector of individual-level control variables, capturing socio-demographic and employment characteristics (such as age, gender, education level, nationality, and region) to account for observable differences across workers. The term δ_t denotes time fixed effects, which absorb unobserved aggregate shocks affecting all individuals simultaneously in a given quarter, such as the macroeconomic and health impacts of the COVID-19 pandemic. Finally, ε_{it} is the idiosyncratic error term.

To evaluate the impact of the reform, I estimate the expected values for four distinct dependent variables concerning working time: actual hours worked, total overtime hours, paid overtime hours, and unpaid overtime hours, conditional on the observable characteristics of the respondents. All baseline regressions are weighted by the population size of each granular group. Because the size of these 90 groups is highly unbalanced, population weights are implemented to prevent small, overrepresented segments from biasing the results (the complete distribution of these groups is provided in Table A.1 in the Appendix).

Under this framework, my causal target parameter is the Average Treatment Effect on the Treated (β_3), which isolates the law's precise impact on those workers effectively exposed to the regulation compared to those less exposed. The causal interpretation and internal validity of the Difference-in-Differences estimator strictly depend on fulfilling a set of structural assumptions. When utilizing microdata derived from an unbalanced panel, the econometric literature requires the satisfaction of four core identification conditions to guarantee that the estimated Average Treatment Effect on the Treated (ATT) reflects an unbiased causal effect. The theoretical framework and the formal definitions of these conditions presented throughout this section closely follow the methodological structure established by Felgueroso et al. (2026).

First, the **parallel trend assumption** requires that if the treatment had never happened, outcomes in the treated group would have followed the exact same trend as in the control group. It is perfectly fine if the treated and control groups have different average overtime hours before the law. The model is still valid as long as the gap between them stays constant over time. In other words, their past trends must be parallel.

I provide evidence in favour of this assumption by estimating an Event Study design using quarterly EPA data. Instead of using a single post-treatment indicator, I interact the treatment dummy with indicators for each quarter before and after the reform, using the quarter immediately preceding the legislation ($t-1$) as the omitted baseline category. This approach allows us to visually examine the pre-reform trends. As shown in Figure 1, almost all pre-reform coefficients are close to zero and not statistically significant. The only clear exception is the second quarter before the reform ($t-2$) for actual working hours. While this jump could just be statistical noise in the data, it might also point to some early adjustment by firms that started changing their schedules, anticipating the new law a few months before it was officially passed. Even with this specific deviation in $t-2$, the overall pre-trend remains reasonably stable for treated and control groups. This confirms that both groups were moving in the same direction before the law was applied for the outcomes studied, proving the validity of my control group and ensuring that my DiD model provides a reliable causal estimation. Furthermore, the post-treatment coefficients capture the dynamic evolution and timing of the policy's impact. Specifically, the estimates reveal an immediate and sustained downward trajectory in extra hours following the mandate.



Actual Hours

Total Extra Hours

Paid Extra Hours

Unpaid Extra Hours

Figure 1: Event Study Estimates for Working Time Variables

Second, the **no anticipation assumption** requires that economic agents did not systematically adjust their behavior in response to the upcoming legislation prior to its official implementation. If companies had preemptively modified schedules or reduced working hours in preparation for the 2019 time-tracking mandate, the pre-reform baseline would be endogenous, potentially biasing the Difference-in-Differences estimates.

A visual inspection of the Event Study estimates (Figure 1) largely attenuates concerns regarding early reactions. Across the overtime models, the pre-reform coefficients remain stable and close to zero. The only notable exception is a statistically significant deviation observed in the second quarter prior to the reform, specifically within the actual working hours model. While this isolated jump could reflect a minor early adjustment by certain firms anticipating the new regulatory environment, or simply statistical noise within the survey, the overall pre-treatment trajectory does not exhibit a continuous or systematic anticipatory trend. Consequently, the visual evidence broadly supports the assumption that the baseline data reliably captures the pre-reform labor market dynamics without severe contamination from early behavioral shifts.

Third, the **stable composition condition** is an extra requirement I must meet because I am using a pooled unbalanced panel, unlike studies that follow the exact same people forever (balanced panels). Since the EPA survey constantly rotates its participants, I do not observe the exact same workers before and after the reform. Because of this, I must ensure that the profile of the workers in each group remains stable over time. If the 2019 time-tracking law caused companies to change the type of workers they hire, any changes I see in overtime hours might just be because of this new workforce, not the law itself. This is known as a composition bias.

To solve this and ensure my results are accurate, I rely on a "Conditional Stable Composition". In practice, this means I explicitly include a large set of control variables in my regressions, such as education level, contract type, age, and nationality. By controlling for these characteristics, I neutralize any shifts in the workforce composition and isolate the true effect of the law.

Table 2: Test of Stable Composition: Labor, Education, and Demographics

Panel A: Labor & Education Characteristics					
	(1) Temp.	(2) Discon.	(3) Low Educ.	(4) Med. Educ.	(5) High Educ.
Treated × Post	-0.027***	0.000	-0.005***	0.004*	0.001
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Panel B: Demographics					
	(1) Male	(2) Immig.	(3) 16-29 yrs	(4) 30-44 yrs	(5) 45-65 yrs
Treated × Post	-0.003	0.004***	-0.007***	0.000	0.007***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)
N	595,661	595,661	595,661	595,661	595,661

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the results of the stable composition tests to verify that the underlying sample characteristics did not change disproportionately for the treated group following the mandate. Education levels are grouped as Low (primary and first stage of secondary), Medium (second stage of secondary and intermediate vocational training), and High (higher vocational training and university).

To formally evaluate this assumption, Table 2 presents a series of regressions where various labor, educational, and demographic characteristics are defined as dependent variables. These variables are regressed against the Difference-in-Differences interaction term. In this framework, a statistically significant coefficient on the interaction term implies that the composition of the treatment group changed differentially relative to the control group following the implementation of the time-tracking legislation.

The results from these tests indicate that while some characteristics remained strictly stable (such as gender, fixed-discontinuous contracts, and the core middle-aged workforce), there are statistically significant compositional shifts in specific variables. Most notably, there is a statistically significant decrease in the probability of holding a temporary contract (with a coefficient of -0.027) within the treated group post-reform. Furthermore, due to the large sample size, minor variations also exhibit statistical significance despite being economically negligible. This includes a slight increase in the share of immigrants, minute fluctuations across the youngest and oldest age brackets, and minimal shifts in educational attainment (a 0.5 percentage point drop in low-educated workers and a 0.4 percentage point rise in medium-educated workers).

The presence of these significant coefficients points to a light compositional bias, suggesting that the underlying profile of the surveyed workers in the treated group experienced slight adjustments after the law's enactment. To effectively address this issue and ensure that the estimated impact of the reform is isolated from these workforce shifts, two empirical strategies are necessary. First, as previously defined, we rely on a "Conditional Stable Composition" approach by explicitly including these specific characteristics as control variables in our baseline regressions. Second, to further validate the robustness of the model, we will conduct a heterogeneity analysis by dividing the sample according to these sensitive traits (e.g., estimating separate effects for temporary versus permanent workers, immigrants versus natives, or across different education levels). Obtaining consistent results across these distinct subgroups will definitively prove that the observed reduction in overtime hours is a direct consequence of the legislation, rather than an artifact of compositional changes in the unbalanced panel.

Finally, the assumption of **exogenous treatment assignment** requires that membership in the treatment group be determined by structural or historical factors preceding the reform and not be the result of strategic decisions by individuals to evade or take advantage of the regulations. In a policy evaluation framework, if workers or firms have the ability to artificially reclassify themselves (by fraudulently changing sectors, provinces, or contract types) in response to the law, assignment to the treatment group becomes endogenous. The empirical design must ensure that the criteria defining the treatment group are robust and cannot be manipulated in the short term by the affected agents.

In order to analyze the fulfillment of this assumption, it is important to define the sample used in this study. As mentioned in the data explanation, I restricted the data to the period between 2018 and 2020, that is, six quarters before the reform was introduced and six ones after the implementation of the reform. Furthermore, self-employed workers are explicitly excluded from the dataset. Because self-employed individuals have complete autonomy over their working hours and schedule flexibility, meaning they determine their own labor supply regardless of the new regulation, they are structurally exempt from the time-tracking mandate. Dropping this group is essential to guarantee the exogenous assignment of the treatment. It ensures that exposure to the policy is strictly imposed from the outside by the legal framework of salaried contracts, rather than being contaminated by an individual's endogenous choice to self-manage their working time.

To ensure the internal validity of the baseline findings, this study incorporates a comprehensive series of robustness checks. As a primary check to rule out pre-existing structural trends, I conduct an in-time placebo test by artificially shifting the implementation date to a pre-reform period (expanding the dataset backwards to include 2017). Re-estimating the baseline model with this fictitious treatment date yields interaction terms that are statistically indistinguishable from zero, firmly validating the causal design.

In addition to the placebo test, I implement three alternative empirical strategies to redefine the treatment assignment, verifying that the conclusions do not rely solely on the baseline metric (the percentage of overtime hours). First, I evaluate exposure based on the absolute volume of overtime hours, assigning the treatment and control groups based on whether the absolute number of extra hours falls above or below the median. Second, I restrict the comparison to the extreme ends of the distribution (the first and fourth quartiles), dropping the middle 50% of the sample to remove the "grey area" near the median. Finally, I exploit institutional differences by using the private sector as the treatment group and the public administration, which was already subject to strict time-tracking prior to 2019, as the control.

The main empirical estimates for all these alternative models are presented in the Robustness section. Furthermore, their corresponding methodological checks, including the Event Study for parallel trends and the stable composition tests, are rigorously replicated and detailed in the Appendix.

4 Results

This section presents the main empirical results of our Difference-in-Differences analysis. Table 3 reports the baseline estimates evaluating the causal impact of the mandatory time-tracking reform on the working time outcomes, using the median pre-reform overtime intensity as the threshold for treatment assignment, as defined in the previous section. Specifically, the table is structured across four columns, each representing a different dependent variable. Column (1) focuses on the total actual hours worked per week. Column (2) examines the overall impact on total extra hours, while columns (3) and (4) decompose this effect by separately analyzing paid and unpaid extra hours, respectively. All specifications incorporate population weights to account for group sample sizes.

Prior to the reform, highly exposed workers (the treated group) consistently worked more hours than the control group across all margins, reporting approximately 2.7 additional actual hours per week. This group also exhibited significantly higher baseline overtime, a gap driven more heavily by unpaid extra hours than by paid ones. While the post-reform period shows a general decrease in actual hours and a slight increase in reported overtime strictly for the control group, the causal impact of the legislation (captured by the interaction term) reveals a significant downward adjustment for the targeted workers.

As a direct consequence of the time-tracking mandate, the treated group experienced a net reduction of 1.064 actual working hours per week, translating to a 3.5% decline compared to the control mean. Crucially, the reform was highly effective in achieving its second core objective: curbing uncompensated labor. While total extra hours for the treated group fell by 77.1% (-0.168 hours per week) relative to the control group mean, this contraction was overwhelmingly driven by a near-complete eradication of unpaid overtime, which dropped by 0.100 hours per week. In comparison, paid extra hours saw a smaller reduction of 55.0% (-0.071 hours per week). This substantial reduction in unpaid overtime is primarily driven by an extensive margin adjustment. Specifically, estimating a logit model on a binary indicator for reporting zero unpaid extra hours reveals that the reform increased the likelihood of workers reporting no uncompensated hours by approximately 57%[1].

Table 3: Main Results: Population-Weighted DiD Estimates

	(1) Actual Hours	(2) Extra Hours	(3) Extra H. Paid	(4) Extra H. Not Paid
Treated	2.679***	0.178***	0.064***	0.117***
	(0.098)	(0.014)	(0.011)	(0.010)
Post	-1.415***	0.107***	0.061***	0.049***
	(0.113)	(0.018)	(0.014)	(0.012)
Treated × Post	-1.064***	-0.168***	-0.071***	-0.100***
	(0.091)	(0.014)	(0.011)	(0.009)
Control Mean	30.638	0.218	0.129	0.094
N	595,661	595,661	595,661	595,661

Notes: This table presents the baseline Difference-in-Differences estimates evaluating the impact of the mandatory time-tracking reform on weekly working time. Columns (1) through (4) display the estimates for actual hours worked, total extra hours, paid extra hours, and unpaid extra hours, respectively. The *treated* variable identifies workers in labor segments with a pre-reform overtime intensity above the median, while *post* indicates the quarters following the law's implementation. The *Treated* × *Post* term captures the DiD estimator (Average Treatment Effect on the Treated). All specifications control for socio-demographic characteristics, include time fixed effects, and are weighted by population size. The 'Control Mean' refers to the pre-reform average of the dependent variable for the untreated group. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Robustness

To verify the consistency of the primary findings, I estimate several alternative specifications as outlined in the methodology section. For each of these alternative strategies, I perform the same rigorous methodological checks applied in the main model, specifically testing the parallel trends assumption and evaluating the stable composition of the groups across labor, educational, and demographic dimensions (detailed results are provided in the Appendix part A.2.).

[1]Full estimation results for this logit specification are available upon request.

As a primary robustness check, Table 4 presents the results of the in-time placebo test, which assumes a fictitious reform date in the third quarter of 2018. As expected, the interaction term is not statistically significant in any of the four specifications. Since there is no statistical significance, the estimated effect of this fake reform is effectively zero for all working time variables. Therefore, we can be confident that the reductions observed in the baseline models are genuinely caused by the 2019 time-tracking legislation, and not by pre-existing trends in the labor market.

Table 4: In-Time Placebo Test (Fake Reform in 2018 Q3)

	(1) Actual Hours	(2) Extra Hours	(3) Extra H. Paid	(4) Extra H. Not Paid
Treated	3.226***	0.112***	0.003	0.113***
	(0.110)	(0.017)	(0.013)	(0.011)
Post	-0.527***	-0.026	0.011	-0.030**
	(0.106)	(0.018)	(0.013)	(0.012)
Treated $\times$ Post	-0.007	0.002	-0.001	-0.002
	(0.093)	(0.015)	(0.012)	(0.010)
Control Mean	30.687	0.234	0.136	0.104
N	497,234	497,234	497,234	497,234

Notes: This table presents the results of the in-time placebo test, artificially shifting the reform to a fictitious date (the third quarter of 2018). The dataset has been extended backwards to 2017 to ensure sufficient pre-treatment periods. The *Treated* $\times$ *Post* interaction captures the placebo effect. Regressions include socio-demographic controls, time fixed effects, and population weights. The 'Control Mean' refers to the average of the dependent variable for the untreated group in the fake pre-treatment period. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Additionally, Table 5 presents the results of the alternative identification strategy based on the absolute volume of overtime hours (referred to as the continuous strategy in the Appendix). This table replicates the exact column layout of Table 3, sequentially evaluating actual working hours, total extra hours, and their paid and unpaid components from left to right.

A direct comparison between these estimates and the baseline results in Table 3 reveals a striking level of consistency, strongly confirming the robustness of the reform's impact. Starting with actual working time, the decrease is slightly more conservative under this alternative assignment (-0.873 hours per week compared to the baseline -1.064), yet it remains highly statistically significant and translates to a 2.8% decline relative to the control group. This overall reduction is driven by a sharp decline in overtime work. The estimated reduction in total extra hours is virtually indistinguishable from the primary model (-0.169 compared to the baseline -0.168). When decomposing this effect, paid extra hours show a remarkably similar reduction (-0.073 compared to -0.071). Most notably, the causal effect on unpaid overtime is exactly identical across both specifications, showing a precise reduction of 0.100 hours per week. Ultimately, these parallel findings demonstrate that the near-complete elimination of unpaid overtime is a structural result of the policy, rather than a mechanical artifact of using a percentage-based metric for the baseline treatment assignment.

Table 5: Reform Effect (High vs Low Intensity, Weighted)

	(1) Actual Hours	(2) Extra Hours	(3) Extra H. Paid	(4) Extra H. Not Paid
Treated	1.522***	0.178***	0.048***	0.136***
	(0.101)	(0.015)	(0.012)	(0.010)
Post	-1.475***	0.063***	0.023	0.029**
	(0.114)	(0.019)	(0.015)	(0.012)
Treated × Post	-0.873***	-0.169***	-0.073***	-0.100***
	(0.091)	(0.014)	(0.011)	(0.010)
Control Mean	31.011	0.236	0.146	0.095
N	595,661	595,661	595,661	595,661

Notes: This table presents a robustness check for the Difference-in-Differences estimates using an alternative treatment assignment based on the absolute volume of overtime hours. The *Treated* × *Post* term represents the causal parameter. All regressions incorporate socio-demographic controls, time fixed effects, and population weights. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Furthermore, Table 6 displays the estimates when redefining the treated and control groups to compare the extreme ends of the pre-reform overtime distribution (1st vs. 4th quartile). By isolating these extremes, we can evaluate how the policy's impact aligns with the baseline findings when concentrated strictly on workers with the most intensive overtime habits. A comparison with the main results reveals striking consistency across all margins. First, the treated group experiences a significant reduction in actual working time of approximately 0.734 hours per week, representing a 2.4% decrease compared to the control group. This drop in actual hours is directly explained by the contraction in overtime. Total extra hours for this heavily treated group fall by 0.164 hours per week, closely mirroring the baseline estimate of 0.168 hours. Decomposing this effect, paid extra hours show a proportional reduction (-0.059 hours compared to the baseline -0.071 hours). Finally, unpaid overtime explains the majority of this overall decline: the reduction of 0.110 hours is perfectly in line with the baseline model (-0.100 hours) and drops by an amount more than double the control mean for this specific group, completely eliminating their unpaid hours.

Table 6: Robustness: 1st vs 4th Quartile (Weighted)

	(1) Actual Hours	(2) Extra Hours	(3) Extra H. Paid	(4) Extra H. Not Paid
Treated	3.638***	0.204***	0.053**	0.159***
	(0.207)	(0.033)	(0.022)	(0.026)
Post	-1.257***	0.006	-0.022	0.014
	(0.160)	(0.027)	(0.021)	(0.018)
Treated × Post	-0.734***	-0.164***	-0.059***	-0.110***
	(0.125)	(0.020)	(0.014)	(0.013)
Control Mean	30.114	0.158	0.113	0.050
N	296,753	296,753	296,753	296,753

Notes: This table presents an additional robustness check comparing the 4th quartile (highest overtime intensity) against the 1st quartile (lowest intensity). All regressions include socio-demographic controls, time fixed effects, and population weights. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As a final analytical exercise, Table 7 presents the estimates for the alternative specification using the sector of employment to define the groups, where the private sector acts as the treatment group and the public sector serves as the control. A direct comparison with the baseline estimates in Table 3 confirms the structural impact of the reform across all variables. Starting with actual working time, treated workers in the private sector experience a significant reduction of 0.573 hours per week, which is a more conservative effect than the 1.064 hours estimated in the main model. Moving to total extra hours, the policy drives a decrease of 0.217 hours per week in the private sector, demonstrating a slightly stronger but highly consistent effect compared to the baseline reduction of 0.168 hours. When evaluating paid extra hours, the reduction of 0.088 hours aligns remarkably well with the 0.071 hours found in the primary specification. Finally, the core finding remains entirely robust: unpaid extra hours fall by exactly 100.0% (-0.138 hours), perfectly mirroring the near-complete elimination of uncompensated labor (-0.100 hours) observed in the baseline model.

Overall, these alternative models prove that the estimated reduction in overtime hours is robust and independent of the methodological choices used to construct the groups. Building on this solid baseline, the next section explores whether this aggregate success varies across different worker profiles.

Table 7: Robustness: Private vs Public Sector (Weighted)

	(1) Actual Hours	(2) Extra Hours	(3) Extra H. Paid	(4) Extra H. Not Paid
Treated	-1.353***	0.057***	-0.010	0.075***
	(0.066)	(0.016)	(0.013)	(0.010)
Post	-0.226**	0.170***	0.066***	0.096***
	(0.091)	(0.027)	(0.022)	(0.016)
Treated × Post	-0.573***	-0.217***	-0.088***	-0.138***
	(0.068)	(0.021)	(0.017)	(0.013)
Control Mean	35.809	0.313	0.180	0.138
N	526,014	526,014	526,014	526,014

Notes: This table evaluates the reform's effect utilizing the sector of employment, with the private sector as the treated group and the public sector as the control. All regressions control for socio-demographic characteristics, time fixed effects, and population weights. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 Heterogeneity

To address the mild compositional bias detected in the Stable Composition Test, I now conduct a comprehensive heterogeneity analysis. By dividing the sample based on the variables that showed significant compositional shifts: contract type, origin, educational attainment, and age, I will try to verify if the main effect of the law holds across different subgroups. This step is necessary to prove my results are driven by the time-tracking legislation itself, and not merely by changes in the workforce composition. To provide a clearer visual comparison, the point estimates for all these subgroups are summarized in two comprehensive graphs: Figure 2 displays the effects on actual hours, and Figure 3 shows the impact on total extra hours. In each figure, the dashed vertical line represents the main coefficient obtained in the baseline regressions. Finally, the full regression tables detailing the exact estimates for all subgroups can be found in the Appendix (see Tables A.2 and A.3).

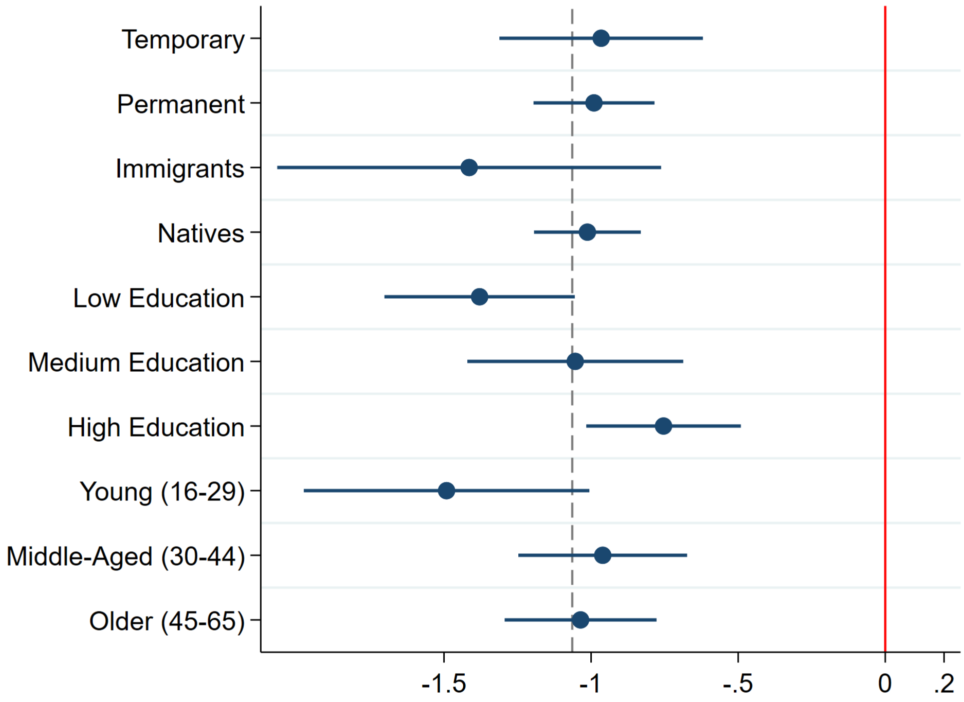


Figure 2: Heterogeneity Analysis: Actual Hours

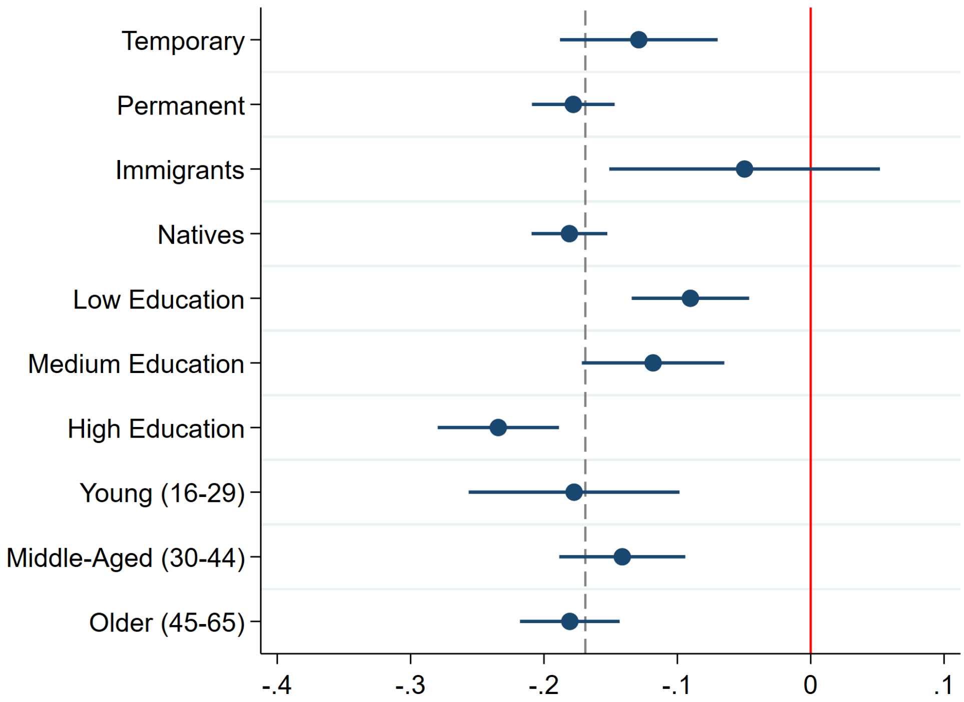


Figure 3: Heterogeneity Analysis: Total Extra Hours

First, I compare the results for temporary workers with those for permanent workers. The goal is to ensure that the policy's impact is consistent across both groups, rather than being an artifact of the decreasing share of temporary contracts. Looking at the plotted estimates, the interaction term is not only negative and significant in both models, but the magnitudes of the effects are also highly comparable. As detailed in Table A.7 in the Appendix, for temporary workers, actual hours decrease by 0.96 hours per week and unpaid extra hours drop by 0.07 hours. Similarly, for permanent workers, actual hours fall by 0.99 hours and unpaid extra hours decrease by 0.10 hours. Since the reform effectively reduces uncompensated overtime for both contract types, I can confidently conclude that the compositional shift in temporary employment does not invalidate my previous findings.

When I conduct this heterogeneity analysis for immigrants and natives, I find divergent results. As clearly visible in the plots, the interaction term for the immigrant subgroup does not show a statistically significant reduction in overtime hours (the confidence intervals cross the zero line). This lack of significance can likely be explained by two main factors. First, from a statistical perspective, the reduced sample size for this specific subgroup decreases the statistical power of the model, leading to larger standard errors that make it difficult to capture the effect. Second, from a labor market perspective, immigrant workers historically face a higher prevalence of informal labor arrangements and lower regulatory compliance by employers. Due to their weaker negotiation power, firms might be more likely to evade the time-tracking mandate, diluting the effective enforcement of the law for this group. For native workers, however, I obtain the exact same robust and significant results as in my baseline models (see Table A.8 in the Appendix), confirming that the policy's impact is concentrated among native workers while highlighting potential enforcement gaps in more vulnerable segments of the workforce.

In addition, I examine the heterogeneous effects across different levels of educational attainment. As previously noted in the compositional checks, there were minor but statistically significant shifts in the shares of low- and medium-educated workers. However, the estimates plotted in the figures confirm that the time-tracking legislation had a significant and negative impact on working hours across the entire educational spectrum (full results in Table A.10 in the Appendix). Actual hours worked decreased by 1.38, 1.05, and 0.75 hours for low, medium, and highly educated workers, respectively. Furthermore, unpaid overtime saw a consistent and significant reduction across all three groups, with the most pronounced absolute drop among highly educated workers (-0.16 hours). These uniform results definitively prove that the baseline findings are robust and driven by the policy itself, rather than by any underlying educational compositional changes in the panel.

Finally, when examining heterogeneity across age groups, the results demonstrate a very consistent reduction in working hours across all cohorts. As illustrated alongside the other demographics in Figures 2 and 3, the negative impact of the reform holds for young workers, middle-aged adults, and older employees alike, with point estimates closely aligned with the baseline effects (represented by the dashed lines). These findings confirm that the legislation's bite on working time is pervasive across the entire lifecycle of the workforce, rather than being concentrated in a specific age bracket. To maintain visual clarity across these diverse demographic subgroups, the graphical illustrations focus exclusively on the two main aggregate dimensions, actual hours and total extra hours, with the paid and unpaid components subsumed within the total. Complete estimates for all dependent variables, however, remain fully detailed in the Appendix, as noted throughout this section.

7 Conclusion

This paper evaluates the causal impact of the mandatory time-tracking legislation implemented in Spain on working hours and uncompensated overtime. Using a Difference-in-Differences framework and exploiting the variation in pre-reform overtime intensity across different labor market segments, this study provides robust empirical evidence on the effectiveness of the policy in the short-run.

The baseline results indicate that the regulation successfully achieved its primary objective. The empirical evidence shows a notable shift for the targeted workers: total extra hours dropped by 77.1% (-0.168 hours per week) relative to the control mean. This steep decline is fundamentally driven by the near-total elimination of uncompensated labor (-0.100 hours per week) after the reform, as a substantial portion of the workforce transitioned to reporting exactly zero unpaid hours. Paid extra hours also decreased significantly by 55.0%. As a direct consequence of this strict control over overtime, the treated group experienced a significant net reduction in actual working time of approximately 1.1 hours per week, translating to a 3.5% decline compared to the control group.

To ensure the internal validity of these findings, the empirical strategy was subjected to rigorous robustness checks. An in-time placebo test firmly validates the causal design, while the event study confirms the parallel trends assumption, collectively ruling out the presence of confounding pre-existing trends. Furthermore, the results remain highly consistent when employing alternative empirical definitions, such as a continuous measure of treatment intensity, comparing extreme quartiles, and using an institutional-based split between the private and public sectors. Across all these alternative models, the eradication of unpaid overtime remains the core driver of the reform's success.

Building on this solid result, the comprehensive heterogeneity analysis demonstrates that the policy's impact is remarkably uniform across the workforce. The reduction in unpaid overtime holds regardless of the workers' contract type (temporary versus permanent), age demographic, or educational attainment, confirming that the baseline findings are not driven by compositional shifts in the unbalanced panel.

However, the analysis also reveals a significant exception regarding immigrant workers. For this subgroup, the legislation did not yield a statistically significant reduction in overtime hours. This lack of significance can be explained from a statistical perspective due to the reduced sample size decreasing the model's power, but it also highlights a critical vulnerability in the law's enforcement. It likely stems from the higher prevalence of the weaker bargaining power that immigrant workers historically face.

In conclusion, the time-tracking mandate has proven to be an effective labor market intervention for the majority of the workforce, successfully curbing unpaid overtime and reducing overall working hours. Nevertheless, the findings suggest that legislative mandates alone may be insufficient for highly vulnerable segments of the labor market. Future policy efforts should complement these regulations with targeted enforcement mechanisms and labor inspections to ensure equitable compliance and protect those workers who lack the negotiation power to enforce their rights.

References

Baker, A., Callaway, B., Cunningham, S., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2025). *Difference-in-Differences Designs: A Practitioner's Guide*. arXiv preprint arXiv:2503.13323.

Conde-Ruiz, J. I., Lahera, J., & Viola, A. (2024). La reducción legal de la jornada de trabajo [The statutory reduction of working time]. *Fedea Policy Papers*, 2024-03. Madrid: Fundación de Estudios de Economía Aplicada (FEDEA).

European Court of Justice (ECJ). (2019). *Judgment of the Court (Grand Chamber) of 14 May 2019. Federación de Servicios de Comisiones Obreras (CCOO) v Deutsche Bank SAE*. Case C-55/18, ECLI:EU:C:2019:402.

Felgueroso, F., García-Perez, J. I., Jansen, M., Jiménez-Martín, S., & Pérez-Gutiérrez, D. (2026). *Does limiting temporary work improve employment stability? Evidence for young workers in Spain*. Estudios sobre la Economía Española, Documento de Trabajo 2026/22. Fundación de Estudios de Economía Aplicada (FEDEA).

Government of Japan. (2018). *Act on the Arrangement of Related Acts to Promote Work Style Reform* (Act No. 71 of 2018). Ministry of Health, Labour and Welfare (MHLW).

Hunt, J. (1999). Has Work-Sharing Worked in Germany? *The Quarterly Journal of Economics*, 114(1), 117-148.

Nold, J., & Backhaus, N. (2022). Jede Stunde zählt? Arbeitszeiterfassung und -konten für verschiedene Überstundengründe und -ausgleichsformen [Every hour counts? Working time recording and accounts for different reasons for overtime and forms of compensation]. *WSI-Mitteilungen*, 75(3), 195-204.

Appendix

A.1. Sample Construction & Treatment Definition

This table presents the distribution of observations across the 90 groups defined by the interaction of economic activity and occupation. These are the exact groups used to calculate the treatment intensity measure for the main analysis. As shown below, there is substantial size variability among them. This significant imbalance in group sizes justifies the use of population weights in my main regressions, ensuring that the empirical results are not distorted by smaller, overrepresented segments and accurately reflect the overall labor market.

Table A.1: Tabulation of `act` × `ocup` Groups

Group	Freq.	Percent	Cum.	Group	Freq.	Percent	Cum.
1	144	0.02	0.02	46	3,431	0.58	25.45
2	291	0.05	0.07	47	8,866	1.49	26.94
3	517	0.09	0.16	48	12,097	2.03	28.97
4	463	0.08	0.24	49	64,796	10.88	39.84
5	599	0.10	0.34	50	140	0.02	39.87
6	3,000	0.50	0.84	51	8,328	1.40	41.27
7	353	0.06	0.90	52	5,668	0.95	42.22
8	1,177	0.20	1.10	53	17,664	2.97	45.18
9	9,969	1.67	2.77	54	1,323	0.22	45.40
10	944	0.16	2.93	55	6,062	1.02	46.42
11	1,333	0.22	3.15	56	7,278	1.22	47.64
12	3,329	0.56	3.71	57	8,892	1.49	49.14
13	3,065	0.51	4.23	58	1,925	0.32	49.46
14	634	0.11	4.33	59	4	0.00	49.46
15	87	0.01	4.35	60	1,523	0.26	49.72
16	10,465	1.76	6.11	61	12,758	2.14	51.86
17	7,371	1.24	7.34	62	3,147	0.53	52.39
18	3,539	0.59	7.94	63	3,014	0.51	52.89
19	1,193	0.20	8.14	64	11,645	1.95	54.85
20	3,566	0.60	8.74	65	11,191	1.88	56.73
21	6,277	1.05	9.79	66	17,505	2.94	59.67
22	3,668	0.62	10.41	67	5,933	1.00	60.66
23	315	0.05	10.46	68	1,277	0.21	60.88
24	19	0.00	10.46	69	1,742	0.29	61.17
25	9,011	1.51	11.97	70	720	0.12	61.29
26	9,996	1.68	13.65	71	16,096	2.70	63.99
27	3,204	0.54	14.19	72	4,217	0.71	64.70
28	836	0.14	14.33	73	2,652	0.45	65.14
29	3,048	0.51	14.84	74	78,794	13.23	78.37
30	4,595	0.77	15.61	75	15,176	2.55	80.92
31	2,636	0.44	16.06	76	18,255	3.06	83.98
32	121	0.02	16.08	77	39,453	6.62	90.61
33	9,249	1.55	17.63	78	681	0.11	90.72
34	6,943	1.17	18.80	79	2,915	0.49	91.21
35	1,285	0.22	19.01	80	2,258	0.38	91.59
36	479	0.08	19.09	81	10,053	1.69	93.28
37	1,961	0.33	19.42	82	762	0.13	93.41
38	2,937	0.49	19.91	83	2,995	0.50	93.91
39	2,533	0.43	20.34	84	4,268	0.72	94.63
40	95	0.02	20.36	85	4,031	0.68	95.30
41	2	0.00	20.36	86	10,693	1.80	97.10
42	18,489	3.10	23.46	87	375	0.06	97.16
43	1,981	0.33	23.79	88	1,311	0.22	97.38
44	3,559	0.60	24.39	89	874	0.15	97.53
45	2,867	0.48	24.87	90	14,728	2.47	100.00

Total Observations: 595,661 (100.00%)

A.2. Descriptive Statistics and Temporal Evolution

Table A.2: Descriptive Statistics and Temporal Evolution by Treatment Status

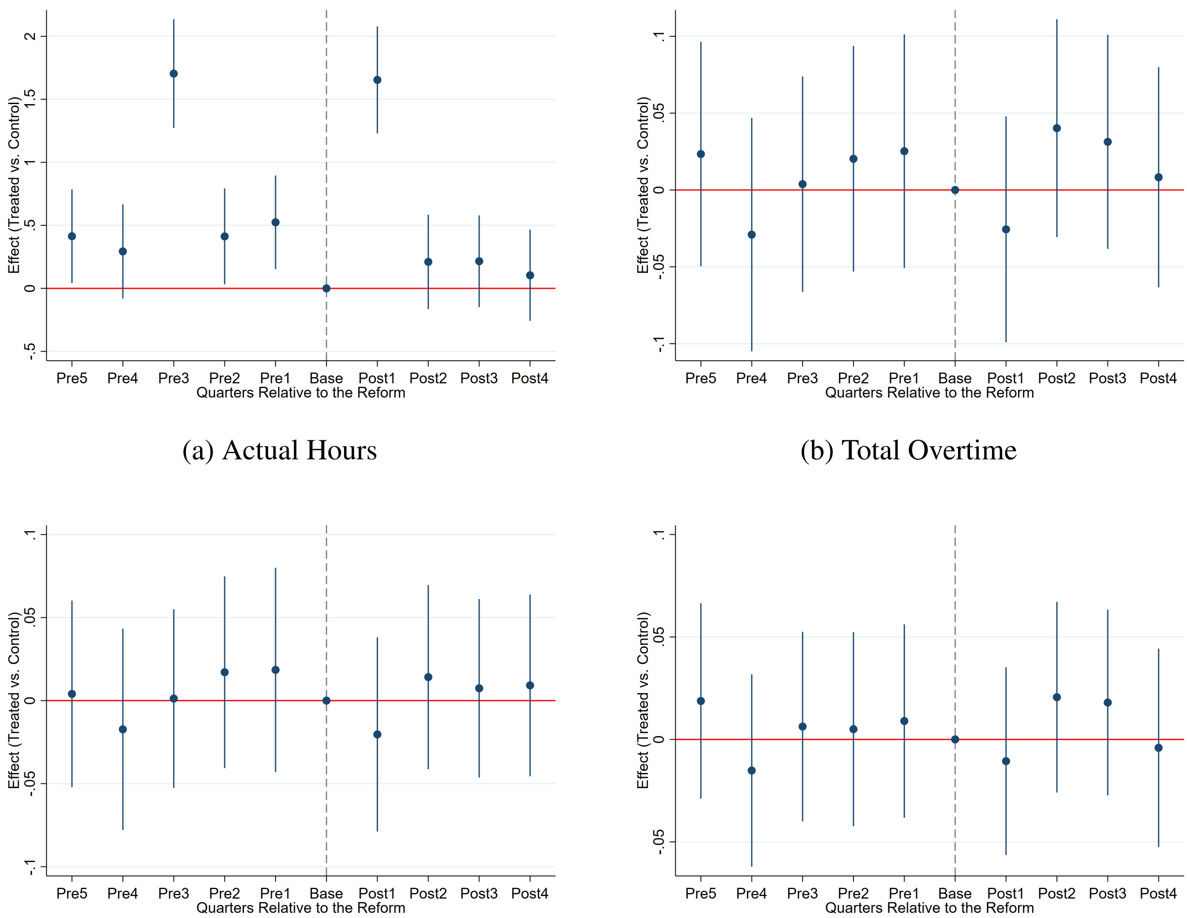
	Control Group			**Treated Group**		
	Pre	**Post**	**Cohen d**	**Pre**	**Post**	**Cohen d**
Working Hours						
Actual Hours	30.638	28.635	-0.136	34.592	31.714	-0.199
Total Overtime Hours	0.218	0.283	0.034	0.465	0.386	-0.035
Paid Overtime Hours	0.129	0.151	0.015	0.258	0.224	-0.020
Unpaid Overtime Hours	0.094	0.136	0.034	0.218	0.169	-0.031
Demographics						
Male	0.381	0.378	-0.006	0.640	0.637	-0.006
Female	0.619	0.622	0.006	0.360	0.363	0.006
Immigrant	0.058	0.053	-0.024	0.076	0.076	-0.001
16 to 29 years	0.106	0.103	-0.011	0.158	0.147	-0.030
30 to 44 years	0.350	0.331	-0.039	0.419	0.401	-0.036
45 to 65 years	0.544	0.565	0.044	0.423	0.451	0.058
Labor Contract						
Temporary	0.261	0.257	-0.009	0.250	0.219	-0.071
Permanent	0.739	0.743	0.009	0.750	0.781	0.071
Fixed-Discontinuous	0.017	0.016	-0.009	0.024	0.023	-0.008
Education Level						
Low Education	0.254	0.236	-0.040	0.384	0.364	-0.042
Medium Education	0.234	0.231	-0.007	0.239	0.238	-0.003
High Education	0.512	0.532	0.041	0.377	0.399	0.045
Sector						
Agriculture	0.053	0.051	-0.009	0.003	0.003	-0.004
Industry	0.037	0.037	-0.001	0.291	0.295	0.008
Construction	0.009	0.009	0.003	0.100	0.102	0.006
Services	0.901	0.903	0.007	0.606	0.601	-0.011

Notes: Education levels are grouped as Low (primary and first stage of secondary), Medium (second stage of secondary and intermediate vocational training), and High (higher vocational training and university). Cohen's d represents the standardized mean difference between the post-reform and pre-reform periods. By convention, an absolute value of $d < 0.10$ indicates a negligible effect size, confirming that the socio-demographic composition of the groups remained stable over time.

A.3. Methodological Validity and Robustness Checks

This section presents a comprehensive set of tests designed to validate the key assumptions underlying the Difference-in-Differences (DiD) framework used in this study. First, I report an in-time placebo test to ensure that the baseline results are not driven by unobserved pre-existing trends. Subsequently, for each alternative identification strategy, I provide two fundamental robustness checks. I present stable composition tests to verify that the socio-demographic and labor characteristics of the groups did not systematically change after the reform, ruling out endogenous sorting. Alongside these, I display Event Study estimates for the four main outcomes to visually inspect the dynamic effects and confirm the parallel trends assumption in the pre-treatment period.

A.3.1. In-Time Placebo Test



(a) Actual Hours

(b) Total Overtime

(c) Paid Overtime

(d) Unpaid Overtime

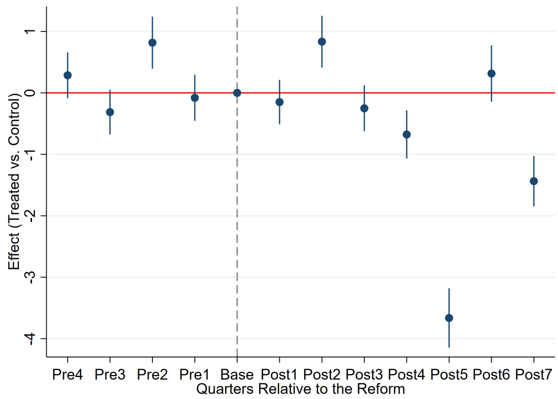
Figure A.1: Event Study Estimates (In-Time Placebo).

Table A.3: Stable Composition Test: In-Time Placebo

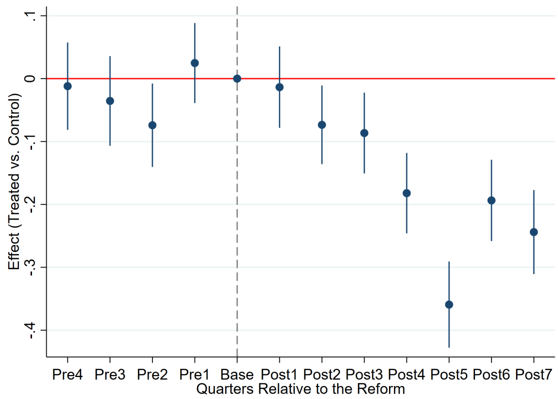
Panel A: Labor & Education Characteristics					
	(1) Temp.	(2) Discon.	(3) Low Educ.	(4) Med. Educ.	(5) High Educ.
Treated × Post	-0.011***	0.001	-0.003	0.000	0.002
	(0.003)	(0.001)	(0.003)	(0.003)	(0.003)
Panel B: Demographics					
	(1) Male	(2) Immig.	(3) 16-29 yrs	(4) 30-44 yrs	(5) 45-65 yrs
Treated × Post	0.000	0.008***	0.003	-0.008**	0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
N	497,234	497,234	497,234	497,234	497,234

Notes: This table reports the Difference-in-Differences estimates to test the stable composition assumption for the In-Time Placebo strategy. The sample is restricted to the pre-reform period (2017Q1-2019Q1), simulating a fake reform in 2018Q3. Panel A presents the results for labor and education characteristics, while Panel B presents the results for demographics. Each column corresponds to a separate regression where the dependent variable is the characteristic indicated in the column header. All specifications include time fixed effects and are weighted by population size. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

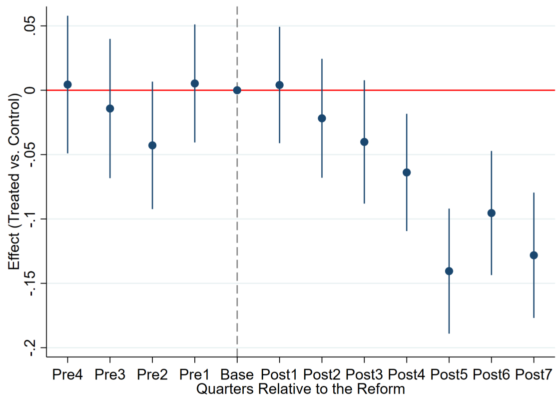
A.3.2. Continuous Intensity Measure



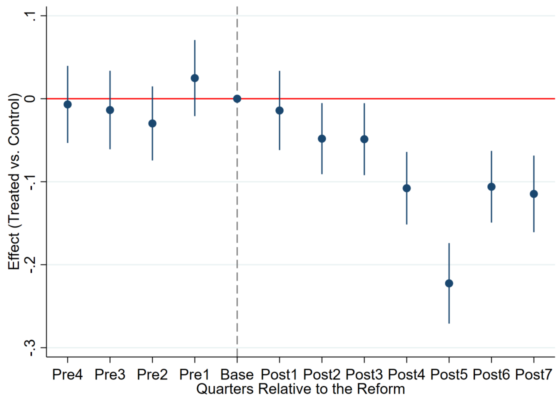
(a) Actual Hours



(b) Total Overtime



(c) Paid Overtime



(d) Unpaid Overtime

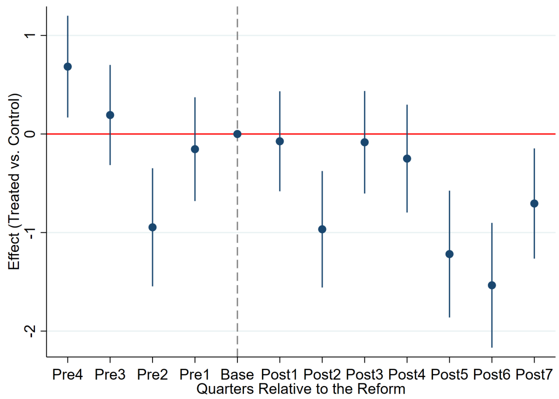
Figure A.2: Event Study Estimates (Continuous Strategy).

Table A.4: Stable Composition Test: Continuous Intensity Measure

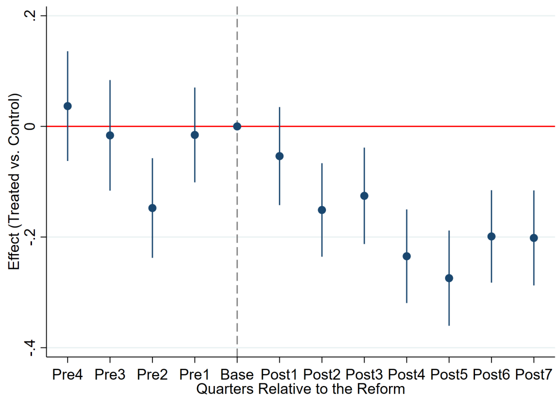
Panel A: Labor & Education Characteristics					
	(1) Temp.	(2) Discon.	(3) Low Educ.	(4) Med. Educ.	(5) High Educ.
Treated × Post	-0.027***	0.000	-0.005***	0.004*	0.001
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Panel B: Demographics					
	(1) Male	(2) Immig.	(3) 16-29 yrs	(4) 30-44 yrs	(5) 45-65 yrs
Treated × Post	-0.003	0.004***	-0.007***	0.000	0.007***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)
N	595,661	595,661	595,661	595,661	595,661

Notes: This table reports the Difference-in-Differences estimates to test the stable composition assumption for the Continuous Intensity strategy. Panel A presents the results for labor and education characteristics, while Panel B presents the results for demographics. Each column corresponds to a separate regression where the dependent variable is the characteristic indicated in the column header. All specifications include time fixed effects and are weighted by population size. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

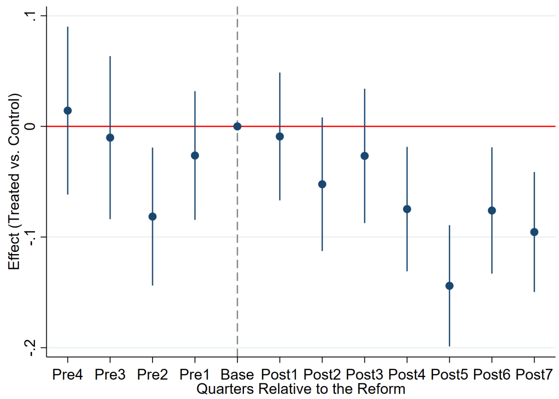
A.3.3. Quartiles Strategy



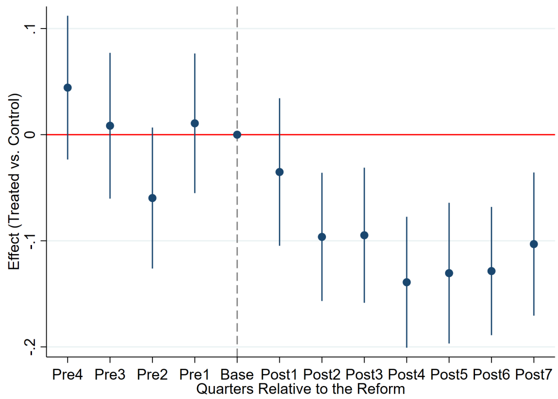
(a) Actual Hours



(b) Total Overtime



(c) Paid Overtime



(d) Unpaid Overtime

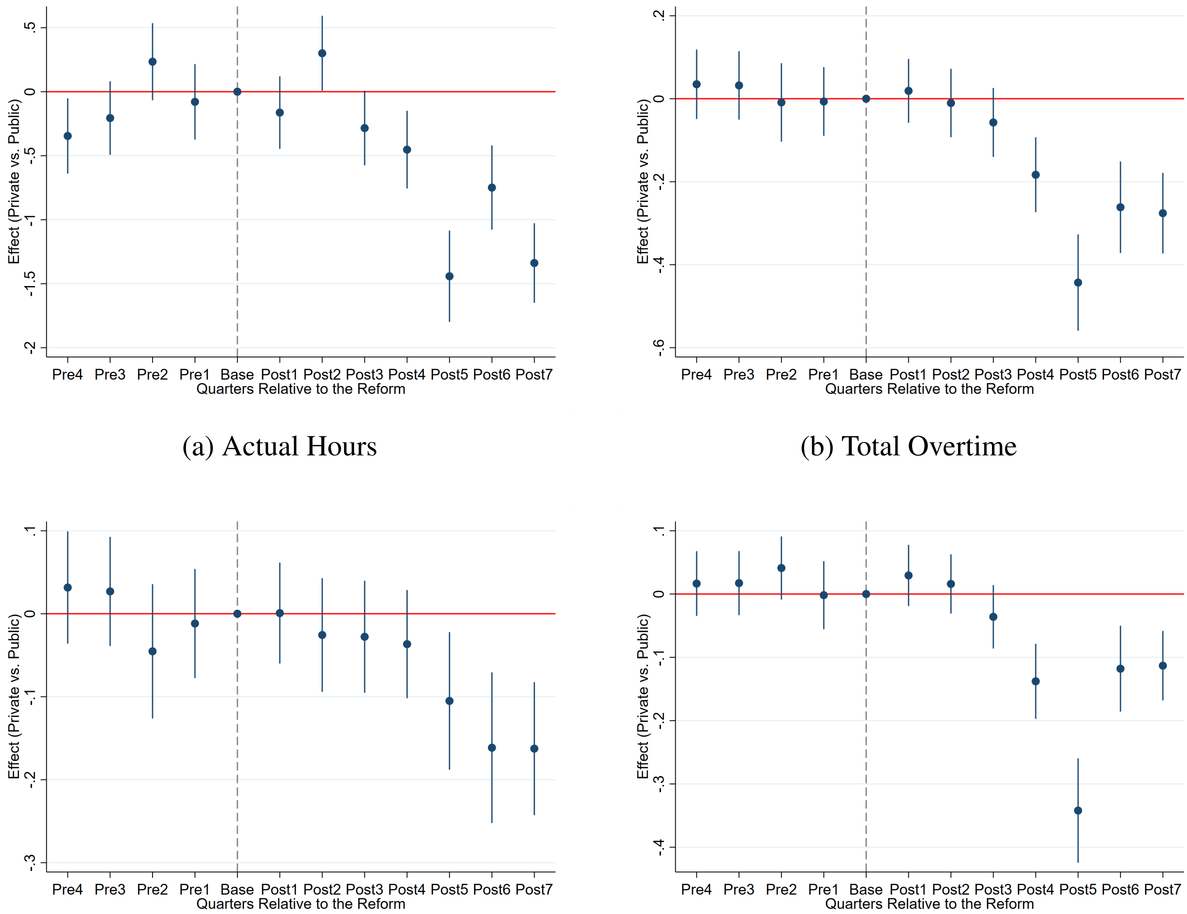
Figure A.3: Event Study Estimates (Quartiles Strategy).

Table A.5: Stable Composition Test: Quartiles Strategy

Panel A: Labor & Education Characteristics					
	(1) Temp.	(2) Discon.	(3) Low Educ.	(4) Med. Educ.	(5) High Educ.
Treated $\times$ Post	-0.026***	0.001	0.003	-0.001	-0.002
	(0.003)	(0.001)	(0.003)	(0.003)	(0.003)
Panel B: Demographics					
	(1) Male	(2) Immig.	(3) 16-29 yrs	(4) 30-44 yrs	(5) 45-65 yrs
Treated $\times$ Post	-0.009***	0.008***	-0.006***	-0.001	0.007**
	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)
N	296,753	296,753	296,753	296,753	296,753

Notes: This table reports the Difference-in-Differences estimates to test the stable composition assumption for the Quartiles strategy. Panel A presents the results for labor and education characteristics, while Panel B presents the results for demographics. Each column corresponds to a separate regression where the dependent variable is the characteristic indicated in the column header. All specifications include time fixed effects and are weighted by population size. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.3.4. Public vs. Private Sector Strategy



(a) Actual Hours

(b) Total Overtime

(c) Paid Overtime

(d) Unpaid Overtime

Figure A.4: Event Study Estimates (Public vs. Private Sector).

Table A.6: Stable Composition Test: Public vs. Private Sector

Panel A: Labor & Education Characteristics					
	(1) Temp.	(2) Discon.	(3) Low Educ.	(4) Med. Educ.	(5) High Educ.
Treated × Post	-0.043***	-0.002**	-0.009***	0.006**	0.003
	(0.004)	(0.001)	(0.003)	(0.003)	(0.003)
Panel B: Demographics					
	(1) Male	(2) Immig.	(3) 16-29 yrs	(4) 30-44 yrs	(5) 45-65 yrs
Treated × Post	0.005	0.008***	-0.012***	0.001	0.010***
	(0.004)	(0.002)	(0.003)	(0.004)	(0.004)
N	526,014	526,014	526,014	526,014	526,014

Notes: This table reports the Difference-in-Differences estimates to test the stable composition assumption for the Public vs. Private Sector strategy. Panel A presents the results for labor and education characteristics, while Panel B presents the results for demographics. Each column corresponds to a separate regression where the dependent variable is the characteristic indicated in the column header. All specifications include time fixed effects and are weighted by population size. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.4. Extended Heterogeneity Results

This section provides the full estimation tables for the heterogeneity analysis discussed in the main text, presenting the detailed coefficients for actual hours and the different extra hour categories across subgroups.

Table A.7: Heterogeneity Analysis: Contract Type (Weighted)

	(1) Actual Hours	(2) Extra Hours	(3) Extra H. Paid	(4) Extra H. Not Paid
Panel A: Temporary Workers ($N = 147,333$)				
Treated	3.045***	0.066**	-0.007	0.074***
	(0.198)	(0.031)	(0.025)	(0.019)
Post	-0.272	0.148***	0.097***	0.055**
	(0.223)	(0.038)	(0.031)	(0.023)
Treated × Post	-0.966***	-0.129***	-0.056**	-0.074***
	(0.176)	(0.030)	(0.025)	(0.017)
Control Mean	29.316	0.268	0.191	0.080
Panel B: Permanent Workers ($N = 448,328$)				
Treated	2.331***	0.215***	0.093***	0.126***
	(0.114)	(0.016)	(0.012)	(0.011)
Post	-1.881***	0.094***	0.048***	0.047***
	(0.131)	(0.020)	(0.015)	(0.014)
Treated × Post	-0.990***	-0.178***	-0.075***	-0.107***
	(0.105)	(0.016)	(0.012)	(0.011)
Control Mean	31.542	0.218	0.120	0.105

Notes: This table presents the heterogeneity analysis of the mandatory time-tracking reform by type of labor contract. Panel A displays the Difference-in-Differences estimates for temporary workers, while Panel B reports the estimates for permanent workers. Columns (1) through (4) display the results for actual working hours, total extra hours, paid extra hours, and unpaid extra hours, respectively. The 'Control Mean' reports the pre-reform average of each dependent variable for the untreated control group within the corresponding subgroup. All specifications incorporate socio-demographic controls, include time fixed effects, and are weighted by population size. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Heterogeneity Analysis: Law Effect by Origin (Weighted)

	(1) Actual Hours	(2) Extra Hours	(3) Extra H. Paid	(4) Extra H. Not Paid
Panel A: Immigrants ($N = 38,894$)				
Treated	2.000***	0.095	0.020	0.073*
	(0.407)	(0.070)	(0.058)	(0.038)
Post	-1.918***	-0.018	0.003	-0.014
	(0.441)	(0.067)	(0.051)	(0.042)
Treated × Post	-1.414***	-0.050	-0.014	-0.037
	(0.333)	(0.052)	(0.042)	(0.030)
Control Mean	31.282	0.301	0.211	0.092
Panel B: Natives ($N = 556,767$)				
Treated	2.726***	0.189***	0.072***	0.120***
	(0.100)	(0.014)	(0.011)	(0.010)
Post	-1.344***	0.121***	0.067***	0.056***
	(0.114)	(0.018)	(0.014)	(0.012)
Treated × Post	-1.012***	-0.181***	-0.077***	-0.107***
	(0.093)	(0.014)	(0.011)	(0.010)
Control Mean	30.899	0.223	0.131	0.099

Notes: This table presents the heterogeneity analysis of the mandatory time-tracking reform by worker origin. Panel A displays the Difference-in-Differences estimates for immigrants, while Panel B reports the estimates for natives. Columns (1) through (4) display the results for actual working hours, total extra hours, paid extra hours, and unpaid extra hours, respectively. The 'Control Mean' reports the pre-reform average of each dependent variable for the untreated control group within the corresponding subgroup. All specifications incorporate socio-demographic controls, include time fixed effects, and are weighted by population size. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Heterogeneity Analysis: Law Effect by Age Groups (Weighted)

	(1) Actual Hours	(2) Extra Hours	(3) Extra H. Paid	(4) Extra H. Not Paid
Panel A: Young Workers (16 to 29 years) ($N = 76,415$)				
Treated	2.171***	0.163***	0.084***	0.078***
	(0.278)	(0.035)	(0.029)	(0.020)
Post	-0.863***	0.143***	0.107**	0.035
	(0.321)	(0.051)	(0.043)	(0.026)
Treated × Post	-1.491***	-0.177***	-0.082**	-0.091***
	(0.248)	(0.040)	(0.034)	(0.021)
Control Mean	30.367	0.254	0.191	0.068
Panel B: Middle-Aged Workers (30 to 44 years) ($N = 223,377$)				
Treated	2.476***	0.174***	0.062***	0.113***
	(0.159)	(0.026)	(0.020)	(0.016)
Post	-1.406***	0.080***	0.069***	0.014
	(0.185)	(0.031)	(0.024)	(0.020)
Treated × Post	-0.960***	-0.141***	-0.061***	-0.084***
	(0.146)	(0.024)	(0.019)	(0.016)
Control Mean	31.123	0.282	0.170	0.120
Panel C: Older Workers (45 to 65 years) ($N = 295,869$)				
Treated	2.763***	0.195***	0.069***	0.131***
	(0.145)	(0.021)	(0.015)	(0.014)
Post	-1.651***	0.111***	0.035**	0.080***
	(0.159)	(0.023)	(0.016)	(0.017)
Treated × Post	-1.036***	-0.181***	-0.075***	-0.111***
	(0.132)	(0.019)	(0.014)	(0.014)
Control Mean	30.936	0.185	0.102	0.088

Notes: This table presents the heterogeneity analysis of the mandatory time-tracking reform by age group. Panel A displays the Difference-in-Differences estimates for young workers, Panel B for middle-aged workers, and Panel C for older workers. Columns (1) through (4) display the results for actual working hours, total extra hours, paid extra hours, and unpaid extra hours, respectively. The 'Control Mean' reports the pre-reform average of each dependent variable for the untreated control group within the corresponding subgroup. All specifications incorporate socio-demographic controls, include time fixed effects, and are weighted by population size. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Heterogeneity Analysis: Law Effect by Education Level (Weighted)

	(1) Actual Hours	(2) Extra Hours	(3) Extra H. Paid	(4) Extra H. Not Paid
Panel A: Low Education ($N = 183,603$)				
Treated	2.132***	0.166***	0.125***	0.041***
	(0.178)	(0.023)	(0.020)	(0.012)
Post	-1.873***	0.007	0.003	0.001
	(0.217)	(0.026)	(0.021)	(0.015)
Treated × Post	-1.379***	-0.090***	-0.058***	-0.032**
	(0.165)	(0.022)	(0.019)	(0.013)
Control Mean	30.437	0.174	0.122	0.057
Panel B: Medium Education ($N = 140,418$)				
Treated	1.925***	0.220***	0.103***	0.121***
	(0.203)	(0.031)	(0.026)	(0.018)
Post	-1.810***	0.025	0.034	-0.010
	(0.238)	(0.033)	(0.023)	(0.024)
Treated × Post	-1.054***	-0.118***	-0.046**	-0.073***
	(0.187)	(0.027)	(0.022)	(0.017)
Control Mean	30.930	0.189	0.130	0.063
Panel C: High Education ($N = 271,640$)				
Treated	2.446***	0.169***	0.033**	0.138***
	(0.155)	(0.024)	(0.017)	(0.018)
Post	-0.994***	0.199***	0.104***	0.103***
	(0.159)	(0.030)	(0.023)	(0.019)
Treated × Post	-0.754***	-0.234***	-0.083***	-0.158***
	(0.134)	(0.023)	(0.016)	(0.017)
Control Mean	31.204	0.282	0.153	0.137

Notes: This table presents the heterogeneity analysis of the mandatory time-tracking reform by education level. Panel A displays the Difference-in-Differences estimates for workers with low education, Panel B for medium education, and Panel C for high education. Columns (1) through (4) display the results for actual working hours, total extra hours, paid extra hours, and unpaid extra hours, respectively. The 'Control Mean' reports the pre-reform average of each dependent variable for the untreated control group within the corresponding subgroup. All specifications incorporate socio-demographic controls, include time fixed effects, and are weighted by population size. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.