JEL Classification: I38, J21, J22, J23
Keywords: policy evaluation, working hours, gender inequality, work-family conflict, female labor participation

Was a One Hour Adjustment in Georgian Public Sector Working Hours 'Family Friendly' and Did It Increase Female Labor Participation?

William APPLEMAN - Charles University in Prague - CERGE-EI

Levan BEZHANISHVILI - Charles University, Institute of Economic Studies at Faculty of Social Sciences, Czech Republic

Zurab ABRAMISHVILI - Ivane Javakhishvili Tbilisi State University

Abstract

This paper evaluates a one-time, immediate policy initiative enacted by the Republic of Georgia that shifted public office operating hours from 10:00-19:00 to 9:00-18:00, which affected the work schedules of government employees. Although the policy affected approximately 200,000 employees, it has never been evaluated; and to our knowledge, nor has any similar policy in any economic literature. In the paper, we examine how the policy impacts gender inequality through female labor participation. Given that the policy did not affect the private sector, we employ a difference-in-differences approach using the National Statistics Office of Georgia Households Incomes and Expenditures Survey from 2013-2016. We find that the policy primarily produces a significant reduction in the average level of working hours of full-time employees with children, directly in line with the prediction following the gender similarity model. We also find a significant increase in average work hour engagement by women without children. However, the placebo effect analysis identifies this as an already existing trend and the short-term analysis indicates that this is an ordinal reaction to the reduction of engagement by full-time employees with children.

https://doi.org/10.32065/CJEF.2022.04.04

The authors are grateful to Fabio Michelucci, Nikolas Mittag, Milan Ščasný, Patrick Gaulé, Randall Filer, Michael McBride and two anonymous referees for all their comments, which helped us improve the paper.

This work was supported by Charles University's PRIMUS/17/HUM/16 and from the European Union's H2020-MSCA-RISE project GEMCLIME-2020 under GA 681228. Responsibility for any errors remains with the authors.

The data that support the findings of this study are openly available in National Statistics Office of Georgia at https://www.geostat.ge/en/modules/categories/128/databases-of-2009-2016-integrated-household-survey-and-2017-households-income-and-expenditure-survey.

The data that support the findings of this study are openly available in Business Information Agency (BIA) at https://www.bia.ge/.

1. Introduction

On August 1, 2014, the prime minister of the Republic of Georgia announced a countrywide initiative to shift the working (operating) hours in the public sector[1] from 10:00-19:00 to 9:00-18:00 (Khunashvili, 2014). There was no parliamentary pushback, no protests by citizens or public employees, no journalistic coverage beyond the announcement, and no related Google search keyword trends. Within a month, it passed through parliament and was enacted on September 1, 2014. This example is as close to a theoretical one-time, immediate policy shift as practically possible. Officially, the rationale of the new policy was to adjust the operating hours of public offices to those more common in 'the modern world'. In fact, the policy was one of several that aligned Georgia more with practices of OECD countries. Unofficially, there is anecdotal evidence that some parliament members also thought the new hours could improve public service efficiency and encourage women to participate more in the labor market.[2] The new policy affected approximately 200,000 public sector workers or 13.4% of the total workforce, a nontrivial amount; and yet, the consequences of the policy have never been studied. Importantly, this is the only policy specifically affecting public employees in the years before and after its implementation.

Work hours have a considerable influence on our personal lives and on a myriad of economic areas. A number of studies address work hours and their relationship to productivity (Golden, 2006), efficiency (Hanse, 1993), types of employment (Wasserman, 2015), wage inequality (Carr, 2011), educational outcomes (Baffoe-Bonnie & Golden, 2007), benefits of flexibility (Bird, 2015), work-life balance (Holly & Mohnen, 2012), intra-household bargaining (Rangel, 2003), gender differences in market and home labor (Goldin, 2014), gender wage gap (Blau & Kahn 2017), impact on health (Dawson et al., 2005), impact on happiness (Galay, 2007), and impact on the environment (Knight et. al., 2013). As far as we are aware, however, no study evaluates the effects of a policy that exogenously shifts the working hours of a major cross-section of workers, eliminating the common self-selection bias issues faced in many work hours studies. This advantage combined with the novelty of the policy render this examination worthwhile and informative to several topics within the working hours literature. Moreover, the dearth of directly related literature appears to be due to the uniqueness of the policy, since there are several adjacent areas of research examining how work hours and work schedules affect economic, physical, emotional, psychological, social, familial, and vocational well-being. The effects of this policy logically, and by intention (of at least some

[1] The policy affected most offices of the government, ministries, the national bank, the national statistical office, among other public offices.

[2] Based on anecdotal evidence gathered from discussions with government officials. It is not surprising that Georgian parliament members would have such concerns in mind, since, despite the recent history of many progressive policy initiatives promoting female labor participation in Georgia, traditional gender roles remain culturally dominant for both men and women (Kachkachishvili et al., 2014). Moreover, as a signatory participant of the 1995 World Women's Conference Platform of Action (Jashi, 2005), Georgia should initiate and assess such policies. Regarding the official rationale, public office operating hours in Georgia were in line with others in and around the Caucasus—where private and public organizations opened at 10:00 (predominantly) or 9:00, while OECD offices mostly opened at 8:00 or 9:00.

members of parliament), impact genders and family types asymmetrically, relating this paper most closely with work-family conflict, gender inequality, and intra-household bargaining and resource allocation literature. Work-family conflict literature identifies two types of conflict, family interference with work (FIW) and work interference with family (WIF). Outcomes of conflicts are informed by two models, the gender similarity model and the gender difference model (Gutek et al., 1991).

A prediction in line with FIW and the gender difference model would be that this policy could relieve familial conflict with (potential) work for mothers. In Georgia, the vast majority of family-related household activities are conducted in the evening and by females.[3] As later working hours were considered a source of personal-professional scheduling conflict for women with household responsibilities, it is understandable why some members of parliament believed the new initiative might be more 'family friendly', i.e. convenient for successfully combining economic and family activities, thus removing barriers for women with families seeking employment in the public sector. Moreover, the policy was put into effect with family schedules in mind; those with children aged 12 or younger were given a half hour of flexibility in their work schedules to relieve the resulting burden from the convergence of their new starting time and the legally-mandated-universal school starting time of 9:00 a.m. (Farulava, 2014), which generally extends to preschool as well as formal childcare.[4] In addition, public office employment in OECD countries generally offers stability, reasonable financial security, and some flexibility, which tends to attract women (Wasserman, 2015; Goldin, 2014; Gicheva, 2013). In Georgia, the public sector holds a greater place in the economic hierarchy than in most OECD countries, with public sector employees earning well above median wages, possessing higher average levels of education, and enjoying a generally esteemed position in society. Thus, it is clear that the policy could make public sector employment even more attractive for mothers and increase female labor participation.

[3] Winett et al. (1982) found that the introduction of flextime programs for working parents in two US federal agencies, which allowed them to shift their work schedules by up to one hour, also led to parents spending more evening time with family. This paper, which still suffers from self-selection bias, and Orpen (1981), which uses randomization in a flextime experiment with 64 female clerical employees to assess effects of flextime on satisfaction and performance, represent the closest examples of anything resembling equivalence to our paper.

[4] All public primary and secondary schools in Georgia are mandated to start at 9:00 a.m. and most also offer late pickup times for working parents. Though not mandated, most private primary and secondary schools follow the same pattern. Preschools, as well as formal and informal childcare, also tend to start at 9:00 a.m. or earlier and are accessible and affordable to the population, with 84.2% of urban children and 67.7% of rural children attending preschool (National Statistics Office of Georgia, 2019). Informal childcare, especially familial, is more common in rural settings, while formal childcare outside of preschool is more common in urban settings. For both settings, formal childcare is more expensive than preschool, but tends to still make economic sense for those with average or better wages. In addition, of note in terms of familial conditions, Georgians tend to marry in their mid-to-late 20s across urban and rural settings (Hakkert, 2017), with an average difference of about 1 year later for urbanites, and tend to start families soon after marriage. Due to the immediate and transient nature of the effects of the policy we study, we do not believe that the policy would reversely impact familial conditions in any statistically significant manner.

On the other hand, a prediction in line with WIF and the gender similarity model would be that this policy could create work-caused conflict for both mothers and fathers, perhaps more so for fathers as they tend to work longer hours on average. Families with young and school-age children are probably less flexible in their daily program than couples without children, single people, or older people with or without older children. For working parents (or those considering entering the labor market), the policy could result in conflicts with their established household itinerary that even the added flexibility and increased evening time with the family would not resolve. Such parents could find it more difficult to participate in the labor market under the new schedule.

Furthermore, though outside of the work-family conflict framework, it is also ambiguous whether those without younger children would find this time shift attractive or not. Some younger, single people might find the possibility of having more free time in the evenings to pursue social activities appealing, while others may be used to sleeping longer in the mornings and find the change objectionable. We hypothesize that the policy disparately impacts affected populations by gender, marital status, family type and size, and along other individual characteristics. Our hypothesis leads us to assume the effects of the policy will be heterogeneous across characteristics and circumstances, informing the main aim of this paper: to determine the dominant effects of the policy, along which dimensions it was most impactful, and related behavioral insights. While the policy could give rise to many compelling research questions that fit our aim, we concentrate on how it may impact gender inequality through female labor participation in government jobs.

Since the policy had no effect on the private sector (where the standard working hours largely remained at 10:00 a.m. - 7:00 p.m.), we are able to employ difference-in-differences (DD) methodology to compare public and private sectors, before and after policy implementation. Given the circumstances and data, DD is the optimal methodology to identify the precise effects of the policy on labor engagement as it separates out all other effects experienced by both control and treatment groups. We find that the policy does not increase female labor participation through an increase in women entering the public sector. In fact, the policy appears to have no significant effect at the extensive margin of employment in either direction.[5] Instead, it primarily leads to a material reduction in the average level[6] of hours worked by full-time employees with children; the outcome in line with the WIF and gender similarity model prediction. At the same time, there is also a significant increase in average work hour engagement by women without children. However, supplemental analyses described in Sections 4 & 5 identify this as an

[5] It is probably worthwhile to note that the lack of press and social discourse around this policy shift in working hours may have also meant that many women outside the labor force or those not already interested in public office jobs may very well not have been aware of the new hours, which would work against the desired encouragement for women to participate more in the labor market and may have directly contributed to this lack of effect at the extensive margin.

[6] The household survey did not request participants to provide an integer of hours worked, rather multiple choice of weekly hours worked intervals (20 hours or less, 21-40 hours, 41-60 hours, and more than 60 hours). Based upon the specific wording to the survey question as well as direct discussions with government officials, GeoStat interviewers, and public employees yielded the opinion that responses to this question principally correspond to contractual hours.

already existing trend and indicate that this is an ordinal response to the reduction of engagement by full-time employees with children. Altogether, we conclude that this increase is a secondary, indirect effect and that the policy did not directly cause an increase in female labor participation. Furthermore, since men with children were most negatively affected and women picked up the gains, the policy may have also indirectly increased overall gender equality. Finally, with regards to the external validity of these findings, it is unclear whether these effects are purely domestic, regional due to cultural similarities of certain countries, or perhaps even global per the intrahousehold division of labor theory of Alberts et al. (2011) further elaborated upon in the next section. We shall not venture to weigh in on this question, but rather proffer the expectation that a disruption of steady state working hours would tend to result in short-term negative repercussions for employees with children per WIF.

2. Literature Review

Gender inequalities in unpaid, household work are known to be directly related to gender inequalities in the paid labor market in terms of participation, engagement, type of employment, vulnerability, career progression, wages, retirement savings, and more (Ferrant et al., 2014). Bearing the majority of responsibility for household duties and the need to coordinate those with paid economic activities results in female 'occupational downgrading', accepting worse conditions, and below-skill-level employment (Hegewisch and Gornick, 2011). Empirical studies of female labor participation are innumerous and results typically point, in one manner or another, to the relationship between household and market labor. Vlasblom and Schippers (2004) identify 'low education' and 'having children' as the most important barriers to female participation in the labor market. Cortes and Pan (2017) conclude that females that anticipate difficulties in balancing career and family are more likely to exit the labor market and specialize at home than their male peers. Herr and Wolfram (2012) claim that an inflexible work environment is a major force driving women to opt out of the labor market at motherhood. Similarly, women might respond to greater occupational time demands by shifting to more family-friendly occupations or by withdrawing from the labor force (Cortes & Pan, 2017). Thus, the time demands of a given occupation seem to, on average, predominantly affect women, who already have a tendency to work less than men, causing women to switch into positions with more flexible time requirements to be able to combine professional and household activities (Wasserman, 2015; Goldin, 2014; Gicheva, 2013).

In the work-family conflict literature, Gutek et al. (1991) were the first to combine research from work-and-family sociologists and development psychologists by bifurcating the work-family conflict into two types of conflict: family interference with work (also known as family-work conflict, FWC) and work interference with family (also known as work-family conflict, WFC). Outcomes of conflicts are informed by two models: the rational view or gender similarity model versus the gender role or gender difference model. The rational view or gender similarity model is predicated on the notion that we have only so much time available to us to split between our work and family roles and as the time spent on them increases, conflict will be perceived regardless of gender. It predicts a convergence of attitudes towards

conflict and the balance of work and family (Keene & Quadagno, 2004). The gender role or gender difference model is based on traditional gender roles resulting in normative differences between the genders, with asymmetric boundaries, expectations, responsibilities, and perceptions of balance. This model predicts that men and women will react to role conflict differently as more time spent in one's gendered domain is perceived as less of a burden (Gutek et al., 1991). In combination, these theories provide testable predictions related to the motivations and behaviors of those affected by an exogenous shock, such as the Georgian policy being evaluated herein, as described in the introduction.

In Georgia, the government enacted initiatives that have been promoting gender equality since 1997 (Jashi, 2005) and there has been a steadily increasing female labor participation rate over the last decade (ILO, 2019). Nonetheless, 89% of the UN's gender relations survey respondents, comprising of both men and women, agree that "a woman's main responsibility is to take care of the family": 50.8% of Georgian respondents were identified as having a 'negative' attitude toward gender equality, while only 3.7% had a 'positive' attitude, and even the 'positive' group maintained a 'patriarchal' pattern of gender-divided household duties (Kachkachishvili et al., 2014). The report concludes that any recent changes in the distribution of household tasks are 'quite superficial' with only a limited amount of actual behavioral and attitudinal modification, while the underlying culturally rooted gender biases have not changed. Georgia is experiencing similar, or more severe, trends as those identified by Sayer (2016).

An examination of why the prevailing attitudes towards gender roles endure reveals a complex, psychological web of attitudes, socialized beliefs, evolutionary differences, and individual thresholds and proclivities that commingle to result in individual, group, and societal standards. For example, in many cultures, the nature of the female gender is perceived as more fluid than that of the male gender. In the context of labor division, this means that it is more acceptable for women to adopt 'masculine' behaviors, such as taking up paid work, than it is for men to adopt 'feminine' behaviors, such as doing unpaid domestic work (Sayer, 2016). By not doing unpaid work, or at least minimizing their involvement in such activities, men may have (perhaps subconsciously) emphasized their masculinity and reinforced their social power (Brines, 1994; Risman, 1999). Extrapolating, it may follow that women performing a greater amount of domestic labor, even under changing socioeconomic conditions, regardless of how much time they spend in paid employment, could persist as a culturally accepted norm. This has, so far, been reflected in what Sayer (2016) finds from five U.S. time use datasets from 1965 to 2012.

Alberts et al. (2011) put forth a compelling theory of domestic labor division that addresses and infuses several single-explanation theories into a more complex, yet rational-based framework. Their theory helps to explain why many contrary phenomena persist in household labor division, including why many full-time employed wives still do a majority of domestic work, why even men that earn less and/or work fewer hours still do not do more domestic work, and why both genders tend to view the currently unequal distribution as equitable. The theory explains that

small differences in traits, informed by evolutionary biological differences and biosocial conditioning,[7] result in disparities in responses to stimuli. Divergent self-organizing systems and response thresholds[8] cause the repetition of minute behaviors that lead to 'expertise' and large behavioral differences that become ingrained over time and across contexts. Moreover, few couples and dyads explicitly discuss domestic labor division and, instead, default to individual response thresholds, social norms, and habits to guide their behavior and only address issues explicitly once discord occurs. According to Alberts et al. (2011) women, on average, have lower innate thresholds for domestic disorder, certain biological characteristics, and different competencies from gendered socialization, which typically lead to higher standards of cleanliness and frequency of household task performance. This puts women at a disadvantage, on average, in the formation and long-term organization of domestic labor. Thus, this theory may substantially explain why the majority of household task responsibility and performance remains with women, despite labor market trends.

Regardless of the underlying cause, the contemporary global labor market is a diverse place, characterized by individual, occupational, local, national, and regional variation in work cultures, work-life balances, standard working hours, and gender-based differences. Moreover, it is clear from the above literature that household characteristics and circumstances affect labor market outcomes, especially impacting women, but there are often conflicting conclusions about the direction and mechanism. The policy being evaluated here, while not revolutionary by any means, imposes a foreign cultural timing onto a significant percentage of the population in an economic ecosystem that was built up, over time, in a local culture. Economic actions cause interactions and externalities. Institutional working hours may have, in part, led to the establishment of specific working hours elsewhere, such as directly related service providers, associated private sectors, schools and childcare facilities, and restaurants, afterwork, and nightlife venues that follow employee schedules. It is reasonable to expect that even a one-hour shift in work hours disrupts a steady-state element of the Georgian society and could cause behavioral adjustments at the individual and/or organizational level. As the data indicates that there were no significant effects at the organizational level, the evaluation of this unique policy may shed light on how individuals and households react to such seemingly minor changes and provide insights into how situational and familial composition affect labor participation behaviors and gender equality, as well as illustrate nuances related to domestic division of labor and intra-familial/intra-household bargaining.

[7] Women, through survival, have developed a better sense of smell as well as more attention and sensitivity to household cleanliness, combined with reinforcement from more time spent in the home due to childbearing (Alberts et al., 2011).

[8] Self-organization systems, evident throughout the living world, explain how local, individual interactions lead to group-level attributes (Camazine et al., 2001). 'Convergent' self-organization is when the behaviors of individuals become more alike. 'Divergent' self-organization is when the behavior of an individual causes the same behavior to be less likely in others and the act of performing the behavior also reduces stimulus-level-causing responses. Response thresholds are 'the perceived stimuli that must exist for an individual to decide to perform a task (Theraulaz, Bonabeau, & Deneubourg, 1998). Like Hrdy's (1999) responsive mothers, individuals with low response thresholds for a specific task are moved to perform the task earlier than individuals who have a higher threshold for the task (Breshers & Fewell, 2001; Robinson & Page, 1989)' [Alberts et al., 2011, page 7].

To assess how the policy impacts labor participation in the affected sector, we turn to the difference-in-differences method using the affected public sector as the treatment group and the unaffected private sector as control. We begin by confirming that the private sector is, in fact, unaffected and that the consequent adjustments are at the individual level. Next, we assess the differences between the employee behaviors in the two sectors following policy implementation. While this paper may be the first to evaluate such a working hour shift policy, it is not the first to use DD methodology to assess outcomes between affected and unaffected sectors. Some recent examples include the specific use of public and private sector employees to evaluate the impact of a Taiwanese pension policy shift to identify the effect employer-sponsored pensions have on household saving (Yang, 2020), the use of sector-specific import tariff increases to estimate their impact on U.S. export growth (Handley et al., 2020), and the use of differences in implementation of anti-smoking regulations amongst sectors and countries across Europe to determine the economic effect on restaurants, bars, and cafes (Pieroni & Salmasi, 2017). In a recent paper closely related to our topic, Angelov et al., (2016) employ DD methodology to assess the long-term effects of entering parenthood on the gender wage gap and female labor participation, though they did not use sectors as an instrument to evaluate a policy. Generally, DD methodology is common in labor economics research and we believe it is appropriate and optimal for the purposes of our analysis.

3. Data

3.1 Primary Dataset

The primary data used in this study is publicly available on the web site of the National Statistics Office of Georgia (GeoStat). In particular, we utilize individual level data from the Households Incomes and Expenditures Survey for the four calendar years 2013, 2014, 2015, and 2016. Every quarter, GeoStat surveys approximately 3,500 Georgian households and aims to have each randomly selected household participate in the survey four consecutive times. The outcome is a piecemeal panel dataset composed of repeated individual observations for up to a one-year history of a household's socio-economic, gender, and geographical characteristics.

As true for any survey dataset, the household budget survey data is expected to contain some measurement error. Each respondent reports detailed information regarding household or private socio-economic and geographical information for the past quarter, which can lead to recall and other inaccuracies while reporting numbers. According to GeoStat documentation as well as direct discussions with data collectors, the collection process uses a best-practice methodological approach supervised by the statistical department and the collected data is a population-representative sample with a small margin of error.[9] All things considered, there seems to be no evidence that the measurement error would not be random.
Table 1 shows and defines the survey variables employed in our analyses.

[9] We concur with the representative assertion. GeoStat provided population-representative weights, which had little variance throughout. We also tested the assertion by running analyses both with and without the weights and there were no meaningful differences in direction, magnitude, or significance.

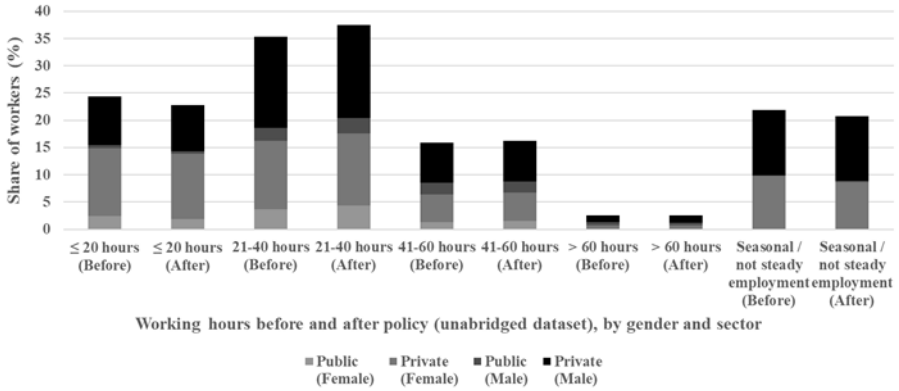
Table 1 GeoStat Household Survey Variables and Their Descriptions

Variables	*Description*
Weekly working hours (intervals)	The number of working hours during the week. Categorical variable: '20 hours and less; 21-40 hours; 41-60 hours; Depends on a period (season); More than 60 hours.' It is important to note that the survey specifically asked: "How many working hours do you usually have per week at your main job excluding breaks?"
Sector	Sector of employment. Categorical variable: 'Private ownership; State ownership; Other'
Activity	Economically active according to the ILO strict criteria. Binary variable: 'Yes; No'
Urban or Rural	Rural\Urban Classification. Binary variable: 'Rural; Urban'
Owner of home	Owner of the dwelling (ownership type). Categorical variable: 'Belongs to the household; Mortgaged; Rented; Used without payment'
Assistance	Whether the household received assistance or any kind of advantage or not. Binary variable: 'Yes; No'
Age	Age of an individual
Family size	Number of household members
Education	Categorical variable: 'Illiterate; Does not have primary education; Lower secondary education; Primary education; Secondary professional program; Higher professional program; Upper secondary education; Vocational program; Bachelor; Master; Doctor. '
Small kids	Number of children (0-7 years old)
Big kids	Number of adolescents (8-15 years old)
Working man	Number of working age men (16-64 years old)
Working woman	Number of working age women (16-59 years old)
Duration in the living place	Duration of living at this address
Dwelling selling price	The amount in local currency that the household would pay to buy a dwelling similar to theirs.
Change in financial condition	Financial condition of the household has changed during the past 12 months (subjective evaluation). Categorical variable: 'improved very much; not changed; slightly improved; slightly worsened; worsened very much'
Attending any professional courses	Whether the household member attended any courses for learning new professions/skills during the past three months. Binary variable: 'Yes; No'
Never worked before	Whether a household member has never worked. Binary variable: 'Yes; No'
Economic condition based on income	Economic condition of the household based on household income (subjective evaluation). Categorical variable: 'Very bad; Bad; Satisfactory; Middle; Good'
Profession	Profession or specialty defined by a diploma, certificate or other document or gained another way. Categorical variable, listed more than 200 professions at the level of 4-digit code adopted to International classification of ISCO-88
Marital status	Categorical variable: 'Single; Married; Non-registered marriage; Divorced; Widowed'
Migration	If a household member moved. Categorical variable: 'From other country; From another region in Georgia; From the same region in Georgia'
Economic condition based on property	Economic condition of the household based on household property (subjective evaluation). Categorical variable: 'Extremely poor; Poor; Middle; Rich; Well-off'
Reason for not applying for assistance	Reason the household has not applied to the Social Service Agency. Categorical variable: "I don't hope to get the assistance; Our family doesn't require social assistance; It's difficult to answer; I don't know where to apply; I can't do it myself and there is nobody to whom I can address for help; I consider it being humiliating for our family"
Special status	Special status of the household member. Categorical variable: 'Chronic patient; Disabled (I group); Disabled (II group); Disabled (III group); IDPs'
Area of dwelling	Total area of the dwelling (in square meters)
Pensioner man	Number of pension age men (65 years and older)
Pensioner women	Number of pension age women (60 years and older)
Mobile phone	Quantity of the owned durable good
Additional activity	Secondary employment

Notes: Variable names adjusted for ease of comprehension. For example, 'Weekly working hours (intervals)' is actually 'TimeDuration'.

The following figures present an examination of our dataset, beginning with a breakdown of the ratio of weekly working hours pooled before and after policy implementation (specifically, using September 1, 2014 for the threshold), delineated by gender and sector.

Figure 1 Weekly Working Hours (Intervals), by Gender, Sector, and Implementation



Within our dataset, 13.7% of all working people are employed in the public sector and 86.1% work in the private sector. Segregating by gender, 56.2% of public sector employees are female and 43.8% are male. Unlike the government sector, the number of men exceeds the number of women working in the private sector. Men account for 53.6% and women 46.4% of workers in the private sector. On average, over the entire period of the dataset, 19.54% of the public sector employees worked 20 hours and less per week, 48.79% worked 21-40 hours per week, and 25.95% worked 41-60 hours per week. Only 1.96% were employed in a seasonal/not steady public sector position and 3.76% worked more than 60 hours per week. Partitioning this information further by whether employees had children gives us the next three figures (for all workers, private sector workers, and public sector workers). In Appendix Tables A60-A67, we provide this and additionally delineated descriptive statistics in table form as well as partition the public sector observation numbers by monthly mean, maximum, minimum, and standard deviation. Further, we provide main variable descriptive statistics by gender and sector and a correlation matrix with the main variables included in our analysis in Appendix Tables A68-A70. Given that it is a representative sample, the tables show a reasonably balanced division amongst the subsample groups.

Regarding the balance between the sectors, Appendix Table A61 shows that the distribution amongst the working hour intervals between the two sectors is fairly similar but diverges most amongst the full-time and seasonal employment figures. In terms of structural differences between the sectors, it is important to recall that in Georgia, public sector employees earn above median wages, considerably more than their private sector counterparts, on average; have much higher average levels of education, most have at least a master's degree; and carry a high level of respect in society. The public sector distribution has less variance than the private sector,

because it simply does not offer many seasonal employment opportunities and mostly does not hire people without a higher education, which immediately removes teen and early adult employees from the variance, who tend to work the fewest hours. When we remove the seasonal workers from the numbers, the distributions become much more similar across the intervals. Prior to the policy, the distribution is almost identical, slightly skewed to greater work hour engagement in the public sector, though this does not reflect top-down organizational differences between the sectors, but rather the natural, bottom-up difference in sector breadth and variance noted above. Moreover, while there is essentially no change in the distribution amongst the intervals in the private sector before and after the policy, we see (here and in corresponding Appendix Table 61) a large increase in the ratio of 21-40-hour-interval workers in the public sector post policy, accompanied by a direct decrease in the ratios just below and above, but especially below. This foreshadows the findings of this paper.

Figure 2 Weekly Working Hours, by Gender, Parental Status, and Implementation

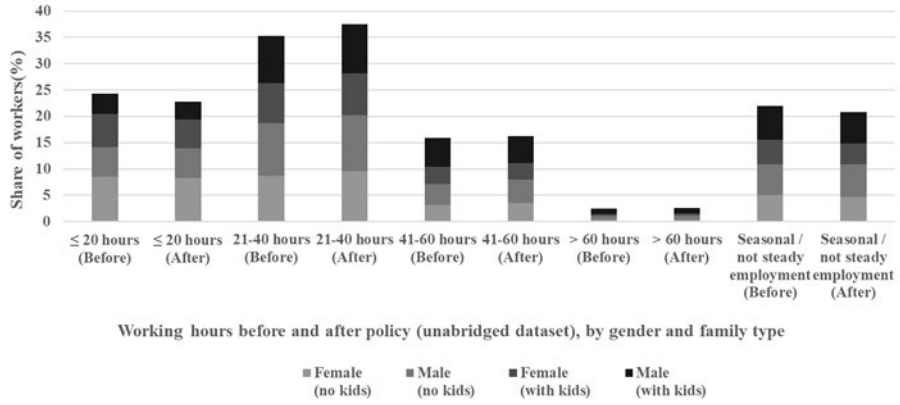


Figure 3 Private Sector Weekly Working Hours, by Gender, Parental Status, and Implementation

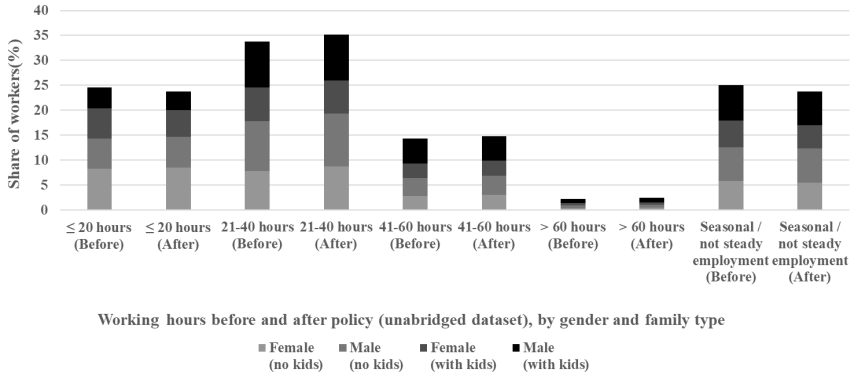
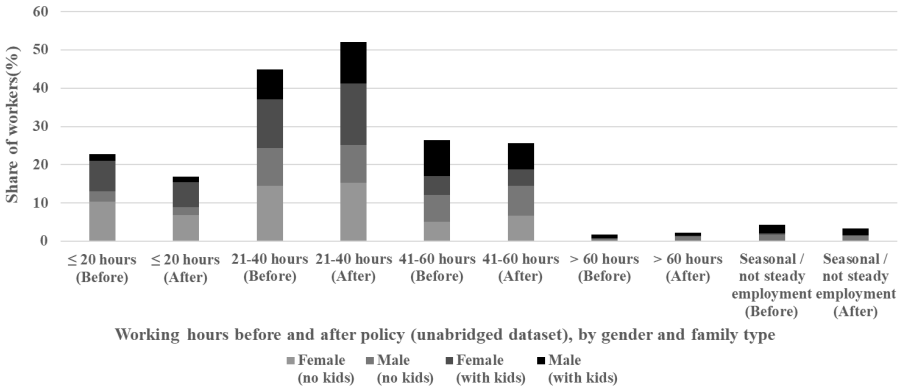


Figure 4 Public Sector Weekly Working Hours, by Gender, Parental Status, and Implementation



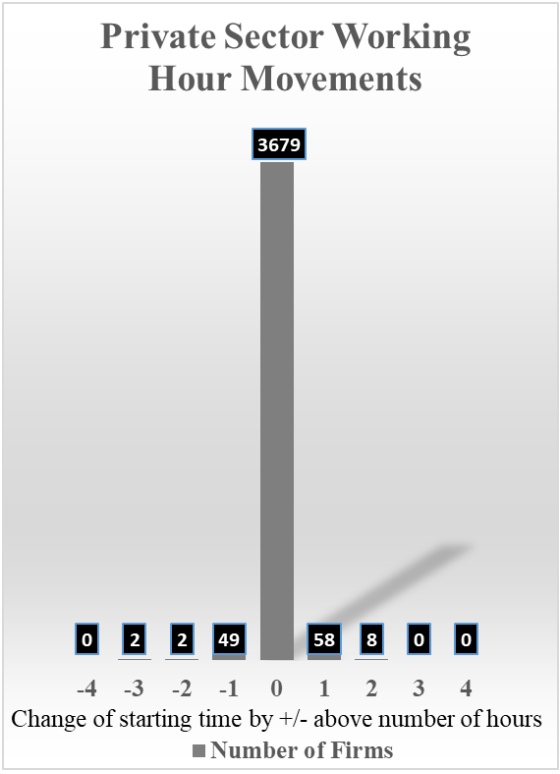
Figures 2, 3, and 4 visually communicate the total number of weekly working hour (intervals) observations in our dataset before and after the policy implementation (specifically, using September 1, 2014 for the threshold) broken down by gender, parental status, and sector. In total, a slim majority of employees working 40 hours or less are females, while employees working overtime hours are mostly male. On the face of the data, it seems that there is a significant increase in the number of female (and male) employees with children working 21-40 hours in the public sector, which some might claim as evidence of 'family friendliness' and increased female labor participation. However, increases are also present for their male, no children, and private sector counterparts, hence the need for the methodology described in the next section to conduct a proper evaluation of the policy. For example, several such regressions without covariates return positive gains, especially for women with small children. However, after including covariates that control for alternative sources of this increased employment, the policy's effect is weakened and becomes statistically insignificant. One notably important revelation is the very small number of observations in our dataset of females working 60 hours or more in the public sector. Such a small sample size is insufficient for reliable inference regarding female labor participation around the 60-hour threshold. While not as impactful to inference, another questionable sample size revealed by the descriptive statistics is the relatively small sample size of men with older children working more than 60 hours.

3.2 Supplementary Dataset

A supplemental, firm-level database is used to check whether the implementation of the government's new policy led the private sector to adjust working hours for their employees. The Business Information Agency (BIA) is a leading data collector of company profiles operating in Georgia. Their database consists of statistical information for more than 45,000 active companies. Each firm's general information (e.g. trademarks, products, registration date, VAT number, business activity, legal address, website, and working hours – which,

specifically, are the firm's standard operating hours) is gathered through the legally-mandated statistical office, with changes double-checked within 6 months with each business directly, and made publicly available on BIA's webpage. We extracted and analyzed the data for a subsample of firms that had observations recorded before and after the policy implementation between 2013 and 2016. We found 3802 firms with observations both before and after the policy implementation between 2013 and 2016. Only 3.2% of those firms changed their business hours after the policy had been applied by the government. Moreover, as evidenced by Figure 5, the changes were normally distributed around the mean and mode of zero change. Additionally, we analyzed the shift of working hours for the placebo threshold of one year before as well as one year after the policy to check that the trend holds for the other periods. The results show that only 4.2% and 3.2% of firms shifted their business-operating hours, respectively, and in a similarly distributed manner. Figure 5 visually illustrates changes in start times of operating hours from our random subsample of BIA data. It visually demonstrates a clear lack of direct effect on private sector working hours from the policy.

Figure 5 Distribution of Private Business Starting Time Movements Post Policy



Notes: The bar chart shows the relative amounts visually, and the actual numbers above the bars, of private sector businesses that changed their starting times after the policy was implemented, and by how many hours (from -4 to +4).

We conclude that the facts presented (the great lack of changed operating hours, the normal distribution of those sporadically changed operating hours, and the placebo comparison to the year before and after yielding very similar amounts of such changes) do support our assertion that the policy we are examining did not cause any direct changes to the working/operating hours of our control group, the entire private sector.

4. Methodology

Having confirmed that the working hours of the private sector were in no way systematically affected by the policy change that directly altered them in the public sector, we now detail how the difference-in-differences method is utilized to determine how the new government policy affected participation in the labor market. According to Angrist and Pischke (2008), the method estimates the effect of the treatment (i.e., an explanatory variable or an independent variable) on the outcome (i.e., the response variable or the dependent variable) by comparing the average change over time in the outcome variable for the treatment group, compared with the average change over time for the control group. We designate the private sector as the control group and the public sector as the treatment group. In formal terms, i denotes the individual, s denotes the sector (either public or private), and t denotes the time period. As the policy was implemented on September 1, 2014, with essentially no notice, we believe that any direct effect of the policy change on labor market participation would not occur before 2 months at the earliest due to established employment notice periods for leaving a position, the time it takes to process and hire a new employee, and the time it takes for managers and employees to assess the policy's actual effects and permanently adjust work hour schedules internally. This assumption, further discussed in section 5.1, is by and large confirmed by the findings of the short-term-effect and September-threshold analyses, which are described at the end of this section. Accordingly, the main analysis time threshold was set as November 1, 2014. In formal terms, this outcome variable takes the form:

$$Y_{ist} = 1, \text{ if an individual is working a specified range of hours per week}$$

$$Y_{ist} = 0, \text{ if an individual is working an alternate range of hours per week}$$

In particular, Y_{ist} equals zero (below) or one (above) across the specific binary extensive margin threshold of 0 hours and more than zero hours (including seasonal / not steady employment), and the following intensive margins (which do not include those working 0 hours nor seasonal / not steady employment): above and below 20 hours, above and below 40 hours, above and below 60 hours, and pairwise[10] amongst the individual weekly working hour values. Linear DD regression equations take two conventional forms (ending up with the same result). We opt for the interaction term form:

[10] The pairwise thresholds represent a supplemental analysis that aims to approximate how local the policy-induced working hour movements are (just a few hours across the nearest threshold or larger jumps) as well as provide an enhanced picture of the movements just around the thresholds.

$$Y_{ist} = \alpha + \beta_1 Treatment_{is} + \beta_2 Time_{it} + \beta_3 Treatment_{is} * Time_{it} + \beta_4 X_{iat} + \varepsilon_{ist} \quad (1)$$

Where $Treatment_{is}$ is a dummy variable that equals one if the observed individual is in the public sector, $Time_{it}$ is a dummy variable that equals one if the time of the observation occurred in November 2014 or later, α is a constant, and X_{ist} is a set of covariates that includes an individual's characteristics and answers to other survey questions that are correlated with the outcome variable. The resulting coefficient, β_3, expresses the post-policy correlation difference between the control and treatment groups, making it the only consequential and relevant coefficient to this research and the only coefficient reported in the output tables. As the weekly working hours replies are in 20-hour intervals, the DD regressions with the constructed thresholds are specifically capturing the changes between the average number of workers below/above a given threshold.[11] We attempt to further distinguish the specific correlation of the policy on Y_{ist} by executing the regression using three sets of covariates[12] that increase the precision of the coefficient and the explanatory power of the regression. Furthermore, we also aim to increase the precision by more accurately defining the treatment threshold.[13] All tables in section 5 and the Appendix display only the coefficients with the full covariate schedules, broken down by increasing particularization of the treatment group.

To support causality inferences of β_3 covariates, we provide parallel trends analyses to assess whether the two groups had similar trends over time prior to the policy implementation, which then diverged due to the effect of the policy on the treatment group. In addition, we check causality by conducting placebo effect analyses, counterfactually changing the time threshold to twelve months prior to the actual change. A resulting lack of a statistically significant β_3 bolsters the notion that

[11] As there are, essentially, no changes along the extensive margin (see section 5.1.1), the constructed threshold regressions are not (or are minimally) capturing new or leaving employees on either side of a given threshold. Moreover, any changes in the average number of employees in a given interval are not captured as part of the DD regression unless they are across a given threshold. That is, a change in average number of employees between the 1-20-hour interval and the 21-40-hour interval is not captured as a difference at the 40-hour threshold.

[12] First, we run the regressions without controls. Next, we add several substantial covariate controls for individual, household, and professional attributes, including age, education, family size, number of working age people in the household, number of children, living in an urban or rural area, length of time living there, owning their own home, several objective and subjective measures of income and wealth, if they are economically active, and if they have ever been unemployed. Finally, we add all remaining covariate controls that had any statistically significant correlation from the DD regression, including marital status, migration history, profession category, additional wealth measures, number of retired family members in the household, and disability status. While only the full covariate results are presented in the body and online appendix of this paper, a full appendix with all results is available by request. Across the regressions of the main thresholds, the covariates that were consistently most correlated with Y_{ist}, which is evident through their statistically significant coefficients (available in Appendix Tables A56 – A59) were urban location, years in this city, wealth and ownership measures, and age.

[13] The baseline is all public employees as treatment and private as control. However, as noted in the introduction, not all public employees were affected by the policy. Therefore, in the second specification, we move the employees from the entirely unaffected public fields, such as public education employees (teachers, school administrators, etc.), to the control group. In the third specification, we move expectedly unaffected public employees to the control group as well. That is, while the expected majority of public employees in specific professions should not be affected, such as dentists, some may happen to be affected by the policy due to certain idiosyncratic peculiarities (such as office location) or the ambiguous nature of certain professions. Hence, they are included only in the final specification.

effects found from the difference-in-differences regressions were specific to the policy change and not just random noise. Consequently, we consider a strongly statistically significant β_3 coefficient that holds in the most stringent control configuration, is part of a parallel trend that diverges post-policy, and does not produce a placebo effect, to be a credible substantiation of a causal effect of the policy on those treated.

Since we employ a two-month lag from the actual initiation of the policy, we further supplement the main analyses with DD regressions of the main thresholds using the September 1, 2014 threshold. Furthermore, we run short-term analyses of the effects for three months, six months, and twelve months from both the November and September thresholds. These results reveal an ordinal nature of the effects of the main analysis, with some of the effects beginning around three to six months after policy initiation only to have the strength of those effects depleted by the end of 2016, while others begin later and grow stronger and more statistically significant through 2016. Lastly, though we control for age and type of location throughout the analyses, we also run partitioned analyses by separating the sample by urban versus rural locations and dividing it in half by median female age (49 years old) and median male age (45 years old) as well as their interaction to assess whether the policy had age-specific and/or location-specific implications. [14] Full result tables are presented in Appendix Tables A1 – A53.

5. Results

In this section, we present and discuss tables that highlight the most significant and relevant regression findings from all three treatment specifications and binary thresholds listed in section 4. Only select subsample groups from the main analyses that give an overview or provide statistically significant results, or their counterparts, are featured herein. Complete full covariate control output tables that exhibit all results for every subsample group and supplemental analysis are presented in the Appendix.

5.1 Main Results

5.1.1 Extensive Margin

The first output table we present is the 0 hours and more than zero hours threshold (the extensive margin between working and not working; including seasonal/not steady employment). As can be seen in Table 2, all of the resulting β_3 coefficients are weak and statistically insignificant. Despite the note at the beginning of section 5 and the practice of including only main and significant results in the tables that follow, we included all output from the combined, with, and without children groups in Table 2 to show the extent of the insignificance at this margin. A detailed analysis of the extensive movements confirms an overall lack of changes at this margin. The analysis uncovered that there were only 303 [311] extensive margin moves out of 5964 [5667] total panel observations and only 102 [102] extensive

[14] An analysis by education was also explored, but since nearly all public employees have higher education degrees and the vast majority have a master's degree or higher, a DD with the private sector population would be biased.

margin moves involving the public sector around the threshold [lagged] of the policy implementation. This resulted in only a 20 [20] net employee gain in the public sector (an insignificant difference). Appendix figures A7 and A8 visualize the extensive margin data points for both the official implementation timing and the lagged threshold used in the DD regressions. These movements are meagre (see Figure A7) and not statistically different from the extensive margin moves from the placebo thresholds of one year prior (see Figure A8). Furthermore, the September, short-term, age, and location analyses all return weak and insignificant results.[15] Therefore, we conclude that the policy did not have a statistically significant effect on the extensive margin of employment.

Table 2 DD Regression Results for Weekly Working Hours (Intervals), Extensive Margin

Subsample	*Gender*	*(1)*	*(2)*	*(3)*	*N*	R^2
All	All	-0.00233	-0.0006	0.000152	85523	0.34
		(0.005)	(0.006)	(0.006)		
	Male	-0.00764	-0.00817	-0.00389	45627	0.32
		(0.009)	(0.009)	(0.01)		
	Female	0.0016	0.006	0.00398	39896	0.38
		(0.006)	(0.007)	(0.008)		
With kids	All	0.00169	0.00352	0.00656	40124	0.31
		(0.008)	(0.009)	(0.009)		
	Male	-0.00514	-0.00712	-0.00172	21979	0.26
		(0.013)	(0.013)	(0.014)		
	Female	0.00675	0.0146	0.02	18145	0.41
		(0.009)	(0.011)	(0.012)		
Without kids	All	-0.00864	-0.00687	-0.00561	45399	0.36
		(0.007)	(0.008)	(0.009)		
	Male	-0.0163	-0.0143	-0.00898	23648	0.37
		(0.012)	(0.012)	(0.013)		
	Female	-0.0012	0.00186	-0.00168	21751	0.36
		(0.008)	(0.009)	(0.011)		

Notes: 10%, 5%, 1%, and 0.1% levels of confidence are indicated by (+), (*), (**), and (***), respectively. Standard errors are in parentheses. Column labels: (1) is the pure sector division of public and private as treatment and control, respectively; (2) adds employees from entirely unaffected public fields, such as public education, into the control group; (3) adds employees with professions where the expected majority would not be affected, into the control group.

5.1.2 Intensive Margin

Further analysis revealed statistically significant movement within the intensive margin of labor participation through the weekly working hours (intervals) variable. We assessed those effects by creating specific binary thresholds using the

[15] The younger [and urban] groups had a few β_3 coefficients at the 10% [and 5%] level for males without kids, especially when not married. However, all significance disappeared in the combined young & urban analysis, which may expose the previous results as spurious or be caused by the reduced sample size, though the latter argument is uncertain with the associated number of observations. Combined with the findings of the detailed analysis, these anomalous results do not change our overall conclusion, but they may indicate that some young, unmarried, urbanite males found the new hours unattractive, as hypothesized in the introduction.

survey's interval responses to the question of how many hours each employed individual works to construct the thresholds of above and below 20 hours, above and below 40 hours, and above and below 60 hours. As there is a lack of women working more than 60 hours and since the 20- and 40-hour thresholds represent standard part-time and full-time working hours with most employment bunched there, those two thresholds are most pertinent herein. We begin with the 20 hours or less versus 21+ hours threshold in Table 3.

Table 3 DD Regression Results for Weekly Working Hours (Intervals), 20-hour Threshold

Subsample	*Gender*	*(1)*	*(2)*	*(3)*	*N*	R^2
All	All	0.0396***	0.0187+	0.0190+	60234	0.15
		(0.009)	(0.01)	(0.011)		
	Male	0.011	0.00611	0.00822	30740	0.13
		(0.013)	(0.013)	(0.014)		
	Female	0.0597***	0.0292*	0.0323+	29494	0.14
		(0.013)	(0.015)	(0.017)		
With kids	Female	0.0335+	0.0102	0.00403	13154	0.14
		(0.019)	(0.022)	(0.025)		
Without kids	Female	0.0854***	0.0414*	0.0575*	16340	0.16
		(0.018)	(0.02)	(0.023)		
Without kids (family size>1)	Female	0.0830***	0.0381+	0.0631**	14363	0.14
		(0.019)	(0.021)	(0.024)		
Without kids (family size>1, married)	Female	0.0933***	0.0518+	0.0821**	9472	0.12
		(0.024)	(0.027)	(0.031)		

Notes: 10%, 5%, 1%, and 0.1% levels of confidence are indicated by (+), (*), (**), and (***), respectively. Standard errors are in parentheses. Column labels: (1) is the pure sector division of public and private as treatment and control, respectively; (2) adds employees from entirely unaffected public fields, such as public education, into the control group; (3) adds employees with professions where the expected majority would not be affected, into the control group.

Regression results for the entire subsample indicate that there is a positive correlation between the DD identified policy effect and weekly working hour interval, but when dividing the subsample by gender, it is evident that correlation is heavily driven by the change in average female working hour engagement. When further dividing that sample into those with and without children, it becomes clear that those who have most increased their engagement are women without children. When again dividing women without children into married and not married subsample groups, it becomes clear that the increase is driven by married women without children. This is an unexpected result, especially for those who believed that changing public office working hours would break down barriers for women with children. Moreover (visible in the unabridged table in Appendix Table A2), at the 20 hour-working-week threshold, there is no significant difference if the children are small (0-7 years old) or big (8-15 years old). According to the output of the short-term analyses, these differences begin to become evident and significant about 8 months post policy initiation and strengthen through the end of 2016.

The supplemental age analysis reveals that these effects were more consistently occurring for older employees, even for men at a much weaker level, across the treatment specifications, including strong, positive effects by singles who

were the sole member of their household, though the sample size may already have become an issue there. By location, the effects are stronger and more consistent for rural female employees, though the subpopulations that were most strongly affected were older, rural women without children, followed by younger, urban women without children. We also witness moderate, negative effects on younger, urban women with children and older, rural women with older children. The refined subpopulation analyses also uncovered some otherwise elusive effects that were averaged out in the larger sample groups: moderate, positive effects of the policy on younger, rural women with children and older, urban women with older children[16] as well as fairly sizable, positive effects for urban males with children, especially older children, and for older, urban males without children.

Table 4 displays the results for the 40 hours or less versus 41+ hours threshold. From the full subsample results, it is evident that the effect is strong at this threshold. Dividing it by gender reveals that both men and women are affected at this threshold, but especially men. This gender difference decreases as the subsample is further reduced to include only those with children. Those with younger children seem most likely to reduce their work engagement across this threshold in general, though men with older children seem more affected than their female colleagues. The lack of an effect on women from the full sample population seems to be due to the countering effect from women without children increasing working hour engagement at this threshold, especially those who are part of a household of two or more people and married.

Further refinements are revealed by the age and location analyses. While women with children were similarly affected across the age groups, urban women with children were much more impacted than their rural counterparts. Urban and rural men with children were similarly negatively impacted, though slightly more so in rural locations. Across locations, older men with children were much more impacted than their younger counterparts. Older men and women with younger children were the most negatively affected at this margin, while of those with older children, only older men were affected and not as strongly. For the younger group, the opposite is true, with the greatest negative effects experienced by women with older kids as well as men with younger kids. Though the positive effects for women without children were universal amongst the partitioned subsamples, the vast majority were experienced by younger women, especially urbanites. Regarding men, only older, urban males without children exhibited positive effects from the policy at this margin.

The short-term and September analyses (from 6 to 14 months post policy implementation) consistently display slightly stronger and more statistically significant results for those with kids at the 40-hour threshold than many of the full-data, November-threshold results above. This indicates that the effects on working hours at this margin are primary and early ordinal results of the policy. It seems these effects at this margin were, on average, greatest and most significant about 12-

[16] The placebo effect output related to the older, urban women with older children returns at a rather significant level, though the placebo trend was in the opposite direction. While this disqualifies this finding as fully credible, it also does not indicate an already occurring trend. Since it is a minor finding, we decided to keep it herein.

14 months after the policy went into effect and then began to decline over time. Given that the policy impacted individual (and by interaction, household) schedules by 30-60 minutes, it seems logical that they would have a transitory nature and be more intense in the short-term and then dissipate as a new steady state is achieved. For women without kids, the effects begin to become significant 12 months post policy and then strengthen.

Table 4 DD Regression Results for Weekly Working Hours, 40-Hour Threshold

Subsample	*Gender*	*(1)*	*(2)*	*(3)*	*N*	R^2
All	All	-0.0289***	-0.0365***	-0.0488***	60234	0.16
		(0.009)	(0.009)	(0.01)		
	Male	-0.0541***	-0.0543***	-0.0596***	30740	0.17
		(0.014)	(0.014)	(0.015)		
	Female	-0.00835	-0.0168	-0.0353**	29494	0.17
		(0.011)	(0.012)	(0.013)		
With kids	All	-0.0898***	-0.105***	-0.120***	27868	0.16
		(0.013)	(0.014)	(0.015)		
	Male	-0.109***	-0.106***	-0.116***	14714	0.17
		(0.02)	(0.021)	(0.022)		
	Female	-0.0682***	-0.0879***	-0.110***	13154	0.15
		(0.016)	(0.019)	(0.021)		
With small kids	All	-0.118***	-0.127***	-0.116***	12630	0.17
		(0.02)	(0.021)	(0.023)		
	Male	-0.140***	-0.122***	-0.107***	6994	0.17
		(0.03)	(0.03)	(0.032)		
	Female	-0.0829**	-0.108***	-0.0898**	5636	0.18
		(0.026)		(0.033)		
With big kids	All	-0.0683**	-0.0861***	-0.139***	9727	0.18
		(0.021)	(0.024)	(0.026)		
	Male	-0.106**	-0.125***	-0.149***	4787	0.2
		(0.037)	(0.038)	(0.041)		
	Female	-0.0302	-0.0309	-0.115***	4940	0.17
		(0.026)	(0.03)	(0.035)		
Without kids	Female	0.0452**	0.0432**	0.0272	16340	0.2
		(0.014)	(0.015)	(0.017)		
Without kids (family size>1)	Female	0.0364**	0.0344**	0.0208	29354	0.17
		(0.012)	(0.013)	(0.014)		
Without kids (family size>1, married)	Female	0.0610***	0.0535**	0.0521*	9472	0.16
		(0.018)	(0.02)	(0.022)		

Notes: 10%, 5%, 1%, and 0.1% levels of confidence are indicated by (+), (*), (**), and (***), respectively. Standard errors are in parentheses. Column labels: (1) is the pure sector division of public and private as treatment and control, respectively; (2) adds employees from entirely unaffected public fields, such as public education, into the control group; (3) adds employees with professions where the expected majority would not be affected, into the control group.

Table 5 DD Regression Results for Weekly Working Hours, 60-Hour Threshold

Subsample	*Gender*	*(1)*	*(2)*	*(3)*	*N*	R^2
All	All	-0.0117**	-0.0108**	-0.00870+	60234	0.03
		(0.004)	(0.004)	(0.005)		
	Male	-0.0181**	-0.0131+	-0.0102	30740	0.03
		(0.007)	(0.007)	(0.007)		
	Female	-0.00588	-0.0068	-0.00534	29494	0.03
		(0.004)	(0.005)	(0.005)		
With kids	All	-0.0189**	-0.0196**	-0.0135+	27868	0.03
		(0.006)	(0.007)	(0.007)		
	Male	-0.0271**	-0.0207+	-0.0104	14714	0.04
		(0.01)	(0.011)	(0.011)		
With small kids	Male	-0.0417**	-0.0394*	-0.0398*	6994	0.05
		(0.015)	(0.016)	(0.017)		

Notes: 10%, 5%, 1%, and 0.1% levels of confidence are indicated by (+), (*), (**), and (***), respectively. Standard errors are in parentheses. Column labels: (1) is the pure sector division of public and private as treatment and control, respectively; (2) adds employees from entirely unaffected public fields, such as public education, into the control group; (3) adds employees with professions where the expected majority would not be affected, into the control group.

At the 60 hours or less versus more than 60 hours threshold, presented in Table 5, the results are somewhat similar to those of the 40-hour threshold, except they are weaker and less statistically significant. Due to the lack of female representation of public employees working more than 60 hours, it is not remarkable that both women with children and women without children have completely insignificant and low magnitude results at this threshold. Men, however, are seemingly quite affected at this margin. In particular, younger men in rural areas with children have the largest reduction in engagement. All effects at this margin with large enough sample sizes for realistic inference are negative and mostly driven by younger workers.

From the September and short-term analyses, we learn that the effects of the policy are even more immediate than at the 40-hour margin, being felt within three months of the commencement of the policy and, therefore, already partially captured in the period prior to the two-month lag of the November threshold, resulting in them being slightly diluted in the main DD comparison. Considering the schedules of those who worked more than 60 hours per week, it is logical that they would be so immediately impacted by this exogenous schedule change. As with the 40-hour margin, the effect seems to peak somewhere around 12 months after the policy went into effect and then dispersed into 2016. The most notable difference between Table 5 and the related September and short-term analyses is that men without children also display a significant, negative effect (amounting to about 2-3%) from the policy at this margin.

5.2 Robustness Checks

As noted in the methodology section, the legitimacy of difference-in-differences regression results rests on certain underlying assumptions, which can be substantiated through parallel trends and placebo effect analyses.

5.2.1 Placebo Effect

Placebo effect analysis helps to confirm that the identified effect is actually directly related to the effect of the policy and not some other cause. This is generally conducted by changing one of the difference points in the DD regression to something that should not be causing an effect similar to the policy. When the resulting β_3 coefficient is statistically insignificant, that supports the contention that statistically significant β_3 coefficients from the actual DD analyses are caused by the policy and not some other phenomenon. In our case, we elected to use the fairly standard placebo threshold of one year prior to the threshold used for the main analysis. The complete results of the main placebo effect analyses are in Appendix Tables A40-A42.

Table 6 displays the output of the placebo effect analysis for the complete sample population at each main threshold as well as all statistically significant findings from the main threshold analyses. Most of the β_3 coefficients from all the placebo analyses are weak and statistically insignificant, confirming that the vast majority of the main analysis results are not caused by some other effect. There are sporadic β_3 coefficients below and in the supplemental placebo analyses that come out as statistically significant, but do not counter the findings and conclusions from the main analysis. These coefficients are from the subsamples of women with young children at the 40-hour threshold, unmarried people at the 60-hour margin, men with older children at the 60-hour threshold, urban women with older children at the 20-hour margin, and younger women with younger kids at the 40-hour threshold.

One major exception to this is the strong and statistically significant positive β_3 coefficients from women without children (including those from families composed of two or more people, both married and unmarried) at the 40-hour threshold. This indicates that women without children were gaining more working hours in public sector jobs than their counterparts in the private sector prior to the policy implementation, ruling out the policy as the explicit cause. Instead, the policy may have aided in the continuation of this trend by providing additional hours for women without children in the public sector to acquire. This interpretation is echoed in the ordinal findings of the short-term analyses.

Table 6 Placebo Analysis Results for Weekly Working Hours (Intervals), Multiple Thresholds

Subsample	*Gender*	*(1)*	*(2)*	*(3)*	*N*	R^2
All (20-hour threshold)	All	0.00222	0.00929	0.00607	25051	0.16
		(0.015)	(0.016)	(0.017)		
	Male	-0.00419	0.00944	0.0103	12834	0.15
		(0.02)	(0.021)	(0.022)		
	Female	0.000653	0.00637	0.00512	12217	0.15
		(0.021)	(0.024)	(0.026)		
All (40-hour threshold)	All	0.0154	0.00634	0.015	25051	0.18
		(0.013)	(0.014)	(0.016)		
	Male	-0.0078	-0.0101	0.0128	12834	0.2
		(0.021)	(0.021)	(0.023)		
	Female	0.0279+	0.0182	0.0197	12217	0.17
		(0.016)	(0.018)	(0.021)		
With small kids (40-hour threshold)	Female	0.0474	0.0479	0.111*	2316	0.2
		(0.041)	(0.046)	(0.052)		
Without kids (40-hour threshold)	Female	0.0567**	0.0668**	0.0483+	6652	0.2
		(0.021)	(0.023)	(0.026)		
Without kids (family size>1) (40-hour threshold)	Female	0.0588**	0.0731**	0.0564*	5858	0.19
		(0.022)	(0.025)	(0.028)		
All (60-hour threshold)	All	2.12E-05	-0.00143	-0.00378	25051	0.03
		(0.006)	(0.006)	(0.007)		
	Male	-0.00248	-0.00576	-0.00584	12834	0.04
		(0.01)	(0.011)	(0.011)		
	Female	0.000567	0.00231	-0.00149	12217	0.03
		(0.006)	(0.007)	(0.008)		
With big kids (60-hour threshold)	Male	0.0559*	0.0735**	0.037	2108	0.07
		(0.025)	(0.026)	(0.028)		
Without kids (family size>1, unmarried) (60-hour threshold)	All	-0.0291*	-0.0325*	-0.0298*	4194	0.05
		(0.013)	(0.014)	(0.015)		

Notes: 10%, 5%, 1%, and 0.1% levels of confidence are indicated by (+), (*), (**), and (***), respectively. Standard errors are in parentheses. Column labels: (1) is the pure sector division of public and private as treatment and control, respectively; (2) adds employees from entirely unaffected public fields, such as public education, into the control group; (3) adds employees with professions where the expected majority would not be affected, into the control group. Time threshold set to one year prior to lagged threshold used in main analysis. All observations up until implementation used to assess placebo effects.

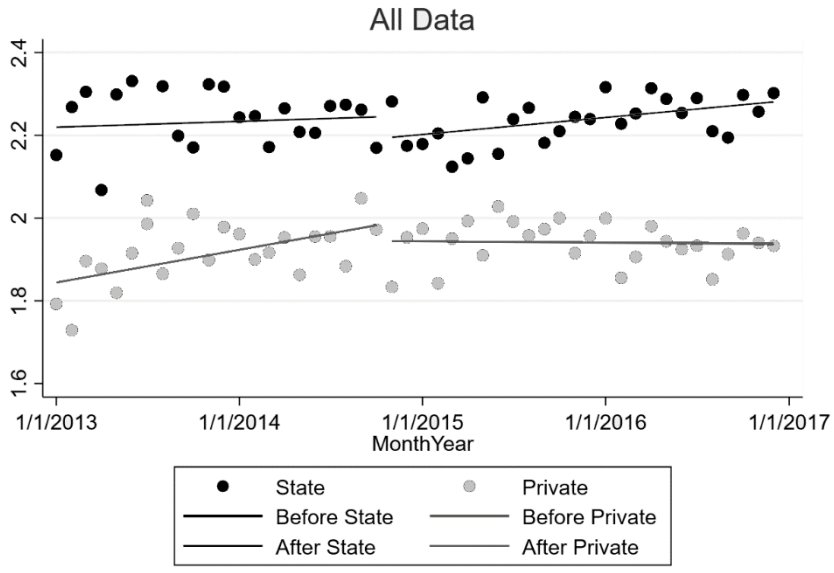
5.2.2 Parallel Trends

Parallel trends analyses assess whether the control and treatment groups were on a trend prior to the implementation of the policy in question and diverged thereafter so that the difference experienced between the groups after implementation can be identified as causal. Figure 6 below is a visualization of mean working hours for the entire sample data, by month, using ordinal integers to represent the survey-generated work-hour-interval bins, and divided into before and after policy implementation periods, each with linear fit lines. Figures 7 and 8 break that down

by gender. The trends are represented by linear best fit lines for the public and private sector groups for the period prior to and post threshold. As the name, parallel trends, suggests, the ideal validation is when the two lines (or the pattern in the data points) prior to the threshold are reasonably parallel to one another to substantiate that the two groups were on a similar trajectory prior to the policy. Post threshold, there should be a level change in the data and/or the lines (or the pattern in the data points) should diverge to confirm that the policy altered their trajectories.

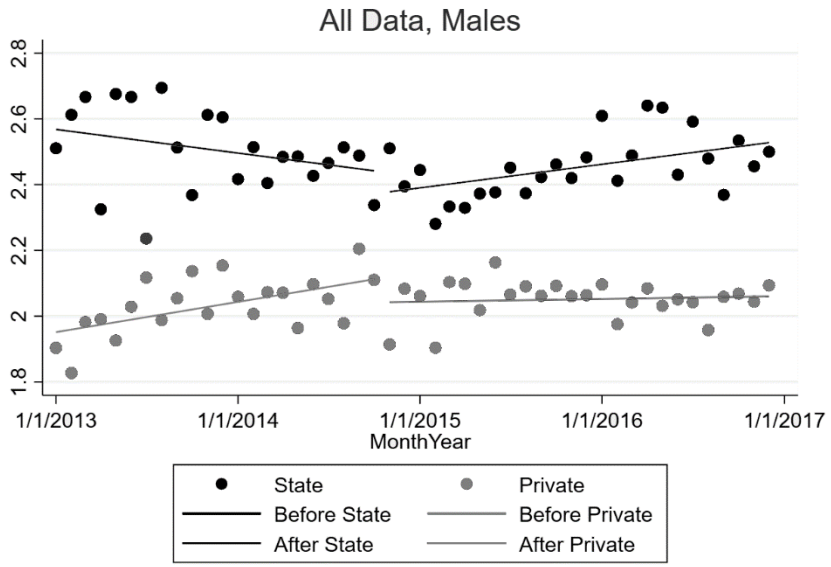
Figure 8 represents a good example of a corroborating parallel trend graph. It is clear that women, overall, between the private and public sectors were following a similar general trend prior to the policy and then diverged thereafter. Figure 6 is a less perfect example but seems to still conform to expectation. Undoubtedly, linear best fit lines are imperfect, and thus latitude on their similarity is expected. Moreover, Figure 6 is almost certainly less perfect than Figure 8 due to the influence of the questionable Figure 7. The diverging trend in Figure 7 is evident, but the pre-policy fitted line for the treatment group seems to be skewed up and to the left by a bunching of some early data points. Sometimes, patterns in the data points, which may not necessarily match the fitted lines due to outliers or bunches, are visually discernible, as is the case with Figure 7. Moreover, there appears to be a similar upwards sloping pattern throughout 2014 to that of the control group. When we examine the trend just one year prior to the threshold, then the pattern does become much more parallel.

Figure 6 Parallel Trend Scatter Plot with Linear Fitted Lines, All Data, All Observations



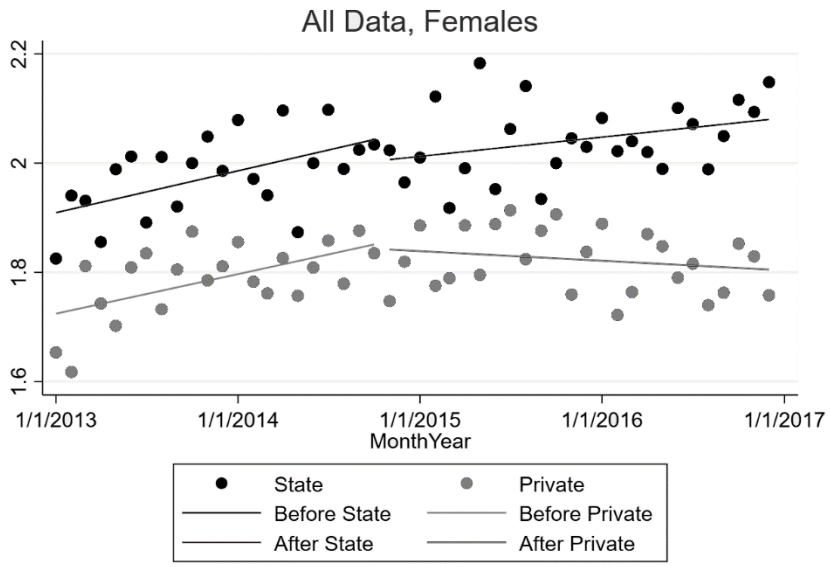
Notes: 'Private' includes all workers in the private sector as well as the certainly unaffected public field workers. 'State' includes all remaining public sector workers. Threshold is set at two months post policy implementation.

Figure 7 Parallel Trend Scatter Plot with Linear Fitted Lines, All Data, Males



Notes: 'Private' includes all workers in the private sector as well as the certainly unaffected public field workers. 'State' includes all remaining public sector workers. Threshold is set at two months post policy implementation.

Figure 8 Parallel Trend Scatter Plot with Linear Fitted Lines, All Data, Females



Notes: 'Private' includes all workers in the private sector as well as the certainly unaffected public field workers. 'State' includes all remaining public sector workers. Threshold is set at two months post policy implementation.

While the overall parallel trends are generally informative, when we examine the subsample populations that constitute the main findings of this paper and focus on their trends around the policy implementation threshold, the picture deviates considerably from the overall trends. For example, the trends of the male subsample groups tend to become generally more corroborating, while the female subsamples deviate. Parallel trend figures for every subsample regression that led to the main findings discussed above are available in Appendix Figures A1-A6. At the 20-hour threshold, both women with and without children have fairly corroborating parallel trend figures. At the 40-hour threshold, the women with children graph, while not perfect, still seems to validate the assumption. The figure for females without children at the 40-hour threshold does not feature very parallel fit lines prior to implementation, though this may not be surprising given the placebo effect analysis outcome. Due to a lack of decisive support for the parallel trends analysis based on the graphs alone, we turn to an alternative method for validation.

We construct a regression based on the DD methodology, but instead of utilizing a single dummy variable for time before and after policy implementation, we create dummy variables for every month in our dataset as well as dummy variables for each month interacting with the single treatment variable. By taking the resulting interaction term coefficients from the period prior to policy implementation and running an F-test on their joint significance, we are able to assess whether they were jointly significant and reject the null hypothesis that they are equal to zero and the groups are the same. As with the visual analysis, this does not represent a perfect confirmation of the parallel trends assumptions but does provide a more rigorous method of assessing whether the parallel trends assumption is broken. See Appendix Table A55 for the F-test significance results analyzing the main subsample groups that constitute the main findings above. Notably, the results show that we cannot reject the parallel trends assumption for women without children at the 40-hour margin and that the parallel trends assumption for men with children at the 60-hour margin is rejected. Due to the latter, the inference in section 5.2 is not substantiated and assumably does not hold. Otherwise, we conclude that the parallel trends visual analysis and the supplemental joint significance F-test analysis support the parallel trends assumption for the main analysis findings discussed below.

5.3 Discussion

An overall depiction of the effect of the policy on labor participation has been revealed from the main and supplemental analyses. Despite the beliefs and intentions of some members of the Georgian parliament, public office working hours do not seem to have been a 'family friendly' barrier to female labor participation and the policy did not cause any statistically significant increase in the extensive margin of employment. Moreover, employees with children reacted to the policy by mostly reducing their working hour engagement. Primarily, the policy negatively affected the ability of full-time employees with children to work the longer hours that they had been working prior to the implementation of the policy. This result echoes the prediction from the WIF conflict type and gender similarity model framework. Secondarily, the engagement of women without children, predominantly married women, substantially increased across both the 20- and 40-hour thresholds. Women

without children were probably able to take up most of those hours given up by parents with children because of having more flexible schedules than their colleagues with children, with married women having flexibility to an even greater extent perhaps due to a more settled-down personal life than their single colleagues. We do not portend to know the exact causes of these behaviors and leave that to the realm of future research.

The ordinal findings of the short-term and September analyses show negative effects on working hour engagement beginning to occur much earlier and dissipating into 2016 and the positive effects beginning to occur later on and accumulating into 2016. These findings indicate that the policy, which caused a 30-60-minute impingement on individual and, by proxy, household schedules, is transient in nature, with stronger effects in the short term that disperse over time as a new steady state is attained.

Delving into the supplemental analyses offers insights of an informative nature. The results indicate that the effects of the policy were not uniform across family types but, as hypothesized in the introduction, were circumstantially disparate, differing in magnitude and direction amongst age- and location-based subsample populations. For example, though the policy mostly resulted in reductions of working hours for parents with children, it seems positive work hour engagement effects at the 20-hour margin were experienced by older, urban and younger, rural women with children as well as urban males with children. This represents the only evidence of any positive, 'family friendly' effects resulting from the policy. Furthermore, women without children may not be the only ones who increased work engagement as a result of the policy; older, urban males without children appear to have done so at the 20- and 40-hour thresholds as well.

In addition, the age and location analyses uncovered further inconsistent patterns of effects that may reveal informative insights into those differences. At the 40-hour margin, the negative impact on women with children was almost exclusive to urbanites, which may reflect a more modern trend in domestic arrangements in urban areas. While men with children were negatively affected at the 40-hour margin, those in rural locations were somewhat more so, which may be related to the respectively greater travel distances and inferior social infrastructure. Older men and women with younger children were the most negatively affected at the full-time employment margin, perhaps reflective of the impact of unanticipated, later-in-life fecundity on families. A number of other conflicting patterns across age- and location-partitioned groups indicates that the opposing hypothesized effects and incentives identified in the introduction all seem to be at play. For example, across age groups, it appears that older people with children bore the greater brunt of the negative effects at the 20-hour threshold, younger women without children experienced most of the positive effects at the 40-hour threshold, and younger men were most negatively impacted at the 60-hour threshold. When including location differences, it appears that the resulting positive work engagement effects were most experienced by younger, urban women and by older, urban men without kids at the 40-hour margin.

5.4 Further Investigation

We attempt to further enhance the perception of the working hour engagement movements by analyzing the working hour intervals pairwise in order to reduce noise from average changes in both directions throughout the entire sample. We also continue to use the methodological setup to further investigate additional subsample groups to see if we can uncover any more circumstantially specific effects of the policy. One circumstance that we conjecture as potentially influential on one's decision to increase or decrease labor participation at work is if they happen to be working in multiple jobs. Another circumstance is related to the composition of a household. Specifically, we hypothesize that married couples with one partner in the treatment group and one in the control group may face a greater strain upon their previously established status quo. Moreover, this may be especially true for couples with a single vehicle. Full results and discussion of these analyses are available in Appendix 1. The results of the pairwise analysis, for the most part, parallel those of the main analysis, implying that the changes are mostly local across the thresholds. No evidence of any effect of multiple employment on the main analysis findings was found with a single discrepancy at the 20-hour threshold, revealing a statistically significant increase for women with children, especially young children (a finding that had been only suggested by the main analysis results, but found at significant levels in this and the pairwise analyses). Regarding mixed sector couples, males had negative effects in terms of work hour engagement, particularly so when the head of the household was in the public sector and when the family had only one vehicle. Females in mixed sector couples in which the spouse was in the public sector showed positive effects in work engagement. Such a combination of results implies gains by women in intra-familial bargaining or a modernization of social norms.

6. Conclusions

On September 1, 2014, the country of Georgia enacted a unique policy moving the working hours of public office employees from 10:00-19:00 to 9:00-18:00, impacting the working hour schedules of all affected employees. While not the official or main reason for implementing such a policy, some members of parliament had believed that the new hours would be 'family friendly', making it easier for women to balance household and professional responsibilities, and thus increase female labor participation. Thanks to access the Georgian government provides to their household data survey, combined with the fact that the policy did not affect the private sector, we were able to implement a difference-in-differences methodology to accurately analyze whether the policy increased female employment and gender equality. This policy affected an estimated 200,000 employees, yet the impact of this policy had never been evaluated. Moreover, we were unable to find any literature evaluating any policy that exogenously adjusted the working hours of a significant portion of employees in an economy. Nevertheless, since the effects of the policy variously impacted employees across multiple characteristics, especially by gender and family type, this study is most closely related to work-family conflict, gender inequality, and intra-household bargaining and resource allocation literature. Based on concepts from the work-family conflict literature, we arrived at two

opposing predictions for the possible effects of the policy on employees with families.

The results discussed in section 5 of this paper reveal that the policy had no significant effect on the extensive margin and, instead, directly and primarily led to a substantial decrease in working hour engagement by full-time employees with children. This result is in accordance with the prediction based on the gender similarity model and WIF conflict type. Although there is some evidence of a modest increase in engagement by part-time employees with children, it does not come close to the magnitude of the negative effect on full-time employees with children. Therefore, we assert that the policy did not directly lead to an increase in female labor participation. While we also found a greater expansion in engagement in the public sector than in the private sector by women without children, the placebo effect analysis discovered that this was a trend already occurring prior to policy implementation and the short-term analyses confirmed that this effect was ordinally second. We infer that those hours gained by married women without children and, to a lesser extent, unmarried women without children, were a subordinate result of the negative effect on working hours of full-time employees with children. Thus, it could be argued that the policy did indirectly increase female labor participation. Furthermore, as the majority of the negative engagement effects fell on male employees and positive effects on female employees, the policy also indirectly improved gender equality by increasing the female side of the gender balance equation of the labor force.

Moreover, there were several additional, informative insights gained into the effect of the policy. As hypothesized in the introduction, the policy caused heterogeneous effects with considerable variance in size and direction that were often strongly informed by circumstance, age, and location. For example, the analyses revealed that male employees with older children appear to be both those that had the largest general negative effect on their engagement when working 40+ hours, especially in rural locations, and the largest general positive effect on engagement for part-time employees working 20 hours or less. Hence, despite the female-focused intentions of certain parliament members, the policy seems to have directly affected male employees on both sides of the spectrum more than female employees. We also find that there were especially negative effects on the engagement of male employees who were part of a mixed sector couple, especially when they had only one vehicle. There are even indications that unmarried men without children had some modest negative effects on engagement from the policy.

Of course, women were certainly affected by the policy as well. The negative effects on urban females with children was substantial, especially for those with younger children. On the more positive side, part-time female employees in a mixed sector couple with zero or one vehicle showed considerably positive effects on their engagement. This may also be true for full-time female employees in couples where the head of the household is in the private sector and the spouse in the public sector, but this result has a questionable sample size. All in all, the additional insights may imply increased female intra-familial bargaining power or that Georgian fathers and husbands (especially in urban areas) have begun to participate more in household duties and are open to more modern feminist outcomes than the UN gender survey

found. Both the former and latter explanation imply an occurring or future evolution in social norms.

Our work contributes to the vast literature on working hours in several dimensions. First, it is the first paper that evaluates such a work hour shift policy. Second, it may contribute to the gender inequality, intra-household, and work-family conflict literatures. And third, to a lesser extent, this unique exogenous policy and the multi-dimensional findings of this study may be useful to those with research areas related to work hours and shifts, such as work-life balance, benefits of flexibility, etc., as well as research bodies dedicated to the field, such as 'The Shift Project'. For instance, the indirect effects of this policy that affected workers differently by their familial conditions may likely provide insights for future research into the myriad work schedule effects on workers and their families, or practical identification of diverse 'family friendly' policies as pursued by Saltzstein et al. (2002). Regarding future research, the policy appears to have revealed evolving social norms and affected the amount of time and manner in which family members spend time together. For example, the policy may have caused WIF spillover leading some families to spend less time together, which may negatively influence family well-being, especially for spouses, as, *ceteris paribus*, the more time spouses spend together, the more satisfying the marriage (Kingston & Nock, 1987). Given that Alberts et al. (2011) find that intrahousehold division of labor may be rather universally human in nature, some of our conclusions may directly extend to counterfactual situations around the world. Nevertheless, we only conjecture that the random disruption of a steady state in working hours will likely result in generally negative consequences for employees with children, at least in the short- to mid-term, probably because disruption of household schedules causes work interference with family conflict for both mothers and fathers. For policymakers considering a similar work hour shift, to ensure fewer negative effects, we would recommend that any such policy be accompanied by even greater flexibility, daycare, and/or other WIF-conflict-reducing support for employees with children.

APPENDIX

This policy evaluation has two appendices. The appendix below offers extended analyses of the further investigations summarized in section 5.4. Please follow the link below to the output appendix with all cited tables, figures, and full covariate results. An unabridged appendix with all results is available by request.
Appendix, A2:
https://www.dropbox.com/s/w747ii45rpxm5s5/GPSWHPS.pdf?dl=0

A1. Further Investigation

A1.1.1. Pairwise Analyses

Given the nature of the methodology employed, only the positive β_3 coefficients from the lowest interval pair and the negative β_3 coefficients from the highest interval pair have undeniable value for interpretation, because only those movements are bounded by absolute frontiers (zero hours and all hours greater than 60). These are presented below. As all other pairwise output is not necessarily capturing movements across the given threshold, those results may only be implicative. Nevertheless, the pairwise analyses may provide some additional insight even at the middle margins and are discussed below. The output tables are presented in Appendix Tables A5-A10. Table 7 examines the movements in the weekly working hours variable from the 20 hours or less interval to/from the 21-40-hour interval.

From this pairwise analysis, just above and below the 20-hour threshold, an enhanced picture of the effects of the policy at this margin has emerged. The β_3 coefficients follow the same pattern as the 20-threshold analysis but have become stronger and more statistically significant. Moreover, this perspective also reveals the positive effects on working hour engagement experienced by parents with children, which are most consistent across the treatment specifications for men with children, especially driven by men with older children. Furthermore, both women without children who are married and unmarried seem to be experiencing positive effects from the policy change, though the effect is more consistent for the married ones.

Table 8 shows a closer view of the 60-hour threshold. This time the results are weak in magnitude and statistically insignificant across the board of all subsample divisions. It may be that any substantial policy-caused effects of the 60-hour threshold are captured in the pairwise analyses of Appendix Tables A7 and A9. There is a noteworthy result for men who make up the whole of their household. The effect appears to be an extreme decrease in working hours across these intervals. However, the sample size is minuscule, which means the result is almost certainly spurious.[17]

[17] The negative placebo effect for single household males at the 60-hour threshold, also with a small sample size, further supports the supposition that the findings for single household males in Appendix Table A7 and Table 8 are spurious.

Table 7 DD Regression Results for Weekly Working Hours, Pairwise, 20 Hours or Less ↔ 21-40 Hours

Subsample	*Gender*	*(1)*	*(2)*	*(3)*	*N*	R^2
All	All	0.0656***	0.0406**	0.0433**	45937	0.11
		(0.012)	(0.013)	(0.015)		
	Male	0.0383*	0.0278	0.0326	22024	0.11
		(0.019)	(0.02)	(0.022)		
	Female	0.0774***	0.0465**	0.0542**	23913	0.10
		(0.016)	(0.018)	(0.02)		
With kids	All	0.0549**	0.0450*	0.0426*	20275	0.11
		(0.017)	(0.02)	(0.022)		
	Male	0.0565*	0.0585*	0.0584+	9846	0.11
		(0.028)	(0.029)	(0.031)		
	Female	0.0577*	0.0368	0.0376	10429	0.11
		(0.023)	(0.027)	(0.03)		
With small kids	Female	0.0781*	0.0471	0.0777	4453	0.11
		(0.037)	(0.043)	(0.047)		
With big kids	Male	0.102*	0.0915+	0.108+	3313	0.13
		(0.05)	(0.052)	(0.056)		
Without kids	Female	0.0914***	0.0480*	0.0721**	13484	0.11
		(0.021)	(0.024)	(0.027)		
Without kids (family size>1)	Female	0.0852***	0.0421+	0.0779**	11742	0.10
		(0.023)	(0.025)	(0.028)		
Without kids (family size>1, married)	Female	0.0830**	0.046	0.0842*	7994	0.09
		(0.028)	(0.032)	(0.035)		
Without kids (family size>1, not married)	Female	0.0872*	0.0508	0.0536	3748	0.16
		(0.04)	(0.045)	(0.048)		
Just singles (family size=1)	Female	0.179*	0.131	0.0262	1742	0.20
		(0.081)	(0.093)	(0.106)		

Notes: 10%, 5%, 1%, and 0.1% levels of confidence are indicated by (+), (*), (**), and (***), respectively. Standard errors are in parentheses. Column labels: (1) is the pure sector division of public and private as treatment and control, respectively; (2) adds employees from entirely unaffected public fields, such as public education, into the control group; (3) adds employees with professions where the expected majority would not be affected, into the control group.

Table 8 DD Regression Results for Weekly Working Hours, Pairwise, 41-60 Hours ↔ More than 60 Hours

Subsample	*Gender*	*(1)*	*(2)*	*(3)*	*N*	R^2
All	All	-0.0247+	-0.0162	-0.00682	14297	0.05
		(0.014)	(0.014)	(0.015)		
	Male	-0.0165	-0.00259	0.00477	8716	0.05
		(0.019)	(0.019)	(0.02)		
	Female	-0.0218	-0.023	-0.017	5581	0.07
		(0.02)	(0.021)	(0.024)		
Just singles (family size=1)	Male	-0.592*	-0.592*	-0.601*	155	0.72
		(0.291)	(0.291)	(0.288)		

Notes: 10%, 5%, 1%, and 0.1% levels of confidence are indicated by (+), (*), (**), and (***), respectively. Standard errors are in parentheses. Column labels: (1) is the pure sector division of public and private as treatment and control, respectively; (2) adds employees from entirely unaffected public fields, such as public education, into the control group; (3) adds employees with professions where the expected majority would not be affected, into the control group.

Appendix Table A6 examines the pairwise intervals of 20 hours or less and 41-60 hours. Here only the positive effects experienced by women without children are significant. While not as strong as the effects experienced by married women, unmarried women from households with two or more members now also exhibit statistically significant effects consistently across treatment specifications. The pairwise analysis between less than 20 hours and more than 60 hours in Appendix Table A7 does not have many β_3 coefficients with statistical significance and does not reveal much new information. A consistent negative effect, though neither strong in magnitude nor statistical significance, seems to be occurring for women with children. However, this is a spurious result given the lack of women in the sample who work more than 60 hours. Moreover, the results from the subsample groups of younger and older children are both insignificant. Another noteworthy result in Appendix Table A7 is that men who are the only members of their household display a distinct increase in working hour engagement across this pair, though the sample size is already rather small and probably also indicates only a spurious outcome of happenstance.

As in Tables 7 and 8, Appendix Table A8 is a pairwise analysis that provides an enhanced depiction of one of the main thresholds: just above and just below the 40-hour threshold. It mostly echoes the 40-hour threshold analysis with a strong negative effect on working hours for all people with children, especially for men, and while the effect is more balanced across genders with small children, it is more pronounced for men with older children. Furthermore, women without children continue to display a strong positive effect, driven by women who are married and part of a household of two or more people.

Similar to the 60-hour threshold, the 21-40 hour and more than 60 hours interval pair in Appendix Table A9 shows only negative effects upon work engagement, driven by men with children, especially those with younger children. However, unmarried men without children in households that are made up of two or more people also display a modest negative effect here. Expectedly, between this pair of intervals, women have almost uniformly insignificant and low magnitude results, with the positive effect women without children have at the lower thresholds completely disappearing in terms of magnitude and significance. While women with older children exhibit a single statistically significant, negative β_3 coefficient at the strictest treatment specification, it is another spurious result due to the small sample size of women working more than 60 hours.

A1.1.2. Multiply Employed

The GeoStat survey asked participants if they held secondary employment. By dividing those who answered ‘yes’ and ‘no’ into two different subsample groups, we then evaluated how each group was affected by the policy. Every main analysis table in the Appendix includes the multiply employed subsample. Appendix Table A54 breaks down the two subsample groups into those with one job and those with more than one job, by threshold and gender. Across the thresholds and genders, the results tend to follow the main results with the singly-employed group having stronger, more statistically significant coefficients and the multiply-employed group displaying much weaker, insignificant results. Furthermore, results from the stricter

treatment-specification groups generally tend to be reflective of those in the pure public/private sector specification. One result that stands out in opposition to both of these trends is that of female workers holding multiple jobs at the 20-hour threshold for the stricter treatment specifications.

Since the effect at the 20-hour threshold is undeniable, not opposed by the placebo analysis, and the sample size large enough, we further explore this group in Appendix Table A11 by subdividing it into the family-type subsample groups used throughout the analyses. However, as the sample sizes tend to become rather small here, we must weigh the results carefully. Women without children holding multiple jobs continue the previous pattern of increasing their working-hour engagement at the 20-hour threshold. The results also provide another example of a small indication that women with children also increased engagement at this threshold. Additionally, there is a consistent result amongst the treatment specifications showing men without children holding multiple jobs give up working hours as a result of the policy, which is the lone example of such a finding at the 20-hour margin and in opposition to the gains found for older, urban males in the age- and location-based analyses. The magnitude and statistical significance are both strong and the placebo analysis finds no opposing results. However, the size of the sample renders the finding plausible but inconclusive.

A1.1.3. Mixed Sector Couples

To evaluate how mixed sector couples may have been affected by the policy, we limited the subsample to only married couples. We identified which couples had one partner in the private sector and one in the public sector.[18] Next, we summed up all the automobiles, trucks, minibuses, and motorcycles into a single variable we dubbed 'vehicle' and divided the mixed sector couples into groups that had zero, one, or more than one vehicle. We also then further divided the mixed sector couples into smaller subsamples by which spouse was in the private sector and which was in the public sector. For the vast majority of the couples in the dataset, the 'head' of the household in a married couple is the husband and the 'spouse' is the wife. Finally, we again divided these subsamples by those who had zero, one, or more than one vehicle. It is presumably not surprising that our sample sizes sometimes dropped far below a minimal level for the central limit theorem to reasonably be in effect. Nonetheless, we present all the results of these analyses as part of every main analysis table in the Appendix.

At the 20-hour threshold (Appendix Table A2), the results of mixed-sector-couples reflect the findings of the corresponding main sample analysis at a generally lower statistical significance. It seems that females in couples without vehicles increase their engagement most at this threshold. While there are not too many divergent results, one that stands out is for men who are part of a couple in which the head of the family is in the public sector and the spouse in the private sector. Their

[18] This may not be exactly treatment versus control, as evidenced by our treatment specifications. However, technical limitations and sample sizes resulted in this division. Moreover, this analysis is beyond the scope of our main research question and we consider this close enough to satisfy curiosity and possibly inspire future research.

hours seem to be severely reduced below the 20-hour threshold because of the policy. However, given the sample size, this result is probably spurious.

Appendix Table A3 may indicate several new insights in addition to those from the 40-hour threshold in the main analysis, though the sample sizes in the majority of the further divided subsample groups tend to be unreliably small. One finding that does seem to come with a large enough sample size for proper inference is that men in mixed sector couples reduce engagement more than their full subsample counterparts (at a substantially increased percent compared to the main analysis), especially for those in couples with just one vehicle. However, there is a modestly statistically significant effect found in the placebo analysis for the mixed sector couple males with just one vehicle, and thus the policy may be exacerbating an underlying trend. The output also indicates that the effect is driven mostly by men in a mixed sector couple in which the head of the family is in the public sector, but here the sample size is already too small to consider this a reliable inference.

One result for women that may be approaching a large enough sample size is the strong, positive effect displayed by women who are part of a mixed sector couple in which the head is in the private sector and the spouse is in the public sector. Furthermore, the negative effect experienced by males in mixed sector couples is driven mostly by men in couples in which the head of the family is in the public sector. Moreover, the considerable increase in female working hours for mixed sector couples is most driven by women in couples with more than one vehicle and in couples in which the spouse is in the public sector with only one vehicle. Of course, these findings come with the consequential caveat that the sample size is very small in the detailed subsamples.

The 60-hour threshold by mixed sector couple analysis in Appendix Table A4 expectedly returns almost no statistically significant β_3 coefficients, except for mixed sector couples with the head of the family in the private sector and the spouse in the public sector, but with a dubious sample size.

A1.1.4. Interpretation of Further Investigations

Supplementing the main threshold analyses with the pairwise analyses both confirmed and enhanced many results from the threshold analyses as well as further revealed new findings. Tables 7, 8, and Appendix Table A8 examine the intervals just below and above each of the thresholds in the previous section. Altogether, they reinforce the conclusions above as well as confirm the existence of the few positive, but weak, 'family friendly' effects on parents at the 20-hour threshold, especially on men with older children. The analysis also indicates that unmarried women without children increased engagement across both the 20- and 40-hour margins, which had not been evident from the full sample threshold analyses. Moreover, the results suggest that the vast majority of the changes across the engagement thresholds were local, meaning that effects on working hours were most commonly to the adjacent interval rather than causing major gains or losses, which seems echoed in the lack of extensive margin movement. Furthermore, such local movements imply that the interval nature of the data is not capturing the full effects of the policy in the intensive margin of working hours, which may indicate an avenue for future research to elaborate further.

The multiply employed analysis found that the only divergence from the main analysis results (for women) occurred at the 20-hour threshold. Results from the pairwise analysis for those holding multiple jobs at the 20 hours or less versus 21-40-hour interval pair further lent support to this finding. It may be that many of this subpopulation who worked 20 hours or less in the public sector held multiple jobs out of necessity, adding or shifting hours to their public sector jobs once it became possible. Further delving into the 20-hour threshold revealed that female workers without children who held multiple jobs continued the previous pattern of increasing their working hours at the 20-hour threshold. However, we know from the placebo effect analysis that this is probably not directly caused by the policy. There is also fairly strong evidence from a potentially large enough sample size showing men without children who held multiple jobs giving up working hours due to the policy. It could be that the new hours conflicted with their other job(s) and, therefore, they reduced their hours in the public sector job to adjust.

The household composition analyses seem to indicate that being part of a mixed sector couple does appear to make a material difference to those affected by the policy. For full-time male employees, especially those who have only one vehicle, the effects are substantially more negative. For part-time female employees, the effect may be moderately more positive for those in a mixed sector couple with one or zero vehicles, though the latter is probably more related to a lack of wealth and income than to transportation difficulties (i.e. indicative of an employee who will work more hours if the opportunity arises). There seems to be a prescient combination of negative effects experienced by men in mixed sector couples (especially with just one vehicle and with the head in the public sector) and positive effects experienced by women in mixed sector couples in which the spouse is in the public sector. Likewise, there is some evidence that full-time female employees in couples in which the spouse is in the public sector experience much stronger positive effects, though it is unclear whether the inference is reliable due to sample size. Altogether, these findings may signify an overall change in social norms or female gains in intra-familial bargaining, perhaps affecting resource distribution and household division of labor. This would be quite contrary to the findings of Kachkachishvili (2014).

REFERENCES

Alberts JK, Tracy SJ, Trethewey A (2011): An Integrative Theory of the Division of Domestic Labor: Threshold Level, Social Organizing and Sensemaking. *Journal of Family communication*, 11(1):21-38.

Angelov N, Johansson P, Lindahl E (2016): Parenthood and the Gender Gap in Pay. *Journal of Labor Economics*, 34(3):545-579.

Angrist JD, Pischke JS (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.

Baffoe-Bonnie J, Golden L (2007): Work-Study: Time Use Tradeoffs, Student Work Hours and Implications for Youth Employment Policy. *Student Work Hours and Implications for Youth Employment Policy* (December 26, 2007):

Bird RC (2015): Why Don't More Employers Adopt Flexible Working Time. *W. Va. L. Rev.*, 118, 327.

Business Information Agency, BIA (2019): Business Information about 49,516 Active Companies. Retrieved from: https://www.https://www.bia.ge/

Blau FD, Kahn LM (2017): The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3):789-865.

Breshers SN, Fewell JH (2001): *Annual Review of Entomology*, 46, 313–340.

Brines J (1994): Economic Dependency, Gender, and the Division of Labor at Home. *American Journal of Sociology*, 100(3):652-688.

Camazine S, Deneubourg JL, Franks NR, Sneyd J, Theraulaz G, Bonabeau E (2001): *Self-Organization in Biological Systems*. Princeton, NJ: Princeton University Press.

Carr MD (2011): Work Hours and Wage Inequality: Evidence from the 2004 WERS. *The Journal of Socio-Economics*, 40(4):417-427.

Cortes P, Pan J (2017): Cross-Country Evidence on the Relationship between Overwork and Skilled Women's Job Choices. *American Economic Review*, 107(5):105-09.

Dawson J, Linsell L, Zondervan K, Rose P, Carr A, Randall T, Fitzpatrick R (2005): Impact of Persistent Hip or Knee Pain on Overall Health Status in Elderly People: A Longitudinal Population Study. *Arthritis Care & Research: Official Journal of the American College of Rheumatology*, 53(3):368-374.

Farulava (2014, September 2): Part of the Public Servants Will Start Working at 09:30am Instead of 09:00am. *liberali*. Retrieved from http://liberali.ge/news/view/11809/sajaro-mokheleta-natsili-mushaobas-0900-saatis-natsvlad-0930ze-daitsyebs

Ferrant G, Pesando LM, Nowacka K (2014): Unpaid Care Work: The missing link in the analysis of gender gaps in labor outcomes. *Boulogne Billancourt: OECD Development Center.*

Galay K (2007): *Patterns of Time Use and Happiness in Bhutan: Is there a relationship between the two?* Institute of Developing Economies.

Gicheva D (2013): Working Long Hours and Early Career Outcomes in the High-End Labor Market. *Journal of Labor Economics*, 31(4):785-824.

Golden L (2006): How Long? The Historical, Economic and Cultural Factors behind Working Hours and Overwork. *Research companion to working time and work addiction*, 36-57.

Goldin C (2014): A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4):1091-1119.

Gutek BA, Searle S, Klepa L (1991): Rational Versus Gender Role Explanations for Work-Family Conflict. *Journal of Applied Psychology*, 76(4):560.

Hakkert R (2017): Population Dynamics in Georgia. *An overview based on the 2014 general population census data*, 60-69.

Handley K, Kamal F, Monarch R (2020): Rising Import Tariffs, Falling Export Growth: When Modern Supply Chains Meet Old-Style Protectionism. *NBER Working Paper*, (w26611):

Hanse GD (1993): The Cyclical and Secular Behavior of The Labor Input: Comparing Efficiency Units and Hours Worked. *Journal of Applied Econometrics*, 8(1):71-80.

Hegewisch A, Gornick JC (2011): The Impact of Work-Family Policies on Women's Employment: A Review of Research from OECD countries. *Community, Work & Family*, 14(2):119-138.

Herr JL, Wolfram CD (2012): Work Environment and Opt-Out Rates at Motherhood across High-Education Career Paths. *ILR Review*, 65(4):928-950.

Holly S, Mohnen A (2012): *Impact of Working Hours on Work-Life Balance* (No. 465): DIW Berlin, The German Socio-Economic Panel (SOEP):

Hrdy SB (1999): *Mother Nature: Maternal Instincts and How They Shape the human Species. New York*: Ballantine Books.

ILO (International Labour Organization): (2019): ILOSTAT database.

Jashi C (2005): Gender Economic Issues: The Case of Georgia. Sida.

Kachkachishvili I, Nadaraia K, Rekhviashvili B (2014): Men and Gender Relations in Georgia".

Keene JR, Quadagno J (2004): Predictors of Perceived Work-Family Balance: Gender Difference or Gender Similarity?. *Sociological Perspectives*, 47(1):1-23.

Kingston PW, Nock SL (1987): Time Together among Dual-Earner Couples. *American Sociological Review*, 391-400.

Khunashvili (2014, August 1): Prime Minister is Taking the Initiative to start Working at 9am in the Public Sector. *Georgian Press*. Retrieved from http://georgianpress.ge/com/news/view/8422

Knight KW, Rosa EA, Schor JB (2013): Could Working Less Reduce Pressures on the Environment? A cross-national panel analysis of OECD countries, 1970–2007. *Global Environmental Change*, 23(4):691-700.

National Statistics Office of Georgia, Employment Indicators (2015): Distribution of Employed by Institutional Sector. Retrieved March 28, 2019 from: https://www.geostat.ge/en/modules/categories/128/databases-of-2009-2016-integrated-household-survey-and-2017-households-income-and-expenditure-survey

National Statistics Office of Georgia (2019): *Georgia Multiple Indicator Cluster Survey 2018, Survey Findings Report.* Tbilisi, Georgia: National Statistics Office of Georgia.

Orpen C (1981): Effect of Flexible Working Hours on Employee Satisfaction and Performance: A Field Experiment. *Journal of Applied Psychology*, 66(1):113.

Pieroni L, Salmasi L (2017): The Economic Impact of Smoke-Free Policies on Restaurants, Cafés, and Bars: Panel Data Estimates From European Countries. *Journal of Policy Analysis and Management*, 36(4):853-879.

Rangel MA (2003): Alimony Rights and Intrahousehold Bargaining: Evidence from Brazil. *UCLA CCPR Population Working Papers*.

Risman BJ (1999): *Gender Vertigo: American Families in Transition*. Yale University Press.

Robinson GE, Page RE (1989): Genetic Basis for Division of Labor in an Insect Society. *The genetics of social evolution*, 61-80.

Sayer LC (2016): Trends in Women's and Men's time use, 1965–2012: Back to the Future? *In Gender and couple relationships* (pp. 43-77): Springer, Cham.

Sayer LC (2005): Gender, Time and Inequality: Trends in Women's and Men's Paid Work, Unpaid Work and Free Time. *Social forces*, 84(1):285-303.

Theraulaz G, Bonabeau E, Denuebourg JN (1998): Response Threshold Reinforcements and Division of Labor in Insect Societies. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1393):327-332.

Vlasblom JD, Schippers JJ (2004): Increases in Female Labor Force Participation in Europe: Similarities and Differences. *European Journal of Population/Revue Européenne de Démographie*, 20(4):375-392.

Wasserman M (2015): Hours Constraints, Occupational Choice and Fertility: Evidence from Medical Residents. MIT.

Winett RA, Neale MS, Williams KR (1982): The Effects of Flexible Work Schedules on Urban Families with Young Children: Quasi-Experimental, Ecological Studies. *American Journal of Community Psychology*, 10(1):49.

Yang TT (2020): The Effect of Workplace Pensions on Household Saving: Evidence from a Natural Experiment in Taiwan. *Journal of Risk and Insurance*, 87(1):173-194.