Follow-up Report for "Washing Machines and Gender Roles. A Pilot Study Intervention"*

Camilo García-Jimeno† Ximena Peña‡

June, 2016

Contents

*Peña: xpena@uniandes.edu.co. García-Jimeno: gcamilo@sas.upenn.edu. We are grateful to Laura Martínez for her leadership role in the project, Diana Ricciulli for her outstanding research assistance, and Camila Uribe, María Angélica Navia, Carolyn Finck and Frank DiTraglia for their valuable suggestions and help. We thank Iquartil, Fundacion Exito, and Haceb for their donations, and CEDLAS, Fundacion Social, CAF, and Universidad de los Andes for the financial support for this project.

†University of Pennsylvania and NBER

‡Universidad de los Andes

1 Introduction

In the last decades a large increase in labor market participation of poor women in developing countries has taken place. In contrast with the trends in developed countries, where a rapid introduction of labor-saving appliances and a redistribution of household chores between genders has taken place[1], it has not been accompanied by a significant fall in female home-production hours[2]. The reasons are not well understood, but its implications have been highly unfavorable for women, and as a result, possibly also to children[3]. In particular, the within-household gender asymmetry in total work hours has widened, which may have led to negative effects on women's and children's health, stress, and educational outcomes.

These patterns in labor-market and time-use outcomes are likely to be associated with the lack of social change regarding gender roles observed in poor households in developing countries. Among other reasons, the rigidity of beliefs about gender roles may be related to the (lack of) access to appliances and other home-production technologies, because in contrast, higher income households in developing countries, which have gained access to them, show larger changes in gender role attitudes.

Although the literature has overlooked this phenomenon and has not directly addressed its causes, the broader literature in labor and family economics has emphasized a set of key drivers of the gender distribution of home production hours. A wide wage gender gap implies a larger opportunity cost of forgone labor supply for men relative to women, which may lead households to allocate more home production time to women[4]. Given the well documented large gender wage gap in Colombia, especially among the poor[5], this may account for part of the dynamic. A second explanation may be related to the determinants of distribution factors (Bourguignon, Browning, and Chiappori, 1995), and in particular, to the differential outside options faced by men and women in poor households, which may themselves lead to asymmetries in the distribution of home chores. For example, if women face tighter mating markets, they may have to acquiesce to a highly uneven distribution of costs and benefits in their current relationships. A vast literature has also documented how men are in the short side of most mating markets, and this is likely to be more so among poor individuals in urban areas[6].

In this research project we suggest that beliefs about gender roles are a key additional driving

[1]Burda et al. (2007); Heising (2011).
[2]Bianchi et al. (2000); Shelton and John (1996); Brines (1993); Marini and Shelton (1993).
[3]Olson (1981); Pfeiffer, Gloyd and Li (2001).
[4]Braun et al. (2008); Lewin-Epstein and Stier (2007).
[5]Badel and Peña (2010); Fernández (2006).
[6]Botticini and Siow (2003); Choo and Siow (2006); Iyigun and Walsh (2007); Peña (2006).

force of the patterns described above, not only to account for the large gender asymmetries in home production but also for how inelastic these have been among poor households in the last decades. This despite important labor market and cultural changes. Beliefs about gender roles can be broadly understood as culturally transmitted beliefs about the appropriate activities, choices and behaviors of the different genders. These are shaped by many forces over time, and the literature in economics has already explored some technological forces explaining variation in gender roles across societies[7].

Alesina et al. (2013), for example, empirically test Ester Boserup's (1970) hypothesis according to which variation in early gender specialization in production stemming from the adoption of the plow in agriculture led to divergent beliefs about gender roles across societies. They find persistent differences in outcomes related to gender roles that are correlated with historical plow usage. Similarly, it is likely that the subsequent specialization in home-production of women has reinforced gender roles consistent with it. We believe that the introduction of a labor-saving technology such as a laundry machine, in addition to training on the appropriate use of the machine, by altering relative costs of effort in a similar way in which the plow did on agricultural production, may also lead to a dynamic of evolving views about gender roles. Alesina et al.'s results suggest that gender roles are responsive to technological changes, but also highly persistent. Evidence of the effects of the introduction of home appliances in the U.S. and Europe, nevertheless, do suggest that views about the appropriate social roles of women may change at much higher frequencies than those suggested by Boserup, but may also make women's labor supply highly wage-inelastic[8].

Although beliefs about gender roles encompass many dimensions, for the purpose of our project here we propose to conceptualize gender roles as beliefs (shared by both men and women) about the comparative advantage of women in home production. This way of thinking about gender roles will allow us to explore very precise short and long-run channels through which a technological shock to households such as the provision of a laundry machine will alter household equilibrium time allocation, investment, and home production choices. In particular, we will emphasize that the existence of such beliefs has both direct and indirect effects on the equilibrium distribution of time across activities between genders. A direct effect arises simply because a household that considers men's home production hours less effective should directly allocate less of those to the man. An indirect effect arises because such an allocation may reduce women's distribution factors. For example, it may reduce market hours and possibly even the productivity of women in the labor market (due to the asymmetry in total hours), translating into lower influence or power of women in

[7]See for example, Alesina et al. (2013) and Greenwood, Sheshadri, and Yorukoglu (2005).

[8]Fernandez (2007); Fernandez and Sevilla-Sanz (2006).

the household decision-making process. Moreover, we will propose the implementation of an RCT that will generate random variation in the home production technology through the allocation of laundry machines. Key about this intervention is that it should not be expected to alter distribution factors.

There is a well established literature studying household behavior that estimates preference and technological parameters in a collective framework[9] where Pareto efficiency within the household is assumed. Most of the empirical literature has relied on household surveys and has exploited variation in distribution factors for identification[10]. As a result, the literature has focused on estimation that holds fixed the Pareto frontier and varies households' location along it. Nevertheless, in any non-experimental setting variation in distribution factors is correlated with technological shocks or prices, requiring strong functional form assumptions for identification. In contrast, we propose exploiting random variation in the location of the Pareto frontier itself holding bargaining power fixed, while additionally allowing for social norms regarding gender roles to endogenously respond to the changes induced by the intervention. Furthermore, under this framework we will be able to allow for public goods within the household, such as child quality. We are unaware of any other RCT proposing such an intervention in the context of the family, nor of any empirical work attempting to trace the links between technology, labor market responses, and cultural norms.

This document describes the impact analysis for the first follow up of the "Washing Machines and Gender Roles" pilot intervention.

2 Background and Context

2.1 Sample and Neighborhood Characteristics

Below we present maps of Bogotá showing the location in the city of our treated neighborhoods and their surroundings. The first map shows all of Bogotá, and the second zooms into the area outlined in red, pointing out each of the neighborhoods. These are not representative of the city, or the poor areas of the city. We chose them as examples of poor and low income areas. We also chose them not to be close to each other to capture differences in the conditions in, say, neighborhoods like Barrio Inglés where household have a good supply of public services, and others like Cazucá where the living conditions are precarious.

[9] Blundel, Chiappori, Meghir (2005); Browning and Chiappori (1998); Chiappori (1997); Udry (1996).
[10] Apps and Rees (1988); Rapoport, Sofer, and Solaz (2009); Couprie (2007); van Klaveren et al. (2008).

Figure 1: Map of Bogotá

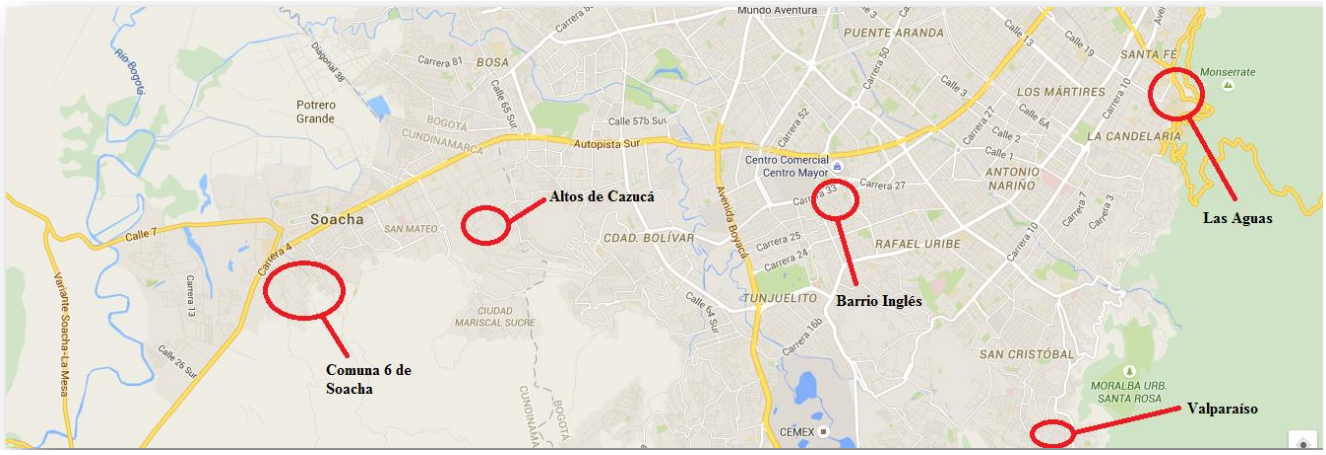

Figure 2: Map of Southern Bogotá and Intervention Areas

Our sample comes from 5 neighborhoods in Bogotá and its vicinity; Las Aguas, Valparaíso and Barrio Inglés and in Bogotá, while Altos de Cazucá and Comuna 6 are both in Soacha. Las Aguas is located in downtown Bogotá. In this neighborhood we also included some families from La Paz, which is right next to it, but above the 'Circunvalar' Avenue. This is an extremely dangerous neighborhood where the police does not regularly enter. Valparaiso, which comprises Juan Rey, Libertadores and La Belleza, is located high in a mountain and although the roads are paved, mobility in the area is far from easy. Barrio Inglés is located behind a big mall, and is the easiest to access. The population in this neighborhood rotates the most. It's highly populated by recent migrants from the Pacific Coast. Altos de Cazucá also has a high percentage of afro population, this time mostly forcefully displaced. In fact, this neighborhood is known as a big recipient of displaced population. Comuna 6 of Soacha is also a neighborhood with a high rate of displaced population. The Comuna includes two different neighborhoods, Santa Isabel/La Cristalina and Florida Alta, and occupies a mountain slope. Families displaced 25 or 30 years ago live down the slope. As new families have arrived in the last decades, they have built their dwellings in ever-higher locations along the slope. The most recent migrants live in the mountain top. Public goods and services are very scarce and their coverage falls as one moves up the mountain slope because the new settlements are illegal. Therefore, there is no running water and no paved roads.

The families in our sample live in the first three socioeconomic strata [11] . However, there are some differences between the neighborhoods, which we illustrate in Table 1. The richest neighborhood is Barrio Inglés, where most of the population is strata 3. One family in Comuna 6 of Soacha is strata 3 and there are no strata 3 families in rest of the neighborhoods. The two poorest neighborhoods, with the bulk of their families classified in strata 1, are Altos de Cazucá and Comuna 6 in Soacha.

There are differences in formal public service availability. However, there is some informality in public service access. Some households without a formal connection to the electric grid establish illegal connections and do not pay for the service. Electricity coverage is high: most households in our sample have electricity, except for 15% of households in Cazucá. Aqueduct coverage is also high, reaching coverage rates above 80%, except in Cazucá where only 53% households have running water. Table 1 also illustrates large differences in sewage coverage across the neighborhoods in our sample.

[11] Socio-economic strata, strata in what follows, is a local measure of poverty, that takes into account the availability of public services and infrastructure. Strata classify dwellings in six groups, from poorest to richest.

Neighborhood Characteristics

VARIABLES	Cazucá	Inglés	Las Aguas	Divino Niño	Valparaiso	Altos de Florida	Neighborhood Differences
Total observations N=194	**N1=32**	**N2=25**	**N3=27**	**N4=45**	**N5=38**	**N6=27**	
Biparental households (%)							
Yes	84,38	64,00	59,26	88,89	86,84	96,30	Pearson chi2(5)=20,33
No	15,62	36,00	40,74	11,11	13,16	3,70	P-Value=0,001***
Average household size	3,97 (1,17)	3,48 (1,16)	3,93 (1,14)	4,00 (1,46)	3,97 (1,50)	4,44 (1,45)	P-Value=0,1672
Average number of children	2,06 (0,98)	1,64 (0,81)	2,19 (1,36)	1,71 (1,20)	2,05 (1,47)	2,30 (1,38)	P-Value=0,3382
Socioeconomic strata (%)							
1	65,63	0	40,74	71,11	18,42	88,89	Pearson chi2(10)=226,44
2	34,38	4,00	59,26	28,89	81,58	7,41	P-Value=0,000***
3	0	96,00	0	0	0	3,70	
Natural Gas (%)							
Yes	50,00	96,00	85,19	75,56	81,58	29,63	Pearson chi2(5)=41,22
No	50,00	4,00	14,81	24,44	18,42	70,37	P-Value=0,000***
Telephone (%)							
Yes	6,25	8,00	44,44	13,33	28,95	3,70	Pearson chi2(5)=25,46
No	93,75	92,00	55,56	86,67	71,05	96,30	P-Value=0,000***
Waste collection (%)							
Yes	46,88	100,00	81,48	75,56	52,63	62,96	Pearson chi2(5)=26,42
No	53,12	0	18,52	24,44	47,37	37,04	P-Value=0.000***
Electricity (%)							
Yes	84,38	100,00	100,00	100,00	100,00	100,00	Pearson chi2(5)=25,98
No	15,62	0	0	0	0	0	P-Value=0,000***
Aqueduct (%)							
Yes	53,12	100,00	85,19	93,33	97,37	74,07	Pearson chi2(5)=38,44
No	46,88	0	14,81	6,67	2,63	25,93	P-Value=0,000***
Sewage (%)							
Yes	28,12	100,00	88,89	73,33	68,42	48,15	Pearson chi2(5)=45,23
No	71,88	0	11,11	26,67	31,58	51,85	P-Value=0,000***
Home Tenancy (%)							
Owned, completely paid	37,50	0	7,41	13,33	7,89	37,04	
Owned, still paying it	56,25	0	0	11,11	7,89	25,93	Pearson chi2(20)=103,66
Leased or subleased	3,13	92,00	81,48	75,56	71,05	37,04	P-Value=0,000***
Under usufruct or other type of tenancy	3,13	4,00	7,41	0	10,53	0	
Unlawfully occupied	0	4,00	3,70	0	2,63	0	

Note: The table presents the distribution of the main household characteristics by neighborhood at baseline. The last column reports pearson chi2 statistics for a joint test of differences across neighborhoods and their associated p-value.

Table 1: Neighborhood Characteristics at Baseline

2.2 Household Characteristics at Baseline

In Table 2 we present household characteristics at baseline, computed on our final working sample. Around 80% of households in our sample are biparental. They also have low income levels: average per capita income is below the Colombian poverty line. Less than 60% of mothers work in the labor market -earning on average nearly USD $130 per month-, 6% are looking for a job, and 36% are "inactive", i.e. they perform household work. 90% of fathers work in the labor market -earning USD $205-, 1.7% are looking for a job, and 7% are mainly involved in domestic activities. There are no statistically significant differences between the treatment and control groups in these variables at standard significance levels.

Members of the families in our sample work many hours per week, both in the labor market and in household production. Women perform most household chores and take care of infants, disabled individuals and the elderly. Mothers in our sample work an average of 22 hours per week in the labor market, and devote 17 hours per week on average to household chores excluding laundry. While at baseline mothers spent 5.5 hours per week doing laundry exclusively, this figure is around 4 hours at follow-up. Mothers also spend 47 hours per week on childcare and 78 on leisure (that is, 11 daily hours of leisure including sleep). . The average father in the sample works 44 hours per week in the labor market, devotes almost 5 hours to home production including laundry, spends 23 hours on childcare, and 98 hours on leisure. This implies men enjoy 14 hours of leisure a day on average, including sleep. These figures show major differences in the time use between men and women in our sample; women enjoy very little leisure and sleep very few hours.

Perceptions about gender roles and female empowerment are key variables in our study. The answers at baseline suggest that many mothers are dissatisfied with their lives within the household. For example, 35% of mother are fed-up from being home-makers, and more than half feel their work as home-makers received little recognition. There are some marginal differences in the questions inquiring about family environment, but these differences are not systematic: mothers in treated households seldom feel their family makes them feel like house work is their sole responsibility, while they also often feel they would like their partners to help them more in household chores.

We also find that female empowerment in the treatment and control groups are equal. There are no significant differences in these dimensions, except our the gender roles index (described below), which was balanced within the initial simple, but now shows a slight imbalance: the treatment group seems to have more progressive gender role perceptions in the final working simple.

There are no differences between treatment and control households (those who won and lost

Household Characteristics at Baseline

Variables	Control	Treatment	Differences Control-Treatment	
Total observations N=145	**N=67**	**N=78**	**Difference[§]**	**P-value**
Biparental households(%)	77,61	82,05	0,014 (0,072)	0,847
Total household labor income n0=67 n1=77	665,791 (389084,7)	709,134,4 (336228)	50744,85 (70489,93)	0,473
Employed mothers (%) n0=66 n1=77	56,06	59,74	0,119 (0,086)	0,169
Mothers looking for a job (%) n0=66 n1=77	7,58	5,19	-0,022 (0,045)	0,625
Inactive mothers (%) n0=66 n1=77	37,88	35,06	-0,122 (0,083)	0,143
Mothers' monthly wage (COP) n0=35 n1=51	357,886 (282289,4)	393,549 (248462,4)	83,048,66 (58856,53)	0,162
Employed fathers (%) n0=53 n1=65	90,57	92,31	0,034 (0,055)	0,537
Fathers looking for a job (%) n0=53 n1=65	1,89	1,54	-0,012 (0,027)	0,641
Inactive fathers (%) n0=53 n1=65	7,55	6,15	-0,022 (0,049)	0,657
Fathers' monthly wage (COP) **n0=44 n1=56**	650,136,4 (253437,9)	592,464,3 (27454)	-66,564,92 (56886,77)	0,245
Gender Roles Index n0=67 n1=77	3,453 (0,576)	3,661 (0,718)	0,246 (0,117)	0,036**
Mothers' Time use				
Laundry time	5,47 (4,15)	5,89 (4,08)	-0,26 (0,68)	0,697
Childcare time	55,11 (27,21)	53,05 (24,97)	0,75 (4,62)	0,871
Home production time - without laundry	18,70 (13,46)	14,81 (8,59)	-2,81 (2,06)	0,174
Leisure time	70,27 (24,30)	74,50 (24,42)	-2,71 (4,15)	0,514
Work time in labor market	18,42 (19,56)	19,73 (19,60)	5,04 (3,42)	0,144
Female Empowerment				
How do you and your partner organize the income you receive?				
I manage all the money and give my partner his part (%)	28,36	26,92		
My partner manages all the money and gives me my part (%)	26,87	12,82		
We share an account and we both take money from there when we need it (%)	32,84	44,87	0,158 (0,228)	0,489
We share an account with part of the money and keep the rest separate (%)	4,48	5,13		
We keep our money separate (%)	7,46	10,26		
Which one of these statements suits best the way you distribute household chores between you and your partner? n0=66 n1=78				
I work a lot more than I'm supposed to (%)	48,48	37,18	0,048 (0,181)	0,79
I work a bit more than I'm supposed to (%)	13,64	21,79		
I approximately work what I'm supposed to (%)	33,33	39,74		
I work a bit less than I'm supposed to (%)	3,03	0		
I work a lot less than I'm supposed to (%)	1,52	1,28		
Family Environment				
In general I'm fed-up of being a home-maker. n0=66 n1=78				
Never (%)	39,39	61,54		
Rarely (%)	30,30	10,26		
Sometimes (%)	15,15	19,23	-0,253 (0,224)	0,261
Frequently (%)	7,58	6,41		
Most of the time(%)	7,58	2,56		
My family usually have excessive demands from me. n0=66 n1=77				
Never (%)	36,36	53,25		
Rarely (%)	34,85	25,97		
Sometimes (%)	18,18	7,79	-0,307 (0,205)	0,137
Frequently (%)	7,58	10,39		
Most of the time(%)	3,03	2,60		
My family makes me feel like taking care of the house is my sole responsibility. n0=66 n1=78				
Never (%)	24,24	43,59		
Rarely (%)	25,76	14,10		
Sometimes (%)	19,70	14,10	-0,463 (0,252)	0,068*
Frequently (%)	21,21	15,38		
Most of the time(%)	9,09	12,82		
I would like my partner to help more in household chores. n0=66 n1=77				
Never (%)	21,21	36,36		
Rarely (%)	4,55	14,29		
Sometimes (%)	27,27	7,79	-0,540 (0,276)	0,052*
Frequently (%)	28,79	25,97		
Most of the time(%)	18,18	15,58		
I like the environment in my home. n0=66 n1=78				
Never (%)	10,61	21,79		
Rarely (%)	6,06	10,26		
Sometimes (%)	10,61	11,54	-0,274 (0,276)	0,323
Frequently (%)	30,30	25,64		
Most of the time(%)	42,42	30,77		
My job as housewife is boring. n0=66 n1=78				
Never (%)	36,36	44,87		
Rarely (%)	27,27	15,38		
Sometimes (%)	16,67	19,23	0,036 (0,263)	0,892
Frequently (%)	10,61	11,54		
Most of the time(%)	9,09	8,97		

* significant at 10%, ** significant at 5%, *** significant at 1%. The sample used in these calculations is the final working sample.

Table 2: Household Outcomes and Characteristics at Baseline

the raffle), which suggests the randomization exercise was adequate. Based on the baseline analysis, we identified a subset of variables that we include as controls in our impact analysis.

3 Theoretical Background

3.1 Basic Model

The detailed nature of the data we collected and will collect in a subsequent round, together with the variation generated by the intervention, will allow us to identify and estimate key elements of a collective model of household behavior with home production and beliefs about gender roles. Here we outline the basic features of the model. Assume a given household has both a husband and a wife $i \in \{h, w\}$. Preferences for each can be represented by a utility function $U_i(c_i, l_i, K)$, where c_i is i's individual consumption, l_i is i's leisure, and K is a public good for the household, which we will think of as children quality[12]. From the survey, we will have a vector proxies for K (health, education, etc.).

We will further assume that $K(H, I, b_h, b_w)$ depends on child investment inputs of parental time b_h and b_w, on expenditure in children I, and on the quality of home care H (cleanliness, food quality, etc.). We will further allow for H to depend on home production time and potentially the availability of a labor-saving technology such as a washing machine. Specifically, assume that

$$H = \left[\alpha T^{\rho} + (1-\alpha)\left(\theta h_h + h_w\right)^{\rho}\right]^{\frac{1}{\rho}} \tag{1}$$

Here $T \in \{0, 1\}$ denotes our Treatment, namely the availability of a washing machine at home, and h_h and h_w is the time allocated to home production by husband and wife. For ρ close to 1, the above technology implies that the washing machine and time are substitutes in production. Our production function (1) also incorporates our way of modeling gender roles. We will assume that each hour of male home production is equivalent to only $\theta \in [0, 1]$ hours of female home production. We will further assume that all members of the household agree on the value of θ, but that their belief may actually be incorrect. Households with a low θ are those where beliefs about gender roles are highly asymmetric[13]. Finally, we follow the

[12] Although K is a public good, notice that we will allow for $\frac{\partial U_h}{\partial K} \neq \frac{\partial U_w}{\partial K}$, so that one gender may perceive higher marginal benefits from the public good than the other.

[13] A collective framework of household decision-making entails efficiency, so that equilibrium choices locate the household in the Pareto frontier. Notice that in a setting where households may have wrong beliefs about θ, equilibrium choices will in general be inefficient. Thus, our model can account for equilibrium outcomes that may be far from the socially desired ones, despite our model remaining within the standard collective framework.

standard collective models (Chiappori, 1988) by assuming the household solves the following problem:

$$max_{\{c_i,l_i,m_i,b_i,h_i,I\}_{i=1,2}} \quad \lambda(\mathbf{z})U_h(c_i,l_i,K) + (1-\lambda(\mathbf{z}))U_w(c_i,l_i,K)$$

subject to equation (1) and:

$$T = m_i + l_i + b_i + h_i \qquad i \in \{h,w\}$$

$$K = K(H,I,b_h,b_w)$$

$$w_h m_h + w_w m_w + y = I + c_h + c_w$$

Here m_h and m_w are time in the labor market, and y is other non-labor income. $\lambda(\mathbf{z})$ is a Pareto weight for the husband, which will depend on a vector of distribution factors $\mathbf{z}$, such as the relative remunerations of husband and wife and any other characteristics that may matter for bargaining between spouses. Although we will be assuming the household solves a static problem[14], we will allow for the post-treatment λ's to change in response to the changes in the time distribution due to the intervention.

Assuming that preferences are separable between K and (c_i, l_i), and given that we will observe children quality, the full time distribution, distribution factors and the treatment status, we will be able to identify the key parameters of this model. We omit that discussion here for lack of space.

The parameter θ is allowed to depend on own treatment (possibly through private household learning), and the household's neighborhood treatment intensity (possibly through social learning). Separately identifying the roles of private vs social learning will rely on the random assignment of both own treatment and neighborhood's treatment intensity. Neighborhoods where intensity is higher should observe smaller within-neighborhood variation in changes in outcomes compared to neighborhoods with smaller treatment intensities.

3.2 Mechanisms

To complement the model above, in Figure 3 we present the mechanisms through which we think the program may generate effects. The washing machine and its use can have a direct effect on the consumption of public services. Our previous research suggests that washing clothes using a laundry machine increases electricity usage and decreases water usage

[14]In this setting this is not a very restrictive assumption since we are being agnostic about the bargaining process itself, and the Pareto weights could implicitly be functions of the min max payoffs or outside options of each spouse in a repeated game.

as compared to washing by hand. This is portrayed in the top left corner of Figure 3. The purple line shows this is a household-level expected impact, and we add the sign of the impact we expect. Despite the increase in electricity usage, anecdotal evidence suggested that the net effect has been to decrease public service expenditures, which can increase household savings.

The second direct impact of the washing machine takes place over the time devoted to washing. We present this channel separately for mothers (red) and fathers (blue). Because it is mothers who do most laundry at baseline, we expect their time devote to washing to decrease. We have no prior about the impact on the father's laundry time. Time saved in laundry could be used by mothers in many ways: childcare, leisure, home production, or market work. If the time freed by the washing machine generates an increase in time devoted to childcare (by either parent), this could translate into improved child development outcomes in the medium run. In fact, although our follow-up survey collected information on these outcomes, here we are not analyzing them. We will wait until the second follow-up 18 months after the intervention to explore this channel.

If the time saved is allocated to leisure, stress levels of mothers could decrease. If time doing laundry is substituted for other home production activities, women could devote more time to market work. In this case, they may generate additional income and possibly increase their empowerment levels. Increased income and empowerment could in turn affect shared gender roles within the family. Gender roles change slowly. As such, we do not expect to see changes in the short-run (this follow-up), but maybe in the second follow-up. In this report we present a basic analysis of the impact of the intervention on gender roles, but deeper analysis will come later in the second follow-up.

Finally, the training session may have some direct impacts as well. First, it can improve usage of the washing machine. Second, it can decrease water consumption. Finally, because it makes mothers and fathers think again about gender roles, it may stir some change in this dimension.

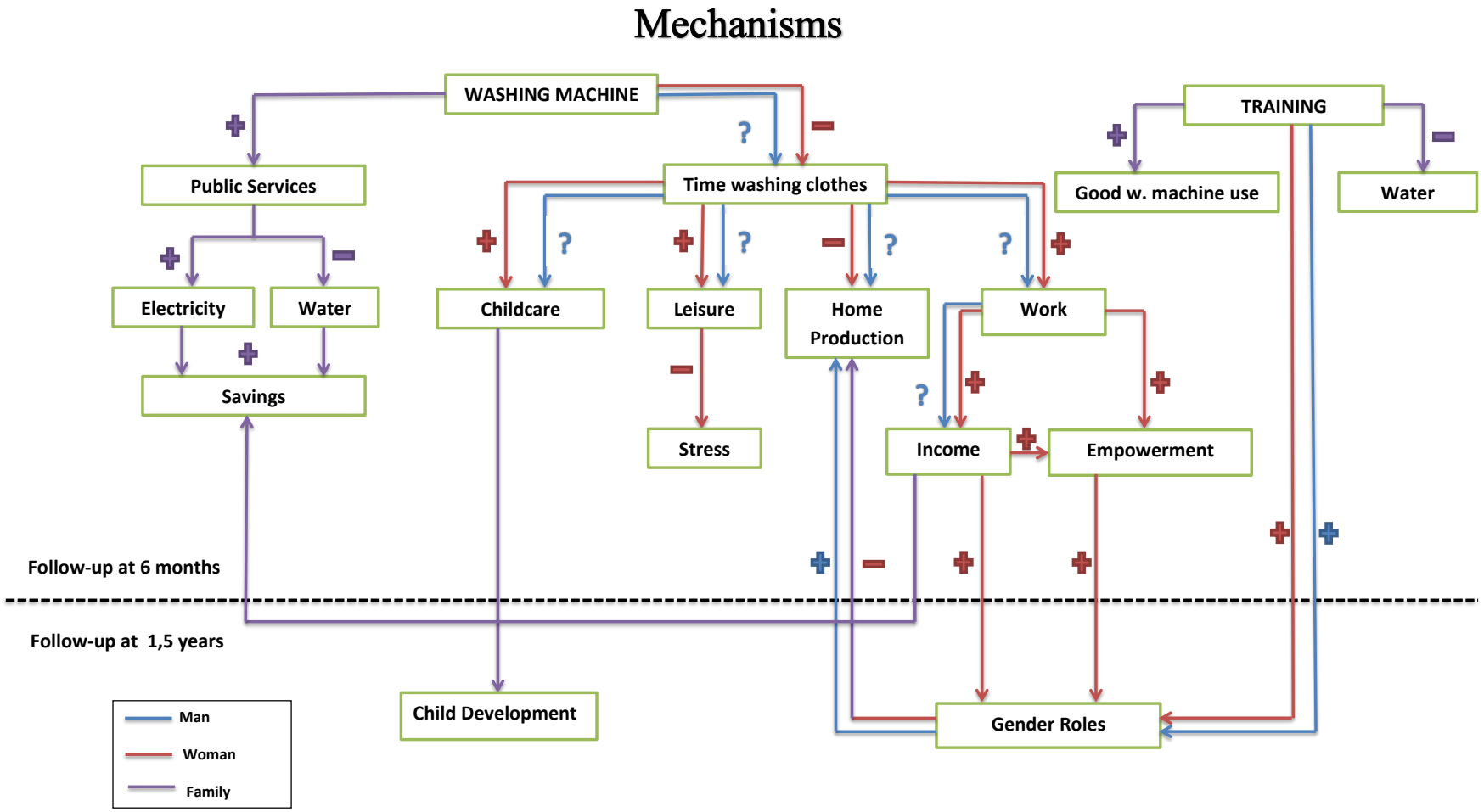

Figure 3: Mechanisms of the intervention

4 Intervention Design and Implementation

4.1 Motivation

The project was motivated by several observations. There is strong demand for labor saving technologies for home production as manifested in the emergence of local laundry machine rental markets. There are very few to no laundromats in Colombia, especially in low income neighborhoods. These households face binding cash and credit constraints preventing them from acquiring appliances such as a laundry machine. The distribution of home production efforts across genders is highly asymmetric and has remained so despite the secular increase in female labor market participation. Finally, both men and women share gender role perceptions that regard home work by men to be inappropriate or unproductive, and that consider men should have stronger claims to authority and decision-making within the family. We intend to use the intervention to explore the relationship between gender role perceptions, family decision-making, and family outcomes. Our main interests are two. First, to asses if and how the introduction of this labor-saving technology alters the distribution and quantity of time allocated to home activities and other health and human capital outcomes of family members. Second, to explore how these changes are related to perceptions about gender roles and the changes in these perceptions induced by the intervention.

4.2 Intervention

Women do most of the household chores, so labor-reducing home appliances benefit them especially. In the case of low income families in developing countries there is a missing market. Washing machines are costly; the price of those used in this project is nearly US$435, while average family income is almost half of that, US$235. Families have little access to credit, and the formal credit they have access to demands very high interest rates. For instance, in June 2016 the rates at which these families can borrow is 30% (the maximum legal rate), while the annual inflation rate is less than 8%. Access to credit is, for the most part, infeasible. Poor families cannot afford a washing machine either. These families also have highly skewed gender role perceptions, making it even less likely they will invest in labor-reducing appliances.

The project at this stage is a proof-of-concept. We eliminate the financial barrier from the families in our sample, to carefully measure the benefits of having a washing machine in the dimensions mentioned above. The intervention consists of selling washing machines to low income families at an almost symbolic price: between 5 to 10% of the market price.

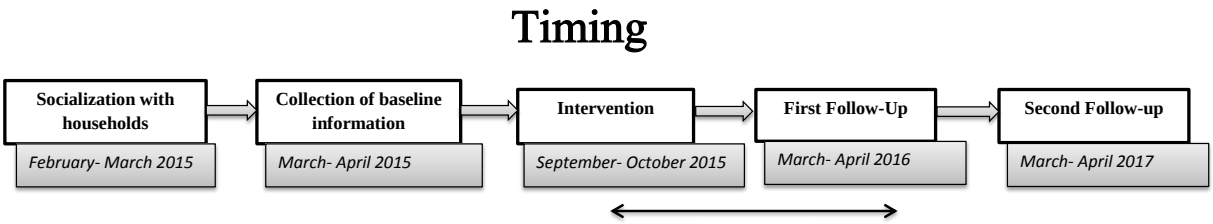

Figure 4: Timing of the Project

The washing machines are new and have a guarantee. We then install the appliances in each treated household, and provide a training session on its adequate use and on tips for how to save by recycling water. We explicitly invited males to participate in the training, which lasted about 20 minutes and was designed following the motivational interviewing literature (see Rollnick, Miller, and Butler (2008)). We made emphasis on engaging male adult household members in using the laundry machine. Instead of telling subjects what to do, this literature emphasizes that training is much more likely to be successful in changing receptiveness and stances of subjects when it leverages their own motivation by leading them to discover and explore by themselves the workings of the treatment. We followed this approach with a protocol designed by a psychologist team member.

The timing of the data collection and the intervention are presented in Figure 4 . We first met with the families in February and March 2015 and then collected the baseline information between March and April of the same year. The laundry machines raffles, installation and training sessions (one per neighborhood), took place in September and October 2015. The first follow-up was collected in March and April 2016, six months after the intervention. We plan to collect a second follow-up 18 months after the intervention, but still need to complete the funding for this second round of data collection.

4.3 Implementation

We first chose foundations with strong presence and links to the community in each of the neighborhoods to work with. We partnered with Fundación Semilla y Fruto in Comuna 6 in Soacha, Fundación Catalina Muñoz in Altos de Cazucá, Fundación Ángel de Luz in Barrio Inglés and Valparaíso, and with the head of the Junta de Acción Comunal (local community leader) in Las Aguas. All of these foundations played a key role in allowing us to make initial contacts within the communities and assembling a list of families fulfilling the project's main eligibility criteria: bi parental families with children under 8 years of age who did not own a washing machine. We devise creative ways of finding out whether the families in fact had no

washing machine (for example, we asked the children, who attend recreation sessions at the foundations' headquarters, to draw the appliances at home).

With the family rosters, the foundations made formal invitations to expositional meetings to all families. There we showed them the questionnaires they would have to answer, and talked to them about informed consent and the timing and characteristics of the project. We indicated that those participating in the study would benefit from a raffle of home appliances at very low cost. This, to prevent families from changing their answers to whatever they believed we might be looking for.

We then collected the baseline survey on the willingly participating families, and subsequently had a public raffle where they were informed that the appliance for winning households would be a washing machine. In Las Aguas we gave the washing machines for free, because we were working with a community leader instead of a foundation. We charged for them in all other neighborhoods, through the foundations themselves. The revenue raise from these raffles was donated to the foundations, which in our understanding has been used for running expenses. Because the charged fee was minimal, the revenue itself was minimal.

The day of the raffle we scheduled appointments for machine instalations and for the the training sessions. However, because families were very eager, many of them installed them before the scheduled appointment. The training sessions were performed by young men who have a similar socioeconomic background to our household sample, and are now attending college. Therefore, the family members could see them as relatable. During the training session families were taught how to recycle water from the washing cycle. We also tried to engage male adult household members in learning how to use the machine.

One important issue worth mentioning is that households could have profitted from the arbitrage opportunity implied by the heavily discounted appliance, making a profit of over one monthly household income. We monitored this possibility carefully - in the follow-up visit, the data collectors had to take a picture of the installed washing machine-. Nevertheless, only 2 household out of a total of 97 sold the machines. These were very particular cases of families with children in prision in desperate need of cash. We interpret this as evidence that low income families do need and appreciate the washing machines but cannot afford them.

5 Evaluation Design

Our experimental design is simple. We use a random lottery to allocate laundry machines to a sample of poor households in six neighborhoods of Bogota, Colombia. In addition to

the laundry machine itself, treated households also received training on its use. The training session, as mentioned above, was designed following the motivational interviewing literature to invite spouses to do laundry. We now track these households and a similar group of control households from the same neighborhoods, collecting baseline and follow-up household survey information on a battery of outcomes.

Our treatment both alters the home-production technology, and generates information to the household. Such information can directly impact views regarding appropriate gender roles within the household. The communication of this information can also reach beyond the household. This possibility is particularly relevant in the context of our intervention because the sample of households available for our study is located in six relatively cohesive neighborhoods, where neighbors know each other and interact on a daily basis (for example within foundations with local presence that provide childcare and nutrition services). In short, our experimental design must deal with the possibility of interference (spillovers).

5.1 Neighborhoods and Households

We chose six spatially disconnected poor neighborhoods in Bogota. We are confident no interference across neighborhoods is possible. Identification of potential spillovers required us to randomize the intensity of treatment offer across neighborhoods. Our budget allowed us to provide around 100 laundry machines, which we distributed across the six neighborhoods according to the randomly determined intensities. We do not have detailed population information for these neighborhoods, but we do have accurate information on the number of blocks per neighborhood. This allowed us to define treatment offer intensity relative to the number of blocks per neighborhood as $\kappa_b = W_b/B_b$, where κ_b denotes intensity in neighborhood b, W_b denotes number of treatment offers, and B_b denotes the number of blocks in the neighborhood. The six intensities were chosen to be evenly spaced and satisfy the budget constraint

$$100 \approx \sum_{b=1}^{6} \kappa_b B_b \tag{2}$$

In a second step, given the randomly chosen treatment intensities we randomly assigned the treatment offers in a public lottery event taking place in each of the neighborhoods. Table 3 describes our sample[15]. The table illustrates the wide range of intensities, ranging from 0.1 households per block in *Barrio Ingles* to 1 household per block in *Las Aguas*. In fact, in both of these neighborhoods the number of treated households is 9, but *Barrio Ingles* has

[15]To maintain homogeneity within a neighborhood we split Comuna 6 de Soacha into two different neighborhoods: Divino Niño and Florida Alta

Neighborhood	Number of Blocks	Intensity	Sample Size	Treatment Offers	Controls
Inglés	92	0.1	25	9	16
Valparaiso	62	0.4	38	25	13
Florida alta	21	0.6	27	12	15
Divino Niño	26	0.7	45	18	27
Cazucá	30	0.8	32	24	8
Las Aguas	9	1.0	27	9	18
TOTAL			194	97	97

Table 3: Intervention Sample. The table reports the neighborhoods, treatment intensities, sample sizes, and resulting number of treatment offers and controls in each neighborhood.

ten times as many blocks as *Las Aguas*. The combined variation in neighborhood sizes and treatment intensities will increase our ability to detect any spillovers despite the sample size of our intervention.

All offered households took up the laundry machine[16]. This is no surprise; the value of this asset is large relative to the average wealth of our sample of households. Nevertheless, we must emphasize that each treated household was charged between US$35 and US$70 for the washing machine, depending on the neighborhood. This is 15-30% of the monthly minimum wage for urban households in Colombia, and implies a large subsidy for the appliance, which has a market price of US$430, more than twice the monthly minimum wage. Thus, we believe it is reasonable to assume all control households would have taken up if offered. As such, we will interpret our difference-of-means estimates as treatment effects on the treated (TOT)[17]. 85% of the households in our sample are bi-parental, this is, a couple in a long term relationship shares the living space. The remainder are mono-parental. The household head in all mono-parental households in our sample is a woman. The inclusion of a subset of mono-parental households will allow us to compare the responses of households where intra-family bargaining must take place, to the responses of households where such bargaining is not present. Section 5.2 below will describe the sample more closely using the baseline information collected in the spring of 2015.

5.2 Data

In this section we briefly describe the information from the baseline and follow-up surveys that we will use in our analysis. We collected information at the household level. Time

[16]However, there is one household who took up the treatment whose washing machine was stolen

[17]Our original sample of households did include a few which did not show up for the lottery, and which we were subsequently unable to contact. These households were eliminated from the study prior to the lottery drawings.

Attrition

Neighborhood	Baseline households	Attrition	Follow-up households	Attrition Rate
Cazucá	32	0	32	0.00%
Barrio Inglés	25	2	23	8.00%
Las Aguas	27	1	26	3.70%
Divino Niño	45	3	42	6.67%
Valparaiso	38	3	35	7.89%
Florida	27	0	27	0.00%
Treatment	97	3	94	3.09%
Control	97	6	91	6.19%
Total	194	9	185	4.64%

Note: Attrition households are those which we were unable to contact for the follow-up survey.

Table 4: Attrition by Neighborhood and Treatment Status

use information for the mother is available both at baseline and follow-up, whereas father's time use information is only available at follow-up. We are following the Pre-Analisys Plan directly regarding data usage.

At baseline, 80% of the households in our sample were bi-parental. Average family size is four, and average number of children per household is two. These three basic characteristics are perfectly balanced between our treatment and control groups. Similarly, household wealth[18] is balanced across treatment and control groups, and almost no household in our sample owns its dwelling.

5.2.1 Sample size, Attrition and Balance

Our attrition rate was low, less than 5%, considering that the follow-up survey was conducted one year after the baseline. From a total of 194 households at baseline, with equal numbers of treatments and controls, of 9 households were not found in the follow-up round: 3 treatments and 6 controls. Given these low numbers, Table 4 suggests that attrition rates were similar between the treatment and control groups and thus attrition does not appear to be related to treatment status. Also, attrition ranged between 0 and 8% across neighborhoods.

[18]We compute household wealth following the ELCA survey methodology, as the first principal component of gas, phone, trash collection, electricity, aqueduct, and sewage service access, ownership of iron, fridge, oven, blender, microwave, water heater, electric shower, TV, radio, sound system, DVD, computer, cellphone, internet access, and cellphone data usage of any family member.

Family composition Transitions

Household Type	Baseline (post-attrition)	Transition Type	T	C	Follow-up (post-attrition)	Working Sample
Biparental	151	Bi->Bi	65	55	123	120
		Bi->Mono	13	18		
Monoparental	34	Mono->Mono	14	17	62	31
		Mono->Bi	2	1		
Total					185	151

Note: Here working sample excludes all households who experienced a family composition transition (either biparental to monoparental or monoparental to biparental) between baseline and follow-up.

Table 5: Transitions in Family Composition

There was a high percentage of changes in family composition between baseline and follow-up. In Table 5, we exclude attritioned households and present the transitions between mono and bi parental status of the households in our sample, discriminating between treatment and control groups. At baseline, we had 151 bi parental households and 34 mono parental, given the emphasis of the intervention in studying the evolution of gender roles. At follow-up, we have 123 bi parental households: 120 are households that were bi parental al baseline and remained bi parental, and 3 were mono parental at baseline and are now bi parental. Of those 120 bi parental households that remained bi parental, 65 were in the treatment groups and 55 were in the control group. In the analysis to follow we will use the families that did not experience transitions in their composition, and so the working sample will be 120 bi parental households, since we exclude the 3 bi parental households at follow-up that were mono parental at baseline. For the case of mono parental households, at follow-up we have 62, which is higher than the 34 at baseline. This is because 33 bi parental households became mono parental, 13 treatments and 18 controls. Excluding those households that faced transitions in the composition, we end up with 31 mono parental households in our working sample. Therefore, the total working sample is 151 households.

We finally turn to the issue of non-compliance, which we illustrate in Table 6. Of the 151 households, we had 1 non-complier in the treatment group, that is, a family that reported having had the laundry machine stolen. We also had five non-compliers in the control group, that is households that did not win the lottery yet purchased the laundry machine between baseline and follow-up. After eliminating these observations, we have a final working sample

Non-Complier Households

	Baseline (post-attrition, post family transitions)	Non-compliers (post-attrition, post family transitions)	Final Working Sample
Treatment	79	1	78
Control	72	5	67
Total	151	6	145

Note: The final working sample excludes all households who experienced a family composition transition between baseline and follow-up, and the households that did not comply with treatment assignment. The one treatment household without a laundry machine in follow up was the victim of burglary of the appliance. The five control households with laundry machine at follow-up purchased the appliance.

Table 6: Non-Complier Households

of 145 households, 78 treatments and 67 control. Therefore, the final working sample excludes the attrition, all households who suffered a family composition transition and non-compliers.

5.2.2 Outcomes

Time Use As described in the model section, for a given household with both a husband and a wife, $i \in \{h, w\}$, total time is divided into i's leisure, l_i, parental time caring for children, b_i, home production time h_i, and time in the labor market, m_i:

$$T = m_i + l_i + b_i + h_i \qquad i \in \{h, w\}$$

Our intervention relies mainly on precise time use measurement to identify each of the categories in the equation above. Both baseline and follow-up surveys allocate a significant part of the instrument to this purpose. Our instrument relied on directly asking about time use for a pre-specified list of twenty-two activities, covering all aspects of daily life[19].

[19]The literature on time use has pointed out the difficulties of relying on time use diaries, which are seldom completed by the family members. This motivated our different approach, given the small sample size.

Home Production Time and Laundry Machine Use The average household allocates 16 hours to home production exclusively[20]. 5.5 of these are spent doing laundry. In our sample, at baseline, all households do laundry by hand or rent washing machines by the hour[21].

Childcare activities Parental time investments are central for children's development and welfare. Activities related to child care take an average of 22 additional hours including personal are routines (feed them, clean them, dress them, put them to bed, etc.) and activities that actively stimulate the children (sing to/with them, read to/with them, etc.). This also includes activities such as playing, taking them to/from school, helping them with homework and taking them to medical check-ups.

Labor Market The time devoted to working outside the household is considered time in the labor market[22]. Labor market characteristics are similar between mothers in treatment and control households. On average, 60% of mothers are employed, 5% are unemployed, and 35% are out of the work force. In contrast, 90% of men report employment, 1.2% report unemployment, and 8% report to be out of the workforce. Most men and women who report to be working do so in informal or part-time jobs.

Health, Education, and Children's Life Quality Our baseline survey also asked about schooling of parents and children, and about health and life quality of children. Around 25% of both mothers and fathers in our sample did not graduate from high school, and 75% school-aged children are enrolled in school. Somewhat surprisingly, our baseline results reveal substantial health concerns for the sample of children in the sample households. Almost half of them had a cold at the time of the interview, 12% had diarrhea, and 8% reported allergies. More strikingly, we find malnourishment rates somewhat higher than those from the ELCA survey, which is representative of poor population at the national level. 18% of interviewed children were found to be under chronic malnourishment, 3.8% under acute malnourishment, and 10% under global malnourishment. All of these characteristics are similar between children in treatment and control households.

[20]The activities included in home production are household chores (not including childcare or leisure) such as washing clothes, ironing, cleaning bathroom, cooking, among others. It also includes caring for elderly or disabled family members.

[21]The machines are delivered and subsequently picked up from their homes. Laundromats do not exist in any of the neighborhoods in our sample.

[22]The time devoted to looking for a job was zero for the bulk of households and therefore we decided not to include it in the estimations

Household Finances At baseline, treatment and control households were indistinguishable in terms of their average monthly income ($US 232), More than half the households in the sample actually fall bellow the official poverty line ($US 72 per household member per month). Less than 30% have loans, and less than 15% have a savings account.

Female Mental Health, Gender Role Perceptions, Female Empowerment Finally, we collected responses from females about their perceptions on gender roles and about their life satisfaction based on several well-known instruments such as the General Social Survey (NORC (2015)) and the Gender Roles Attitude Scale (Zeyneloglu and Terzioglu (2011)). Here we asked mothers about their stress, worries, interest in their daily activities, depression, family pressure and family environment. To capture gender role beliefs, we asked ten questions/statements, with categorical answers based on which we will compute measures of gender role perceptions. We asked the following ten questions: i) Can working mothers develop as good a relationship with their children as stay-home mothers? ii) Is it likely that children in pre-school age will suffer if their mother works? iii) Is family life hurt by the excessive attention men give to their jobs? iv) Should both men and women contribute to household income? v) Is the duty of men to earn a living, and of women to take care of the home and family? vi) Should family planning decisions be made by both spouses? vii) Do you believe women are better at home chores than men? viii) Should the husband make decisions on the wife's life choices? ix) Should the man be the household head? x) Should the husband have more schooling than the wife? Answers to these questions in the baseline exhibits significant variation, but no systematic differences between treatment and control households. For example, 30% of women believe men should be the household head, and 65% believe women are better at home chores.

We measure female empowerment using 2 standard questions from the literature. First, How do you and your husband organize the income generated by both? The possible answers are: I administer all the money and give my couple his share, my couple administers all the money and gives me my share, we pull our incomes together and each spends as needed, we pool together a fraction of our income and keep a fraction to ourselves, we each handle our money separately. Second, Which of the following statements best describes the distribution of domestic chores between you and your husband? a) I work a lot more than I should b) I work a little more than I should c) I work what I should d) I work a little less than I should e) I work a lot less than I should.

Finally, we use two questions to measure whether women feel that their families appreciate their work at home, which we call the Family Environment Perception Index. This is composed of 2 questions, and in both questions mothers have to tell how often do they feel

like a particular statement. The two statements are: "In general I am fed up with being a homemaker", and "My family often has excessive demands". The response options are: never, rarely, sometimes, frequently, most of the times.

5.2.3 Data collection procedures and quality control

The information collection took place in the household, where parental permissions to participate in the study and household questionnaires were collected. The team of information collectors was trained for one week and assessed for reliability. Prior to field-work, the supervisor contacted the Foundations and the households to schedule a visit. Data collection was paper-based at baseline, and computer-based at follow-up. The filed supervisor centralized the digital information in her computer and transmitted it to the headquarters in Bogotá every night. Weekly reports were given to the research team during field work. All of the information was electronically tracked by a control map designed for the study. The quality control software generated periodic indicators that allowed tracking of coverage which allowed for timely corrections or appropriate adjustment to work in the field.

Various types of supervision were implemented during field work. First, there was a supervisor working always alongside the surveyors. We also performed indirect data verification through random phone calls to households,to confirm information. When necessary, the interview was repeated. Once the complete data set was received by the research team, it was assessed for consistency and completeness. Minor inconsistencies or missing data were reported to our field work provider and they would re-contact the household by phone to complete the missing or verify the inconsistent information. This iteration process took about one month. After this time, the research team initiated the statistical analysis.

6 Impact Analysis

We now turn to the estimation of the causal (direct and spillover) effects of the intervention on the array of outcomes we consider might be affected by the provision of a laundry machine, and to suggest some key mechanisms relating the intervention with gender role perceptions and subsequent outcomes. As mentioned in the experimental design section, households in each of the neighborhoods in our intervention know each other, and interact frequently. As such, an intervention that intends to explore changes in perceptions or beliefs must allow for the possibility of interference across treated and control households, possibly taking place through communication and social learning. The Impact Analysis follows exactly what we

proposed in the Pre-Analisys Plan. However, due to the short period of time between data collection and this report, we present only the first stage of the analysis, which already provides answers to the basic evaluation questions, and delay for the next draft the more nuanced analysis.

The measurement of most of our outcomes is straightforward except for gender role perceptions, which are based on the answers to ten questions as described in section the previous section. Answers to these questions were recorded on a 1 to 5 scale, which we sort so that higher values in each answer denote less asymmetry in beliefs regarding gender roles. Because each question attempts to capture different features of gender role perceptions, we believe it is natural to model the answers to these questions y_{ir}, $r \in \{1, 2, ..., 10\}$ as being generated by a latent factor model for categorical data[23]. We computed a gender roles perception index π_i^t as the predicted factor, both for the baseline ($t = 0$) and for the follow-up ($t = 1$). We estimated π_i^t to capture gender role perceptions in the empirical exercises outlined below.

In this draft we impute missing covariate values (for baseline values) with the average of the non-missing observations and account for the imputation with a dummy variable. In the next draft we plan to check whether item non-response is correlated with treatment status. If we find that it is not we will continue to apply the above imputation method.

6.1 Inference

The small sample size in our intervention makes asymptotic-based inference inapplicable. Thus, we use exact randomization inference to assess the statistical significance of our treatment on the different outcomes under consideration. The potential for interference between subjects in our experimental setting does not allow for standard exact randomization inference procedures. We will borrow methods from Rosenbaum (2007), who develops exact tests for inference based on randomization, that allow for arbitrary interference between treated

[23]We assume the answer to each question is generated from a latent variable $\tilde{\pi}_i$ with a Gaussian prior. As such, the categorical factor model we estimate takes the form

$$p\left(\tilde{\pi}_i\right) = \mathcal{N}(0,1)$$

$$p(\mathbf{y}_i|\tilde{\pi}_i, \mathbf{\Lambda}) = \prod_{r=1}^{10}\prod_{c=1}^{4} \mu_c^{\mathbf{1}\{y_{ir}=c\}} \tag{3}$$

where

$$\mu_c = \frac{e^{\lambda_c \tilde{\pi}_i + \lambda_{0c}}}{1 + \sum_{k=1}^{3} e^{\lambda_k \tilde{\pi}_i + \lambda_{0k}}}, \qquad c = 1, ..., 3$$

$$\mu_4 = \frac{1}{1 + \sum_{k=1}^{3} e^{\lambda_k \tilde{\pi}_i + \lambda_{0k}}}$$

and control subjects, and for covariates. These methods fit our experimental design perfectly: They rely on randomized treatment assignment and perfect compliance[24], features both shared by our intervention. Here we outline briefly the nature of the exact randomization tests we will employ.

The hypothesis tests rely on comparing the realized assignment of treatment and control across subjects to a hypothetical assignment where the treatment is withheld from all the same subjects (the "uniformity trial"). Under this hypothetical uniformity trial, every subject is as likely to present a higher value of the outcome than any other subject, making the distribution of many statistics based in this trial known a priori. We can then compare the observed values of such statistics using the actual randomization assignment against this null distribution[25].

We will undertake two classes of hypothesis tests. First we will test for the absence of an effect. The null hypothesis here is that the outcome for every subject is the same as its counterfactual outcome under the uniformity trial. Second, we will test for "exceedance", namely, tests under the null hypothesis that the outcomes of treated subjects do not exceed a given quantile of the outcomes of control subjects (for example, the median outcome of the control group). The test statistics used for both hypothesis tests can be used to construct measures of the magnitude of the average effects after an appropriate normalization.

We first introduce some notation. $b = 1, 2, ..., 6$ indexes our neighborhoods, and I_b is the number of households in neighborhood b participating in the intervention. In neighborhood b, J_b households were chosen at random for treatment, and the remaining $I_b - J_b$ are our control households in the neighborhood. $T_{ib} = 1$ denotes assignment to treatment for household i. $\mathbf{T}_b \equiv (T_{1b}, ..., T_{I_b b})$ is the vector of treatment assignments. Randomization implies that $\begin{pmatrix} I_b \\ J_b \end{pmatrix}^{-1}$ is the probability of any possible treatment assignment where J_b households receive treatment out of the I_b in the experiment.

6.1.1 Test of no effect

Here we ask: Are treated households more likely to have higher values of the outcome than control households relative to what would have been observed in a uniformity trial? Define

[24]Even though between baseline and follow-up some control households purchased a laundry machine, we excluded them from our final working sample.

[25]Such a test has no power against the possibility that interference is perfect, i.e., if the intervention changes the value of the outcome by exactly the same amount to treated and control subjects (this is also called "no primary effect"). Given the nature of our intervention we consider this possibility extremely unlikely.

the statistic

$$\tau(\mathbf{T}) = \sum_{b=1}^{6} \sum_{i=1}^{I_b} \sum_{j=1}^{I_b} T_{ib}(1 - T_{jb}) f_{ijb}(\mathbf{T}) \tag{4}$$

where

$$f_{ijb}(\mathbf{T}) = \begin{cases} 1 & if\ y_{ib}(\mathbf{T}) > y_{jb}(\mathbf{T}) \\ 0 & if\ y_{ib}(\mathbf{T}) \leq y_{jb}(\mathbf{T}) \end{cases}$$

and $y_{ib}(\mathbf{T})$ denotes the counterfactual outcome that household i would have experienced under randomization assignment $\mathbf{T}$. For $\mathbf{T} = \mathbf{U}$, the uniformity trial, the null distribution of $\tau(\mathbf{U})$ is known, and for a sample size above ≈ 20, it is very well approximated by a Normal distribution with mean

$$\mu_U = \frac{1}{2} \sum_{b=1}^{6} J_b(I_b - J_b)$$

and standard deviation

$$\sigma_U = \left(\frac{1}{12} \sum_{b=1}^{6} J_b(I_b - J_b)(I_b + 1) \right)^{1/2}.$$

Now define the statistic $F = \tau(\mathbf{T}) - \tau(\mathbf{U})$. F is the number of times that treated households exceeded control households in the actual randomization but not in the uniformity trial, minus the number of times that treated households would exceed control households in the uniformity trial but not in the actual randomization. Observe that if the treatment had no effect, $F = 0$, and if all treated households had responses above control households, then $\mathbb{E}[F] = \frac{1}{2} \sum_{b=1}^{6} J_b(I_b - J_b)$. We can now make confidence statements about F using the observed value of $\tau(\mathbf{T})$ and the known distribution of $\tau(\mathbf{U})$:

$$\begin{aligned} &\mathbb{P}(F > q_\alpha) = 1 - \alpha \\ &= \mathbb{P}(\tau(\mathbf{U}) < \tau(\mathbf{T}) - q_\alpha) \\ &= \Phi\left(\frac{\tau(\mathbf{T}) - q_\alpha - \mu_U}{\sigma_U} \right) \end{aligned}$$

which we can easily invert to obtain any desired confidence intervals. For example, p-values for the test of no effect can be easily computed as

$$p - value = \Phi\left(\frac{\tau(\mathbf{T}) - \mu_U}{\sigma_U} \right)$$

6.1.2 Test of Treatment Responses Above Quantiles of the Control Outcomes

Here we ask: Are treated households likely to have values of the outcome above the median value of the outcome for control households relative to what would have been observed in a uniformity trial? Define the statistic

$$\varsigma^k(\mathbf{T}) = \sum_{b=1}^{6} \sum_{i=1}^{I_b} T_{ib} s_{ib}^k(\mathbf{T}) \tag{5}$$

where

$$s_{ib}^k(\mathbf{T}) = \begin{cases} 1 & if\ y_{ib}(\mathbf{T}) > q_b^k(\mathbf{T}) \\ 0 & if\ y_{ib}(\mathbf{T}) \leq q_b^k(\mathbf{T}) \end{cases}$$

and $q_b^k(\mathbf{T})$ is the k-th order statistic value of the outcome among control households under randomization assignment $\mathbf{T}$[26]. For $\mathbf{T} = \mathbf{U}$, the null distribution of $\varsigma(\mathbf{U})$ is known, and for a sample size above ≈ 20, it is well approximated by a Normal distribution with mean

$$\mu_U^k = \frac{1}{2} + \sum_{i=1}^{6} J_b \left(1 - \frac{k_b}{I_b - J_b + 1}\right)$$

and standard deviation

$$\sigma_U^k = \left(\sum_{i=1}^{6} \frac{J_b k_b (I_b - J_b - k_b + 1)(I_b + 1)}{(I_b - J_b + 1)^2 (I_b - J_b + 2)}\right)^{1/2}.$$

Now define the statistic $S^k = \varsigma^k(\mathbf{T}) - \varsigma^k(\mathbf{U})$. S^k as a magnitude of the treatment effect, is the number of times treated households exceeded the k-th order statistic among control households in the actual randomization but not in the uniformity trial, minus the number of times that treated households would exceed control households in the uniformity trial but not in the actual randomization. We can now make confidence statements about S^k using the observed value of $\varsigma(\mathbf{T})$ and the known distribution of $\varsigma(\mathbf{U})$:

$$\begin{aligned} &\mathbb{P}(S^k > q_\alpha) = 1 - \alpha \\ &= \mathbb{P}(\varsigma^k(\mathbf{U}) < \varsigma^k(\mathbf{T}) - q_\alpha) \\ &= \Phi\left(\frac{\varsigma^k(\mathbf{T}) - q_\alpha - \mu_U^k}{\sigma_U^k}\right) \end{aligned}$$

[26]For example, if we are interested in the median outcome of control households, then $k_b = (I_b - J_b + 1)/2$ for $I_b - J_b$ odd.

which we can easily invert to obtain any desired confidence intervals.

Although our treatment-control covariate balance is satisfactory (see our baseline report), we will perform covariance adjustment to dampen chance imbalances in observables. This can be accommodated easily within this exact randomization inference framework, by applying the tests describe above to the residuals from regressions of our outcomes on the covariates $\mathbf{x}_{ib}$.

6.2 Exact Randomization Inference Test Results

Tables 7-9 presents out main results based on the exact randomization inference tests of no effect, that allow for interference across households within neighborhood (Rosenbaum (2007)). For each of our outcomes of interest, the first column of Tables 7 and 9 presents the test statistic $\tau(\mathbf{T})$ computed according to equation (4), which we contrast against the known null distribution of the uniformity trial as described in section 6.1. A possible realization of this null distribution counts the number of times that, within each neighborhood, each household labeled as treated would exhibit a value of the outcome above that of each household labeled as control, when the intervention actually has no effect. In this scenario, the outcome of a treated household would be above the value of a control household with a 50-50 chance. The second and third columns of the table present the mean and standard deviation of this null distribution, which is approximately normal as described in section 6.1. All of the tests are one-sided, but based on our prior beliefs about the likely effects of the intervention on different outcomes, as described in the Mechanisms section, we perform either right-sided or left-sided tests. When we expect the effect to be positive, naturally we perform a right-sided test, where $\tau(\mathbf{T})$ counts the number of times treated households exhibit outcomes above control households under our intervention. When we expect the effect to be negative, we perform a left-sided test, in which case $\tau(\mathbf{T})$ counts the number of times treated households exhibit outcomes below control households under our intervention. Column 4 in reports the nature of the test for each outcome. Column 5 then presents the p-value of the test, and columns 6 and 7 indicate the sample size for both treated and control groups across all neighborhoods.

The results presented in this draft correspond only to households that were bi parental at baseline and remained biparental at follow-up and that did not face transitions in marital status, that is, the households in the final working sample, as described above. This is because we are very interested in estimating the effects on gender roles, which can be studied more directly in bi parental households.

6.2.1 Mothers' outcomes

Take the results for time devoted to laundry, shown in Panel A in Table 7. We expected the intervention to have a negative effect on this outcome, to decrease the time it takes to wash clothes, so we report results of a left-sided test. In this case the test statistic is $\tau(\mathbf{T}) = 268$. The null distribution is such that on average, given the sample sizes of treatments and controls, we would expect treatments to have higher values than controls in their same neighborhood 220.5 times, with a standard deviation of 28.3. Thus, we find that our intervention is 2 standard deviations above the mean of the null distribution of no effect, with a corresponding p-value of 0.05. This is solid evidence of a negative effect on the time devoted to laundry. Our descriptive statistics of the differences between treatment and control households at follow-up suggest that the impact size is 2.5 hours per week, given a baseline mean of 5.5 hours. Therefore, having a laundry machine cuts the time spent doing laundry almost in half: the amount of time saved is not marginal.

Leisure time for mothers appears to decrease as well. The test statistic is $\tau(\mathbf{T}) = 266$, and the null distribution has a mean of 220.5 and a standard deviation of 28.3. The corresponding p-vale is 0.05 and thus our intervention is 1.5 standard deviations from the mean. We consider this to be good evidence of a negative effect on mothers' leisure time (left-handed test). This is surprising because the expected result in the Mechanisms section was that it may increase.The ratio of childcare time to home production time for mothers has a test statistic of $\tau(\mathbf{T}) = 255$, the null distribution has a mean of 217 and a standard deviation of 28.4, which yields a p-value of 0.09, providing solid evidence that after the intervention mothers spend relatively more time doing childcare as opposed to house work than without the intervention.

There are the variables that are statistically significant at standard confidence levels in Panel A. However, given the sample size, we believe that outcomes with p-values below 0.15 still provide good evidence of program effects. However, we prefer to let the reader decide for themselves and therefore do no include stars.

Let us discuss the variables that have have p-values above 0.15 regarding Mothers' time use. Home production time (p-value 0.15) decreases for mothers in the treatment groups as compared to those in the control group. Time working in the labor market increases (p-value 0.15), which coincides with what was presented in the Mechanisms section.

We find no effect on the time devoted to childcare, on home production excluding laundry, in the ratio between leisure and total time or in the ratio of market work to home production time.

In Panel B of Table 7 we present the Mothers' labor market characteristics. The program

EXACT RANDOMIZATION INFERENCE TEST OF NO EFFECT ALLOWING FOR INTERFERENCE WITHIN CLUSTERS

PANEL A: MOTHERS' TIME USE	(1)	(2)	(3)	(4)	(5)	(6)	(7)
OUTCOME	τ	μ_u	σ_u	Test	p-value	T	C
Laundry time	268	220.5	28.30	Left-sided	0.047	60	50
Childcare time	242	220.5	28.30	Right-sided	0.224	60	50
Home production time	250	220.5	28.30	Left-sided	0.149	60	50
Home production time without laundry	237	220.5	28.30	Right-sided	0.280	60	50
Leisure time	266	220.5	28.30	Left-sided	0.054	60	50
Work time in labor market	215	188	25.50	Right-sided	0.145	58	46
Ratio of leisure to all time	207	188	25.50	Left-sided	0.228	58	46
Ratio of childcare time to home production time	255	217	28.04	Right-sided	0.088	60	49
Ratio of market work time to home production time	206	184.5	25.22	Right-sided	0.197	58	45
PANEL B: MOTHERS' CHARACTERISTICS							
OUTCOME	τ	μ_u	σ_u	Test	p-value	T	C
Employed	255	220.5	28.30	Right-sided	0.111	60	50
Studying	215	220.5	28.30	Right-sided	0.577	60	50
PANEL C: MOTHERS' MENTAL HEALTH, EMPOWERMENT AND GENDER ROLE PERCEPTIONS							
OUTCOME	τ	μ_u	σ_u	Test	p-value	T	C
Family Environment Perceptions index	216	207.5	27.39	Left-sided	0.378	59	49
Stress Index	203	220.5	28.30	Left-sided	0.732	60	50
How money is administered	243	217	28.04	Right-sided	0.177	60	49
Fairness in homework distribution	267	219	28.20	Right-sided	0.044	60	49
Gender roles PCA index	209	220.5	28.30	Right-sided	0.658	60	50
PANEL D: HOUSEHOLD-LEVEL OUTCOMES							
OUTCOME	τ	μ_u	σ_u	Test	p-value	T	C
Price of electric bill	112	143	21.04	Right-sided	0.930	49	41
Price of aqueduct bill	115	115	18.40	Left-sided	0.500	38	35
Household labor income	258	220.5	28.30	Right-sided	0.092	60	50
Household saves	286	220.5	28.30	Left-sided	0.010	60	50
Father does laundry	198	220.5	28.30	Right-sided	0.787	60	50

Note: The first column reports the value of the test statistic τ for the exact test of no effect. The second and third columns report the mean and standard deviations of the approximatly normal distribution of the test statistic under the null hypothesis of no effect. The fourth column indicates the test side. The fifth column reports the p-value for τ under the null hypothesis. Finally, columns six and seven report the number of treated and control households used for the test. All mothers' time use outcomes are measured in daily minutes. Laundry time corresponds to daily minutes mothers spend washing clothes for household members without payment. Childcare time refers to daily minutes used up by mothers in childcare (feeding, bathing and clothing, playing with children under 5 years, taking or pick up children from medical appointments or kindergarten, homework help and spending time together). Home-production time includes daily minutes spent by mothers on housework- including laundry- and caring for elderly or disable relatives. Leisure time covers daily minutes mothers spend on recreation or entertainment (video games, board games or gambling, taking household members to different cultural activities; reading books, magazines or newspaper; watching TV or listening to radio as a single activity; browsing the internet; sleeping; eating; praying and resting without doing anything else). Work time in the labor market includes daily minutes looking for a job or doing errands to establish a business, and working on a job. Ratio of leisure to all time is the daily minutes spend on childcare as a fraction of total time devoted to other categories. Ratio of childcare time to home production time is the number of minutes spent on childcare per minute spent on housework. Ratio of market work time to home production time is the number of minutes spent on market work per minute spent on home production including laundry. Employed is a dummy equal to one if the mother worked at least one hour in an income-generating activity, worked as a family helper, worked at least one hour, looked for a job, or did not work but had a job from which she received income. Study is a dummy equal to one if the mother is studying. Price of electric and aqueduct bills correspond to the households' payment for each service during the last billing period. Household labor income aggregates all wages earned by family members participating in the labor market. Household saves is a dummy equal to one if the family saves part of their income. Father does laundry is a dummy equal to one if the father washes clothes. The Family Environment Perception Index is a standardized measured that aggregates the mother's emotional conditions and burnouts level responses: intense and persistent disorder, loss of energy, low motivation and extreme irritability produced by family the environment. Higher values denote worse family environment perceptions. The Stress Index measures the mother´s depression and anxiety conditions. Higher values are associated with more stress. How money is administered indicates how earnings are managed in the household between spouses. Higher values denote more control of household finances by the mother. Fairness in homework distribution measures the mother's perception of fairness in chore distribution. Higher scores denote a more even distribution. Gender Roles PCA index is the principal component from the factor model that uses the mother's answers to ten questions about gender roles. More egalitarian views on gender are associated with higher values for the index.

Table 7: Test of No Effects - Mothers' and Household Outcomes

appears to have an effect on whether mothers are employed (p-value 0.11), but not on whether they are studying. Panel C of the same Table displays the variables related to mothers' mental health, empowerment and genre role perceptions. We find a significant and positive impact on the perception of fairness in home production time (p-value 0.04). This signals that female empowerment may be increasing in response to the intervention.

Finally, in Panel D, we present some household-level outcomes. We expected the intervention to have a positive effect on household income, so we report results of a right-sided test. In this case the test statistic is $\tau(\mathbf{T}) = 258$. The null distribution is such that on average, given the sample sizes of treatments and controls, we would expect treatments to have higher values than controls in their same neighborhood 220.5 times, with a standard deviation of 28.3. Thus, we find that our intervention is 1.5 standard deviations above the mean of the null distribution of no effect, with a corresponding p-value of 0.09. Given the sample size, we consider this to be good evidence of a positive effect on household income. We also have a significant decrease in household savings due to the intervention (p-value 0.01). This is surprising since we expected a positive impact if at all on household savings. This is an interesting dimension that we will explore in the following versions of this report.

6.2.2 Exceedence Results

Table 8 then presents the results of our exact randomization inference tests of exceedence, that allow for interference across households within neighborhood (Rosenbaum (2007)). For each of our outcomes of interest, the first column present the test statistic p-value associated with the test statistic $\varsigma^{25}(\mathbf{T})$ computed according to equation (5), resulting from contrasting it against the known null distribution of the uniformity trial as described in section 6.1. A possible realization of this null distribution counts the number of times that, within each neighborhood, each household labeled as treated would exhibit a value of the outcome above the 25th percentile of the distribution for that outcome among the control households in the neighborhood, when the intervention actually has no effect. Column 1 also reports the value of the 25th quantile of the empirical distribution among all controls, to provide a sense of the magnitude of any effects. Analogously to Table 7, depending on our prior beliefs about the effect of the intervention we considered either right-sided or left-sided tests. This information is reported in column 4, and columns 5 and 6 report the sample sizes of treated and control households. Columns 2 and 3 similarly report p-values and for the 25th, 50th and 75th percentile tests of no exceedence, with the associated values of the quantiles of the empirical distributions among all controls.

Again, let us describe the results presented in Panel A of Table 8. Consider the results for

EXACT RANDOMIZATION INFERENCE TESTS OF NO EXCEEDENCE ALLOWING FOR INTERFERENCE WITHIN CLUSTERS

PANEL A: MOTHERS' TIME USE	(1)	(2)	(3)	(4)	(5)	(6)
OUTCOME	Quantiles and p-values			Test	T	C
	25	50	75			
Laundry time	120	180	240			
	1.000	0.072	0.000	Left-sided	60	50
Childcare time	480	750	1080			
	0.154	0.248	0.239	Right-sided	60	50
Home production time	300	360	480			
	1.000	0.202	0.001	Left-sided	60	50
Home production time without laundry	120	180	300			
	0.200	0.096	0.239	Right-sided	60	50
Leisure time	730	870	1,060			
	1.000	0.540	0.004	Left-sided	60	50
Work time in labor market	0	138	540			
	0.265	0.073	0.046	Right-sided	58	46
Ratio of leisure to all time	0.4	0.544	0.767			
	1.000	0.442	0.008	Left-sided	58	46
Ratio of childcare time to home production time	1.25	2	3.18			
	0.057	0.440	0.426	Right-sided	60	49
Ratio of market work time to home production time	0	0.33	1.2			
	0.257	0.046	0.018	Right-sided	58	45
PANEL B: MOTHERS' CHARACTERISTICS						
OUTCOME	Quantiles and p-values			Test	T	C
	25	50	75			
Employed	0	1	1			
	0.253	0.125	0.116	Right-sided	60	50
Studying	0	0	0			
	0.931	0.934	0.772	Right-sided	60	50
PANEL C: MOTHERS' MENTAL HEALTH, EMPOWERMENT AND GENDER ROLE PERCEPTIONS						
OUTCOME	Quantiles and p-values			Test	T	C
	25	50	75			
Family Environment Perceptions index	-0.977	0.008	0.599			
	1.000	0.695	0.000	Left-sided	59	49
Stress Index	0	2	3			
	1.000	0.356	0.003	Left-sided	60	50
How money is administered	1	2	2			
	0.148	0.176	0.364	Right-sided	60	49
Fairness in homework distribution	1	2	3			
	0.122	0.121	0.195	Right-sided	60	49
Gender roles PCA index	3.23	3.54	4.14			
	0.591	0.660	0.720	Right-sided	60	50

EXACT RANDOMIZATION INFERENCE TESTS OF NO EXCEEDENCE ALLOWING FOR INTERFERENCE WITHIN CLUSTERS (Cont.)

PANEL D: HOUSEHOLD-LEVEL OUTCOMES

OUTCOME	Quantiles and p-values			Test	T	C
	25	50	75			
Price of electric bill	8,500	15,500	30,000			
	0.883	0.607	0.453	Right-sided	49	41
Price of aqueduct bill	6,000	20,000	40,000			
	1.000	0.922	0.126	Left-sided	38	35
Household income	600,000	670,000	900,000			
	0.084	0.601	0.475	Right-sided	60	50
Household saves	0	0	0			
	1.000	0.356	0.004	Left-sided	60	50
Father does laundry	1	1	1			
	0.380	0.951	0.957	Right-sided	60	50

Note: Columns one, two, and three report the values of the 25th, 50th, and 75th percentiles of the empirical distribution of the outcome variable at baseline, and the corresponding p-value of the test statistic ς for the exact test of no exceedence of that quantile. P-values are ureported under each quantile value. The fourth column indicates the test side. Finally, columns six and seven report the number of treated and control households used for the test. All mothers' time use outcomes are measured in daily minutes. Laundry time corresponds to daily minutes mothers spend washing clothes for household members without payment. Childcare time refers to daily minutes used up by mothers in childcare (feeding, bathing and clothing, playing with children under 5 years, taking or pick up children from medical appointments or kindergarten, homework help and spending time together). Home-production time includes daily minutes spent by mothers on housework- including laundry- and caring for elderly or disable relatives. Leisure time covers daily minutes mothers spend on recreation or entertainment (video games, board games or gambling, taking household members to different cultural activities; reading books, magazines or newspaper; watching TV or listening to radio as a single activity; browsing the internet; sleeping; eating; praying and resting without doing anything else). Work time in the labor market includes daily minutes looking for a job or doing errands to establish a business, and working on a job. Ratio of leisure to all time is the daily minutes spend on childcare as a fraction of total time devoted to other categories. Ratio of childcare time to home production time is the number of minutes spent on childcare per minute spent on housework. Ratio of market work time to home production time is the number of minutes spent on market work per minute spent on home production including laundry. Employed is a dummy equal to one if the mother worked at least one hour in an income-generating activity, worked as a family helper, worked at least one hour, looked for a job, or did not work but had a job from which she received income. Study is a dummy equal to one if the mother is studying. Price of electric and aqueduct bills correspond to the households' payment for each service during the last billing period. Household labor income aggregates all wages earned by family members participating in the labor market. Household saves is a dummy equal to one if the family saves part of their income. Father does laundry is a dummy equal to one if the father washes clothes. The Family Environment Perception Index is a standardized measured that aggregates the mother's emotional conditions and burnouts level responses: intense and persistent disorder, loss of energy, low motivation and extreme irritability produced by family the environment. Higher values denote worse family environment perceptions. The Stress Index measures the mother´s depression and anxiety conditions. Higher values are associated with more stress. How money is administered indicates how earnings are managed in the household between spouses. Higher values denote more control of household finances by the mother. Fairness in homework distribution measures the mother's perception of fairness in chore distribution. Higher scores denote a more even distribution. Gender Roles PCA index is the principal component from the factor model that uses the mother's answers to ten questions about gender roles. More egalitarian views on gender are associated with higher values for the index.

Table 8: No Exceedance Test - Mothers' and Household Outcomes

laundry time in table 8. For this outcomes we performed a left-sided test. The 75th percentile of the distribution of controls is 240. The p-value in this case is 0.00, suggesting that treated households are systematically below the 75th percentile of control households' laundry time. The 50th percentile of the distribution of controls is 180, and the p-value of the test of no exceedence is 0.07; treated households are systematically below the median of the control household's laundry time. However, the p-value for the 25th percentile is 1, and thus, we cannot reject that the effect of the intervention did not lead treated households to increase their income above the 75th percentile of the income distribution for controls.

Consider the results for household income in table 8. For this outcome we performed a right-sided test. The 25th percentile of the distribution of controls is COP$600,000. The p-value in this case is 0.08 only, suggesting that treated households systematically exceed the 25th percentile of control households' income. The 75th percentile of the distribution of controls is COP$900,000, and in this case the p-value of the test of no exceedence is 0.47. Thus, we

EXACT RANDOMIZATION INFERENCE TEST OF NO EFFECT ALLOWING FOR INTERFERENCE WITHIN CLUSTERS

PANEL A: FATHERS' TIME USE	(1)	(2)	(3)	(4)	(5)	(6)	(7)
OUTCOME	τ	μ_u	σ_u	Test	p-value	T	C
Laundry time	215	193.5	25.47	Left-sided	0.199	57	47
Childcare time	244	193.5	25.47	Right-sided	0.024	57	47
Home production time	223	193.5	25.47	Right-sided	0.123	57	47
Home production time without laundry	225	193.5	25.47	Right-sided	0.108	57	47
Leisure time	214	193.5	25.47	Left-sided	0.211	57	47
Work time in labor market	159	163	22.29	Right-sided	0.571	56	42
Ratio of leisure to all time	195	163	22.29	Left-sided	0.076	56	42
Ratio of childcare time to home production time	76	55	10.84	Right-sided	0.026	36	20
Ratio of childcare time to all time	182	163	22.29	Left-sided	0.197	56	42
PANEL B: FATHERS' CHARACTERISTICS							
OUTCOME	τ	μ_u	σ_u	Test	p-value	T	C
Employed	211	220.5	28.30	Right-sided	0.632	60	50
Labor income	133	110	17.15	Right-sided	0.090	42	34

Note: The first column reports the value of the test statistic τ for the exact test of no effect. The second and third columns report the mean and standard deviations of the approximatly normal distribution of the test statistic under the null hypothesis of no effect. The fourth column indicates the test side. The fifth column reports the p-value for τ under the null hypothesis. Finally, columns six and seven report the number of treated and control households used for the test. All fathers' time use outcomes are measured in daily minutes. Laundry time corresponds to daily minutes fathers spend washing clothes for household members without payment. Childcare time refers to daily minutes used up by fathers in childcare (feeding, bathing and clothing, playing with children under 5 years, taking or pick up children from medical appointments or kindergarten, homework help and spending time together). Home-production time includes daily minutes spent by fathers on housework- including laundry- and caring elderly or disable relatives. Leisure time covers daily minutes fathers spend on recreation or entertainment (video games, board games or gambling; taking household members to different cultural activities; reading books, magazines or newspaper; watching TV or listening to radio as single activity; browsing the internet; sleeping; eating; praying and resting without doing anything else). Work time in labor market includes daily minutes looking for a job or doing errands to stablish a business and working on a job. Ratios of leisure and childcare to all time are the daily minutes spend on leisure or childcare as a fraction of total time devoted to other categories. Ratio of childcare time to home production time is the number if minutes spent on childcare per minute spent on housework. Employed is a dummy equal to one if the father worked at least one hour in an income-generating activity, worked as family helper, worked at least one hour, looked for a job, or did not work but had a job from which he received income. Study is a dummy equal to one if father is studying.

Table 9: Test of No Effects - Fathers' Outcomes

cannot reject that the effect of the intervention did not lead treated households to increase their income above the 75th percentile of the income distribution for controls.

6.2.3 Fathers' Outcomes

On Table 9 we present the exact randomization inference test of fathers' outcomes. Note that father's time use information is only available at follow-up and therefore we cannot calculate quantiles for the baseline distribution of fathers' outcomes. Thus, here we omit the results of the exact tests of no exceedence for fathers[27]. Even though we do not have baseline information on this variable to check whether the treatment and control groups were balanced, our finding that both groups were balanced in time use of mothers and in most other categories makes us confident that the differences found at follow-up in father's variables can be attributed to the intervention.

[27]Results are available upon request.

As in the case for mothers, in Panel A we present fathers' time use. Let us take the results for time devoted to childcare. We expected the intervention to have a positive effect on this outcome if at all, so we report results of a right-sided test. In this case the test statistic is $\tau(\mathbf{T}) = 244$. The null distribution is such that on average, given the sample sizes of treatments and controls, we would expect treatments to have higher values than controls in their same neighborhood 193.5 times, with a standard deviation of 25.5. Thus, we find that our intervention is over 2 standard deviations above the mean of the null distribution of no effect, with a corresponding p-value of 0.02. We also find an increase in the ratio of childcare to total time (p-value 0.03). This is solid evidence of a positive effect on the time fathers devoted to childcare. Fathers' home production time (p-value 0.12) and home production time excluding laundry (p-value 0.11) both increase. In Panel B we find a significant increase in the father´s labor income (p-value 0.09). These results are surprising and we will investigate this further via the qualitative analysis.

6.3 Direct and Spillover Effects

In the following sections we present the next steps in the analysis that will be included in the next version of this report.

6.3.1 Using Variation Within and Across Neighborhoods

Based on our experimental design we can estimate treatment on the treated (TOT) effects, and average spillovers on the treated based on the following specification:

$$y_{ib} = \beta T_{ib} + \phi T_{ib} I_b + \mathbf{x}_{ib}\boldsymbol{\delta} + \eta_b + \varepsilon_{ib} \tag{6}$$

In equation (6) y_{ib} denotes the change in an outcome for household i in neighborhood b between baseline and follow-up. T_{ib} denotes randomly assigned treatment (provision of a laundry machine), I_b denotes (randomly assigned) intensity of treatment in neighborhood b, $\mathbf{x}_{ib}$ is a vector of household characteristics, the η_b are neighborhood fixed effects, and ε_{ib} denotes mean-zero unobservables. β captures the average difference in outcome y for a treated household relative to a control household, while ϕ captures the average slope along the neighborhood treatment intensity gradient for *treated households*. Equation (6) will be estimated on the full sample of treated and control households in all six neighborhoods.

The spillover effects (if any) on control households (which may differ from those on treated households) cannot be recovered from equation (6). Nevertheless, we can recover them esti-

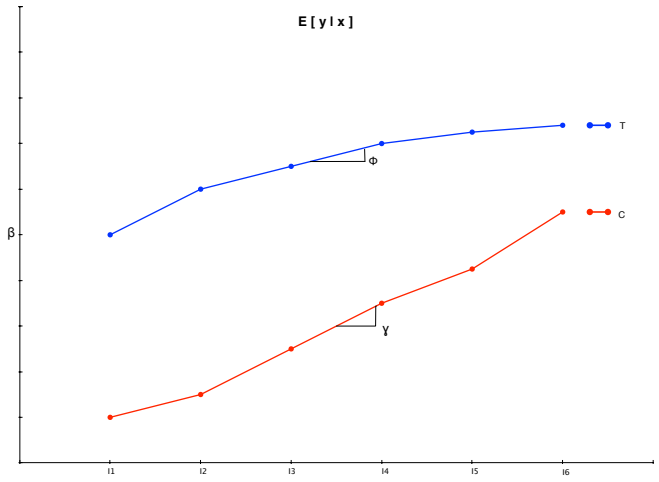

Figure 5: Direct and Spillover Effects. The Figure plots hypothetical conditional means for Treated and Control households along the neighborhood treatment intensities gradient. The average slope for the Treatment households conditional mean captures ϕ. The average slope for the Control households conditional mean captures γ. The difference in intercepts between both conditional means captures β.

mating the following specification:

$$y_{ib} = \gamma I_b + \mathbf{x}_{ib}\boldsymbol{\delta} + \bar{\mathbf{x}}_b\boldsymbol{\psi} + \varepsilon_{ib} \tag{7}$$

where (7) will be estimated on the sample of control households only. Here γ captures the average slope along the neighborhood treatment intensity gradient for control households. Because we are unable to include neighborhood fixed effects in this case, besides controlling for household characteristics $\mathbf{x}_{ib}$, we also control for neighborhood characteristics $\bar{\mathbf{x}}_b$. Figure 5 illustrates graphically the objects of interest we intend to recover based on equations (6) and (7), where the neighborhood treatment intensities appear sorted from smallest to largest. The figure depicts the alternative hypotheses that $\beta > 0$, $\gamma > \phi > 0$.

6.3.2 Matching

We will alternatively implement a matching methodology to estimate the same key objects of interest. First, we will use within-neighborhood variation by pairing every treated household $i \in T$ with a control household $i' \in C$ in its same neighborhood that satisfies a distance condition on its covariates of the form

$$\|\mathbf{x}_{ib} - \mathbf{x}_{i'b}\| < \epsilon$$

Using the sample of all pairs of households we can estimate the following specification:

$$y_{ib} - y_{i'b} = \beta + \phi I_b + (\mathbf{x}_{ib} - \mathbf{x}_{i'b})\,\boldsymbol{\delta} + \varepsilon_{ii'b} \tag{8}$$

Finally, we can use cross-neighborhood variation by finding pairs of control households $i, i' \in C$ from different neighborhoods satisfying a similar covariate-distance condition. Using this matched sample we can estimate the following specification:

$$y_{ib} - y_{i'b'} = \gamma(I_b - I_{b'}) + (\mathbf{x}_{ib} - \mathbf{x}_{i'b'})\,\boldsymbol{\delta} + \varepsilon_{ii'bb'} \tag{9}$$

6.4 Heterogeneity

6.4.1 Distribution Factors

The effectiveness of an intervention that alters the relative costs and benefits of female effort in home production such as the one we exploit here could induce heterogeneous responses across households with varying characteristics. In particular, we conjecture that the similarity of couples along characteristics $\mathbf{z}_{ibg}$, $g \in \{Male, Female\}$, such as education, employment status or age, frequently associated with distribution factors in the intra-household bargaining literature, may induce differential responses to treatment[28]. To test for such effects we will estimate specifications of the form

$$y_{ib} = \beta_0 T_{ib} + (\mathbf{z}_{ibm} - \mathbf{z}_{ibf})\,T_{ib}\boldsymbol{\beta}_1 + \beta_2 m_{ib} T_{ib} + \phi T_{ib} I_b + \mathbf{x}_{ib}\boldsymbol{\delta} + \eta_b + \varepsilon_{ib} \tag{10}$$

on the full sample of treated and control households. In equation (10), m_{ib} is an indicator variable for mono-parental households, which constitute 15% of all households in the study. For these households we adopt the convention that $\mathbf{z}_{ibm} - \mathbf{z}_{ibf} = 0$.

6.4.2 Gender Roles Perceptions

Responses to the intervention also may depend on baseline beliefs related to gender roles shared by the household members. For example, if changes in the distribution of home chores across genders requires experimentation with the newly introduced technology, and if in households with strongly entrenched views against male home labor men never experiment with the technology, changes in outcomes may be distinct from those in households

[28] We do not have a large enough sample size to detect differential effects of spillovers along similarity characteristics of couples.

where initial gender role perceptions lead to male experimentation. We will exploit our baseline measures of gender role perceptions summarized by the index π^0_{ib}, and will estimate specifications of the form

$$y_{ib} = \beta_0 T_{ib} + \beta_1 \pi^0_{ib} + \beta_2 \pi^0_{ib} T_{ib} + \beta_3 \pi^0_{ib} m_{ib} + \beta_4 m_{ib} T_{ib} + \beta_5 \pi^0_{ib} m_{ib} T_{ib} + \phi T_{ib} I_b + \mathbf{x}_{ib}\boldsymbol{\delta} + \eta_b + \varepsilon_{ib} \quad (11)$$

on the full sample of treated and control households. We expect treated mono-parental households to be informative about the effects of the treatment in the absence of intra-household conflict between adults.

6.5 IV for the Gender Time Distribution

Our discussion related to gender role perceptions also suggests using our experimental variation to implement an instrumental variables strategy with the purpose of assessing the importance of changes in gender role perceptions on the distribution of home production efforts across genders. Here we focus on one specific outcome, $\Delta_{ib} \equiv \frac{f_{ib}}{f_{ib}+m_{ib}}$, the share of all home production hours allocated to the female at follow-up. We will estimate a specification of the form

$$\Delta_{ib} = \alpha + \beta \left(\pi^1_{ib} - \pi^0_{ib}\right) + \mathbf{x}_{ib}\boldsymbol{\delta} + \eta_b + \varepsilon_{ib} \quad (12)$$

using T_{ib} as an instrument for $(\pi^1_{ib} - \pi^0_{ib})$, on the sample of bi-parental treated and control households. Thanks to the randomization, $cov(T_{ib}, \varepsilon_{ib}) \approx 0$, so that the key assumption for the validity of this specification is that treatment status is excludable from equation (12) (i.e., that it has an effect on the distribution of home hours across genders *only* through its effect on gender role perceptions). This assumption would not be tenable for other outcomes such as the number of hours in different activities, or any health or labor market outcome which could be directly affected by the availability of a laundry machine. In the case of the distribution of hours of home production between male and female partners, in contrast, the change in the time budget constraint induced by the availability of the laundry machine should have no effect on the *distribution* of hours between genders unless it induces changes in household perceptions about male home work. This, of course, requires that the availability of the laundry machine not alter distribution factors, and is the reason why during the intervention, we emphasized the collective ownership of the laundry machine to all household members.

7 Preliminary conclusions

The project at this stage is a proof-of-concept. We eliminate the financial barrier from the families in our sample, to carefully measure the benefits of having a washing machine in the dimensions mentioned above.

The results presented in this draft correspond to the first follow-up, 6 months after the intervention. In this draft of the report we focus on preliminary findings of impacts. The results suggest that despite the small sample size, we find evidence that the intervention is having significant impacts. For instance, the time devoted to laundry decreased, and this time is apparently being used by mother to work more, being employed more often and increasing the share of market to home work. Household income is rising, probably as a results of increased fathers' income and increased female labor participation. Female empowerment is also rising.

There are some interesting and puzzling results, such as the significant increase in fathers' time devoted to childcare and decreased leisure, or the decrease in household savings. We will investigate these findings in the qualitative component.

We will collect another follow-up 18 months after the intervention, to assess the impacts on the variables presented here, as well as on child development.

In the future we intend to design a financially sustainable version of this intervention so that we can create the missing market without turning to government subsidies.

References

Acemoglu, D., C. Garcia-Jimeno, and R. O'Keefe-O'Donovan (2016): "Strategic Interactions in the Estimation of Spillover Effects," University of Pennsylvania.

Alesina, A., P. Giuliano, and N. Nunn (2013): "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics*, 128(2), 469–530.

Apps, P., and R. Rees (1988): "Taxation and the Household," *Journal of Public Economics*, 35, 355–369.

Badel, A., and X. Peña (2010): "Decomposing the Gender Wage Gap with Sample Selection Adjustment: Evidence from Colombia," *Revista de Analisis Economico*, 25(2), 169–191.

Baird, S., A. Bohren, C. McIntosh, and B. Ozler (2015): "Designing Experiments to Measure Spillover Effects," PIER Working Paper.

Basu, K. (2006): "Gender and Say: A Model of Household Behavior with Endogenously Determined Balance of Power," *Economic Journal*, 116, 558–580.

Becker, G. (1991): *Treatise on the Family*. Harvard University Press, Cambridge, MA.

Bianchi, S., M. Sayer, and J. Robinson (2000): "Is Anyone Doing the Housework? Trends in the Gender Division of Household Labor," *Social Forces*, 79(1), 191–228.

Blundell, R., P.-A. Chiappori, and C. Meghir (2005): "Collective Labor Supply with Children," *Journal of Political Economy*, 113(6), 1277–1306.

Boserup, E. (1970): *Women's Role in Economic Development*. George Allen and Unwin Ltd., London, UK.

Botticini, M., and A. Siow (2003): "Why Dowries?," *American Economic Review*, 93, 1385–1398.

Bourgignon, F., M. Browning, and P.-A. Chiappori (1995): "The Collective Approach to Household Behavior," Working Paper No. 95-04, DELTA.

Brines, J. (1993): *The Exchange Value of Housework*. University of Washington Press, Seattle, WA.

Brown, H., H. Lewin-Epstein, and M. Baumgartner (2008): "Perceived Equity in the Gendered Division of Household Labor," *Journal of Marriage and Family*, 70(5), 1145–1156.

Browning, M., and P.-A. Chiappori (1998): "Efficient Intra-Household Allocations: A General Characterization and Empirical Tests," *Econometrica*, 66, 1241–1278.

Browning, M., P.-A. Chiappori, and Y. Weiss (2011): "Family Economics," .

Burda, M., D. Hamermesh, and P. Weil (2007): "Total Work, Gender and Social Norms," NBER.

Chiappori, P.-A. (1988): "Rational Household Labor Supply," *Econometrica*, 56, 63–90.

——— (1997): "Introducing Household Production in Collective Models of Labor Supply," *Journal of Political Economy*, 105, 191–209.

Choo, E., and A. Siow (2006): "Who Marries Whom and Why?," *Journal of Political Economy*, 114, 175–201.

Couprie, H. (2007): "Time Allocation within the Family," *Economic Journal*, 117, 278–305.

Fernandez, C., and A. Sevilla-Sanz (2006): "Social Norms and Household Time Allocation," .

Fernandez, G., M. Dallo, C. Duran, F. Caperchione, S. Gutierrez, and J. Dapueto (2010): "Cuestionario sobre Calidad de Vida Pediátrica (PedsQL) versión 4.0: fase inicial de la adaptación transcultural para Uruguay," *Archivos de Pediatría del Uruguay*, 81(2), 91–99.

Fernandez, M. (2006): "Determinantes del Diferencial Salarial por Género en Colombia, 1997-2003," *Desarrollo y Sociedad*, 58, 165–208.

Fernandez, R. (2007): "Alfred Marshall Lecture: Women, Work, and Culture," *Journal of the European Economic Association*, 5, 305–332.

Gonzalez, M., T. Landero, R. Hernandez, and J. M. de la Rubia (2009): "Cuestionario de Burnout para amas de casa (CUBAC): evaluación de sus propiedades psicométricas y del Modelo Secuencial de Burnout," *Universitas Psychologica*, 8(2), 533–544.

Greenwood, J., A. Sheshadri, and M. Yorukoglu (2005): "Engines of Liberation," *Review of Economic Studies*, 72, 109–133.

Heckman, J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261–294.

Heising, J. P. (2011): "Who Does More Housework: Rich or Poor? A Comparison of 33 Countries," *American Sociological Review*, 76(1), 74–99.

Iyigun, M., and R. Walsh (2007): "Building the Family Nest: Premarital Investments, Marriage Markets, and Spousal Allocations," *Review of Economic Studies*, 74, 507–535.

Kroenke, K., R. Spitzer, J. Williams, and B. Lowe (2009): "An ultra-brief screening scale for anxiety and depression: the PHQ–4," *Psychosomatics*, 50(6), 613–621.

Lewin-Epstein, N., and H. Stier (2007): "Policy Effects on the Division of Housework," *Journal of Comparative Policy Analysis: Research and Practice*, 9(3), 235–259.

Marini, M., and B. Shelton (1993): "Measuring Household Work: Recent Experience in the United States," *Social Science Research*, 22(4), 361–382.

Murphy, K. (2012): *Machine Learning. A Probabilistic Perspective*. the MIT Press, Cambridge, MA.

NORC (2015): "General Social Survey," .

Olson, J. (1981): "The Impact of Housework on Child Care in the Home," *Family Relations*, 30(1), 75–81.

Peña, X. (2006): "Assortative Matching and the Education Gap," Borradores de economia, Banco de la Republica.

Pfeiffer, J., S. Gloyd, and L. Li (2001): "Intrahousehold Resource Allocation and Child Growth," *Social Science and Medicine*, 53, 83–97.

Rapoport, B., C. Sofer, and A. Solaz (2009): "Household Production in a Collective Model: Some New Results," *Journal of Population Economics*, 24(1), 23–45.

Rollnick, S., W. Miller, and C. Butler (2008): *Motivational Interviewing in Health Care*. The Guilford Press, New York, NY.

Rosenbaum, P. (2007): "Interference Between Units in Randomized Experiments," *Journal of the American Statistical Association*, 102(466), 191–200.

Shelton, B., and D. John (1996): "The division of household labor," *Annual Review of Sociology*, 22, 299–322.

Udry, C. (1996): "Gender, Agricultural Production, and theTheory of the Household," *Journal of Political Economy*, 104, 1010–1046.

van Klaveren, C., B. van Praag, and H. M. van der Brink (2008): "A Public-Good Version of the Collective Household Model: an Empirical Approach with an Application to British Household Data," *Review of Economics of the Household*, 6, 169–191.

Varni, J. (2015): "The PedsQL. Measurement Model for the Pediatric Quiality of Life inventory," .

Zeyneloglu, S., and F. Terzioglu (2011): "Development and psychometric properties gender roles attitude scale," *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 40.