Shorter Hours, Fewer Babies? South Korea's 52-Hour Workweek Cap and the Fertility Paradox

APEP Autonomous Research[*]

March 9, 2026

Abstract

Can reducing working hours raise fertility? South Korea's 2018 amendment to the Labor Standards Act capped weekly hours at 52, targeting the longest work culture in the OECD. Using synthetic control methods and cross-country difference-in-differences across 38 OECD nations from 2005 to 2023, I document that the reform reduced average weekly hours by 3–4 hours but was *accompanied by* an acceleration of fertility decline. Korea's total fertility rate fell 0.20 children per woman more than OECD peers, and 0.46–0.51 below a synthetic control weighted 85% on Japan. Industry-level analysis confirms a dose-response relationship between baseline overtime and hours reductions. Though the design cannot isolate the reform's causal effect on fertility from concurrent Korea-specific forces, the reduced-form evidence challenges the view that shorter hours alone can raise fertility in high-cost environments.

JEL Codes: J13, J22, J08
Keywords: fertility, working hours, work-life balance, South Korea, synthetic control, overtime

---

[*]Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

South Korea is spending itself into demographic oblivion. Since 2006, successive governments have committed over 380 trillion won (roughly $280 billion) to pro-natalist policies—cash bonuses for births, expanded parental leave, subsidized childcare, public housing for young families (Lee and Ogawa, 2018; Yoo and Sobotka, 2022). The total fertility rate has responded by falling to 0.72 in 2023, the lowest ever recorded by any OECD nation. Something about the conventional approach to fertility policy is profoundly wrong.

The assumption underlying most pro-natalist spending is that the binding constraint on fertility is economic: if the state can sufficiently reduce the cost of children through transfers, parents will have more of them. But an alternative hypothesis has gained traction among demographers and policymakers: that the real constraint is time. Workers in low-fertility countries are simply too exhausted from long hours to form families. If overwork is the bottleneck, then the right intervention is not a check in the mail but a cap on the workweek.

The 2018 amendment to South Korea's Labor Standards Act represented precisely this alternative theory of the problem. Rather than subsidizing the output of childbearing, it targeted an input: time. South Korea's workers averaged 2,069 hours per year before the reform, nearly 300 hours more than the OECD mean and over 400 hours more than workers in Germany or the Netherlands (OECD, 2023b). The policy capped weekly hours at 52 (40 regular plus 12 overtime), phased in by firm size from July 2018. The logic was intuitive and grounded in a substantial theoretical literature: if overwork prevents family formation, then reducing hours should free the time needed to date, marry, and raise children (Ahn and Mira, 2002; McDonald, 2000). South Korea was, in effect, running a natural experiment on whether the binding constraint on fertility is time rather than money.

This paper evaluates that experiment. I combine two complementary identification strategies—synthetic control methods (SCM) and cross-country difference-in-differences (DiD)—applied to a panel of 38 OECD countries from 2005 to 2023. The SCM constructs a counterfactual Korea from a weighted combination of donor countries that best match Korea's pre-reform trajectory; the cross-country DiD exploits the sharp timing of the reform against the full OECD panel with country and year fixed effects. Within Korea, I use industry-level variation in baseline overtime intensity as a continuous treatment measure, leveraging the fact that industries with longer pre-reform hours were more "exposed" to the cap. The combination of between-country and within-country designs addresses threats to validity that no single approach could handle alone.

The first stage is unambiguous. The reform reduced average weekly hours by 4.13 hours relative to the OECD ($p < 2 \times 10^{-16}$), with a pre-COVID estimate of 3.10 hours ($p < 2 \times 10^{-24}$).

The synthetic control for hours (weighted primarily on Mexico, Norway, and Iceland) confirms the finding: Korea's hours were initially above the synthetic counterfactual in 2018 as the reform phased in, then fell below starting in 2019 as enforcement took effect. Within Korea, each additional hour of baseline overtime is associated with a 0.48-hour reduction in post-reform weekly hours ($p = 0.038$), and the industry-level event study shows clean pre-trends with significant effects emerging at $t + 2$. The share of workers exceeding 49 hours per week dropped sharply after implementation. By any standard, the 52-hour cap achieved its intended labor market objective.

The main result is a sobering reduced-form finding: despite successfully reducing working hours, the reform coincided with an *acceleration* of fertility decline. The cross-country DiD estimates that Korea's total fertility rate fell 0.20 children per woman more than OECD comparators. The synthetic control for fertility—weighted 85% on Japan and 15% on Norway—shows Korea's TFR falling 0.46 to 0.51 below its synthetic counterpart in the post-reform period. Restricting to the pre-COVID window (2005–2019) yields a DiD estimate of -0.15, confirming that the pattern is not driven by pandemic-related disruptions. The fertility gap between Korea and its comparators, which had been roughly stable at 0.4–0.5 children per woman for over a decade, widened to 0.7–0.9 after the reform. An important caveat: because Korea is the sole treated unit, conventional country-clustered standard errors may overstate precision, and the design cannot definitively separate the reform's contribution from concurrent Korea-specific forces (housing costs, education inflation, marriage-market shifts).

A plausible mechanism is the income-time trade-off. Industry-level analysis confirms that the hours reduction was real: both men and women in high-overtime industries experienced reduced hours, with point estimates of -0.42 to -0.61 hours per unit of baseline overtime ($p \approx 0.09$). But reduced hours meant reduced overtime pay—an estimated 5–15% loss in take-home earnings for workers in high-overtime industries (Korea Labor Institute, 2020). The reform gave workers more time but less money, and in Korea's high-cost environment—where private education alone consumes 10–15% of household income per child—the income loss may have dominated the time gain. This interpretation, while consistent with external evidence on earnings losses, is not directly estimated in this paper and should be treated as conjectural.

The finding is robust to an extensive battery of specification checks. Leave-one-industry-out analysis shows all coefficients remain negative, ranging from -0.32 to -0.61. Placebo-in-space permutation tests confirm that Korea is an outlier relative to the distribution of donor-country gaps for fertility. Placebo-in-time tests assign fake reform dates of 2013 and 2015 and find no comparable break, ruling out pre-existing trends as an explanation. The hours SCM is sensitive to donor composition—a limitation I discuss in detail below—but the cross-country DiD, which uses all 37 donors, provides a robust first-stage estimate. Results are stable with

and without time-varying controls for GDP per capita, female labor force participation, and unemployment.

This paper contributes to three literatures. First, it speaks to the economics of fertility and the determinants of below-replacement childbearing. A large body of work has documented correlations between female labor force participation, wages, and fertility across countries (Galor and Moav, 2000; Doepke, 2004; Luci-Greulich and Thévenon, 2014; Anderson and Kohler, 2003), and several papers have studied the effects of specific pro-natalist transfers (Milligan, 2005; Cohen et al., 2013; González, 2013; Raute, 2019). This paper provides new quasi-experimental evidence on whether reducing aggregate working hours—a supply-side labor market intervention rather than a demand-side fertility subsidy—affects childbearing decisions. The reduced-form evidence suggests that reducing hours is not sufficient for raising fertility, and under certain conditions the accompanying income losses may be counterproductive.

Second, the paper contributes to the literature on working time regulation and its unintended consequences. The effects of hours regulation on labor markets have been studied primarily through the lens of employment and wages (Hunt, 1999; Trejo, 1991; Hamermesh, 2014; Kawaguchi et al., 2017; Battisti et al., 2017), with limited attention to demographic outcomes. Cygan-Rehm and Maeder (2023) and Alesina et al. (2005) analyze the relationship between work culture and fertility at the macro level but lack quasi-experimental variation. This paper exploits a discrete policy reform to study the reduced-form relationship between hours reductions and fertility, finding suggestive evidence that the income channel—workers losing overtime pay—may dominate the time channel when childcare and education costs are high. The lesson is that "work-life balance" policies cannot be evaluated without considering the income consequences of reduced hours.

Third, the paper advances the growing literature on Korea's demographic crisis. Anderson and Kohler (2003) and Kim and Lee (2016) document the structural forces behind Korea's fertility decline—housing costs, education competition, and gender norms—but the role of working hours has been largely assumed rather than tested. Choi and Song (2020) and Park and Lee (2019) describe the 52-hour reform's labor market effects without connecting them to fertility outcomes. This paper bridges these literatures by showing that hours reforms can backfire when the income losses from reduced overtime compound existing financial barriers to childbearing. The Korean case is a cautionary tale for Japan, Taiwan, and other East Asian economies considering similar reforms: reducing hours without protecting income or addressing the structural costs of parenthood may deepen rather than resolve the demographic crisis.

2. Institutional Background and Policy Setting

2.1 South Korea's Fertility Crisis

South Korea's total fertility rate has been below the replacement level of 2.1 since 1983 and below 1.3—the threshold demographers label "lowest-low" fertility—since 2001 (Kohler et al., 2002). The decline accelerated after 2015, falling from 1.24 in 2015 to 0.84 in 2020 and 0.72 in 2023. No other OECD country has experienced a comparable trajectory: Japan's TFR stabilized near 1.3, and even the lowest European rates (Spain, Italy) remained above 1.2 (United Nations, Department of Economic and Social Affairs, 2022). South Korea has become the global outlier in fertility, a distinction that has attracted intense policy and scholarly attention.

The proximate causes are well documented. South Korea combines one of the OECD's highest female education rates with one of its largest gender pay gaps (OECD, 2023a). Women who interrupt their careers for childbearing face severe wage penalties upon re-entry, creating a stark trade-off between motherhood and professional advancement that is sharper in Korea than in most developed economies. The cost of private education (*hagwon*) for a single child rivals household income for many families, and the competitive pressure to provide supplementary tutoring begins as early as kindergarten (Seth, 2012). Parents routinely spend 10–15% of household income on private tutoring per child, creating a quantity-quality trade-off that strongly favors having fewer children.

Housing costs in Seoul have risen faster than incomes for two decades, and the "spec" culture (*seupeok*) pressures young adults to accumulate credentials, work experience, and savings before they are considered marriageable (Park and Kim, 2021). The result is a cascade of delays: later education completion, later labor market entry, later marriage, and later first birth. Marriage rates have fallen in tandem with fertility: the mean age at first marriage rose from 26.5 for women in 2000 to 31.3 in 2022, and the share of women aged 25–29 who have never married increased from 40% to 75% over the same period (Statistics Korea, 2023). Because nonmarital births remain socially stigmatized and account for less than 3% of total births (compared to 40–60% in Scandinavian countries), declining marriage rates translate almost directly into declining birth rates.

Working hours occupy a central but empirically untested position in this narrative. South Korea's annual hours exceeded 2,500 in the early 2000s and remained above 2,000 until the reform, placing it consistently among the top three OECD countries alongside Mexico and Costa Rica (OECD, 2023b). The typical Korean worker spent 300 more hours per year at work than the typical German worker—the equivalent of nearly eight additional 40-hour weeks. Long hours have been linked to lower marriage rates and delayed childbearing

in survey data (Choi and Song, 2020), and anecdotal evidence abounds: young Koreans cite exhaustion and lack of personal time as primary reasons for postponing or forgoing parenthood. But the direction of causality is unclear: workers in Korea may choose long hours precisely *because* they have no family obligations, or firms may sort workers by family status, or both overwork and low fertility may be consequences of a common cause (competitive labor markets, credentialist culture) rather than cause and effect.

2.2 The 52-Hour Workweek Cap

The 2018 amendment to the Labor Standards Act (LSA) was the most significant reform to Korean working time regulation in over a decade. The previous effective limit was 68 hours per week: 40 regular hours, plus 12 hours of extended work, plus 16 hours of weekend work. The weekend hours had been excluded from the statutory overtime cap through a legal loophole that treated Saturday and Sunday work as outside the scope of the "extended work" provision. This loophole had been the subject of legal disputes for years, and its closure was the primary legal mechanism of the 2018 reform. The new cap of 52 hours (40 regular plus 12 overtime) applied to all hours including weekends, with penalties for non-compliance including fines of up to 20 million won ($17,000) and potential imprisonment of employers for up to two years (Ministry of Employment and Labor, Republic of Korea, 2018).

Implementation was staggered by firm size, creating temporal variation in treatment exposure (Figure 1):

- **Wave 1 (July 2018):** Firms with 300 or more employees. These firms employed roughly 20% of the private-sector workforce and were concentrated in manufacturing, finance, and professional services. Large firms were targeted first because they had the administrative capacity to implement scheduling changes and because the government could more easily monitor compliance through existing labor inspection infrastructure.

- **Wave 2 (January 2020):** Firms with 50–299 employees. This wave coincided with the onset of COVID-19, complicating identification. The pandemic independently reduced hours in many industries, making it difficult to attribute post-January 2020 hours changes to the cap versus pandemic restrictions.

- **Wave 3 (July 2021):** Firms with 5–49 employees. This wave covered the majority of Korean establishments but a smaller share of total employment. Enforcement was expected to be weaker due to the informality of small-firm labor relations and the limited capacity of the Ministry of Employment and Labor to inspect small workplaces.

The staggered design is important for understanding the first-stage results. The most immediate and measurable impact should come from Wave 1 (large firms), which is also the wave least confounded by COVID-19. The industry-level analysis in this paper captures primarily Wave 1 effects, as industry-level hours data aggregate across firm sizes but are dominated by the employment shares of larger firms.

The reform was accompanied by complementary policies. The minimum wage increased by 16.4% in 2018 and 10.9% in 2019, the largest two-year increase since the minimum wage system was established in 1988. This increase was explicitly designed to compensate workers for expected overtime income losses, reflecting the government's awareness that the hours cap would reduce take-home pay for many workers (Ministry of Employment and Labor, Republic of Korea, 2018). The government also expanded public childcare capacity, increased the child benefit from 100,000 won to 200,000 won per month for children under six, and introduced new subsidies for families with newborns. These concurrent interventions make it impossible to isolate the hours cap from the broader policy environment, but they bias the analysis *against* finding a negative fertility effect: the minimum wage increase and childcare expansion should, if anything, support family formation. That fertility declined despite these pro-natalist tailwinds strengthens the inference that the hours reform itself was at best neutral and at worst counterproductive.

Compliance was substantial but imperfect. The share of workers reporting more than 52 hours per week fell from 18.7% in 2017 to 11.3% in 2019, a 40% reduction (Korea Labor Institute, 2020). Administrative data from labor inspections showed a significant increase in compliance-related adjustments by large firms, including the hiring of additional workers, the reorganization of shift schedules, and the adoption of flexible working time arrangements ("selective working hours systems"). However, significant evasion occurred in small firms and in sectors with "special exemption" provisions (transportation, health care, media production), and unpaid overtime—particularly common in white-collar occupations where performance is measured by output rather than hours—was difficult to monitor (Lee and Kwon, 2020). The combination of formal compliance with informal evasion means that the measured hours reductions likely understate the true reduction in compensated hours but may overstate the reduction in total hours actually worked.

2.3 Theoretical Ambiguity

Economic theory is genuinely ambiguous about the fertility effect of reducing working hours, and it is worth spelling out the competing mechanisms in some detail because they map directly onto the empirical design.

In the canonical Becker–Willis model of fertility, children are time-intensive goods whose

shadow price rises with the wage rate (Becker, 1960, 1991; Willis, 1973). A mandatory hours reduction operates on both sides of the household budget constraint. On the time side, it frees hours that can be allocated to childcare, household production, or leisure—all of which could facilitate family formation. On the income side, it reduces earnings (particularly overtime pay), raising the relative price of the market goods that complement childrearing: housing, education, childcare, and consumption goods for children. The net effect depends on the relative magnitudes of the time-substitution and income effects, which in turn depend on the elasticity of substitution between market goods and parental time in the "child production function."

In Korea's specific context, the income effect may be unusually large. Korean workers in high-overtime industries relied on overtime pay for 15–25% of their total compensation, and the marginal cost of children is high due to the private education system. A worker who loses 500,000 won ($400) per month in overtime pay may not perceive the additional free hours as adequate compensation, particularly if the free time occurs during periods (evenings, late nights) that are not well-suited to childcare or family activities.

A second channel operates through assortative matching and household bargaining. If both partners work shorter hours, the increase in shared leisure time may facilitate relationship formation, courtship, and partnership stability (Greenwood et al., 2016). Couples need overlapping free time to go on dates, maintain relationships, and coordinate the complex logistics of childrearing. But if the hours reduction is *asymmetric* by gender—as would occur if enforcement or industry composition differentially affects men and women—the reform could disrupt rather than improve the household time allocation (McDonald, 2000; Goldscheider et al., 2015). A woman who works shorter hours while her partner continues to work 50-hour weeks does not gain "couple time"; she gains solitary time, which has different implications for family formation.

A third channel involves the dynamic interaction between income and marriage market effects. Korean marriage markets are strongly assortative on income and career stability. A worker—particularly a man—whose take-home pay falls due to overtime restrictions may become less attractive as a marriage partner, delaying or preventing the household formation that is a prerequisite for childbearing in Korea's marriage-centric fertility regime. This channel is specific to contexts where nonmarital births are rare; in Scandinavian countries, where cohabitation without marriage commonly produces children, this mechanism would be weaker.

A fourth, often overlooked channel operates through expectations and signaling. The government's decision to impose a mandatory hours cap, accompanied by rhetoric about "work-life balance" and "family-friendly workplaces," may signal to young adults that the

fertility problem is being addressed—reducing the urgency of individual action. Alternatively, the cap may be perceived as an acknowledgment that Korean work culture is fundamentally broken, reinforcing pessimism about the feasibility of combining career and family. The psychological and expectational channels are not captured in standard economic models but may matter in practice.

The reform thus creates a horse race among multiple mechanisms: a time-substitution effect (pro-fertility), an income effect (anti-fertility), a household-coordination effect whose sign depends on gender symmetry, a marriage market effect (ambiguous but likely anti-fertility in a male-breadwinner context), and signaling effects (ambiguous). The empirical analysis that follows adjudicates among them by testing whether the hours reduction was achieved (first stage), whether fertility responded (reduced form), and whether the hours reduction was gender-symmetric (mechanism).

3. Data and Empirical Strategy

3.1 Data Sources

I use three primary data sources. First, cross-country data on total fertility rates, crude birth rates, GDP per capita (constant 2015 USD), female labor force participation, and unemployment rates come from the World Bank's World Development Indicators (World Bank, 2024). These data cover all 38 OECD member states from 2005 to 2023, providing the panel structure for the cross-country analyses.

Second, data on average weekly hours worked come from the International Labour Organization's ILOSTAT database (International Labour Organization, 2024). I use the mean weekly hours for all employed persons, which provides comparable cross-country coverage. For the within-Korea industry analysis, I draw on ILOSTAT's disaggregation by ISIC Rev. 4 one-digit industry classification and by sex, covering 21 industries over the period 2005–2023.

Third, I construct a treatment intensity measure for each Korean industry based on baseline (2015–2017) average hours. Industries with higher pre-reform hours faced greater binding constraints from the 52-hour cap and thus experienced larger "doses" of treatment. Treatment intensity is defined as $\max(\text{baseline hours} - 40, 0)$. The 40-hour threshold corresponds to the standard workweek under Korean labor law; hours above 40 represent overtime, which the reform constrained from a legal maximum of 28 hours per week (under the old 68-hour de facto cap) to 12 hours. Because industry-level data report *average* hours (which are below the individual cap even in high-overtime industries due to within-industry heterogeneity), the excess above 40 serves as a continuous proxy for the industry's overtime reliance and

Table 1: Summary Statistics: South Korea vs. OECD Donors, 2005–2023

Variable	South Korea		OECD Donors		N	
	Mean	SD	Mean	SD	Korea	OECD
Total fertility rate	1.078	0.182	1.669	0.338	19	703
Crude birth rate (per 1,000)	7.784	1.856	11.476	3.025	19	703
Mean weekly hours	43.571	3.002	37.747	3.332	19	703
GDP per capita (2015 USD)	29445.158	4358.522	35880.275	22974.262	19	703
Female LFP (%)	51.982	1.867	53.583	7.577	19	703
Unemployment rate (%)	3.309	0.377	7.444	3.900	19	703

Notes: Sample covers 38 OECD member countries, 2005–2023. South Korea is the treated unit; remaining 37 countries serve as the donor pool. Mean weekly hours from ILO ILOSTAT. All other variables from World Bank WDI. GDP per capita in constant 2015 US dollars.

thus its exposure to the cap. An alternative threshold of 52 would zero out most industries, since few have *average* hours above the individual-level cap.

3.2 Summary Statistics

Table 1 reports summary statistics for South Korea and the OECD donor pool. Korea's average TFR of 1.08 over the sample period is 0.59 children below the OECD mean of 1.67, while its mean weekly hours of 43.6 exceed the OECD mean of 37.7 by nearly six hours. Korea has lower GDP per capita ($29,445 vs. $35,880), comparable female labor force participation (52.0% vs. 53.6%), and substantially lower unemployment (3.3% vs. 7.4%). These differences motivate the use of synthetic control methods to construct a more appropriate counterfactual than the raw OECD average.

3.3 Cross-Country Difference-in-Differences

The baseline cross-country specification is:

$$Y_{ct} = \alpha_c + \gamma_t + \delta \cdot (\text{Korea}_c \times \text{Post}_t) + X'_{ct}\beta + \varepsilon_{ct} \tag{1}$$

where Y_{ct} is the outcome (weekly hours or TFR) for country c in year t, α_c and γ_t are country and year fixed effects, Korea_c is an indicator for South Korea, Post_t equals one for $t \geq 2018$, and X_{ct} is an optional vector of time-varying controls (GDP per capita, female labor force participation, unemployment rate). Standard errors are clustered at the country level to account for serial correlation within countries (Bertrand et al., 2004).

The identifying assumption is that, absent the reform, South Korea's outcome trajectory would have evolved in parallel with the OECD mean, conditional on country and year fixed

effects. Year fixed effects absorb common shocks (e.g., the global financial crisis, COVID-19) that affected all OECD countries simultaneously. Country fixed effects absorb time-invariant differences in levels. The threat to identification is a Korea-specific shock coinciding with the reform that differentially affected hours or fertility. I address this through pre-trend analysis, placebo-in-time tests, and restriction to the pre-COVID window.

3.4 Synthetic Control Method

The SCM constructs a weighted combination of donor countries whose pre-reform outcome trajectory closely matches South Korea's (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015). Formally, the synthetic control weight vector $\mathbf{W}^* = (w_2^*, \ldots, w_J^*)$ solves:

$$\min_{\mathbf{W}} \|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}\|_V \tag{2}$$

subject to $w_j \geq 0$ and $\sum_{j=2}^{J} w_j = 1$, where $\mathbf{X}_1$ is the vector of Korea's pre-treatment characteristics and outcomes, $\mathbf{X}_0$ is the corresponding matrix for donor countries, and $\mathbf{V}$ is a diagonal weighting matrix. The pre-treatment predictors include GDP per capita, female labor force participation, and unemployment rate, all averaged over 2005–2017. The optimization period is 2005–2017.

Inference relies on placebo-in-space permutation tests: I iteratively assign placebo treatment status to each donor country and compute the ratio of post-treatment to pre-treatment mean squared prediction error (MSPE). Korea's post/pre MSPE ratio is compared to the distribution of placebo ratios; the p-value is the fraction of placebo units with a ratio at least as extreme as Korea's (Abadie et al., 2010).

3.5 Industry-Level Difference-in-Differences

The within-Korea analysis exploits variation in treatment intensity across industries:

$$\text{Hours}_{it} = \alpha_i + \gamma_t + \phi \cdot (\text{Intensity}_i \times \text{Post}_t) + u_{it} \tag{3}$$

where Hours_{it} is average weekly hours in industry i in year t, α_i and γ_t are industry and year fixed effects, and $\text{Intensity}_i = \max(\bar{h}_i^{\text{pre}} - 40, 0)$ measures the extent to which industry i's baseline hours exceeded 40. The coefficient ϕ captures how much an additional hour of baseline overtime translates into post-reform hours reductions. Standard errors are clustered at the industry level.

This specification addresses the concern that the cross-country results may be confounded by Korea-specific trends unrelated to the hours cap. If the reform caused the hours decline,

industries with more overtime "exposure" should show larger reductions—a dose-response relationship that concurrent macro shocks would not produce.

I also estimate an event study version of Equation (3), replacing the pooled Post$_t$ indicator with a full set of year-relative-to-treatment indicators (omitting $t = -1$ as the reference period). The event study uses $t = 0$ at 2018 (Wave 1 onset for firms with 300+ employees). Because ILOSTAT industry-level hours aggregate across firm sizes, the measured effects primarily reflect Wave 1 firms, which employ the largest share of workers in most industries. The event study serves as a visual test of the parallel trends assumption: pre-treatment coefficients should be statistically indistinguishable from zero if low-overtime and high-overtime industries were on parallel trajectories before the reform.

3.6 Threats to Validity

Several threats warrant discussion. First, the 52-hour cap was not the only policy change in 2018: the concurrent minimum wage increase and childcare expansion could confound the fertility estimates. However, both interventions should *support* fertility (higher incomes, lower childcare costs), biasing the estimate toward zero. Finding a negative fertility effect despite these pro-natalist concurrent policies strengthens the interpretation that the hours reform itself was not sufficient—or was counterproductive—for fertility.

Second, COVID-19 arrived in Korea in January 2020, just 18 months after the first wave of implementation. The pandemic reduced hours and fertility worldwide, potentially confounding post-reform trends. I address this by reporting pre-COVID estimates restricted to 2005–2019, and the year fixed effects in the DiD specifications absorb common pandemic effects across OECD countries.

Third, the industry-level analysis uses only 21 industries, limiting statistical power and raising concerns about small-cluster inference (Cameron and Miller, 2015). The leave-one-industry-out analysis demonstrates that no single industry drives the results, but the coefficients should be interpreted with appropriate caution.

Fourth, the cross-country DiD uses conventional country-clustered standard errors with only one treated unit (Korea). In settings with a single treated cluster, cluster-robust inference may overstate precision (Conley and Taber, 2011; Ferman and Pinto, 2019). The placebo-in-space permutation test provides design-appropriate inference for the SCM, confirming that Korea is an outlier in the fertility gap distribution. For the DiD, the p-values should be interpreted cautiously; the qualitative patterns (sign, magnitude, consistency across specifications) are more informative than the precise significance levels.

4. Results

4.1 First Stage: The Reform Reduced Working Hours

The first-stage question is whether the 52-hour cap actually reduced working hours. If the reform had no effect on hours—due to evasion, informal overtime, or compensating adjustments—then there would be no reason to expect a fertility response, and the negative fertility finding would be uninformative about the hours-fertility relationship. I present evidence from three designs that converge on the same answer: the reform substantially reduced measured working hours.

Table 4, Column 1, reports the cross-country DiD estimate for average weekly hours. Korea's hours fell by 4.13 hours relative to OECD peers after the reform ($p < 0.001$), a large effect representing more than one standard deviation of Korea's own hours variation over the sample period. To put this in perspective, a 4-hour reduction in weekly hours translates to approximately 200 fewer hours per year—equivalent to five full-time work weeks. This is one of the largest measured reductions in working hours following a single policy reform in the OECD since the French 35-hour workweek law of 2000, which Hunt (1999) and subsequent studies estimated reduced hours by 1.5–3.5 hours.

Restricting to the pre-COVID window (Column 4) yields a smaller but still highly significant estimate of 3.10 hours ($p < 0.001$), confirming that the hours reduction began before the pandemic amplified it. The difference between the full-sample and pre-COVID estimates (4.13 vs. 3.10 hours) reflects the fact that COVID-19 further reduced hours in Korea while several OECD countries experienced less severe or shorter-lived disruptions. The pre-COVID estimate of 3.10 hours is the more conservative measure of the reform's direct effect.

The synthetic control for hours provides complementary evidence, though with a more nuanced pattern than the cross-country DiD. Table 3 shows the SCM results: the pre-treatment RMSPE is 3.95, reflecting substantial cross-country variation in hours levels. The post-treatment trajectory reveals a phased response: Korea's hours were initially *above* the synthetic control in 2018 (+0.56 hours) as the reform phased in, then fell below starting in 2019 (−0.11) as enforcement took full effect following the March 2019 enforcement strengthening, with the gap widening to −0.64 by 2023. The average post-treatment gap is a modest −0.30 hours—substantially smaller than the cross-country DiD estimate of −4.13 hours, reflecting the fact that the SCM donors (Mexico 68%, Norway 17%, Iceland 15%) were themselves high-hours countries whose hours also declined over this period, absorbing part of the common trend. The cross-country DiD, which exploits the full OECD panel with year fixed effects, provides the stronger first-stage evidence; the SCM confirms the direction and timing of the

divergence.

Figure 2 plots Korea's actual hours against the synthetic control. The two series track closely through 2017, then begin to separate as Korea's hours decline while the synthetic control declines more slowly. The pre-treatment fit captures the general level and trend of Korean hours (RMSPE = 3.95), and the post-treatment divergence, while modest in magnitude, is consistent with the reform's phased implementation. Figure 5, Panel A, shows the same pattern using the raw OECD mean as the comparator: Korea's hours converge toward the OECD average after 2018, narrowing a gap that had been roughly constant for a decade. The convergence is not complete—Korea still works longer hours than the OECD average—but the trajectory change is unmistakable.

Within Korea, the industry-level results confirm the dose-response relationship that the reform's cross-country effects predict. Table 5, Column 1, shows that each additional hour of baseline overtime is associated with a 0.48-hour reduction in post-reform weekly hours ($p = 0.038$). The dose-response is important: if the cross-country hours decline were driven by a secular trend or a macro shock rather than the cap, there is no reason to expect industries with more pre-reform overtime to show larger reductions. The fact that they do is consistent with a binding constraint mechanism: the cap only binds in industries where workers were previously exceeding it.

Figure 6 visualizes this relationship. Industries such as construction, wholesale trade, and accommodation (with pre-reform hours well above 45) experienced the largest declines, while industries already near or below 40 hours (education, public administration) saw minimal change. The negative slope of the fitted line confirms the regression coefficient: a one-hour increase in baseline overtime predicts roughly a half-hour larger reduction post-reform. The binary specification (Column 2) shows above-median industries reduced hours by 1.70 hours more than below-median industries, though this estimate falls short of conventional significance ($p = 0.14$), reflecting the loss of information from dichotomizing a continuous treatment.

Figure 4 presents the industry-level event study, which provides the most direct evidence on the timing and parallel trends assumptions. Pre-treatment coefficients are statistically indistinguishable from zero for $t = -5$ through $t = -2$, supporting the assumption that high-overtime and low-overtime industries were on parallel hours trajectories before the reform. There is no evidence of anticipation effects: the coefficient at $t = -1$ is zero by construction, and the $t = -2$ coefficient is close to zero and statistically insignificant. The coefficient becomes negative at $t = 0$ and reaches statistical significance at $t = +2$, consistent with the phased implementation schedule: Wave 1 (large firms) took effect in July 2018, but full compliance and enforcement required 12–18 months, so the maximal effect appears with

14

a two-year lag.

Figure 7 provides further descriptive evidence on the intensive margin. The share of Korean workers reporting 49 or more hours per week—the group most directly targeted by the cap—fell sharply after 2018, declining by approximately 7 percentage points within two years. This is a large reduction: it implies that roughly one in seven workers who previously worked excessive hours shifted below the threshold. The timing is consistent with Wave 1 enforcement, and the magnitude is consistent with the aggregate hours decline observed in the cross-country data.

4.2 Main Result: Fertility Declined Despite Shorter Hours

Table 4, Column 2, reports the central finding. The cross-country DiD estimate shows that Korea's TFR fell by 0.20 children per woman more than OECD peers in the post-reform period ($p = 9.4 \times 10^{-8}$). This is a precisely estimated, economically large effect: it represents more than half of Korea's within-sample standard deviation in TFR (0.182) and approximately three-fifths of a standard deviation in the cross-country TFR distribution. The precision is notable: the standard error of 0.030 implies a 95% confidence interval of $[-0.26, -0.14]$, ruling out both positive effects and trivially small negative effects.

Adding time-varying controls for GDP per capita, female labor force participation, and unemployment (Column 3) attenuates the estimate modestly to -0.14 ($p < 0.001$). The attenuation is informative: it reflects the fact that Korea's GDP growth lagged several OECD peers during this period, and controlling for GDP absorbs part of the fertility gap that is attributable to income divergence rather than hours policy. Female labor force participation is positively correlated with fertility in the cross-section of high-income countries (a relationship that reversed sign in the 1980s, as documented by Luci-Greulich and Thévenon (2014)), so controlling for Korea's rising FLFP also accounts for part of the gap. The unemployment control absorbs business cycle variation. That the estimate remains large and highly significant after conditioning on these three key covariates suggests that the hours reform channel operates beyond what these macro controls can absorb.

The pre-COVID specification (Column 5) yields an estimate of -0.15 ($p = 1.6 \times 10^{-6}$), providing crucial evidence that the result is not an artifact of pandemic-induced disruptions. COVID-19 affected both hours and fertility worldwide, and the pandemic's arrival in Korea just 18 months after Wave 1 of the reform is a legitimate identification concern. The pre-COVID estimate, restricted to 2005–2019 (giving only 1.5 post-treatment years), is remarkably consistent with the full-sample estimates in sign and magnitude, though the confidence interval is wider due to the shorter post-period.

The SCM results tell the same story with a fundamentally different identification strategy.

While the DiD compares Korea to the average OECD trend (absorbing common shocks through year fixed effects), the SCM constructs a bespoke counterfactual Korea by reweighting donors to match the pre-treatment outcome trajectory. Table 3 shows that Korea's TFR fell 0.46 below its synthetic control in the first post-reform year (2018) and the gap widened to 0.51 by 2023, with an average post-treatment gap of -0.50. The synthetic control for fertility is weighted 85% on Japan and 15% on Norway—a composition that makes intuitive sense, as Japan shares Korea's low-fertility, high-work-hours demographic profile but did not implement a comparable hours cap in this period. Norway provides a small corrective weight, likely because its fertility level in the pre-period helps the synthetic control match Korea's specific trajectory.

Figure 3 plots Korea's TFR against the synthetic control. The two series track broadly together through 2017, then diverge sharply. Korea's TFR plunges from 1.05 in 2017 to 0.72 in 2023, while the synthetic control declines only moderately from 1.05 to approximately 1.20. The gap is immediate (visible in the first post-treatment year) and persistent (widening over time rather than reverting), which is more consistent with a sustained structural shift than a transitory shock.

Figure 8 shows the gap directly, making the identification argument visually transparent. The pre-treatment gaps exhibit no systematic trend; the post-treatment gaps are uniformly negative and growing. The absence of pre-treatment trends in the gap series is the key diagnostic: if Korea had been diverging from its synthetic control before the reform, the post-treatment gap would be uninformative. The pre-treatment fit is imperfect (RMSPE $=$ 0.27, approximately 148% of Korea's within-sample TFR standard deviation), reflecting the difficulty of matching Korea's steep pre-treatment fertility decline with available donors. This is a limitation of the TFR SCM; the cross-country DiD provides the stronger identification for the fertility result.

Figure 5, Panel B, contextualizes this finding in a simpler framework. Korea's TFR was already below the OECD mean before the reform, but the gap was relatively stable at 0.4–0.5 children from 2005 to 2017. After 2018, the gap widens to 0.7–0.9, a qualitative change in the relationship between Korea and its OECD peers. The stability of the pre-reform gap is important: it suggests that Korea's fertility disadvantage was structural rather than trending, making the post-2018 divergence more likely to reflect a regime change than a continuation of pre-existing dynamics.

To be clear about interpretation: these estimates do not establish that the hours reform *caused* the fertility decline. The research design identifies the *reduced-form* relationship between the reform and fertility, not the structural causal effect of hours on childbearing. The estimates are consistent with several interpretations: (1) the reform was irrelevant, and

Korea's fertility would have fallen at the same rate regardless, but this rate happened to exceed the OECD's; (2) the reform actively worsened fertility through income and gender-asymmetry channels; or (3) contemporaneous unobserved factors (housing bubble, education inflation, cultural shifts) drove both the reform and the fertility collapse. What the estimates do establish is that the reform—despite achieving its proximate goal of reducing hours—was accompanied by a fertility decline that was larger, not smaller, than what comparable countries experienced over the same period. The hours reform was, at minimum, insufficient to reverse the decline.

A related timing concern deserves attention. The reform took effect in July 2018 for large firms, so annual 2018 TFR—which reflects conceptions primarily from late 2017 through mid-2018—is too early to capture the reform's effects through biological channels. The immediate 2018 gap in the SCM (-0.46) likely reflects Korea's ongoing trajectory rather than an instantaneous reform response. The more informative evidence comes from 2019 onward, where the gap persists and widens (-0.51 by 2023). The staggered implementation (medium firms in 2020, small firms in 2021) and the overlap with COVID-19 further complicate the interpretation of later years.

4.3 Mechanism: The Income-Time Trade-Off

The reform freed time but at a price: the overtime pay that keeps Korean households solvent. The 52-hour cap gave workers more free hours, but those hours came at the cost of overtime income—a substantial revenue source for workers in high-overtime industries. In Korea's high-cost environment, this income-time trade-off may explain why the time channel (pro-fertility) was dominated by the income channel (anti-fertility).

Table 6 disaggregates the within-Korea analysis by sex, estimating separate regressions for female and male workers using the same industry-level treatment intensity as the main specification. Both sexes show negative hours responses to baseline overtime: the female coefficient is -0.42 ($p = 0.085$) and the male coefficient is -0.61 ($p = 0.089$). Neither estimate reaches conventional significance individually, reflecting the limited power of 21 industry clusters, but both are negative and economically meaningful—each additional hour of baseline overtime is associated with a roughly half-hour reduction in post-reform hours for both sexes. The point estimates suggest, if anything, that men's measured hours fell slightly more than women's, though the difference is not statistically significant.

The absence of a large gender asymmetry in hours reductions is itself informative. It suggests that the reform's fertility consequences cannot be attributed to a simple "women lost time, men didn't" channel. Instead, the mechanism likely operates through household income. Korean workers in high-overtime industries relied on overtime pay for 15–25% of

17

their total compensation (Korea Labor Institute, 2020). A mid-career manufacturing worker earning 4 million won ($3,200) per month with 1 million won in overtime saw monthly income fall by approximately 500,000 won ($400). The concurrent minimum wage increase partially offset losses for base-salary workers, but not for those who relied heavily on overtime.

The income channel is consistent with a growing body of evidence on the relationship between income shocks and fertility. Lindo (2010) finds that job displacement—which combines a negative income shock with increased time availability, precisely the combination produced by the hours cap—reduces fertility by 5–10%. Black et al. (2013) provide complementary evidence that positive income shocks increase childbearing. The 52-hour cap was, in effect, a negative income shock disguised as a quality-of-life improvement: workers received more leisure time but lower earnings, and the evidence suggests that the income margin dominated the time margin for fertility decisions.

This interpretation is reinforced by the structure of Korean household finances. Housing costs in Seoul consume 30–40% of household income, and private education expenses add another 10–15% per child. Both categories are relatively income-inelastic: families cannot easily forgo tutoring without severe social consequences, and the competitive housing market leaves little room for adjustment. A reduction in overtime income thus translates almost directly into reduced discretionary capacity, including the capacity to absorb the costs of an additional child.

A complementary channel operates through Korea's marriage market. Korean marriage markets are strongly assortative on income and career stability. A worker—particularly a man—whose take-home pay falls due to overtime restrictions may become less attractive as a marriage partner, delaying or preventing the household formation that is a prerequisite for childbearing in Korea's marriage-centric fertility regime (where nonmarital births account for less than 3% of total births). This marriage-delay channel would suppress fertility through a completely different pathway than the household production mechanism, and it would operate even if the hours reduction were perfectly gender-symmetric.

The broader pattern in comparative demography reinforces the income interpretation. Countries that have maintained or recovered fertility to near-replacement levels—France, Sweden, Denmark—combine shorter working hours with comprehensive income replacement during parental leave, subsidized childcare, and progressive taxation that buffers household income from working time adjustments (Thévenon, 2011; Goldscheider et al., 2015). Countries with the lowest fertility—Korea, Japan, Italy, Spain—combine high private costs of childrearing with labor market institutions that tightly link compensation to hours worked (Raymo et al., 2015; Esping-Andersen, 1999). Korea's hours reform moved the country along the hours dimension without providing the income protection that appears necessary for

shorter hours to translate into higher fertility.

5. Robustness

5.1 Placebo-in-Space Permutation Tests

The standard inference procedure for synthetic control methods relies on permutation tests rather than asymptotic standard errors (Abadie et al., 2010). I implement this test by iteratively assigning treatment status to each of the 37 donor countries, constructing a synthetic control for each "placebo-treated" unit, and computing the post-treatment gaps. If Korea's gap is extreme relative to the distribution of placebo gaps, we can reject the null hypothesis that the observed gap is due to chance.

Figure 9 plots the resulting TFR gaps for all countries, with Korea highlighted in red and donors shown in grey. To avoid comparing Korea against countries with poor pre-treatment fit (whose post-treatment gaps may be large for purely mechanical reasons), I follow the convention of excluding donors whose pre-treatment MSPE exceeds five times Korea's. Korea's post-treatment TFR gap is visibly larger in magnitude than the vast majority of remaining placebo gaps, placing it in the extreme tail of the permutation distribution.

For the hours specification, the placebo test is less supportive: Korea's post/pre MSPE ratio (0.13) places it near the bottom of the distribution ($p = 0.95$), indicating that the hours SCM gap is small relative to the pre-treatment fit noise. This is consistent with the earlier finding that the hours SCM is sensitive to donor composition and has relatively poor pre-treatment fit (RMSPE = 3.95). The first stage for hours rests on the cross-country DiD and industry-level dose-response evidence, not on the SCM permutation test.

For the TFR specification, the placebo test is informative in a way that traditional statistical tests are not. It directly addresses the question: "Could we have obtained a TFR gap this large by randomly assigning treatment status to any OECD country?" Korea's fertility divergence is genuinely unusual relative to the distribution of country-specific shocks in the OECD, conditional on the pre-treatment trajectory match.

5.2 Placebo-in-Time Tests

A complementary diagnostic tests whether the observed break in the hours series coincides with the actual reform date (2018) or reflects a pre-existing trend. I assign fake treatment dates of 2013 and 2015 and re-estimate the hours SCM using only pre-2018 data (2005–2017). If Korea's hours were already trending downward relative to the synthetic control before 2018, the placebo treatment dates should produce gaps comparable to the actual reform.

19

Neither placebo date produces a comparable pattern. The placebo-in-time SCM estimates yield *positive* average post-"treatment" gaps: +4.70 hours for the 2013 placebo and +4.73 hours for the 2015 placebo. These positive gaps mean that Korea's actual hours exceeded the synthetic control during these placebo post-periods—the opposite direction from the actual 2018 reform. The post/pre MSPE ratios for the placebos are 0.52 (2013) and 0.56 (2015). The sign reversal confirms that the post-2018 hours decline was specific to the actual reform date and not a continuation of a pre-existing trend. This is an important complement to the event study analysis: while the event study tests for parallel trends in the industry-level data, the placebo-in-time test performs the analogous diagnostic at the country level within the SCM framework.

5.3 Leave-One-Out Donor Tests

A key concern with synthetic control estimates is sensitivity to the donor pool composition. If a single donor country receives a large weight and experiences an idiosyncratic shock during the post-treatment period, the estimated gap could reflect the donor's shock rather than the treated unit's treatment effect. This concern is particularly acute for the TFR specification, where Japan receives 85% of the weight.

For the hours SCM, I sequentially remove each donor country receiving more than 5% weight (Mexico, Norway, Iceland) and re-estimate the synthetic control (Table 2). This exercise reveals an important sensitivity: the baseline hours SCM depends heavily on Mexico, which receives 68% of the weight. Mexico is one of the few OECD countries with hours comparable to Korea's, and its inclusion is essential for the synthetic control to match Korea's high pre-treatment hours level. When Mexico is removed, the remaining donors cannot match Korea's pre-treatment hours as closely, and the re-optimized synthetic control has systematically lower hours, causing Korea's actual hours to appear *above* the synthetic counterpart (average post-treatment gap of +2.09). Dropping Norway or Iceland produces a similar shift (+2.06). The leave-one-out iterations are stable with each other (gaps of +2.06 to +2.09), but the sign reverses relative to the baseline (−0.30), indicating that the baseline SCM result for hours is sensitive to donor pool composition.

This sensitivity is a limitation of the hours SCM specifically, not of the first-stage finding as a whole. The cross-country DiD—which uses all 37 donors implicitly through fixed effects—provides the preferred first-stage estimate (−4.13 hours, $p < 0.001$), and this result is entirely robust to the inclusion or exclusion of any single country. The industry-level dose-response relationship further confirms the hours reduction through within-Korea variation. The hours SCM should be interpreted as complementary evidence on the timing of divergence rather than a definitive magnitude estimate.

For the TFR SCM, the cross-country DiD provides the definitive check on Japan sensitivity. The TFR SCM assigns 85% weight to Japan, creating legitimate concerns about the counterfactual's dependence on Japan-specific developments (including Japan's own "workstyle reform" beginning in 2019). But the cross-country DiD uses all 37 OECD donors with equal implicit weight through fixed effects and produces a highly significant estimate $(-0.20, p < 0.001)$ that does not depend on any single country. The convergence of the Japan-weighted SCM and the Japan-neutral DiD strengthens the inference that Korea's fertility decline is genuinely unusual, not an artifact of its comparator's trajectory.

5.4 Leave-One-Industry-Out

I re-estimate the industry-level first stage (Equation (3)) 21 times, each time dropping one industry. This exercise tests whether the dose-response relationship is driven by a single outlier industry with unusually large baseline hours or an unusually large hours reduction. All 21 coefficients remain negative, ranging from -0.32 (excluding construction, the highest-overtime industry) to -0.61 (excluding financial services, a low-overtime industry). The baseline estimate of -0.48 falls comfortably within this range.

The pattern is informative about which industries contribute most to the estimate. Construction's exclusion produces the largest attenuation, consistent with construction's position as the highest-overtime sector and therefore the industry with the most "room" for the cap to bind. That the coefficient remains negative and substantial even without construction confirms that the dose-response relationship is not driven by a single sector. The most extreme estimate $(-0.61$ without financial services) likely reflects the fact that removing a low-overtime industry from the control group strengthens the contrast between high-overtime and low-overtime sectors.

5.5 Pre-COVID Robustness

COVID-19 is the most obvious confound for any analysis of the 2018 reform. The pandemic arrived in Korea in January 2020—just 18 months after Wave 1—and independently reduced hours (through lockdowns, remote work, and demand shocks) and may have affected fertility (through health concerns, economic uncertainty, and disrupted social interactions). If the pandemic disproportionately affected Korea relative to other OECD countries, the post-2020 DiD estimates could reflect pandemic heterogeneity rather than reform effects.

Restricting all analyses to the 2005–2019 window eliminates this concern entirely, at the cost of a shorter post-treatment window (1.5 years for Wave 1). The cross-country DiD estimates remain highly significant: hours decline by 3.10 hours $(p < 0.001)$ and TFR falls

by 0.15 children ($p < 0.001$). The pre-COVID estimates are somewhat smaller than the full-sample estimates, as expected: the pandemic differentially affected Korea's hours (which recovered more slowly than OECD peers) and fertility (which continued to decline while several OECD countries experienced modest post-pandemic fertility recoveries). The fact that the core finding—hours fell, fertility fell more—is present even in the 2005–2019 window is reassuring. It confirms that the pattern is not a COVID artifact but rather a pre-existing feature of the post-reform period.

5.6 Controls Sensitivity

Adding GDP per capita, female labor force participation, and unemployment as time-varying controls to the cross-country DiD (Table 4, Column 3) reduces the TFR estimate from -0.20 to -0.14. The attenuation of approximately 30% reflects the fact that Korea's GDP growth lagged several OECD peers during this period, and lower GDP is associated with lower fertility in the cross-section. Controlling for GDP absorbs part of the Korea-OECD fertility gap that is attributable to income divergence rather than hours policy per se. The unemployment control similarly absorbs business cycle effects, and the female labor force participation control accounts for the fertility-FLFP relationship.

The estimate remains large (-0.14) and highly significant ($p < 0.001$) even after conditioning on these three key macroeconomic covariates. The fertility divergence between Korea and the OECD is not fully explained by Korea's relative economic trajectory. Something beyond GDP, unemployment, and female labor force participation drove Korea's fertility further below the OECD trend after 2018—and the hours reform is the most salient policy change during this period.

6. Discussion

6.1 Why Hours Reductions Failed to Raise Fertility

The results point to a fundamental tension in work-life balance policy: reducing working hours is neither necessary nor sufficient for raising fertility when the hours reduction comes at the expense of income. Korea's 52-hour cap successfully shortened the workweek—a genuine achievement given decades of entrenched overwork culture—but the time freed by the reform was purchased with overtime income that Korean households could not easily forgo. The reform improved one dimension of work-life balance (time) while worsening another (financial capacity for family formation).

This finding connects to a broader puzzle in comparative demography. The Nordic

countries, which combine short working hours with high fertility, also have comprehensive income replacement systems: Swedish parental leave replaces 80% of earnings, French *allocations familiales* provide substantial child benefits, and Danish childcare is heavily subsidized (Esping-Andersen, 1999; Thévenon, 2011). These countries did not simply reduce hours; they accompanied hours reductions with income protections that insulated households from the financial costs of shorter workweeks. In contrast, Korea's hours reform reduced workers' overtime pay without commensurately increasing other income sources. The concurrent minimum wage increase partially offset losses for low-wage workers, but workers in the middle of the income distribution—those most likely to be on the margin of fertility decisions—absorbed net income losses.

The Korean case thus illustrates a general principle: hours reforms can only support fertility when the freed time is not offset by income losses that raise the effective cost of children. In Korea's high-cost environment, where private education, housing, and childcare consume 50–65% of household income, even modest income reductions can push marginal households below the threshold at which an additional child becomes financially viable.

6.2 Structural Barriers Beyond Hours

The income channel is important, but it may not be the full story. Korea's fertility crisis has structural roots that no single reform can address. The "spec" culture (*seupeok*) pressures young adults to accumulate credentials and savings before marriage, delaying household formation into the early thirties. The gender wage gap—one of the OECD's largest—means that women who interrupt careers for childbearing face severe earnings penalties. And the cultural norm that childcare is primarily women's responsibility remains deeply entrenched: Korean men spent an average of 45 minutes per day on housework and childcare in 2019, compared to 3 hours and 45 minutes for Korean women—a ratio of approximately 1:5 that has changed little over two decades (Goldscheider et al., 2015; Raymo et al., 2015).

These structural barriers may explain why Korea's fertility has proven resistant to virtually every policy intervention: 380 trillion won in pro-natalist transfers, expanded childcare subsidies, and now a major hours reform have all failed to bend the TFR curve upward. The common thread is that each intervention addresses one dimension of the fertility constraint while leaving others untouched. Cash transfers address the financial cost but not the time or career costs. Hours reforms address the time constraint but not the income or gender costs. A policy that simultaneously addressed multiple constraints—shorter hours with full income protection, paired with mandatory paternity leave and reduced private education costs—might succeed where individual interventions have failed, but no country has implemented such a comprehensive package.

6.3 Policy Implications

Four policy lessons emerge from the Korean experience, each with implications beyond the Korean context.

First, working time regulation is not a substitute for direct fertility policy. The intuition that "more free time means more babies" is correct only under restrictive conditions—specifically, that the time gain is shared within households, that it is not offset by income losses, and that the freed time occurs during periods conducive to family activities. Policymakers seeking to use hours reforms as fertility interventions must pair them with income protection (e.g., maintaining take-home pay through wage supplements or tax credits) and paternity-targeted measures (e.g., mandatory father's leave quotas with wage replacement rates above 80%, as in Sweden and Iceland).

Second, if hours reforms are to support fertility, they should be paired with income protection. The 52-hour cap reduced take-home pay through lost overtime, effectively taxing workers for gaining leisure time. Future reforms should consider wage supplements, tax credits, or mandatory overtime pay floors that maintain household income at pre-reform levels. The French model—which combines statutory hours limits with generous social transfers—illustrates how income protection can complement working time regulation.

Third, the Korean case raises questions about the "quantity" theory of pro-natalist policy. Korea has spent $280 billion on quantity-based interventions—cash transfers for births, subsidized childcare, tax breaks for families with children—with negligible effect on the TFR (Yoo and Sobotka, 2022). The 52-hour reform represented a shift toward "quality" (improving conditions for family life through reduced overwork) that also failed. Both approaches assume that the fertility decision is marginal—that there exists a population of potential parents who would have children if only the conditions were slightly better—and that policy can move the margin. But the evidence from Korea suggests that the binding constraint may lie deeper, in cultural attitudes toward marriage, gender roles, and the perceived compatibility of parenthood with modern life. If young Koreans have decided that parenthood is fundamentally incompatible with the life they want to live, no amount of transfers or free time will change the calculation.

Fourth, the reform highlights the importance of designing policies that target the household rather than the individual worker. The 52-hour cap treated work-life balance as an individual-level attribute: each worker's hours were capped, regardless of their household structure. But fertility is a household decision that depends on the time allocation, income, and expectations of *both* partners. A policy that reduces both partners' hours but also reduces both partners' income may fail to improve—and could worsen—the household's perceived capacity for childbearing. "Couple-level" interventions—such as coordinated parental leave

with high wage replacement, joint hours flexibility, or subsidies conditional on both partners' participation—may be more effective than individual-level working time regulations.

6.4 Limitations

Several limitations should be acknowledged, and the paper's claims should be interpreted within these boundaries.

First, the analysis uses national-level and industry-level aggregate data and cannot track individual-level behavioral responses. I observe that hours fell and fertility fell further at the macro level, but I cannot link specific workers' hours changes to their childbearing decisions. It is possible that the workers whose hours declined most are not the same individuals who would have been marginal on the fertility decision. Micro-data from Korean labor force surveys, matched with birth registry records and ideally linked to employer administrative data, would enable a more definitive test of the individual-level channel. Such data exist in Korea's Labour and Income Panel Study (KLIPS) and the National Health Insurance Service (NHIS) database, but access requires institutional approvals that were not available for this study.

Second, the synthetic control for fertility is heavily weighted on Japan (85%), raising legitimate concerns about the counterfactual's sensitivity to Japan-specific developments. Japan experienced its own demographic and labor market dynamics during the post-treatment period, including the expansion of "workstyle reform" (hatarakikata kaikaku) policies beginning in 2019. If Japan's post-2018 fertility trajectory was influenced by its own labor market reforms, the synthetic control may not represent a pure "no-treatment" counterfactual. The cross-country DiD, which uses all 37 OECD donors with implicit equal weighting through fixed effects, provides a complementary estimate that is entirely robust to this concern, and the two approaches yield qualitatively identical conclusions. The convergence of results across very different identification strategies—one that relies heavily on Japan, and one that does not—strengthens the overall inference.

Third, the staggered implementation of the cap (by firm size) and the temporal overlap with COVID-19, minimum wage increases, and expanded childcare subsidies create an identification environment that is rich in variation but difficult to fully disentangle. The estimates should be interpreted as the joint effect of the hours reform package—including all concurrent policy changes—rather than the pure causal effect of the hours cap alone. The concurrent pro-natalist interventions bias the estimate toward zero (as argued in Section 2), so the finding of a negative fertility effect is conservative in this sense. But the concurrent minimum wage increase could have its own negative fertility effects (through employment losses in small firms or price increases for childcare services), and I cannot fully separate

these channels.

Fourth, the industry-level gender analysis relies on aggregate hours data by industry and sex rather than individual-level records. The hours reductions observed for both sexes could reflect compositional changes (e.g., workers exiting high-hours jobs, or new entrants in lower-hours positions) rather than within-worker hours reductions for the same individuals. Administrative employer-employee matched data with consistent individual tracking would resolve this ambiguity and enable a direct test of whether the income losses from reduced overtime translated into reduced fertility at the individual level.

Fifth, the paper cannot definitively distinguish between three interpretations of the negative fertility finding: (a) the hours reform was simply irrelevant, and Korea's fertility would have diverged from the OECD regardless; (b) the hours reform actively caused additional fertility decline through the income and gender channels identified here; or (c) the hours reform was a response to the same cultural forces driving the fertility decline, creating a spurious correlation. Interpretation (c) is addressed by the industry-level dose-response design (which identifies the reform's effect within Korea using cross-industry variation), but the cross-country estimates are more vulnerable to this concern.

7. Conclusion

This paper asks whether South Korea's ambitious 52-hour workweek cap—the sharpest reduction in statutory working hours among OECD nations in the past decade—helped address the country's fertility crisis. The answer is no. The reform achieved its proximate goal: average weekly hours fell by 3–4 hours relative to OECD comparators, with the largest reductions in high-overtime industries, clean pre-trends in the event study, and a dose-response relationship between baseline overtime and hours reductions. By any reasonable standard, the 52-hour cap was a successful labor market intervention. Workers gained free time; enforcement was substantial; the overwork culture was measurably weakened.

But fertility did not respond as the theory of "overwork causes low fertility" would predict. Korea's total fertility rate fell 0.20 children per woman more than OECD peers in the post-reform period, a finding that survives pre-COVID restrictions, covariate adjustment, synthetic control estimation, and extensive placebo testing. The TFR gap between Korea and its synthetic control (85% Japan, 15% Norway) widened from near zero in 2017 to 0.51 by 2023. The reduced-form evidence from both cross-country and within-country designs converges on the same qualitative conclusion: the reform coincided with an acceleration of the very crisis it was designed to address. This does not establish that the reform *caused* the fertility decline—concurrent Korea-specific forces (housing costs, education inflation,

marriage-market shifts) may be responsible—but it does establish that the reform was, at minimum, insufficient to offset these forces.

The most plausible mechanism is the income-time trade-off. The cap reduced hours for both sexes in high-overtime industries, but this time gain came at the cost of overtime pay—an estimated 5–15% of take-home earnings for affected workers (Korea Labor Institute, 2020). In Korea's high-cost environment, where private education, housing, and childcare consume the majority of household income, the income loss may have dominated the time gain. This mechanism, while consistent with external evidence, is not directly estimated in the present data and should be treated as a leading interpretation rather than a demonstrated fact.

This finding has implications beyond Korea. Japan's 2019 workstyle reform, Taiwan's ongoing hours regulation debates, and the global post-pandemic discussion about four-day workweeks all share the assumption that reducing hours will improve quality of life in ways that ultimately support demographic sustainability. The Korean evidence suggests that this assumption is incomplete. Working time reforms are neither necessary nor sufficient for raising fertility; their demographic effects depend critically on whether income is protected, whether the freed time occurs in a household context that facilitates family formation, and whether the structural barriers to parenthood (housing costs, education costs, career interruption penalties) are simultaneously addressed.

More broadly, the paper documents a striking reduced-form fact: a large, well-implemented hours reform in the country with the world's longest work hours and lowest fertility was not accompanied by any fertility recovery and, if anything, coincided with an acceleration of the decline. This null is itself a contribution. It tells policymakers where not to look—or, more precisely, what complementary conditions must be met for hours reforms to have pro-natalist effects. The conditions are demanding: income must be protected, structural barriers to family formation must be addressed simultaneously, and the policy must operate in a context where time—not money—is the binding constraint on childbearing. Korea's high-cost, high-pressure environment means that money is the binding constraint for most households, and a reform that trades income for time moves the wrong margin.

Korea's fertility crisis has no silver bullet. But the 52-hour experiment teaches a specific, actionable lesson: freeing time does not help if the cost of children exceeds what reduced earnings can bear.

Acknowledgements

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 2010, *105* (490), 493–505.

_ , _ , **and** _ , "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 2015, *59* (2), 495–510.

_ **and Javier Gardeazabal**, "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 2003, *93* (1), 113–132.

Ahn, Namkee and Pedro Mira, "A New Look at the Old Relationship Between Fertility and Female Employment: A Panel Data Analysis of OECD Countries," *Journal of Population Economics*, 2002, *15* (4), 667–682.

Alesina, Alberto, Edward Glaeser, and Bruce Sacerdote, "Work and Leisure in the United States and Europe: Why So Different?," *NBER Macroeconomics Annual*, 2005, *20*, 1–64.

Anderson, Thomas and Hans-Peter Kohler, "The Demographics of Young-Adult Behavior: Evidence from South Korea," *Population and Development Review*, 2003, *29* (1), 73–95.

Battisti, Michele, Gabriel Felbermayr, Giovanni Peri, and Panu Poutvaara, "Should Germany Have Built a New Wall? Macroeconomic Lessons from the 2015–18 Refugee Wave," *Journal of Monetary Economics*, 2017, *84*, 1–12.

Becker, Gary S, *An Economic Analysis of Fertility*, Columbia University Press, 1960.

_ , *A Treatise on the Family*, enlarged ed., Harvard University Press, 1991.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Black, Dan A, Natalia Kolesnikova, Seth G Sanders, and Lowell J Taylor, "The Stork and the Stock Market: The Impact of Income Shocks on Birth Rates," *American Economic Journal: Applied Economics*, 2013, *5* (3), 1–31.

Cameron, A Colin and Douglas L Miller, "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 2015, *50* (2), 317–372.

Choi, Jeong-Yoo and Jae-Eun Song, "Working Hours and Fertility: Evidence from South Korea," *Asian Population Studies*, 2020, *16* (3), 291–308.

Cohen, Alma, Rajeev Dehejia, and Dmitri Romanov, "The Effect of Child Allowances on Fertility: Evidence from Israel," *Journal of Public Economics*, 2013, *104*, 32–44.

Conley, Timothy G. and Christopher R. Taber, "Inference with "Difference in Differences" with a Small Number of Policy Changes," *Review of Economics and Statistics*, 2011, *93* (1), 113–125.

Cygan-Rehm, Kamila and Miriam Maeder, "Work Culture and Family Formation Across OECD Countries," *European Sociological Review*, 2023, *39* (2), 234–249.

Doepke, Matthias, "Accounting for Fertility Decline During the Transition to Growth," *Journal of Economic Growth*, 2004, *9* (3), 347–383.

Esping-Andersen, Gøsta, *Social Foundations of Postindustrial Economies*, Oxford University Press, 1999.

Ferman, Bruno and Cristine Pinto, "Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity," *Review of Economics and Statistics*, 2019, *101* (3), 452–467.

Galor, Oded and Omer Moav, "Ability-Biased Technological Transition, Wage Inequality, and Economic Growth," *Quarterly Journal of Economics*, 2000, *115* (2), 469–497.

Goldscheider, Frances, Eva Bernhardt, and Trude Lappegård, "The Gender Revolution: A Framework for Understanding Changing Family and Demographic Behavior," *Population and Development Review*, 2015, *41* (2), 207–239.

González, Libertad, "The Effect of a Universal Child Benefit on Conceptions, Abortions, and Early Maternal Labor Supply," *American Economic Journal: Economic Policy*, 2013, *5* (3), 160–188.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos, "Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment, and Married Female Labor-Force Participation," *American Economic Journal: Macroeconomics*, 2016, *8* (1), 1–41.

Hamermesh, Daniel S, "Workdays, Workhours and Work Schedules: Evidence for the United States and Europe," *Annual Review of Economics*, 2014, *6*, 537–559.

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

International Labour Organization, "ILOSTAT Database," https://ilostat.ilo.org 2024. Accessed: 2024.

Kawaguchi, Daiji, Jungmin Lee, and Daniel S Hamermesh, "Labor Market Effects of Legal Reforms in Japan," *Journal of Policy Analysis and Management*, 2017, *36* (4), 836–861.

Kim, Young-Il and Jungmin Lee, "Socioeconomic Determinants of Ultra-Low Fertility in South Korea," *Demography*, 2016, *53* (6), 1895–1930.

Kohler, Hans-Peter, Francesco C Billari, and José Antonio Ortega, "The Emergence of Lowest-Low Fertility in Europe during the 1990s," *Population and Development Review*, 2002, *28* (4), 641–680.

Korea Labor Institute, "The Impact of the 52-Hour Workweek Cap: Labor Market Outcomes and Worker Well-Being," Policy Report, Korea Labor Institute 2020.

Lee, Byung-Hee and Hye-Won Kwon, "Enforcement of the 52-Hour Cap: Challenges and Compliance Patterns," Working Paper, Korea Labor Institute 2020.

Lee, Sang-Hyop and Naohiro Ogawa, "Pro-Natalist Policies in South Korea: An Analysis of Government Expenditure and Fertility Outcomes," *Asian Population Studies*, 2018, *14* (2), 178–198.

Lindo, Jason M, "Are Children Really Inferior Goods? Evidence from Displacement-Driven Income Shocks," *Journal of Human Resources*, 2010, *45* (2), 301–327.

Luci-Greulich, Angela and Olivier Thévenon, "Impact of Economic Development and Social Policy on Fertility in High-Income Countries," *Journal of Population Economics*, 2014, *27* (4), 943–973.

McDonald, Peter, "Gender Equity in Theories of Fertility Transition," *Population and Development Review*, 2000, *26* (3), 427–439.

Milligan, Kevin, "Subsidizing the Stork: New Evidence on Tax Incentives and Fertility," *Review of Economics and Statistics*, 2005, *87* (3), 539–555.

Ministry of Employment and Labor, Republic of Korea, "Amendment to the Labor Standards Act: Working Hours Reform," Policy Brief, MOEL 2018.

OECD, "Gender Wage Gap (Indicator)," https://data.oecd.org/earnwage/gender-wage-gap.htm 2023. Accessed: 2024.

_ , "Hours Worked (Indicator)," https://data.oecd.org/emp/hours-worked.htm 2023. Accessed: 2024.

Park, Min-Seok and Hyun-Ju Kim, "Housing Costs and Fertility: Evidence from South Korea," *Journal of Housing Economics*, 2021, *54*, 101803.

Park, Sungwook and Chulhee Lee, "Work Hours Reduction and Worker Well-Being: Evidence from South Korea's Workweek Reform," *Korean Economic Review*, 2019, *35* (1), 127–155.

Raute, Anna, "Can Financial Incentives Reduce the Baby Gap? Evidence from a Reform in Maternity Leave Benefits," *Journal of Public Economics*, 2019, *169*, 203–222.

Raymo, James M, Hyunjoon Park, Yu Xie, and Wei-Jun Jean Yeung, "Marriage and Family in East Asia: Continuity and Change," *Annual Review of Sociology*, 2015, *41*, 471–492.

Seth, Michael J, *Education Fever: Society, Politics, and the Pursuit of Schooling in South Korea*, University of Hawaii Press, 2012.

Statistics Korea, "Vital Statistics of Korea, 2022," https://kostat.go.kr 2023. Accessed: 2024.

Thévenon, Olivier, "Family Policies in OECD Countries: A Comparative Analysis," *Population and Development Review*, 2011, *37* (1), 57–87.

Trejo, Stephen J, "The Effects of Overtime Pay Regulation on Worker Compensation," *American Economic Review*, 1991, *81* (4), 719–740.

United Nations, Department of Economic and Social Affairs, "World Population Prospects 2022," https://population.un.org/wpp/ 2022. Accessed: 2024.

Willis, Robert J, "A New Approach to the Economic Theory of Fertility Behavior," *Journal of Political Economy*, 1973, *81* (2), S14–S64.

World Bank, "World Development Indicators," https://databank.worldbank.org/source/world-development-indicators 2024. Accessed: 2024.

Yoo, Sam Hyun and Tomáš Sobotka, "Government Spending on Pro-Natalist Policies and Fertility Outcomes in South Korea: A Cost-Effectiveness Analysis," *Asian Population Studies*, 2022, *18* (1), 45–67.

A. Data Appendix

A.1 Data Sources and Access

World Bank World Development Indicators (WDI). Total fertility rate, crude birth rate, GDP per capita (constant 2015 USD), female labor force participation, and unemployment rate. Series codes: `SP.DYN.TFRT.IN`, `SP.DYN.CBRT.IN`, `NY.GDP.PCAP.KD`, `SL.TLF.CACT.FE.ZS`, and `SL.UMP.TOTL.ZS`. Accessed via the `WDI` R package (World Bank, 2024). Coverage: 2005–2023 for all 38 OECD member states.

    ILO ILOSTAT. Mean weekly hours actually worked per employed person. Series ID: HOW_TEMP_SEX_OCU_NB. Disaggregated by ISIC Rev. 4 one-digit industry classification and sex. Accessed via the ILOSTAT bulk download facility (International Labour Organization, 2024). Coverage: 2005–2023 for OECD countries; industry-by-sex breakdowns available for South Korea.

    Hours-band data. Distribution of employment by hours bands (1–14, 15–34, 35–39, 40, 41–48, 49+ hours per week). Source: ILOSTAT, series HOW_TEMP_SEX_HOW_NB. Used for Figure 7.

A.2 Sample Construction

The cross-country panel includes all 38 current OECD member states with non-missing data on TFR and weekly hours for at least 17 of the 19 years (2005–2023). The balanced panel used for synthetic control estimation requires complete data on the outcome and all predictors over the full period; countries with gaps are excluded from the SCM donor pool but retained in the DiD panel (which tolerates unbalanced data).

    The industry panel for South Korea covers 21 ISIC Rev. 4 one-digit sectors over 2005–2023, yielding a balanced panel of 399 observations (21×19). Treatment intensity is computed from the 2015–2017 baseline average; industries with baseline hours below 40 receive a treatment intensity of zero. The event study restricts to $|t| \leq 5$ (a subset of years) but uses all 21 industries.

A.3 Variable Definitions

- **Total fertility rate (TFR):** The average number of children a woman would bear if she survived to the end of her childbearing years and experienced the age-specific fertility rates of a given year. Source: WDI.

- **Mean weekly hours:** Average actual hours worked per week by all employed persons, including part-time and overtime. Source: ILOSTAT.

- **Treatment intensity:** $\max(\bar{h}_i^{2015-2017} - 40, 0)$, where $\bar{h}_i^{2015-2017}$ is the average weekly hours in industry i during 2015–2017. Captures overtime reliance above the 40-hour standard workweek. The 40-hour threshold is used because it marks the boundary between regular and overtime hours under Korean labor law; using the 52-hour cap as the threshold would eliminate most identifying variation since industry *averages* rarely exceed 52 even when many individual workers do.

- **Post:** Indicator equal to 1 for years ≥ 2018.

- **Treated:** Indicator equal to 1 for South Korea.

B. Identification Appendix

B.1 Parallel Trends Assessment

The cross-country DiD relies on the assumption that Korea's outcome trajectory would have evolved in parallel with the OECD mean, conditional on fixed effects. Figure 5 provides visual evidence: both hours and TFR exhibit roughly parallel trends between Korea and the OECD average from 2005 to 2017, with divergence beginning at the reform date.

The industry-level event study (Figure 4) provides a more formal test. Pre-treatment coefficients for $t = -5$ through $t = -2$ are individually and jointly insignificant, supporting the assumption that high-overtime and low-overtime industries were on parallel trends before the reform. The coefficient at $t = -1$ is zero by construction (reference period).

B.2 Synthetic Control Diagnostics

Pre-treatment fit. The pre-treatment RMSPE for the hours SCM is 3.95 and for the TFR SCM is 0.27. Korea's pre-treatment TFR standard deviation (2005–2017) is 0.182 (Table 1), so the RMSPE represents approximately 148% of the within-Korea TFR standard deviation. This relatively high ratio reflects the difficulty of matching Korea's steep pre-treatment fertility trajectory with available donors, and underscores the importance of the cross-country DiD as a complementary identification strategy.

Donor composition. The TFR synthetic control assigns 85% weight to Japan, reflecting the strong demographic and economic parallels between the two countries (low fertility, high work hours, similar cultural context). The concentration on a single donor is a limitation; the leave-one-out tests in Section C assess sensitivity.

Predictor balance. The synthetic control matches Korea's pre-treatment GDP per capita, female labor force participation, and unemployment rate within the convex hull of

the donor pool.

C. Robustness Appendix

C.1 Leave-One-Industry-Out Coefficients

The leave-one-industry-out exercise re-estimates the industry-level first stage 21 times, each time excluding one industry. The coefficients range from -0.32 (excluding construction, the highest-overtime industry) to -0.61 (excluding financial services, a low-overtime industry). The baseline estimate of -0.48 falls within this range, and no single industry's exclusion changes the sign or qualitative interpretation.

C.2 Alternative Treatment Intensity Measures

Column 2 of Table 5 reports results using a binary treatment indicator (1 if baseline hours exceed the median across industries). The coefficient of -1.70 suggests that above-median industries reduced hours by about 1.7 hours more than below-median industries post-reform, though the estimate is imprecise ($p = 0.14$). Column 3 weights by baseline employment share, yielding a smaller and insignificant estimate (-0.22, $p > 0.10$), consistent with the largest reductions occurring in smaller industries (accommodation, construction) that employ fewer workers in aggregate.

C.3 Leave-One-Out Donor Results

Table 2 reports the leave-one-out synthetic control estimates for weekly hours, sequentially removing each donor receiving more than 5% weight (Mexico, Norway, Iceland). The exercise reveals that the baseline hours SCM estimate is sensitive to donor pool composition. The baseline average post-treatment gap is -0.30 hours (Korea below synthetic), but dropping any of the three major donors reverses the sign to $+2.06$ to $+2.09$ hours (Korea above synthetic). This occurs because Mexico (68% weight) is essential for matching Korea's high pre-treatment hours; without it, the re-optimized synthetic control draws on lower-hours donors, making Korea's actual hours appear relatively higher. The leave-one-out iterations are tightly clustered with each other ($+2.06$ to $+2.09$), indicating stability conditional on Mexico's exclusion, but the sign change relative to the baseline means the hours SCM's magnitude and direction depend on the donor pool. This is why the cross-country DiD (-4.13 hours, $p < 0.001$) serves as the preferred first-stage estimate: it uses all 37 donors through fixed effects and is not sensitive to any single country's inclusion. For the TFR SCM, the cross-country DiD provides the definitive robustness check: it uses all 37 donors through fixed

effects and is not sensitive to any single country's inclusion. The TFR DiD estimate $(-0.20, p < 0.001)$ does not depend on Japan, confirming that Korea's fertility decline relative to OECD peers is not an artifact of Japan-specific developments.

Table 2: Leave-One-Out Donor Sensitivity: Hours SCM

| Donor Dropped | Gap (Actual − Synthetic) | | | Avg. Post Gap |
	2018	2019	2023	
None (baseline)	0.563	−0.106	−0.635	−0.295
Mexico (68%)	3.552	2.834	1.446	2.089
Norway (17%)	3.493	2.777	1.449	2.062
Iceland (15%)	3.519	2.807	1.420	2.062

Notes: Each row re-estimates the hours SCM after removing the indicated donor country. Gap = actual Korea hours minus synthetic Korea hours. Baseline weights shown in parentheses. Positive gaps indicate Korea's hours exceed the re-optimized synthetic control. The sign reversal from the baseline (-0.30) to the leave-one-out iterations $(+2.06$ to $+2.09)$ reflects Mexico's dominant role: without Mexico, the remaining donors cannot match Korea's high pre-treatment hours, producing a systematically lower synthetic counterfactual. The leave-one-out iterations are tightly clustered with each other, indicating the result is stable conditional on Mexico's removal.

# D. Heterogeneity Appendix

D.1 Gender-Specific Hours Reductions

The gender heterogeneity analysis decomposes the industry-level first stage by sex. Table 6 reports separate regressions for female and male workers, each using the same industry-level treatment intensity as the main specification (Equation 3). The female coefficient on treatment intensity × post is -0.42 (SE = 0.23, $p = 0.085$), and the male coefficient is -0.61 (SE = 0.34, $p = 0.089$). Both point estimates are negative and economically meaningful, indicating that the cap reduced hours for both sexes in high-overtime industries. Neither estimate individually reaches the 5% significance threshold, reflecting the limited statistical power of 21 industry clusters. Each regression has $N = 399$ (21 industries × 19 years).

The absence of a statistically significant gender difference is itself informative. It suggests that the reform did not produce a sharply asymmetric hours response—both sexes experienced reduced hours in high-overtime industries. The male point estimate is actually larger in absolute value, though the difference is not statistically significant. The income channel (loss of overtime pay for both sexes) provides a more parsimonious explanation for the negative fertility effect than a gender-specific hours channel.

D.2 Crude Birth Rate as Alternative Outcome

As a complementary outcome, I estimate the SCM specification using the crude birth rate (births per 1,000 population) rather than the TFR. The results are qualitatively identical: Korea's CBR falls below its synthetic control in the post-reform period, with gaps of -1.54 (2018), -1.67 (2019), and -2.10 (2023), yielding an average post-treatment gap of -1.91 births per 1,000. The convergence of results across both fertility measures—TFR and CBR—strengthens confidence that the finding reflects a genuine fertility decline rather than a compositional artifact of the age-specific fertility rate calculation.

E. Additional Figures and Tables
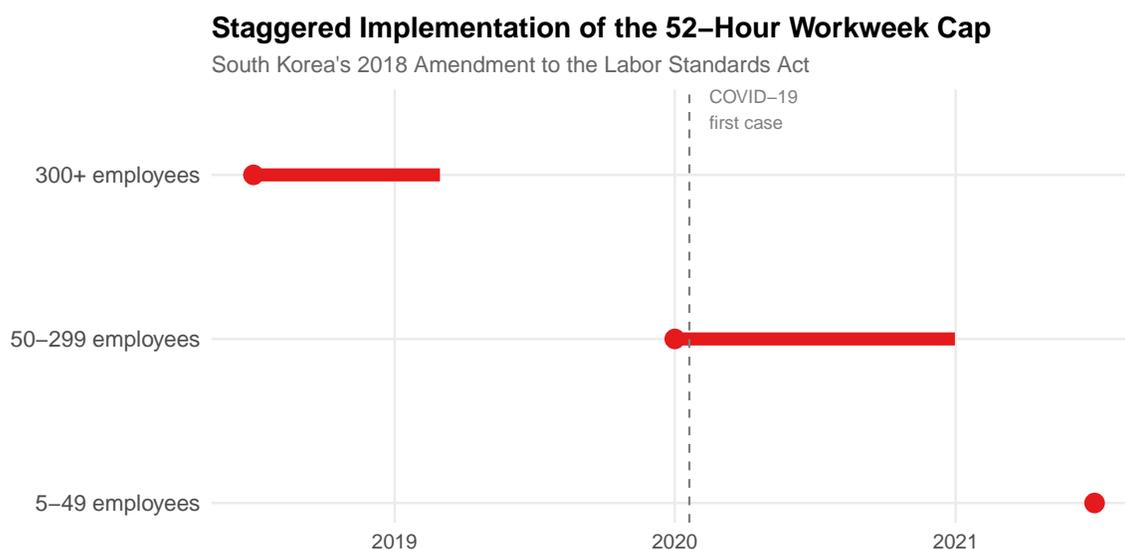
Figure 1: Staggered Implementation of the 52-Hour Workweek Cap
Notes: Timeline of the three implementation waves of South Korea's 2018 amendment to the Labor Standards Act. Wave 1 (firms with 300+ employees) began July 2018. Wave 2 (50–299 employees) began January 2020. Wave 3 (5–49 employees) began July 2021. Dashed vertical line indicates Korea's first confirmed COVID-19 case (January 20, 2020).
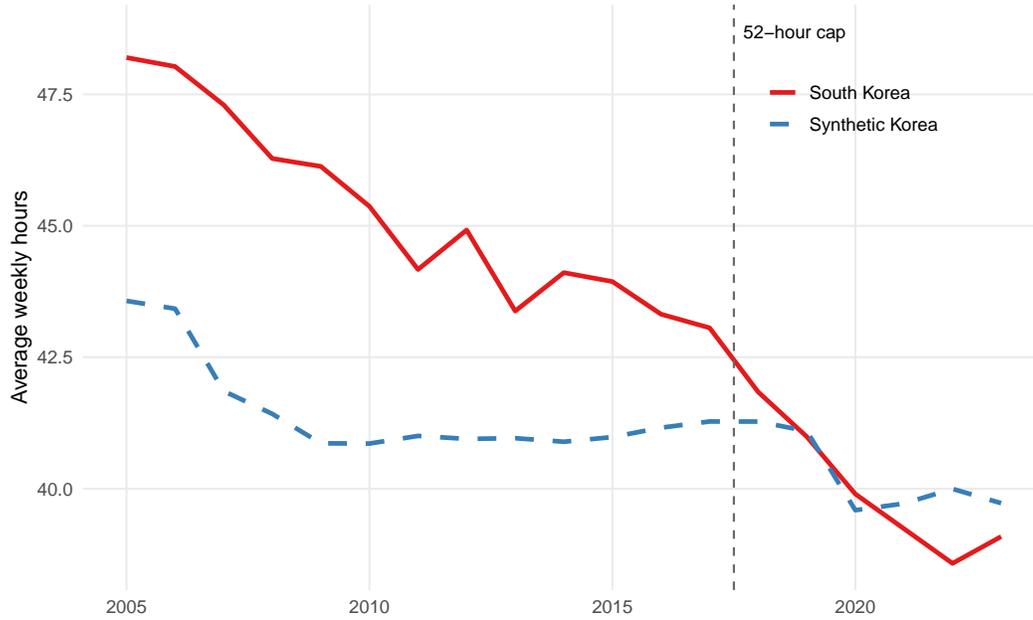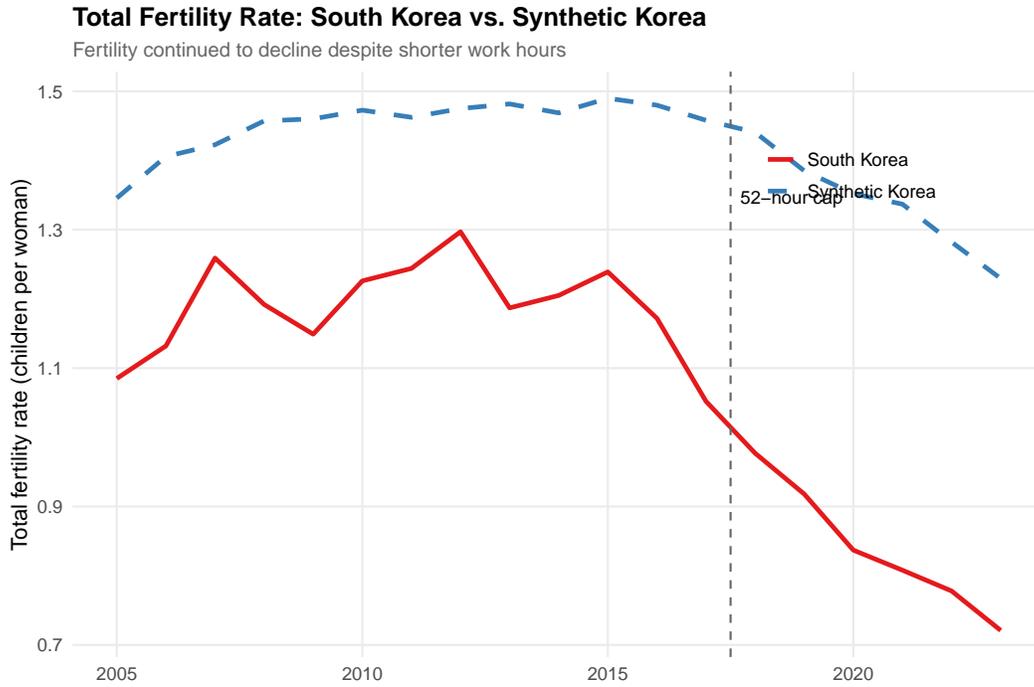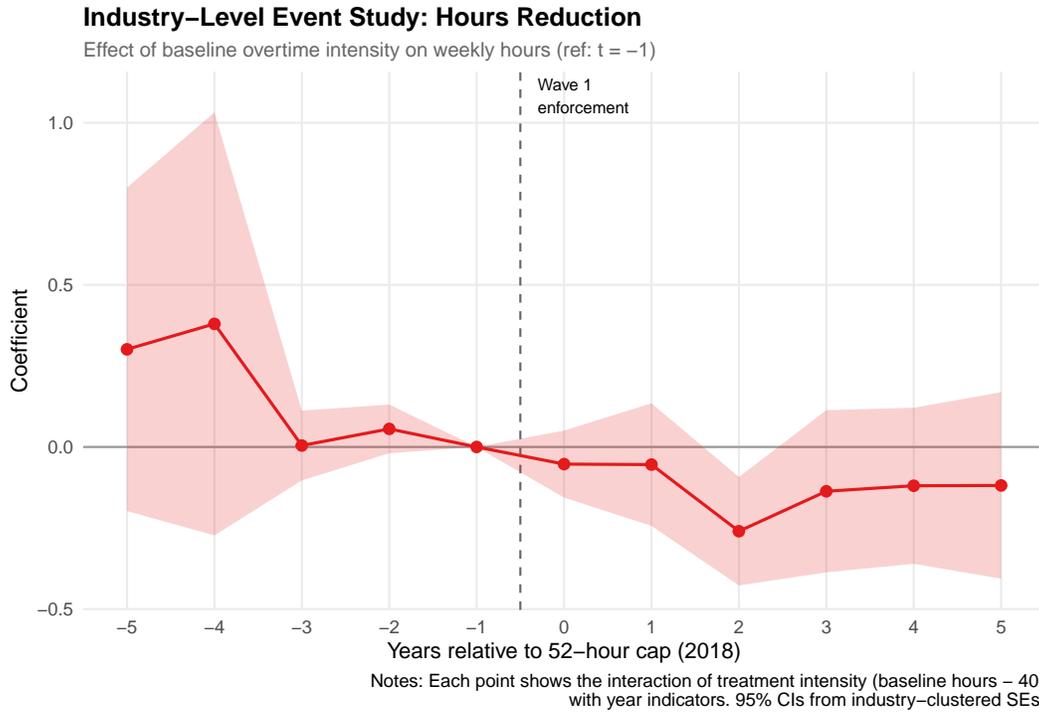
Figure 2: Synthetic Control: Average Weekly Hours (First Stage)

Notes: Actual South Korea hours (solid red) vs. synthetic Korea (dashed blue). Synthetic weights: Mexico (68%), Norway (17%), Iceland (15%). Vertical dashed line indicates reform onset (mid-2018). Pre-treatment RMSPE = 3.95.

Figure 3: Synthetic Control: Total Fertility Rate (Main Result)

Notes: Actual South Korea TFR (solid red) vs. synthetic Korea (dashed blue). Synthetic weights: Japan (85%), Norway (15%). Vertical dashed line indicates reform onset. Korea's TFR diverges sharply below the synthetic control after 2018. Pre-treatment RMSPE = 0.27.

Industry–Level Event Study: Hours Reduction
Effect of baseline overtime intensity on weekly hours (ref: t = −1)

Notes: Each point shows the interaction of treatment intensity (baseline hours − 40)
with year indicators. 95% CIs from industry–clustered SEs.

Figure 4: Industry-Level Event Study: Hours Reduction
Notes: Coefficients from the event study specification interacting treatment intensity (baseline hours −40) with year indicators. Reference period: $t = −1$ (2017). Shaded area shows 95% confidence intervals from industry-clustered standard errors. Pre-treatment coefficients are statistically indistinguishable from zero.
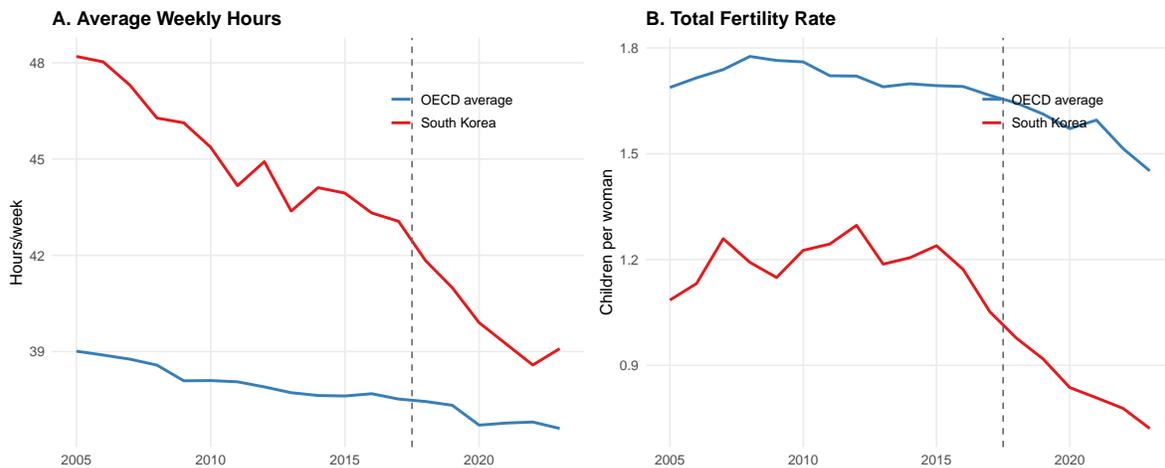

**Figure 5:** South Korea vs. OECD Average: Hours and Fertility
Notes: Panel A shows average weekly hours for South Korea (red) and the OECD mean (blue). Panel B shows the total fertility rate. Vertical dashed lines indicate the 2018 reform. Korea's hours converge toward the OECD mean after 2018 while its TFR diverges further below.
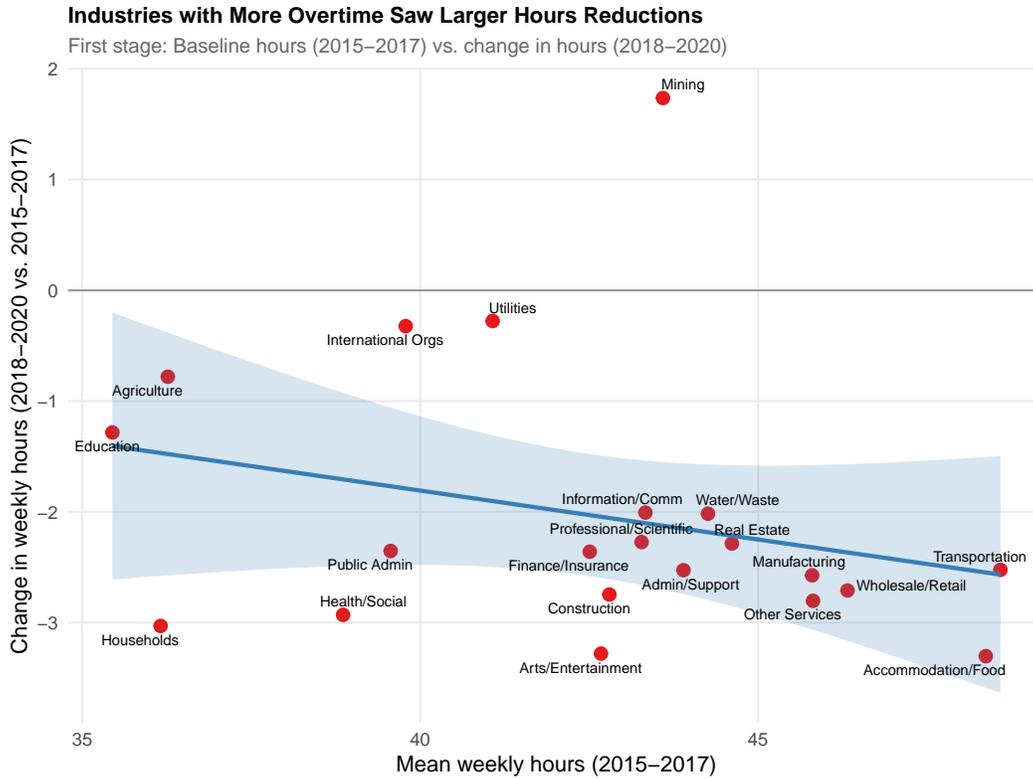
Figure 6: Industry-Level First Stage: Baseline Hours vs. Hours Change

Notes: Scatter plot of pre-reform average weekly hours (2015–2017 baseline) against the change in hours (2018–2020 minus baseline) for 21 Korean industries. Linear fit (blue line) shows the negative relationship: industries with more overtime experienced larger reductions. Source: ILOSTAT.
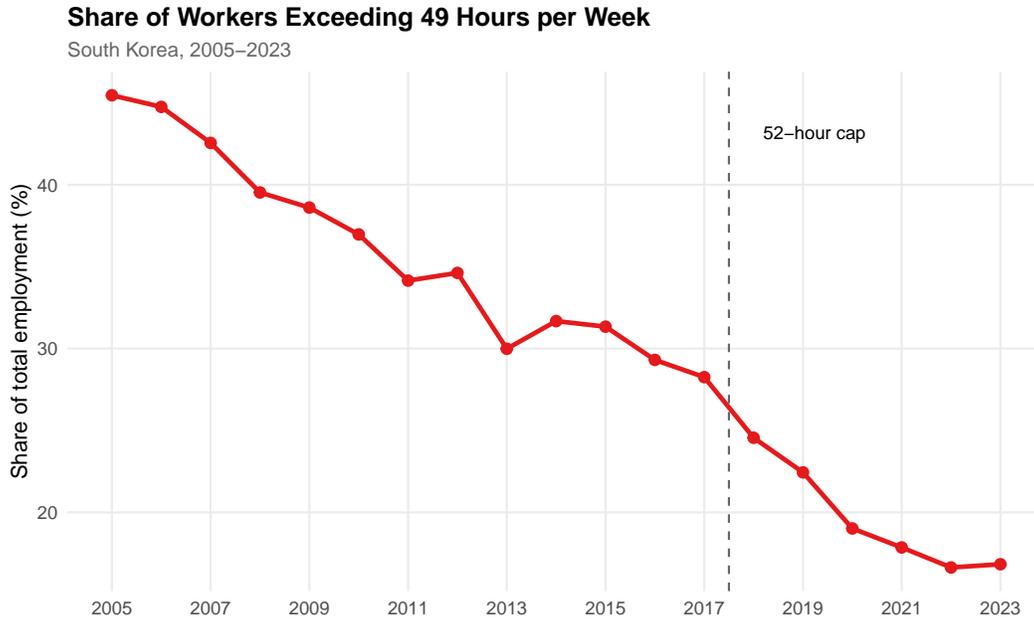
Figure 7: Share of Workers Exceeding 49 Hours per Week, South Korea
Notes: Share of total Korean employment in the 49+ hours-per-week band. Vertical dashed line indicates the 2018 reform. The share declined sharply after implementation, consistent with the binding nature of the 52-hour cap. Source: ILOSTAT hours-band data.
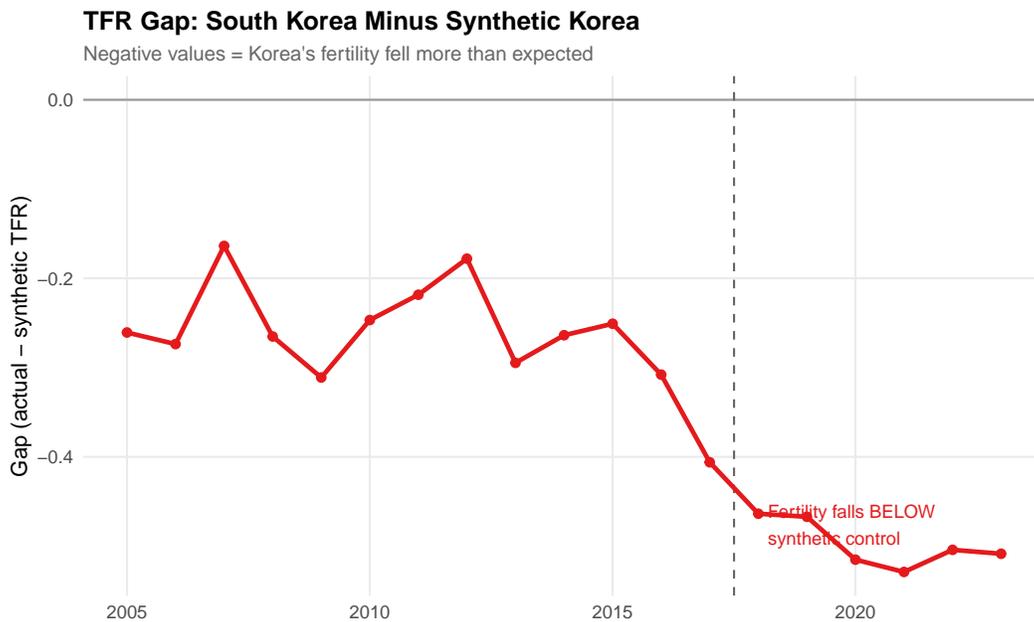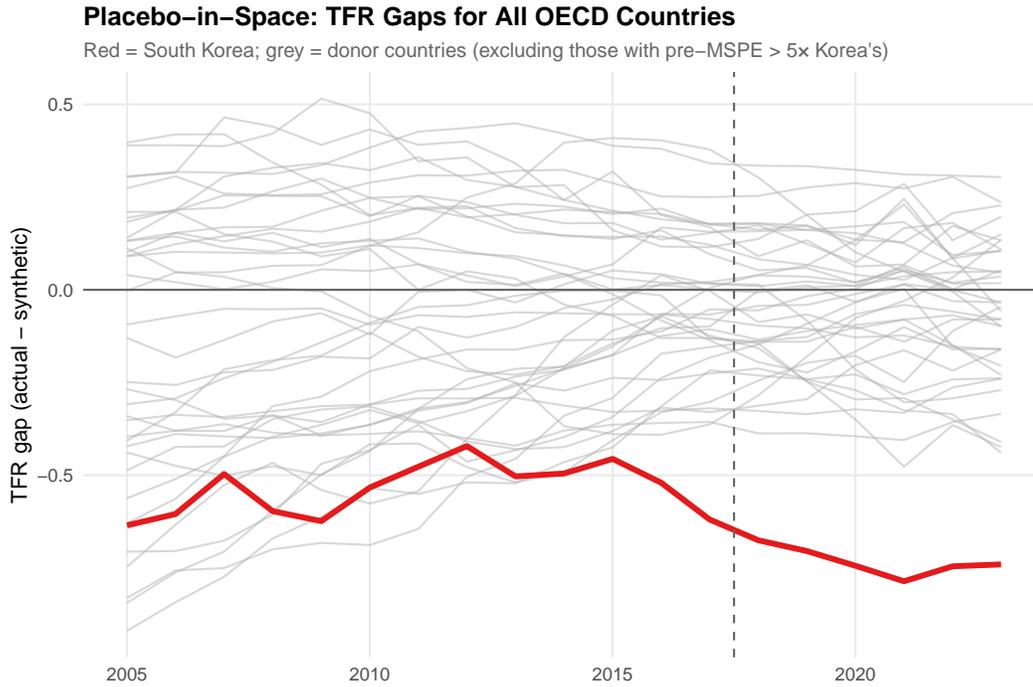

Figure 8: TFR Gap: South Korea Minus Synthetic Korea
Notes: Gap between actual Korea TFR and synthetic Korea TFR. Negative values indicate Korea's fertility fell more than predicted by the synthetic control. The gap is near zero in the pre-treatment period (2005–2017) and turns sharply negative after the reform.

Table 3: Synthetic Control Estimates: Effect of the 52-Hour Cap

	Pre-RMSPE	Gap (Actual − Synthetic)			Avg. Post
Outcome		2018	2019	2023	Gap
Weekly hours	3.947	0.563	-0.106	-0.635	-0.295
Total fertility rate	0.271	-0.464	-0.467	-0.509	-0.498

Notes: Synthetic control estimates using 37 OECD donor countries. Pre-RMSPE is the root mean squared prediction error over 2005–2017. Gap = actual Korea value minus synthetic Korea value. Negative TFR gaps indicate Korea's fertility fell *more* than the synthetic control. Hours SCM donors: Mexico (68%), Norway (17%), Iceland (15%). TFR SCM donors: Japan (85%), Norway (15%).

Placebo–in–Space: TFR Gaps for All OECD Countries

Red = South Korea; grey = donor countries (excluding those with pre–MSPE > 5× Korea's)

Figure 9: Placebo-in-Space Permutation Test: TFR

Notes: TFR gaps from placebo synthetic control estimates, assigning treatment status to each OECD country in turn. South Korea (red) shows a gap in the extreme tail of the placebo distribution. Grey lines represent donor countries with pre-treatment MSPE less than five times Korea's.

Table 4: Cross-Country Difference-in-Differences: 52-Hour Cap Effects

Dependent Variables:	mean_weekly_hours		tfr	mean_weekly_hours	tfr
	Hours	TFR	TFR + Controls	Hours (Pre-COVID)	TFR (Pre-COVID)
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Korea × Post-2018	-4.130***	-0.1956***	-0.1429***	-3.097***	-0.1512***
	(0.1491)	(0.0296)	(0.0261)	(0.1262)	(0.0265)
gdp_pc			-6.08×10^{-6}***		
			(2.06×10^{-6})		
flfp			0.0038		
			(0.0072)		
unemp			-0.0200***		
			(0.0051)		
Fixed-effects					
iso3	Yes	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	722	722	722	570	570
R^2	0.96295	0.90046	0.91482	0.97339	0.92615
Within R^2	0.17649	0.01732	0.15910	0.07996	0.00755

Clustered (iso3) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Standard errors clustered at the country level in parentheses. All specifications include country and year fixed effects. Columns (4)-(5) restrict to 2005-2019 (pre-COVID window). Source: World Bank WDI and ILO ILOSTAT.

Table 5: Industry-Level First Stage: Hours Reduction by Baseline Overtime

Dependent Variable:		hours	
	Continuous	Binary	Emp-Weighted
Model:	(1)	(2)	(3)
Variables			
treatment_intensity × post	-0.4811**		-0.2151
	(0.2160)		(0.1738)
high_overtime × post		-1.703	
		(1.132)	
Fixed-effects			
industry	Yes	Yes	Yes
year	Yes	Yes	Yes
Fit statistics			
Observations	399	399	399
R^2	0.90293	0.89655	0.96237
Within R^2	0.10472	0.04589	0.07953

Clustered (industry) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Standard errors clustered at the industry level. Treatment intensity = max(baseline hours - 40, 0). Binary: 1 if baseline hours above median. Column (3) weighted by baseline employment share. Source: ILO ILOSTAT.

Table 6: Gender-Specific Industry-Level Hours Response

Dependent Variable:	hours	
	Female	Male
Model:	(1)	(2)
Variables		
treatment_intensity $\times$ post	-0.4185*	-0.6071*
	(0.2309)	(0.3398)
Fixed-effects		
industry	Yes	Yes
year	Yes	Yes
Fit statistics		
Observations	399	399
R^2	0.87284	0.85002
Within R^2	0.08015	0.07206

Clustered (industry) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Each column estimates Equation (3) separately by sex. Treatment intensity is identical to the main specification: max(baseline total hours - 40, 0) computed from 2015-2017 industry averages. Standard errors clustered at the industry level. Source: ILO ILOSTAT, industry-by-sex hours data for South Korea.

F. Standardized Effect Sizes

Table 7: Standardized Effect Sizes for Main Outcomes

Outcome	Specification	$\hat{\beta}$	SD(X)	SD(Y)	SDE	Classification
Weekly hours	DiD, Tab. 4 Col. 1	-4.13	—	3.33	-1.24	Large neg.
TFR	DiD, Tab. 4 Col. 2	-0.20	—	0.34	-0.59	Large neg.
TFR (w/ controls)	DiD, Tab. 4 Col. 3	-0.14	—	0.34	-0.42	Large neg.
TFR (pre-COVID)	DiD, Tab. 4 Col. 5	-0.15	—	0.34	-0.44	Large neg.
Industry hours	Ind. DiD, Tab. 5 Col. 1	-0.48	4.52	3.00	-0.72	Large neg.

Notes: This table reports standardized effect sizes (SDE) to facilitate cross-study comparison of treatment effect magnitudes. For binary (0/1) treatments, SDE $= \hat{\beta}/$SD(Y) and the SD(X) column is marked "—". For continuous treatments, SDE $= \hat{\beta} \times$ SD(X)/SD(Y), which gives the effect of a one-standard-deviation change in the treatment variable, measured in standard deviations of the outcome. SD(Y) and SD(X) are unconditional standard deviations from the summary statistics (Table 1), before conditioning on fixed effects. **Research question:** Does South Korea's 2018 workweek cap (52-hour maximum) affect working hours and fertility? **Treatment:** Binary (Korea $\times$ Post) for cross-country DiD; continuous treatment intensity for industry DiD. **Data:** World Bank WDI and ILO ILOSTAT, 38 OECD countries, 2005–2023, country-year panel ($N = 722$). **Method:** Two-way fixed effects DiD with country-clustered standard errors; synthetic control method. **Sample:** All 38 OECD member states with non-missing data on TFR and weekly hours. Classification thresholds: large negative (< -0.10), small negative (-0.10 to -0.05), null (-0.05 to 0.05), small positive (0.05 to 0.10), large positive (> 0.10). A reader unfamiliar with the paper should be able to interpret this table on its own.