The 52-Hour Work Week in South Korea: Implementation and Effects

By

KIM, Ki Sung

THESIS

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC POLICY

2023

The 52-Hour Work Week in South Korea: Implementation and Effects

By

KIM, Ki Sung

THESIS

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC POLICY

2023

Professor Rhee, Inbok

The 52-Hour Work Week in South Korea: Implementation and Effects

By

KIM, Ki Sung

THESIS

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC POLICY

Committee in charge:

Professor Rhee, Inbok, Supervisor

Professor Merfeld, Joshua D.

Professor Park, Jinseong

Approval as of August, 2023

Abstract

Despite exhibiting a continuous and downward trend in working hours, South Korea nonetheless maintains one of the highest working hours in the OECD. To combat this, previous administrations phased in the 40-hour work week from 2004 to 2011, with mixed results. More recently, an overtime cap reform began phasing in since 2018, essentially confining the maximum weekly working hours to 52 for workers falling within its scope.

This paper addresses the question of whether or not working hours were actually reduced by this policy. In other words, did the reform have the desired effect? If this is not the case, then it is hard to imagine that the perceived benefits of reduced working hours (work-life balance, physical and mental health, etc.) have been gained in tandem. Preliminary research as regards this policy generally conclude that working hours were indeed reduced, but I find that may not be the case.

Panel data from the Korea Labor Institute was utilized, and various statistical methods were deployed (i.e. fixed effects, difference-in-differences, propensity matching, regression discontinuity). Taken as a whole, the results show that working hours were not significantly reduced by this policy. This implies that similar mistakes were made as in the implementation of the 40-hour work week and/or new mistakes were made in the context of a perpetually changing labor environment.

Limitations and directions for future research are discussed, but more importantly, the results of this paper highlight the importance of proper implementation of a well-intentioned policy, as well as having a proper backdrop for such a policy to be carried out effectively. Furthermore, people who fall within the scope of such a policy must be in a position where they are able to benefit from it. In other words, if there are no viable alternatives to overwork, people will continue to do so.

Introduction

Even the most conservative estimates contend that over half of the working-age population is employed in one way or another. This estimate for the developed world stands even higher at roughly 67% (OECD 2021). With work being a major aspect in so many people's lives, it is no wonder that policies have been implemented throughout history in multifaceted ways to ameliorate the working experience. One such tool at policymakers' disposal is working hours. The importance of working hours is exemplified by the fact that the first ever Convention held by the International Labor Organization (C001) was in regard to setting working hours standards for industry.

Arguments have been made addressing the detrimental physical effects of long working hours (Bambra et al 2017), even in the context of major diseases such as stroke and heart disease (Virtanen et al 2018). Long working hours also lead to negative mental health outcomes and strained family and social relationships (Joyce & Hewitt, 2016). Conversely, comparative studies suggest that countries with shorter average working hours exhibit better work-life balance for their workers (Golden & Pahnke, 2016). Other studies (e.g. Kondo & Sugawara, 2021) suggest that reduced working hours may have a positive effect on job creation and employment. Taken as a whole, it seems that when working hours are confined to within reason, there are positive outcomes for both people and the economy.

Bearing this in mind, South Korea has maintained one of the highest working hours in the OECD. Out of every six South Korean workers, one works in excess of 55 hours a week, which is more than twice the OECD average. This is still an improvement from 20 years ago, in which this figure was closer to one in three workers. Some may argue that this is either a direct and/or indirect cause of many deleterious effects on mental and physical health, among other things. For instance, South Korea has the third highest fatal workplace injury rate in the OECD (Hijzen & Thewissen 2020).

Several legislative measures have been implemented in Korea in an attempt to combat the deep-seeded and perpetual issue of long working hours in the country, such as the nominal "40 Hour Week" phased in from 2004 to 2011, which set the normal work week to 40 hours. In this paper, I look into the most recent working time reform, the 52-Hour Overtime Cap Reform, which as the name suggests, effectively sets the maximum actual hours worked to 52 hours a week. Given the purported objective of such a reform, I seek to answer the question of whether or not working hours actually decreased for those affected.

To measure the potential reduction in working hours, various permutations of regression analysis were used, culminating in the well-known difference-in-differences (DID) method. This study allows for the use of this method because much like the previous 2004-2011 reform, the reform in question was also implemented in a phase-in structure. This allows for setting a treatment group relative to a control group within the same time frame to check for significant results. Additionally, I make brief use of the regression discontinuity (RDD) method, which has exploded in popularity in the economic literature in the past 20 years. Using multiple empirical methods allows for an additional check of the data results and identification of potential discrepancies.

In summary, it is unclear whether the cap on weekly overtime led to a significant reduction in hours worked. The results of DID go as far as to suggest that working hours may have actually increased for the treatment group, although significance goes away once propensity score matching is added. Conversely, results from the RDD analysis contend that there was indeed a reduction in working hours for the treatment group. However, the author cannot ignore the fact that the "lack" of observations in the vicinity of the cutoff point may potentially skew analysis. Due to the higher robustness of the former part of the analysis, I conclude that the overtime cap did not necessarily exhibit the desired effect for reasons that will be discussed later in the paper.

While this study did not lead to a unitary, concrete answer to the research question, it nonetheless contributes to the existing literature by suggesting the importance of utilizing a variety of methods in testing hypotheses. It bears credence to studies on the previous "40 Hour Week" reform, in that haphazard implementation of a well-intentioned policy can lead to mitigation of desired effects, or furthermore to unintended consequences (Choi 2012, Park & Park 2019). It also adds to existing

studies on the reform in question, as many of these have an insufficient time frame in the data, specifically regarding years of actual reform implementation (e.g. Shim & Kim 2021, Park & Ko 2020). Thus, many of the results of previous studies on this matter should be interpreted in a preliminary light.

This paper is outlined as follows. After a brief historical background for context, previous literature will be analyzed. Methods and results will be shown, followed by a discussion section and concluding remarks. The discussion section will address various policy implications, difficulties in implementing such a policy, and ways to improve efficacy of such policies moving forward.

Previous Studies

There exists a healthy library of previous literature to refer to as regards working time (e.g. too much of it, reductions in it. etc.). Landers, Rebitzer & Taylor (1996) looked into high incentives towards overwork in the legal industry, which echo the aforementioned "presenteeism" and its perceived value in Korean society. As managers do not directly perceive employees, the quality of work and time spent at work become fallaciously integrated when it comes down to decisions such as employee promotion. As employees (especially in highly competitive industries such as law) become increasingly aware of this, a socially inefficient equilibrium is formed, whereby each individual is incentivized to work as long as possible for no particular reason.

Choi (2012) also mirrors this notion that higher working hours alone do not lead to higher productivity. He uses a fuzzy regression discontinuity design to show that, upon implementation of the 2003 reform in South Korea (from 2004 to 2011), the perceived loss in competitiveness that employers were so disconcerted about was in no way as severe as they expected because labor productivity also increased during this time. Park & Park (2019) also conducted a study on this reform, concluding that productivity in the manufacturing sector increased by 1.5% during the seven-year period and that actual working hours were indeed reduced, but not as much as expected. They also found no significant effects on employment despite increased labor costs, which can again be seen in relation to increased productivity during this time.

There are also studies on the effects of working time on personal well-being. Hamermesh, Kawaguchi & Lee (2017) looked into working hours reform in Japan and found that life satisfaction of both the worker and spouse increased, suggesting that there were positive externalities stemming from the reform. Furthermore, there did not seem to be a subsequent loss in income. Results from past studies unfortunately have to be taken with a grain of salt because of the rapidly changing nature of society. As the composition of the workforce is fundamentally different now, so too might the results from new working time reforms. For instance, the 2017 reform in Korea mainly affects those who work a lot of overtime in general, which may not necessarily have been the case in the 2003 reform.

A preliminary study on the 52-Hour Cap Reform was done by Hijzen & Thewissen (2020) on behalf of the OECD. They deploy a linear probability model and set the treatment group to firms with 300 or more employees, as this is the first group to be affected by the reform. The share of employees working in excess of 52 hours decreased by 26% for the treatment group and by 11% for the control group. The results here may be taken with a grain of salt as well. First, while an LPM is a useful analytical tool when dummy variables are involved, it may not be immune to endogeneity problems. Also, as the sample is restricted to those working overtime, it introduces potential for further bias. Furthermore, the 11% reduction in the control group may imply that working hours are on a downward trend regardless of the reform (as we may see later in the paper). Lastly, upon second glance at the time frame of this study's data, the grace period for firms with 300 or more employees is still active and may skew the results in some way. That being said, the authors make clear that this is a preliminary study, and as such the results should not be taken as once-and-for-all.

Other studies on the reform in question reach similar conclusions using a variety of statistical tools, (Shim & Kim 2021, Park & Ko 2020), though the latter recognizes unintended consequences of the reform whereby precarious employees in the treatment group saw no reduction in working hours during the reform period. The authors also state the reform also saw an increase in employment in

precarious jobs, and imply that the vestigial expectation of employees to work long hours (tacit or otherwise) mitigates desired effects.

Historical Context

Despite the deeply seeded Korean notion of forgoing personal well-being for the sake of the economy's competitiveness, South Korea furthermore exhibits relatively low labor productivity when compared to their OECD counterparts (Jones 2019). A primary contributor to long hours coupled with low productivity is the notion of "presenteeism" that is still omnipresent in the country, stemming from long history of military rule. Leaving the office early or even on time is considered almost to be rude, especially when supervisors or managers are present (Choi 2012).

For historical context, the Labor Standards Act (hereinafter LSA) was first enacted in 1953, setting the normal workweek to 48 hours and the normal workday to 8 hours. The LSA was revised in 1989 shortly after the democratic transition to set the normal workweek to 44 hours. The current 52-hour cap reform to be discussed in this paper, essentially a cap on overtime, builds on a previous reform that reduced maximum normal working hours from 44 to 40 between 2004 and 2011.

Despite the passage of the 40-hour workweek in 2003, the issue of long working hours is still prevalent in Korea. One of the reasons is that, contrary to the convenient "40-hour week" mantra, the Ministry of Employment and Labor at the time accepted the interpretation that a week is 5 days, thus allowing a maximum of 68 hours per 7 days (40+12+8+8). The case regarding the legitimacy of this interpretation had been drifting for a long time in the Supreme Court, and prior to its ruling, the National Assembly declared an amendment stipulating a maximum of 52 hours over 7 days, as reflected in an amendment to the LSA. This implies that the 40-hour workweek was not implemented in actuality until the 52-hour cap reform on overtime. The Supreme Court took the National Assembly decision as a cue and came to its own ruling (refer to the Supreme Court's decision *2011da112391* on June 21, 2018).

The effects of the 2003 reform were at best questionable. Choi (2012) notes that the hours actually worked were not reduced dramatically, as Article 50 of the LSA specifically allows for an upper bound of 12 hours of overtime if it is previously agreed upon with an employer. Perhaps the worst of these inadvertent consequences of the 2003 reform is the one described above, the increase in the effective maximum working hours via increasing the scope of overtime. Before 2004, workers could supply 64 weekly hours, i.e. 44 normal and 12 overtime during the week, and 8 overtime hours during the weekend. The 2003 reform effectively increased maximum hours to 68, i.e. 40 normal and 12 overtime during the week, and 16 overtime hours during the weekend (Hijzen & Thewissen 2020).

The 52-Hour Overtime Cap Reform

It is safe to say that South Korea has undergone immense pressure, especially in recent years, to keep its working hours in line with international standards. Following the Candlelight Vigils of 2016 and the impeachment of the conservative president, Moon Jae-in of the Democratic Party was sworn in after the subsequent election. As he ran on a platform of "a society that respects labor," his administration has focused on a variety of issues such as minimum wage, working hours, and transition to a more official status of non-regular employees. Another source of potential pressure may have stemmed from the EU-Korea Free Trade Agreement, which has been in effect since 2015 but is currently undergoing review at the request of the EU. To form an FTA with the EU, the EU requires that the signatory party's standards are in line with those of its member states. For instance, as per the EU Working Time Directive, the maximum allowable number of hours worked in any given week is 48. Although 52 is still more than 48, the reform can be seen as an effort by the Moon administration to more closely align South Korea's working hours standards with those of the West.

According to a 2017 amendment to the LSA, a week is now defined as seven consecutive days, including holidays (Article 2(1) subparagraph 7). The reform, then, essentially caps the

maximum allowable hours worked per seven-day period to 52 instead of 68. Violation (non-compliance) of the amendment can lead to up to three years in prison and/or a fine of up to 20 million KRW. According to many employers, this measure was seen as much too drastic. To assuage their concerns, the administration set up the reform as a gradual phase-in, much like the implementation of the 2003 reform. From July 2018, the reform would apply to all firms with 300 or more employees, with the grace period ending in 2019. From January 2020, it would apply to all firms with 50 or more employees, with the grace period ending in 2021. On paper, the final phase-in for all firms with 5 or more employees has been in effect since July 2021, but the grace period has been extended indefinitely due to increasing concerns voiced by employers at small & medium-sized enterprises.

Data and Methods

Data used in this paper are from the Korean Labor & Income Panel Study (KLIPS), an annual set of panel data released by the Korea Labor Institute, the key research hub for labor and employment policy in South Korea. The number of individuals surveyed is 27,107. Years selected are from 2015 to 2020 inclusive. I chose this data range to prevent capturing any potential policy lag effects from the 2004-2011 reform. I also executed a scraping process to increase the probability that any given individual has the same job over time, as changes to working hours could potentially stem from job changes.

To do this, I set up counter variables for industry category and firm size for individuals through time. The count (number) increases whenever the criterion in question remains the same year by year. For those individuals whom both industry category and the level of firm size (e.g. 0-50, 50-100, 100-200 employees, etc.) remain largely the same throughout the time frame of this paper, it would be safe to include them in the analysis as the probability they kept the same job is significantly higher. In doing so, this lowers the total number of observations in the data from 112,846 to 41,830, which is still more than sufficient for the purposes of this study.

As in the Hijzen & Thewissen paper, I set firms with 300 or more employees as the treatment group, and test whether the reform had a statistically significant impact on said group. I implement a variety of research methods, namely standard OLS, OLS with an interaction term, fixed effects with an interaction term (difference-in-differences), propensity score matching, and regression discontinuity design. As expected, the dependent variable is the number of hours worked per week.

H_0: The 2017 amendments to the LSA and subsequent reform do not have a statistically significant impact on the treatment group. Working hours for the treatment group did not decrease after the reform.

H_1: The 2017 amendments to the LSA and subsequent reform have a statistically significant impact on the treatment group. Working hours for the treatment group decreased after the reform.

Arguments for and against using OLS estimators are widely known at this point, and at the very least they are useful for assessing the general situation. That KLIPS is panel data means that fixed effects estimators can and should be utilized. Furthermore, the phased-in nature of this reform makes it an ideal candidate for a difference-in-differences (hereinafter DID) study. DID is useful in that it essentially isolates the effect of the independent variable on the dependent variable by comparing effects averaged out over time, for both treatment and control groups. The three regression models used in this paper are as follows:

$$p\_hours_i = \beta_0 + \beta_1 size_i + \beta_2 reform_i + \gamma \boldsymbol{X}_i + \varepsilon_i \qquad for\ i = 1,2,\dots n \qquad (1)$$

$$p\_hours_i = \beta_0 + \beta_1 size_i + \beta_2 reform_i + \beta_3 interaction_i + \gamma \boldsymbol{X}_i + \varepsilon_i \quad (2)$$
$$for\ i = 1,2, \dots n$$

$$p\_hours_{it} = \beta_0 + \beta_1 size_{it} + \beta_2 reform_{it} + \beta_3 interaction_{it} + \gamma \boldsymbol{X}_{it} + \varepsilon_{it} \quad (3)$$
$$for\ i = 1,2, \dots n$$
$$and\ for\ t = 1, 2, \dots T$$

The p_hours variable is the dependent variable, namely average hours worked per week. (1) is the standard OLS model, with **X** being a matrix of control dummies (educational level, job status, marital status, region of residence, and existence of a union). The size variable is a dummy variable equal to 1 if the firm that individual i works for employs 300 people or more, and 0 otherwise. The reform variable is also a dummy, equal to 1 if the year is 2019 or 2020, and 0 otherwise. (2) adds an interaction term, which is a product of the size and reform variables. In equation (3), β_3 is effectively the average treatment effect for the treated (ATT).

This study also makes use of the regression discontinuity design (hereinafter RDD) model. The first widely-known use of RDD is Thistlehwaite & Campbell (1960), in which education psychologists find a potential relationship between recognition and educational outcomes. While this novel research method went largely unnoticed at the time, in the past 20 years the use of RDD has exploded, perhaps due to its simplicity in assumptions, and more importantly, its elimination of selection bias (Cunningham 2021). Another example of an ingenious use of RDD is Lee (2001), a paper that examines the advantages of incumbency.

Consider a variable (e.g. firm size) assigning units to a treatment. This variable can, both directly *and* via the treatment, affect another variable (e.g. working hours). In this case, because firm size affects both treatment assignment and working hours, it is a confounder. If we then observe hours worked in the neighborhood of a given threshold (in this case, 300), this effectively cuts the relationship between the two variables at that point. Herein lies another benefit of RDD; it also gets rid of omitted variable bias (Cunningham 2021).

For RDD to work, the expected number of hours worked has to be a smooth function of firm size at the threshold in absence of the treatment. This is to ensure that the whatever is affecting the treatment is the only thing that could have caused a jump at the threshold, of which I can then find the local average treatment effect (as firm size "approaches" 300 from either side). In other words, in absence of the reform, there is no reason for an employee at a firm with 300 or more employees to all of a sudden start working less overtime. It follows that in the RDD model, the size variable is no longer a dummy, but rather the continuous original size variable in the panel data.

Results

Summary statistics are shown in Tables 1 and 2. Mean working hours is slightly above 43. The mean firm size of 68 suggests the prevalence of small & medium-sized enterprises in Korea. The firm size variable exhibits a highly skewed right tail.

There are also noticeable differences in the demographic composition between the treatment and control groups. The treatment group has a higher proportion of individuals in regular employment (97%) than those in the control group (73%). There is also a wide differential in unionized workplaces (56% vs. 7%), those regularly working overtime (33% vs. 14%), and those with a college degree (57% vs. 28%). Those in the treatment are also more likely to be married, which can be attributed to increased job stability in larger firms.

<Table 1> Comparison (treatment vs. control, averages)

Variable	Control	Treatment
Job status	.734	.972
Married	.651	.756
Region (urban)	.442	.544
Union	.066	.560
Overtime	.137	.326
University	.277	.570

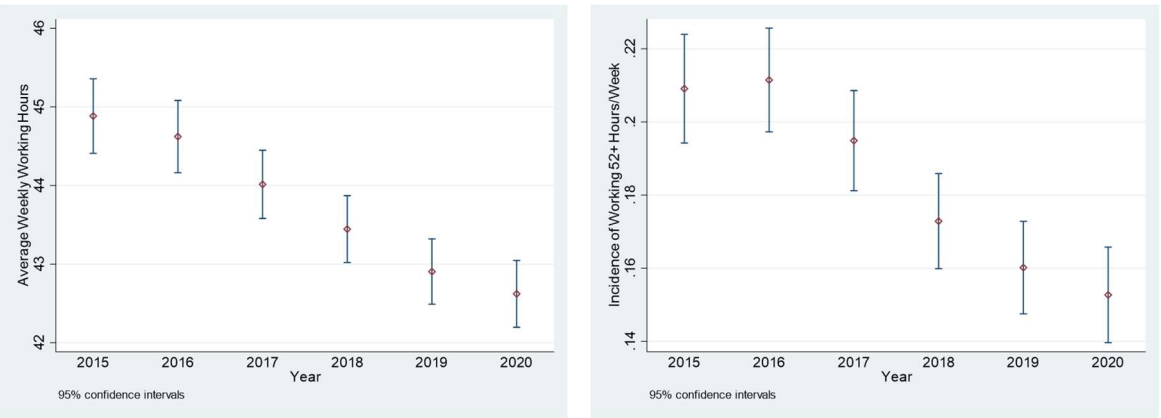
<Table 2> Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Average Weekly Working Hours	18687	43.744	12.523	1	155
Firm Size	4288	67.58	458.188	1	13000

Graphs showing working hours trends in Korea over time can be seen below in <Figure 1>. Within the time range of this study (2015-2020), it can be seen that average working hours has been on a gradual downward trend. As regards incidence of working in excess of 52 hours a week, this has been on a downward trend since 2016. This can be interpreted as a net positive in itself, and implies that progress is being made towards compliance to OECD and other global labor standards. At the same time, it renders inference more difficult in the sense that it is hard to say with certainty on whether or not certain interventions have the desired effect.

Effectiveness of the reform aside, it does seem that South Korea's working hours are on a downward trend in general. This could be due to a variety of factors, such as increased societal cognizance regarding the importance of a work-life balance. In previous generations, South Koreans who overworked were heralded as good members of society. Nowadays, however, there are growing movements such as *gwarosa* ("death by overwork") that draw attention to needless pain and despair spawned by working too much. It should be noted that for these graphs, I did not use the scraped version of the data, but rather the full sample for 2015-2020.

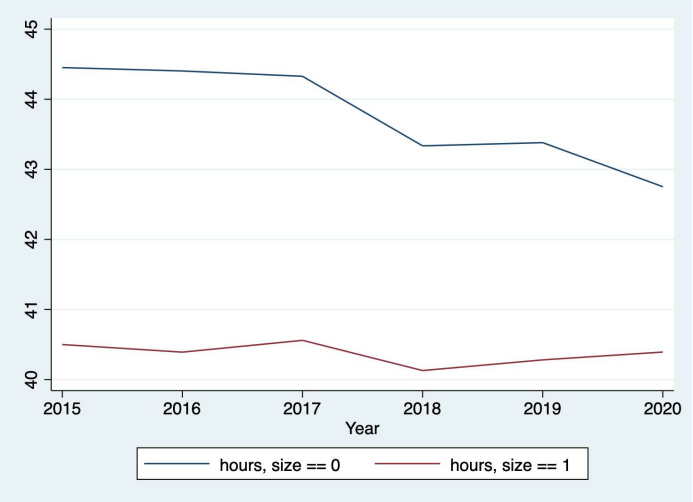
<Figure 1> Working Hours Trends (by year)



Any study that makes use of DID warrants a discussion on parallel trends. The parallel trends assumption is essentially a robustness check on the validity of the study, and contends that in absence of the intervention in question, both the treatment and control groups would have moved (or not moved) in the similar direction. Ideally, we would want extended parallel trends to hold, as this would imply a constant bias between groups. In our case, as evident in the figure below, this stronger version of the assumption likely does not hold, given the trend lines between 2016 and 2018. However, there does seem to be linear time-varying confounding, which means that the weaker parallel trends assumption, i.e. the parallel trends-in-trends assumption, holds (Egami & Yamauchi 2021). For the purposes of this study, the grounds for DID seem to be sufficient.

To further ensure robustness of analysis, I have also combined propensity score matching with the DID. Matching is a technique commonly deployed when the parallel trends assumption appears not to hold (Khandker et al 2010). It essentially matches individuals in the treatment with equivalent untreated individuals (essentially a counterfactual), and measures the difference in outcome between them. This effectively reduces confounding error, leading to a more robust estimate of the ATT. Propensity score matching was first introduced in 1983 by Rosenbaum & Rubin, and despite heavy promotion, some authors (e.g. LaLonde 1986) have taken issue with the method, stating it increases inefficiency and bias. I have chosen to show the DID results both with and without matching as a supplementary robustness check to ensure all bases are covered.

<Figure 2> Parallel Trends Assumption



<Table 3> Regression Results

	(1) OLS	(2) OLS+control	(3) OLS+interact	(4) DID	(5) DID+match
interaction			1.065**	1.185***	1.185
			(.458)	(.281)	(.956)
size	-3.367***	-2.43***	-2.824***	-.491	-.491
	(.195)	(.276)	(.324)	(.355)	(1.213)
reform	-.98***	-.913***	-.955***	-1.086***	-1.086***
	(.318)	(.313)	(.325)	(.212)	(.18)
educ		-2.644***	-2.645***	4.149*	4.149**
		(.31)	(.31)	(2.356)	(1.927)
job_status		3.692***	3.693***	6.055***	6.055***
		(.523)	(.523)	(1.68)	(.72)
married		-.256	-.26	-1.024	-1.024
		(.354)	(.354)	(.954)	(.799)
region		-.159	-.158	-.486	-.486
		(.309)	(.309)	(.65)	(1.092)
union		-1.594***	-1.596***	-.218	-.218
		(.353)	(.353)	(.439)	(.716)
_cons	44.092***	42.235***	42.251***	38.92***	38.92***
	(.196)	(.588)	(.589)	(1.55)	(1.076)
Observations	4287	4287	4287	4287	4287
R-squared	.006	.037	.037	.035	.035

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

<Table 3> displays regression results. In the standard OLS model (with and without basic controls), both the size and reform variables were significant. These results should not be taken too seriously, as zero conditional mean is likely not satisfied here. In two out of three models in which the interaction term is present (OLS and DID), the interaction term is positive and significant. Furthermore, in the DID model, the size variable is no longer significant. When DID is coupled with matching, the interaction term is no longer significant at the 5% level.

The perplexing result that the reform may have gone as far as to contribute to increased working hours in the treatment goes against all intuition and expectations. Perhaps the reader believes that, due to my aforementioned scraping of the data, too many observations were potentially deleted, leading to a strange output. This is incorrect, as the data yield the same results without the scraping (See Appendix).

Perhaps something more fundamental is going on, e.g. some external factor is leading employees in larger firms to work longer hours, for example, due to financial necessity. If something is affecting the treatment group in this way (relatively more so than the control group), then it could imply a violation of parallel trends in the first DID model. It is safe to say that 2020 was a strange year for most. If one works in a larger firm, the higher the likelihood of that person working from home. Perhaps these employees are logging higher hours just for the sake of it.

One will notice that marital status and region of residence (urban vs. rural) does not have a significant impact on hours worked in South Korea. This may have been different in the past, but does not seem to be the case now. Union presence and education level (whether one attended university or

not) had a significant impact in the OLS models, but significance dwindled in the DID model. Job status maintained high significance throughout.

<Table 4> Regression Results for exempt industries

	(1) p_hours
size	-.759
	(1.446)
reform	.091
	(.403)
interaction	-.88
	(1.119)
educ	-2.909***
	(.598)
job_status	2.878**
	(1.228)
married	.878
	(.941)
region	-2.262**
	(1.011)
union	-.029
	(.638)
_cons	42.38***
	(1.328)
Observations	1093
R-squared	.011

Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

<Table 4> above shows the result of running the model on only the five exempt industries. This result is in line with what we would expect, as in no significant effect from the reform. According to Park & Ko (2020), however, workers in the previously exempt 21 industries did not start getting "treated" until July 2019. Thus, as an additional check, I ran the model again with these industries dropped from the analysis. These results can be found in the appendix.

<Table 5> Regression Discontinuity Design Results

Outcome: Average weekly working hours. Running variable: Firm size.

p_hours	(1) Post reform	(2) Pre reform	(3) Post reform, exempt industries
RD_Estimate	-9.09***	-.039	1.266
	(3.451)	(1.574)	(5.255)
Observations	6057	9470	445

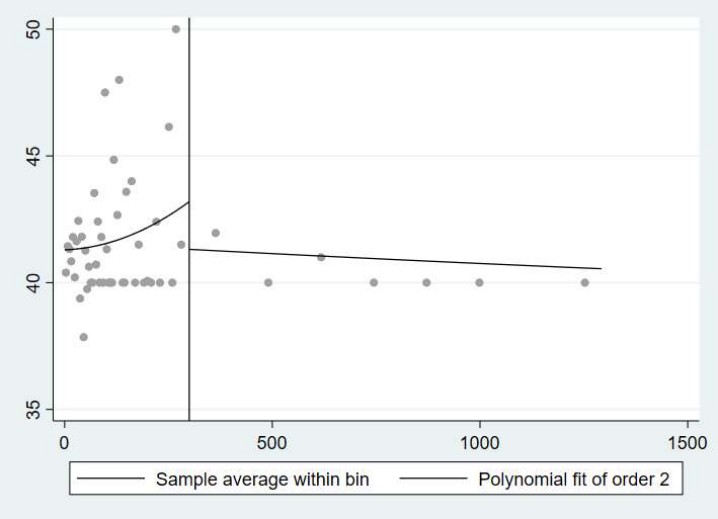
Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

The RDD results suggest the opposite of those of the regression models. <Table 5> shows results with the same dependent variable as before, average weekly working hours. The post-reform coefficient for the treatment group is significant at the 1% level. The effect on the treatment group prior to the reform is not significant, indicating that the reform may be having the desired effect. The reform does not seem to have a statistically significant effect on those in the treatment group who are in exempt industries, as expected.

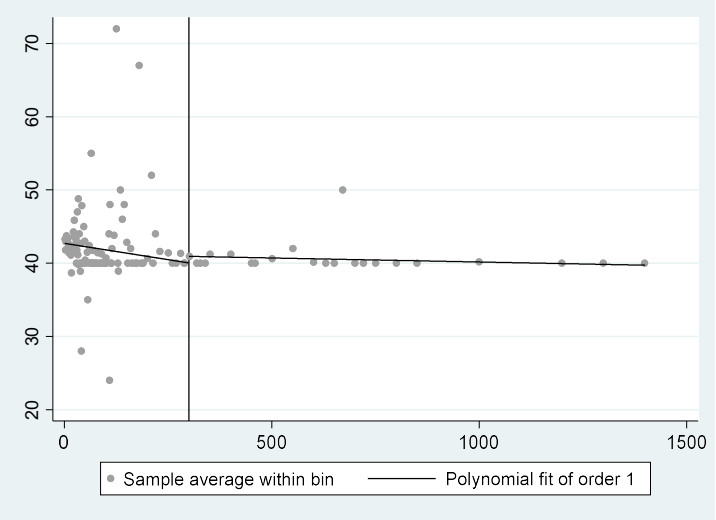
This result, albeit promising, is not conclusive. The RDD model I ran was of polynomial order 2, but there are some who contend that higher-order polynomials should not be used in RDD (Gelman & Imbens 2016). Running the model with polynomial order 1 yields no significance. As can be seen in the graph, the "jump" is hard to notice because there are so many data points around the 40-hour mark. That being said, there is a conspicuous drop in instances of excessive work hours after the cutoff point.

Additional RDD results can be found in the appendix.

<Figure 3> Regression Discontinuity



<Figure 4> Regression Discontinuity, pre-reform



Discussion

There are a number of reasons that potentially explain the rather inconclusive results of this study. The factors to be discussed below may also be interacting with one another in various ways to further cloud the results. One of the main takeaways of this study is that working hours in Korea have been on a downward trend prior to the reform. It can be argued that policymakers are (consciously or otherwise) formulating policies to adjust to societal changes, as opposed to the usual notion of policies inducing behavioral changes in society. Similarly, it is possible that workers and firms began changing their working hours patterns shortly after the announcement of the reform prior to actual implementation. An announcement of a policy in itself has the potential to change the behavior (either directly or indirectly) of workers and employers alike, if it is interpreted as a signal or nudge.

On the more pessimistic (but just as likely) end of the spectrum, there is the distinct possibility of non-compliance by firms. There were already relatively high minimum wage hikes during the Moon administration, which were seen as more than enough of a setback by employers. This compounded pressure may have led some of them to decide that the "benefits" of non-compliance were worth the risk of penalties associated with it. The DID result that showed increased working hours in the treatment group relative to the control may be due to employers at small & medium-sized enterprises having more flexibility as regards employee recruitment and working hours. For example, it may be easier for them to utilize job sharing schemes that can in turn reduce per-employee hours.

Article 59 of the LSA was amended in 2017, reducing the number of industries exempt from working time reform from 26 to 5. The most notable remnants are the healthcare and freight transport industries. The reader will note that this is the classic adverse selection problem at play. Even though this is a massive improvement, industries still exempt from the overtime cap reform are exactly the ones most in need of it. It seems that policymakers failed to learn the lessons from the 2003 reform, whereby too many exemptions and loopholes hindered its potential success.

Due to the reform's implementation, firms have turned to yet another loophole, whereby special overtime work is allowed when permission is granted by the Ministry of Employment and Labor. Indeed, the number of such authorized cases in 2021 was seven times that of 2019 (Kim 2022). It seems that more and more firms have been exploiting this application system to circumvent the reform. As suggested by Landers, Rebitzer & Taylor (1996), there needs to be a fundamental way to disincentivize working long hours, as merely introducing laws just paves the way for future loopholes to be found. In other words, a reform will not necessarily have the desired effect if individuals still need to work longer to be financially stable or even just to scrape by. A reform has to be coupled with viable alternatives for both businesses and workers.

This study also warrants a discussion on wage. In Korea, in particular in the manufacturing industry, the base rate (as a part of the ordinary wage) is set artificially low to raise profits for firms, an anachronistic byproduct from the days of pro-business military juntas. As further explanation, the base pay is what is used to calculate bonuses, severance pay, etc., thereby increasing firm profits the lower this is. This creates a situation in which, despite overtime pay not being calculated into bonuses and severance, the worker is still highly incentivized to work long hours out of financial necessity.

Wage issues are harder to resolve than working hours issues. As wage is determined by tripartite cooperation, namely labor, management, and the government, it is not something the government has direct levers over. This is to say that the government placing a cap on overtime hours is at best a bandage. Complementary measures will be required to tackle long working hours at its core. Without them, workers may be forced to find a second job or side hustle to maintain previous household income levels, which is in direct contradiction to what the reform is trying to accomplish.

Potential limitations to this study are as follows. As the mean firm size in the sample is roughly 67 with a long right tail, the KLIPS data will not have sufficient observations of firms with 300 or more employees. As future KLIPS data are released, further studies can look into the second phase-in (firms with 50 or more employees), which will potentially yield much more robust conclusions. There could also be unobserved heterogeneity present. For instance, workers who prefer

longer hours can move to smaller firms. This is, however, highly unlikely in the case of South Korea due to massive wage gaps between large and small firms (Hijzen & Thewissen 2020).

It is possible that the size variable in the RDD is not sufficiently continuous. In other words, it would have been ideal if there were more observations closer to values such as 299 or 301, as this would have yielded much more robust results in line with the model's continuity requirement. As mentioned, there is also a high possibility of non-complying firms, especially in a country like South Korea. Future studies may want to consider a fuzzy RDD model to account for this discrepancy.

Concluding Remarks

Unfortunately, for the DID model, the null could not be rejected. In fact, the results went so far the other way that some would say it rejects the null but in the other direction, as in the reform made the situation worse. This would echo certain sentiments regarding the 2003 reform, as it inadvertently increased the maximum allowable weekly working time. For the RDD model, the null was rejected at the 5% level, but significance broke down as controls were added. When using heteroskedasticity-robust errors and fitting the model to a lower polynomial, the null can be rejected at the 1% level. Further, the reform seems to have reduced the incidence of individuals working more than 52 hours a week for the treatment group, which is definitely a positive sign.

I conclude with the following statistic. Even when the 2017 reform is fully underway, two in five workers will still be exempt from the legislation. Regardless of different opinions on the Moon administration, it cannot be said with confidence that this reform passes with flying colors. Even if it was indeed reducing working hours, discriminating by workplace size still contradicts OECD protocol (Hijzen & Thewissen 2020). Despite the attempted phase-ins, it does not seem like employers in the final target group are going to budge anytime soon, further hindering the reform's questionable successes up to this point. There is also a chance that the reform will exist on paper only moving forward, as the new Yoon administration has heralded that the working hours cap may be enforced at the monthly level instead of the weekly level. For all practical intents and purposes, this implies that an individual can potentially work up to 88 hours in a single week (Kim 2022).

The silver lining from this study is that, whether the reform is effective or not, the overwhelming trend in South Korea still seems to be towards reduced working hours. In all five regression models, the reform variable was significant, implying this downward trend through time. The underlying task ahead for South Korean policy makers is to find ways to disincentivize long working hours at the core, thereby reducing the necessity for the general public to put in long hours each and every day, something past generations were all too accustomed to in years past.

References

Bambra, C., Lunau, T., Van der Wel, K. A., Eikemo, T. A., Dragano, N., & Krokstad, S. (2017). Work, health, and welfare: The association between working conditions, welfare states, and self-reported general health in Europe. *Scandinavian Journal of Work, Environment & Health*, 43(5), 425-435.

Choi, H. (2012, September). *Effects of a reduction in hours of work on labor productivity and labor costs in South Korea: Evidence from a regression discontinuity design* (Thesis). Lund University.

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

Egami, N., & Yamauchi, S. (2021). Using multiple pre-treatment periods to improve difference-in-differences and staggered adoption designs. *Political Analysis*, 31(2). https://doi.org/10.48550/arXiv.2102.09948

Gelman, A., & Imbens, G. W. (2016). Why High-Order polynomials should not be used in regression discontinuity designs. *Journal of Business and Economic Statistics*, 37(4). https://doi.org/10.1080/07350015.2017.1366909

Golden, L., & Pahnke, K. E. (2016). Working time and work-life balance in the EU and USA: A comparison of European and American time-use data. *Journal of Labor Research*, 37(4), 397-417.

Jones, R. S. (2019, May). *Korea's economic growth prospects under the income-led strategy*. KEI. https://keia.org/wp-content/uploads/2020/05/kei_koreas_economy_2019_jones_190522_final.pdf

Joyce, K., & Hewitt, P. L. (2016). The effects of working hours on health and well-being: A review of the literature. *Journal of Public Health*, 38(4), e417-e426.

Hamermesh, D., Kawaguchi, D., & Lee, J. (2017). Does labor legislation benefit workers? Well-being after an hours reduction. *Journal of the Japanese and International Economies*, 44(C). https://doi.org/10.1016/j.jjie.2017.02.003

Hijzen, A., & Thewissen, S. (2020, August). *The 2018–2021 working time reform in Korea: A preliminary assessment* (No. 248). OECD.

Khandker, S., Koolwal, G., & Samad, H. (2010). *Handbook on impact evaluation: Quantitative methods and practices*. World Bank. https://openknowledge.worldbank.org/handle/10986/2693 License: CC BY 3.0 IGO.

Kim, M. Y. (2022, June 24). Milestone "labor market reform." *Labor Today*, 2–3.

Kondo, J., & Sugawara, S. (2021). The effects of reduced working hours on employment and labor productivity: Evidence from a field experiment. *Journal of Labor Economics*, 39(1), 173-207.

Labor Standards Act 1997 (Law No. 5309). § 50, 59. (South Korea).

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*. 76(4): 604–620. JSTOR 1806062

Landers, R. M., Rebitzer, J., & Taylor, L. J. (1996). Rat race redux: Adverse selection in the determination of work hours in law firms. *American Economic Review*, 86(3). https://EconPapers.repec.org/RePEc:aea:aecrev:v:86:y:1996:i:3:p:329–48

Lee, D. S. (2001). The electoral advantage to incumbency and voters' valuation of politicians' experience: A regression discontinuity analysis of elections to the U.S. House. *NBER Working Papers*, 8441. https://doi.org/10.3386/w8441

Park, S., & Ko, H. (2020). Intended and unintended consequences of the 52-hour workweek policy in South Korea: Focusing on precarious workers. *Asian-Pacific Economic Literature*, 32(2), 103-118.

Park, W., & Park, Y. (2019). When less is more: The impact of the regulation on standard workweek on labor productivity in South Korea. *Journal of Policy Analysis and Management*, 38(3). https://doi.org/10.1002/pam.22136

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the Propensity Score in observational studies for causal effects. *Biometrika*. 70(1): 41–55. doi:10.1093/biomet/70.1.41

Shim, Jae-sun, Kim, Hohyun. (2021). The Effect of the 52-hour Workweek Policy on Long-time Workers. *Korean Management Consulting Review*, 21(3), 251-259.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6). https://doi.org/10.1037/h0044319

Virtanen, M., Jokela, M., Nyberg, S. T., Madsen, I. E. H., Lallukka, T., Ahola, K., . . . Kivimäki, M. (2018). Long working hours and health: A systematic review and meta-analysis. *The Lancet*, 392(10164), 294-313.

Appendix

<Table A1> shows results from running the same five regression models as in the main text, but while maintaining the entire sample from 2015 to 2020. It should be noted that the DID model (without matching) might not be appropriate as far as analysis is concerned, as the non-scraped data include all of the individuals who were changing jobs throughout the time period, and may thus cloud subsequent inference. Nevertheless, the results (and implications) are similar.

In the matching portion of this analysis, i.e. (5), the education (university) control was omitted as the balancing property was not satisfied when it was included.

<Table A1> Regression Results in the absence of data scraping

	(1) OLS	(2) OLS+control	(3) OLS+interact	(4) DID	(5) DID+match
interaction			2.231***	1.562***	1.558*
			(.434)	(.378)	(.842)
size	-1.109***	-1.42***	-2.339***	-1.093***	-1.076
	(.217)	(.232)	(.259)	(.418)	(.841)
reform	-1.553***	-1.499***	-1.555***	-1.087***	-1.065***
	(.177)	(.17)	(.174)	(.149)	(.129)
educ		-2.603***	-2.602***	3.603***	
		(.164)	(.164)	(1.377)	
job_status		6.498***	6.502***	5.609***	5.748***
		(.251)	(.251)	(.728)	(.352)
married		-.462**	-.468**	-.035	.005
		(.182)	(.182)	(.692)	(.544)
region		.453***	.448***	-.541	-.463
		(.171)	(.171)	(.578)	(.598)
union		-.956***	-.988***	-.046	-.017
		(.225)	(.225)	(.35)	(.465)
_cons	42.533***	38.629***	38.656***	37.427***	38.301***
	(.117)	(.28)	(.281)	(.756)	(.512)
Observations	15527	15527	15527	15527	15266
R-squared	.005	.074	.074	.035	.034

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

<Table A2> Additional Regression Results

	(1) p_hours	(2) p_hours
size	-.493	
	(.326)	
reform	-1.17***	-1.37***
	(.193)	(.217)
interaction	1.567***	
	(.52)	
size2		-1.291***
		(.395)
interaction2		1.688***
		(.346)
_cons	42.463***	42.627***
	(.087)	(.104)
Observations	9884	9884
R-squared	.009	.011

Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

<Table A2> displays additional regression results. As explained prior, for the 21 previously exempt industries, workers were not officially subject to the reform's implementation for roughly half of 2019. (1) displays the result with 26 industries (exempt and formerly exempt) dropped. Results are basically identical to those of the main paper. (2) shows the result for when the treatment in the model includes firms with 50 or more employees (the second stage phase out). This result may be unreliable as there is only one year in which the target group is affected, but I have added it for reference purposes. It should be noted that the data scraping process was not applied to this part of the analysis.

<Table A3> Additional RDD Results

p_hours	(1) Post reform	(2) Pre reform	(3) Post reform	(4) Pre reform
RD_Estimate	-9.513**	2.285***	-7.436*	.004
	(4.396)	(.76)	(4.446)	(.92)
Observations	4436	5448	4436	5448

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

These are the RDD results with the 26 industries dropped. As mentioned in the main text, these include both currently and previously exempt industries. (1) and (2) are the post-reform and pre-reform result with polynomial order 2. The result is similar to that of the main paper. (3) and (4) are the same respectively, but with polynomial order 1. As in the main paper, some of the significance is lost.