Conditional Cash Transfers and Intra-Household Time Allocation: Evidence from Urban Mexico

Andrea Flores[†]
Washington University in St. Louis

May 16, 2019

Abstract

Conditional Cash Transfer (CCT) programs have become an increasingly popular instrument for poverty alleviation in developing countries, particularly in Latin America. A salient feature shared by most of these programs is its gendered focus as the transfers are placed directly in the hands of women, thereby impacting the decision making process in recipient households. Using the official evaluation data from the urban component of the Oportunidades program in Mexico and implementing both, a difference-in-differences and a matching difference-in-differences approach, I investigate the impact of the program on intra-household time allocation in poor urban households. The results suggest there is a significant increase in women's weekly leisure hours in two-parent beneficiary households mostly stemming from a significant reduction in weekly hours devoted to home production.

1 Introduction

Conditional cash transfer (CCT) programs are a form of social assistance generally designed to increase investments on children's human capital in poor households by conditioning the transfers on observable outcomes such as school attendance or regular health checkups. These programs are based on the premise that poor households are not able to invest enough in human capital accumulation, health and nutrition, therefore trapping them in a vicious cycle of poverty. As developing countries face the pressing challenge of poverty alleviation, these programs have become the flagship policies for breaking the intergenerational transmission of poverty. Particularly in Latin America and the Caribbean, as of 2011 there were 18 CCT programs in operation with countries such as Bahamas, Barbados, Belize and Suriname in the process of designing one (Stampini and Tornarolli 2012). Despite the wide heterogeneity among these programs in terms of their design, implementation and evaluation, a common feature is that they explicitly target the transfers to women as a way to improve their position at home. Such a gendered focus is unsurprising as policymakers become increasingly aware of the connection between female empowerment and policy effectiveness suggested by the existence of evidence that key development outcomes are dependent upon women's ability to negotiate effectively for favorable resource allocations (Doss 2013; Duflo 2003; Duflo 2012).

In the particular case of Mexico, a significant amount of work done on Progresa/Oportunidades has been focused on its impacts in rural areas. This focus on the rural implementation of the program is based both on its popularity and the experimental nature of its data. To give a brief overview of this strand

[†]Department of Economics, Washington University in St. Louis, Campus Box 1208, One Brookings Drive, St. Louis, MO 63130. Email: andrea.flores@wustl.edu.

of the literature, this work focuses on outcomes regarding food consumption (Attanasio and Lechene 2002; Attanasio and Lechene 2014), school enrollment and performance (Todd and Wolpin 2006; Dubois et al 2011; Behrman and Parker 2001; Skoufias 2005), child labor (Skoufias et al, 2001), fertility (Todd and Wolpin 2006; Darney et al) and household time allocation (Parker and Skoufias, 2000; Skoufias and Di Maro, 2006). Work on the urban component of the program is more scarce and is focused on the impact of the program on poor urban households' food consumption (Angelucci and Attanasio 2013; Angelucci, Attanasio and Di Maro, 2012), savings and in-kind transfers (Angelucci, Attanasio and Di Maro, 2012) as well as the methodological issues of the evaluation data (Angelucci, Attanasio and Shaw 2005).

Focusing on the work addressing the impact of Progresa on intra-household time allocation, Parker and Skoufias (2000) explore the impact of Progresa on work, leisure and time allocation for children and adults in rural households. While they find evidence of an increase in the amount of time children dedicate to school, they find some weak evidence that the program slightly reduced women's participation in domestic work but no overall significant impact of the program on leisure time of either men or women. Similarly, Skoufias and Di Maro (2006) report no significant effect of the program on adult labor supply and leisure time. Both papers construct a residual measure of leisure as the difference between 24 hours and the time spent on all 18 activities reported in the time use module available only in the June 1999 round. However, the authors underline a weakness of the data used: the reference period of one day used in the data is not entirely ideal as for some individuals, the survey may refer to a day which could not be considered typical in terms of normal activities. Using this same data, Rubio-Codina (2010) finds no overall impact on adult's leisure but observes some cross-substitutability of home production hours between mothers and teenage daughters as her results suggest a significant increase in the home production hours of prime-age mothers.

Given the gendered nature of the *Oportunidades* program, an important strand of the literature that is relevant to the analysis of CCTs is the growing targeting literature. This strand is comprised of a growing number of empirical studies supporting the notion that programs targeting benefits to a particular household member ultimately has an impact on the allocation of resources within the household. Among these studies we find Duflo (2003) and Lundberg, Pollak and Wales (1997) who find evidence of changes in households' consumption structure when benefits are targeted to female heads. That is, who gets to receive a benefit in the household matters: it affects the balance of power within the household and this is reflected in the observed household outcomes. The latter constitutes a rejection of the standard unitary framework's implication that targeting is ineffective since income is pooled at the household level for decision-making purposes.

Within the context of the evalutation of CCTs, Attanasio and Lechene (2002) reject the unitary model's income pooling hypothesis using the Progresa rural evaluation data. they take the responses provided by women to five different questions about who makes important decisions in the household with a particular focus on the response to "who decides how to spend the wife's extra money?" The authors analyze the before and after responses to these questions within program beneficiaries, within potential beneficiaries in control villages and within non-program beneficiaries in treatment villages. [1] They argue

[1] Within program beneficiaries, the proportion of women responding that they decide increased from 18% to 30%, while for the rest of the questions, the proportion of women responding that the decisions were made jointly decreased, typically in favor of her making the decisions. Within the potential beneficiaries in control villages, they found a similar increase in the proportion of women who declared that they decided how to spend their extra income and that the proportion of women responding that the decisions were made jointly increased, but typically in favor of her husband making the decisions. A similar result was found among nonpoor households in treatment villages. The authors claim that they cannot exclude Hawthorne-like and/or spillover effects behind these results.

that their results show that a change in the income share of women seems to be related to changes in the decision-making process of the household, suggestive of a failure of the unitary model. Using this same data, Attanasio and Lechene (2014) argue that the collective household model is better than the unitary model at predicting the effect of exogenous increases of household income on household consumption and do not find enough evidence to reject the efficiency of household decisions. Focusing on the impact of the urban implementation of the *Oportunidades* program on the consumption of high-protein food, Angelucci and Attanasio (2013) find that the Engel curve is misspecified and fails to predict the change in food consumption observed after the transfer, which suggests an increase in the bargaining power of women. After ruling out the possibility of this misspecification arising from changes in nutrition and health information by looking at female-headed households, the authors interpret their results as a failure of the unitary model, from which the estimated Engel curves are derived.

This paper brings together both the targeting and the CCT impact evaluation literature by using the urban evaluation data of *Oportunidades* to investigate the effect of increasing women's share of household nonlabor income on the composition of time allocation within two-parent poor urban households. Acknowledging the shortcoming of the dichotomization of time between leisure and market work criticized by Apps and Rees (1997), the main object of interest in this analysis is a relatively rich measure of leisure that resembles the one constructed by Parker and Skoufias (2000) and Skoufias and Di Maro (2006) for the rural component of the program. Contrary to these papers, however, one of the main advantages of this dataset is that the time use module takes a week as the reference period and it was collected in each of the three waves available. Hence, I am able to observe changes not only between treatment and control groups but also across time, which allows for the implementation of a difference-in-differences strategy and its matching counterpart.

The results of this analysis are of particular interest given concerns raised in the literature that placing *both* the benefits and the conditionalities' burden on women might perpetuate their roles as child-bearers and care-givers and policymakers' concern that giving extra cash to households might discourage their adult members to engage in market work. The paper proceeds as follows: Section 2 describes the urban component of the program, Section 3 describes the data, Section 4 discusses the theoretical framework through which the program could impact time allocation, Section 5 describes the empirical strategies implemented, Section 6 presents the results and Section 7 provides an overall discussion of the analysis.

2 Description of the Progresa/Oportunidades Program

In the case of Mexico, Progresa was initially implemented in rural areas by the Zedillo administration in 1997. The program intervenes simultaneously in the areas of education, health and nutrition (Skoufias and di Maro 2006). The selection of beneficiaries into the program was performed in two stages. The first stage consisted of geographic targeting, selecting 506 rural areas in 7 of the 31 states to be randomly assigned to control or treatment groups. Once eligible communities were identified, the second stage consisted of household selection through a discriminant analysis performed using census data. The analysis was based on comparing the household marginality index against a local cutoff and selecting those households falling below this cutoff. Once a household is deemed eligible, the program targets the female family head, conditioning the receipt of the health and nutrition component of the transfer to regular health clinic checkups and the educational component to school attendance. The program evaluation was administered by the International Food Policy Research Institute, which yielded results that were, in general, positive (Parker and Todd 2017). As its initial implementation was deemed a success in key areas such as school enrollment and child health outcomes, the next step was to increase the scale and scope of the program. Following this new objective, the program was then expanded to include semi-urban and urban localities. With such expansion and change of federal administration came the change of the program's name to Oportunidades in 2002 (Levy 2006). This urban implementation differs from its rural counterpart mainly in two ways: the evaluation design and the targeting and incorporation of beneficiaries (Angelucci and Attanasio 2009).

In terms of the evaluation design, given feasibility issues, the nature of the evaluation design is quasi-experimental. Using the 2000 census and the INEGI's 2000 National Survey of Household Income and Expenditure, and taking city blocks as the unit of analysis, the program was intended to be first offered in city blocks with the highest incidence of poverty. The administration proceeded to compute of a propensity score at the city block level which predicted the city block's probability of being part of the intervention. Upon the identification of a representative sample of intervention blocks, these propensity score was then used to match these to a sample of blocks in control areas on the basis of similar propensity scores (SEDESOL 2005).

In terms of the targeting and incorporation of beneficiaries, we have that household's uncertainty regarding their eligibility status might have impacted the program's take-up rate. While rural households knew their eligibility status at the beginning of the program, urban households had to first approach local registration offices to inquire about their eligibility in terms of an estimated poverty index. Even though the administration invested a significant amount of resources in announcing the availability of the program and the location of registration offices, Angelucci and Attanasio (2009) argue that this might have impacted potentially eligible households' program participation at least during the first two years as they might have simply ignored the program's existence to begin with or because they were too uncertain about their eligibility status.

Similarly to its urban counterpart, however, the urban implementation retained its three main areas of intervention: health, nutrition and education. In the education component, grants are available for primary, secondary and high school students with the amount of these grants increasing with the grade achieved and higher for girls upon the start of secondary school. There are additional transfers provided by the program for the acquisition of school supplies at the beginning of the school year. Moreover, the urban implementation provides a savings plan called *Jovenes con Oportunidades* for high school students which is composed by a grant that grows every year starting on ninth grade and is paid to the students until graduation as a way to foster higher completion rates among poor urban students. In the nutrition component, households receive a transfer for food purchases plus an additional transfer under *Apoyo Alimentario Complementario* to compensate for increases in international food prices. In the health component, beneficiary households receive a basic health package which is composed of free preventive consultations and informational health talks.

3 Data

Given that the focus of this paper is on the urban component of the Oportunidades program, I obtain the data from the PROSPERA External Evaluation datasets provided by the program's administration. Particularly, I focus on the sociodemographic module of the Urban Evaluation Surveys (ENCELURB) to obtain information regarding household consumption and intra-household time allocation decisions for the period of time comprised by 2002-2004. This section provides a description of the ENCELURB with a few warnings about the methodological issues posed by the evaluation design, an overview of the key features of the data available for each of the waves provided, and relevant descriptive statistics.

3.1 Methodological Issues Regarding the Data

The non-random implementation of the urban implementation is a feature of this component of the program that must be taken into consideration when estimating the impacts of the program on intra-household time allocation decisions. Moreover, as Angelucci, Attanasio and Shaw (2005) and Angelucci and Attanasio (2013) mention, there is an oversampling of participating households in the evaluation samples collected by the program administration. In order to make the observed sample more representative, these authors create an additional weight using the administrative data regarding the transfers made to beneficiary households. This is going to be an issue I will have to revisit as I perform further robustness checks on the initial estimates obtained, especially for the identification of the average intention to treat parameter that takes city block assignment to intervention and control zones as the treatment variable. Another issue that must be taken into consideration throughout the analysis is the low program participation in the urban localities, contrary to what was observed in program's rural component. This might be related to the way in which potential beneficiaries were selected in urban areas and the uncertainty of program eligibility that applicants faced at the time they visited the program module for the first time.

3.2 The ENCELURB

The ENCELURB data was gathered in three waves. The first wave captured baseline information and was gathered in the fall of 2002, once beneficiary households had been determined but prior to the provision of any benefits. The second wave captured the first follow up information, being gathered in the fall of 2003. The third wave captured the second follow up information, being gathered on the fall of 2004. In the following subsection, I provide a description of the relevant information contained in the three different waves and how the dataset structure in the three waves varies across time.

3.2.1 ENCELURB 2002

As aforementioned, this wave contains baseline data. Originally, the individual table corresponding to this wave contains 76,002 observations and the household table contains 17,201 observations. Once these two tables are merged, the 76,002 individual observations correspond to 15,700 household observations.

The household table contains remarkably detailed information about household expenditures, breaking it down into different food and durable good items. It particularly distinguishes expenditures on adult clothing from expenditures on children clothing. Further information related to household expenditures include rent, utilities and transportation. Moreover, the table also provides thorough information on in-kind and other monetary transfers received by the household during the survey's reference year. Additional useful information found in this table directly relates to the Oportunidades program: whether anyone in the household visited a program registration office, how frequently he or she visited said registration office, the time and money spent in these visits as well as a self-reported variable of selection and subsequent incorporation into the program. The table also provides lagged information of income, occupation and employment of the household head and his or her spouse in 1999, 2000, and 2001. There is also information of self-reported decision-making power.[2] Other topics include the household percep-

[2] The questions for these measures ask: If your child gets sick, who decides when to take him/her to the hospital? If your child does not want to attend school one day, who decides whether she/he has to go or not? Whenever you need to make a decision regarding children clothing purchases, who decides whether or not to make the purchase? Who makes the decisions regarding important aspects that affect all household members? When there is extra income in the household, does the recipient

tions of violence within the community, incidence of domestic violence and whether the respondent was a victim of violence (hitting) at the hands of his/her parents when she/he was a child.

The individual table contains information about each individual's educational background, schooling status (whether currently attending school or not), use of health services, education and health-related expenditures, employment status, occupation and earnings. It also provides through information about the amount of time children spend doing homework, whether they receive any help for completing their homework and the amount of money spent in tuition, uniforms and transportation to attend school. It further provides information about individual time use along the different activities: household care, child care, food preparation, procurement of goods for the household, trash disposal, and carrying water[3].

3.2.2 ENCELURB 2003

This is the first follow-up wave provided for evaluation. Originally, the individual table corresponding to this first follow-up contains 77,764 observations and the household table contains 18,041 observations. Once these two tables are merged, there are 77,764 matched individual observations corresponding to 16,149 households. The household table follows a similar structure to the baseline one, with the exception that there is no previous labor supply information of the household head and spouse. Further main differences include the omission of information about the self-reported measures of decision-making power available in the baseline data. This poses a problem if trying to do an informal analysis of decision-making power using these questions as Attanasio and Lechene (2002) did using data from rural households. Starting in this wave, however, there is data about children's future earnings expectations in the sense that they ask the respondent what they think their child's income might be at different education levels. The structure of the individual table is also very similar to the previous wave with the exception that it now incorporates information about health talk attendance and questions about body image perceptions. In the time-use component of the table, it now contains information about the number of weekly hours spent watching television, reading and helping elderly or sick people.

3.2.3 ENCELURB 2004

This is the second follow-up wave provided for evaluation. Originally, the individual table corresponding to this second follow-up contains 72,421 individual observations and the household table contains 17,023 observations. Once these two tables are merged,there are 72,421 individual observations corresponding to 15,056 household observations. The household table follows a very similar structure to the previous follow-up wave with a few exceptions. One of these exceptions is that there is now information regarding household expenditures on women's clothing and men's clothing, which was previously unavailable. However, checking the quality of this data, there is a considerable amount of missing values,

of this extra income get to decide how to spend it? The response to these questions, except for the last one can be only the father, only the mother, both in agreement, another man in the household, another women in the household or not applicable. For the last question, the response is either yes or no (or not applicable). Further questions ask: If the extra income is received by the husband, in what would this extra income be primarily spent? If the extra income is received by the wife, in what would this extra income be primarily spent? The responses can be: food, education or child health, adult clothing and shoes, child clothing and shoes, alcohol and entertainment, furniture and other household items, transportation expenditures, other or not applicable.

[3] The questions for these measures ask: During the last week, how many hours did you spend on [relevant time use category] in total? The answers are coded in a way that if the individual performed less than an hour, a value of 0 is immediately assigned and a value of 98 for those who did not perform the activity at all. A value of 99 indicated that the respondent did not know and this was then recoded to a missing value. The observations corresponding to a value of 98 were recoded to 0 in order to understand that these observations correspond to no involvement in the activity.

which would thwart any attempt to make use of that information. Another difference with the previous wave is that in this one, respondents are asked to show the documentation they have from the program (if any). The individual table also follows basically the same structure as the previous wave.

3.3 Sample Description

Given that the focus of this paper is on household dynamics and time allocation within the household, I focus on nuclear families. That is, households with at most two adults living maritally or in the form of cohabitation with any number of children younger than 25 [4]. From an original sample of 76,002 individual observations in 2002, this sample restriction further reduces the sample to 39,272 individual observations corresponding to 7988 household observations.

Table 1 describes the number of households according to their poverty classification (poor, almost poor, non-poor) across intervention and control blocks in the ENCELURB 2002. It is worth mentioning that the poverty classification used in the following tables and throughout this paper is the one calculated in the Sociodemographic module of the ENCELURB 2002 and not on the poverty classification provided in the Screening survey at the beginning of the program. For their single counterparts, the resulting sample consists of 5,047 individual observations corresponding to 1,331 household observations.

Table 1: Distribution of Households according to their Poverty Classification across Intervention and Control Blocks, 2002

	Two-Parent		**Single-Parent**	
Poverty Classification	Control	Intervention	Control	Intervention
Poor (eligible)	2,015	3,168	217	655
Almost poor (almost eligible)	529	1,114	58	209
Non-poor (ineligible)	138	1,024	21	171
Total	2,682	5,306	296	1,035

In the case of 2003, we have that the sample restriction imposed by the analysis to be performed in this paper reduces the sample size from 77,631 individual observations to 40,764 individual observations from 8,259 households, out of which 1,102 have missing observations regarding the poverty classification of the 2002 Sociodemographic module of ENCELURB, leaving 7,157 households which we can use for the analysis. For their single counterparts, the sample size reduces to 5,741 individual observations corresponding to 1,507 household observations, out of which 5,204 individual observations (1,355 households) do not have missing values for the 2002 poverty classification. Table 2 shows how these households are distributed across intervention and control zones according to their poverty (marginality) classification.

It is worth mentioning that out of the 7,988 two-parent households observed in 2002, only 6,062 were observed in 2003 as well. In the case of one-parent households, out of the 1,331 households observed in 2002, only 887 were observed in 2003 as well. This yields the following distribution of the resulting households across intervention and control areas according to their poverty (marginality) classification.

[4] This age restriction is based on the ages specified in the 2002 ENCELURB questionnaire of the target respondents of the education component of the sociodemographic module. It also makes sense since at this point, individuals are expected to have completed at least their undergraduate studies, and no further significant investments in education are expected from the parents. This can be modified if needed.

Table 2: Distribution of Households according to their Poverty Classification across Intervention and Control Blocks, 2003

	Two-Parent		Single-Parent	
Poverty Classification	Control	Intervention	Control	Intervention
Poor (eligible)	1,730	2,874	221	672
Almost poor (almost eligible)	465	1,028	51	220
Non-poor (ineligible)	125	935	19	172
Total	2,320	4,837	291	1,064

Table 3: Distribution of Matched Households according to their Poverty Classification across Intervention and Control Blocks, 2002-2003

	Two-Parent		Single-Parent	
Poverty Classification	Control	Intervention	Control	Intervention
Poor (eligible)	1,509	2,413	142	447
Almost poor (almost eligible)	412	850	32	149
Non-poor (ineligible)	106	772	13	104
Total	2,027	4,035	187	700

In the case of 2004, the sample restriction here used reduces the original sample of 72,421 (17,023) individual (household) observations to 36,738 (7,390) individual (household) observations, out of which 31,936 (6,377) individual (household) observations have valid information regarding their poverty classification. For their single counterparts, the sample restriction reduces the original sample to 5,399 (1,428) individual (household) observations, out of which 4,830 (1,272) individual (household) observations have valid information regarding their 2002 poverty classification.

Table 4: Distribution of Households according to their Poverty Classification across Intervention and Control Blocks, 2004

	Two-Parent		Single-Parent	
Poverty Classification	Control	Intervention	Control	Intervention
Poor (eligible)	1,621	2,521	217	634
Almost poor (almost eligible)	411	898	51	198
Non-poor (ineligible)	118	808	14	158
Total	2,150	4,227	282	990

Moreover, when focusing only on those households observed throughout the three waves, we have that out of the 6,062 two-parent households matched between 2002 and 2003, only 4,842 of these households are observed in 2004 as well. This leaves a total of 25,537 individual observations matched throughout the 2002-04 period. In the case of single-parent households, we have that out of the 887 observed throughout 2002 and 2003, only 644 are observed in 2004 too. This leaves a total of 2,363 individual observations matched throughout the 2002-04 period. Table 5 shows the distribution of these households across intervention and control zones according to their poverty classification.

Table 5: Distribution of Matched Households according to their Poverty Classification across Intervention and Control Blocks, 2002-2004

	Two-Parent		**Single-Parent**	
Poverty Classification	Control	Intervention	Control	Intervention
Poor (eligible)	1,251	1,930	110	333
Almost poor (almost eligible)	334	668	22	100
Non-poor (ineligible)	79	580	6	70
Total	1,664	3,178	138	503

3.3.1 Geographic Distribution of Households in the Final Sample

Taking a closer look at the geographic distribution of households in the final sample, approximately 50% of two-parent households in control zones are located in the central states of the country while approximately 36% of the two-parent households in intervention zones are located in the southern states and 30% located in the eastern states. For single-parent households, we can see that approximately 35% of households in intervention zones are located in the southern states and approximately 36% of households are located in the eastern states. On the other hand, approximately 44% of single-parent households are located in the central states.

Table 6: Distribution of Matched Households according to their Region across Intervention and Control Blocks, 2002-2004

	Two-Parent		**Single-Parent**	
Region	Control	Intervention	Control	Intervention
North	106	132	16	17
Central	847	653	62	89
East	424	976	34	186
West	18	254	4	33
South	269	1,163	25	178
Total	1,664	3,178	138	503

It could also be useful to look at the geographic distribution of poverty in Mexico and to look at how the above information links with such distribution and the program's targeting. Using the 2000 poverty figures from the National Council for the Evaluation of Social Development Policy (CONEVAL by its spanish acronym), it is possible to identify five regions of marginality[5] in Mexico. As can be seen in Figure 1, these five regions are demarcated by the five ranges of poverty incidence established by the CONEVAL. It is noticeable from Figure 1 that the highest incidence of poverty in Mexico is concentrated in the southern region, while the lowest incidence of poverty is concentrated at the north. Furthermore, it can be seen that Mexican states are not evenly distributed across these five regions. Hence, using the data from the CONEVAL where it is possible to retrieve the actual percentage documented for each state in 2000, I was able to assign each individual state to the corresponding range. It is worth mentioning that the CONEVAL's data for 2000 is derived from the 2000 National Survey of Household Income and

[5]It is worth mentioning that at the time of data collection, the CONEVAL reported three different dimensions of poverty: nutritional poverty, poverty of means and wealth poverty. Given the focus of the program on nutrition, health and education, I focused on the poverty of means as this is defined as "the lack of monetary capacity to afford the value of a basic food basket, cover medical and educational expenses even after devoting the totality of household income only to the latter" (CONEVAL, 2000).

Expenditures (ENIGH) provided by the Mexican National Institute of Statistics and Geography (INEGI), which was in turn used by the Oportunidades' administration for the initial discriminant analysis used to target poor city blocks. This allowed me to check the distribution of households across the five regions demarcated using the same data used for program targeting.

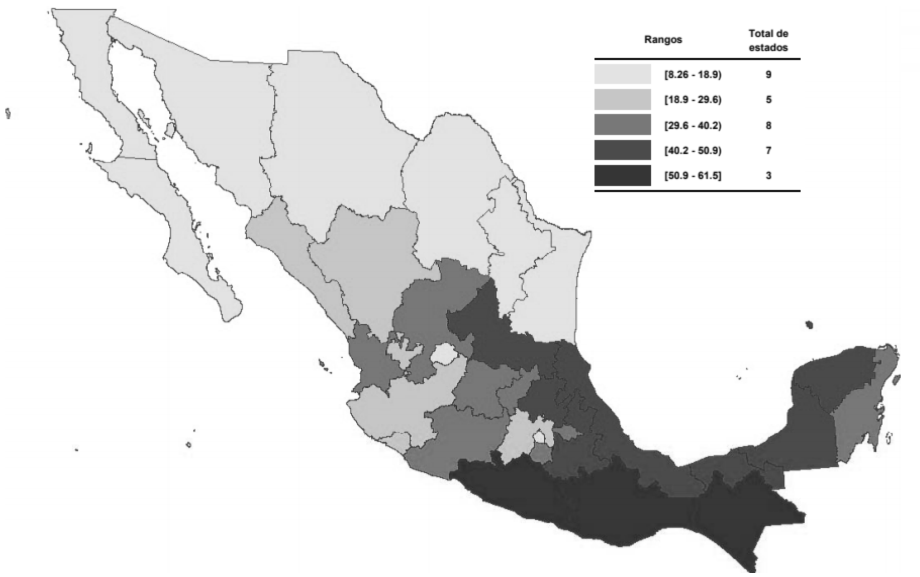

Figure 1: Demarcation of five major regions according to the percentage of state population living in poverty, 2000. *Source:* CONEVAL using the XII Population and Household 2000 Census and the 2000 National Survey of Household Income and Expenditure, ENIGH

Table 7 shows the distribution of households in intervention and control blocks across the five regions. It is noticeable that around 66% of two-parent households in intervention blocks are located in high to extreme poverty areas, while approximately 53% of two-parent households in control blocks are located in low to very low poverty areas. The same pattern can be seen for single-parent households. This is unsurprising as it can be seen in Table 5 that a relatively high proportion of households in intervention zones are located in the southern and eastern states while a higher proportion of households in control zones are located in the central states. This is consistent with the way the program administration selected intervention city blocks, since it was targeting the program to those city blocks with the highest incidence of poverty as mentioned in Section 2.

Table 7: Distribution of Matched Households according to their Poverty Region across Intervention and Control Blocks, 2002-2004

	Two-Parent		**Single-Parent**	
Poverty Region	Control	Intervention	Control	Intervention
Extreme	150	884	14	130
High	374	1211	35	227
Moderate	253	560	18	99
Low	811	391	58	30
Very Low	76	132	16	17
Total	1,664	3,178	138	503

The dispersion of households across different regions in the country might be an issue to keep in mind for the implementation of the intended empirical strategy and a further consideration for robustness checks as it was also pointed out by Angelucci and Attanasio (2013) and Angelucci and Attanasio (2009). This will be revisited in Section 5 and Section 7 in the discussion of the assumptions imposed by the empirical strategy and the potential sources of bias. Furthermore, such geographic dispersion might be behind the systematic socioeconomic differences observed between households in intervention and control city blocks and described in the next section of this subsection.

3.3.2 Descriptive Statistics for Final Sample

Table 8 shows some basic descriptive statistics for the households in the final sample. It is worth mentioning that education refers to the total number of years completed by the individual. The categories for children and children in school correspond to the number of children present in the households by age groups 0-5, 6-12, 13-15 and 16-20 for the former and for the number of children in the households who are in school for the different age groups for the latter. The income shock variable indicates whether the household suffered of some loss that would require its members to adjust their expenditures. Such loss could be caused either by a household member's death, job loss, business loss or natural disaster loss during the survey's reference period (12 months). This table also reports information about the total number of people living in the household as well as indicators of inadequate housing. Such indicators of inadequate housing are in the form of dirt floors and the absence of running water in the family's dwelling.

Table 8: Means of Socioeconomic Indicators of Eligible Households in Control and Intervention City Blocks, 2002

		Intervention		Control		
		Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Diff**
Women	Age	31.767	8.077	31.954	7.691	-0.186
	Education (years)	6.876	1.757	7.114	1.721	-0.238***
	Employed 2002	0.369	0.483	0.268	0.443	0.101***
	Employed 2001	0.238	0.426	0.176	0.381	0.0615***
	Employed 2000	0.23	0.421	0.15	0.357	0.0800***
	Employed 1999	0.195	0.397	0.137	0.344	0.0587***
Men	Age	35.428	9.451	35.311	8.694	0.118
	Education (years)	6.748	1.714	6.942	1.684	-0.193***
	Employed 2002	0.984	0.124	0.98	0.142	0.00478
	Employed 2001	0.971	0.168	0.969	0.172	0.00141
	Employed 2000	0.967	0.177	0.967	0.18	0.000832
	Employed 1999	0.966	0.182	0.954	0.209	0.0115
Household	Household size	5.146	1.523	5.325	1.546	-0.178***
	Number of children	3.053	1.454	3.23	1.489	-0.177***
	Number of children in school	1.675	1.384	1.743	1.392	-0.0683
	Poverty Incidence	45.971	12.726	34.975	13.166	11.00***
	Income Shock	0.283	0.45	0.202	0.402	0.0801***
	Home Ownership	0.725	0.447	0.732	0.443	-0.00766
	Bank Account	0.006	0.079	0.008	0.089	-0.00179
	Rooms	1.376	0.685	1.56	0.883	-0.184***
	No water connection	0.904	0.295	0.830	0.376	0.0738***
	Dirt floor	0.533	0.499	0.372	0.484	0.161***

We can see in Table 8 that there are systematic differences among eligible households and their members across intervention and control city blocks. On average, women living in intervention city blocks tend to be less educated and more likely to have been employed in 1999-2002 and men living in intervention city blocks tend to be less educated than their control counterparts. Moreover, on average, households located in intervention city blocks are smaller, comprised of fewer children, in states with a higher incidence of poverty, more likely to have suffered an income shock in 2001, reside in dwellings with fewer rooms, inadequate water connection systems and floors made out of dirt. These systematic differences between these comparison groups then provides further motivation for the implementation of a matching difference-in-difference estimator that matches each household in intervention city blocks with an observably similar household in control city blocks. This is further discussed in Section 5.

3.4 Time Allocation: Defining Leisure Based on Time-Use Variables in the ENCELURB

Following Aguiar and Hurst (2007) and Aguiar, Hurst and Karabarbounis (2012), I define the following time-use groups to be analyzed for households across comparison groups in the empirical analysis: market work, core home production, procurement of goods and services and child care. Moreover, following Aguiar and Hurst (2007), it could be possible to further dissect child care into three components: primary, educational, and recreational. In the individual datasets of the ENCELURB, it is possible to find information about the number of weekly hours parents spend helping their children with homework, i.e. educational child care. However, this information is based on the responses provided by the children in the household and not by the parents themselves, as is the case with the other time-use categories. Below, I define the major time-use groups according to the information provided in the ENCELURB.

- Market Work: Primary job work hours and secondary job work hours
- Core Household Production: Food preparation, household care (doing laundry, dusting, ironing, doing dishes, vacuuming and maintenance), trash disposal and carrying water
- Procurement of goods and services: shopping for household items
- Child care: in all three dimensions discussed by Aguiar and Hurst (2007). In this group, it is possible to obtain information from the dataset about the educational subgroup.

To compare alternative measures of leisure used in the literature, I focus on two leisure definitions. I define Leisure Definition 1 as

$$L_1^i = \bar{T} - M^i \tag{1}$$

where I have set $\bar{T}$, the total weekly time endowment available to all individuals in the household, at 112 hours to allow for individuals to spend 8 hours on sleep and personal care and M^i denotes the amount of weekly hours spent in market work. That is, Leisure Definition 1 is the component of total time endowment (after accounting for sleep and personal care) that is not spent in market work.

On the other hand, the level of disaggregation of the time use data provided by the ENCELURB allows me to construct a richer definition of leisure, called Leisure Definition 2, defined as

$$L_2^i = \bar{T} - M^i - H^i - Ch^i \tag{2}$$

where M_i refers to weekly market hours, H_i to weekly total home production hours, and Ch_i to weekly child care hours.

4 Theoretical Framework

As mentioned in Section 1, the main shortcoming of the standard unitary model for analyzing targeted benefits such as CCTs is that it ignores the collective nature of the household and the individuality of its members by positing that household outcomes can be described as stemming from the decisions of a single agent. Hence, under this framework, who gets to receive a particular benefit should have no impact on the allocation of resources within the household. However the implications of this theoretical framework have been systematically rejected by the growing targeting literature (Thomas 1990; Lundberg, Pollak and Wales 1997; Duflo 2014).

Given that the research question at hand acknowledges the individuality of household members, a more adequate and popular alternative conceptual framework to analyze the impact of the *Oportunidades* cash transfer on the intrahousehold time allocation is the collective household model. This model assumes that household outcomes are Pareto efficient irrespective of the bargaining mechanism behind these outcomes. Hence, following Blundell, Chiappori and Meghir (2005), we can consider a collective model of a household consisting of two adult decision-makers (where A denotes the wife and B denotes the husband) and children with no bargaining power of their own. Hence, the allocation of time and consumption between the two decision-makers can be described as the result of the following maximization problem.

$$\max_{l^A,l^B,h^A,h^B,c^A,c^B,c^H} \lambda(w^A,w^B,y^A,y^B,z)U^A(l^A,c^A;Q)+(1-\lambda(w^A,w^B,y^A,y^B,z))U^B(l^B,c^B;Q) \tag{3}$$

s.t.

$$c^A+c^B+c^H=w^Am^A+w^Bm^B+y^A+y^B \tag{4}$$
$$Q=F(h^A,h^B,c^H) \tag{5}$$
$$h^i+m^i+l^i=T\ (i=A,B) \tag{6}$$

where T is the total time endowment available to each decision-maker in the household, m^i refers to the amount of time allocates to market work, h^i corresponds to the amount of time devoted to household production (which within this framework includes child care for simplicity), and l^i corresponds to the amount of time individual devotes to leisure. Moreover, y^i denotes the non-labor income of individual i and λ denotes the Pareto weight of the wife, which is a function of prices, non-labor income and distribution factors (variables that affect the household allocation of resources through their effect on the Pareto weight). Moreover, we have that Q is a domestic good that is publicly consumed in the household and takes as inputs home time (h^A and h^B) and expenditures on other public goods (c^H), with the production function $F(h^A,h^B)$.[6] Furthermore, this domestic good is assumed to be not marketable[7], which then implies that its price, $P(w^A,w^B,y^A,y^B,z)$ is endogenous and is, in general, a function of

[6] Q can be an unobservable good. For instance, Cherchye, de Rock and Vermeulen (2012) consider the case in which Q could be the utility of having a child or the utility of having a clean home.

[7] If we were to consider the case in which the output of home production is marketable, the separability implication of this assumption implies that the demand side is divorced from the production side, and therefore, we can see that changes in non-labor income would only affect the demand side (in this case, being leisure of particular interest) within this framework in the same way it did in the absence of home production hours. More importantly, we can see that the only way in which home production hours can be affected is through technological changes in the household's production function, through changes in the individuals' wages or through changes in the market price of the domestic good.

wages, non-labor income and the decision process itself.[8,9] Household income can be allocated to a Hicksian composite good whose price can be normalized to one and which can be used for the private expenditures of the two adult members (c^A and c^B) and for the public goods used in the production of the domestic good (c^H).

The presence of public goods introduces externalities into the model and requires for these to be dealt with at the level of the household in order for the household outcomes to be efficient. Once these externalities have been addressed, then the adult household members can proceed to choose their private consumption optimally. This then leads to the notion of the *conditional sharing rule*. The intuition is that the household problem can be broken down into two main stages. In the second stage, typically referred to as the intra-household allocation stage, each household member solves the following problem privately

$$\max_{l^i, c^i} U^i(l^i, c^i; Q) \tag{7}$$

s.t.

$$c^i + w^i l^i \leq \rho^i \tag{8}$$

where ρ^i is the conditional sharing rule of member i, which is defined as the part of residual household nonlabor income allocated to individual i after accounting for expenditures on the domestic good. Both ρ^i and Q are taken as given (fixed) at this stage as both are outcomes of the first stage of the household's problem, which is typically referred to as the resource allocation stage. Defining $V^i(w^i, \rho^i, Q)$ for $i \in \{A, B\}$ as the indirect utility functions corresponding to the individual programs in the intra-household resource stage, the household's resource allocation stage then consists of solving the following maximization problem

$$\max_{\rho^A, \rho^B, Q} \lambda(w^A, w^B, y^A, y^B, z)V^A(w^A, \rho^A, Q) + (1 - \lambda(w^A, w^B, y^A, y^B, z))V^B(w^B, \rho^B, Q) \tag{9}$$

s.t.

$$\rho^A + \rho^B + P(w^A, w^B, y^A, y^B, z)Q = y^A + y^B \tag{10}$$

$$\rho^A + \rho^B = y^A + y^B - P(w^A, w^B, y^A, y^B, z)Q \tag{11}$$

We can then turn to the household production process. The optimal choice of inputs is then dictated by productive efficiency (i.e. cost minimization) upon the optimal consumption of the domestic good agreed upon in the resource allocation stage. Hence, we can take the cost function $P(w^A, w^B, y^A, y^B)Q$ and apply Shephard's lemma to obtain the conditional factor demands for the inputs h^A, h^B and c^H.

Within this framework, we can then think about what could be the impact of an increase in y^A on both l^A and l^B, which is the primary focus of this paper. We know that given the existence of the conditional sharing rule, the observable demand for leisure of each adult in the household has the following form

$$l^A = \tilde{l}^A(w^A, \rho^A(w^A, w^B, y^A, y^B, z)) \tag{12}$$

$$l^B = \tilde{l}^B(w^B, \rho^B(w^A, w^B, y^A, y^B, z)) \tag{13}$$

[8] In this case, P is referred to as the "imputed price" of the domestic good at the equilibrium level of public expenditures (Apps and Rees, 1997).

[9] The implicit price, P, becomes a function of only wages under the assumption of the linear homogeneity of the home production function, F (Apps and Rees 1997; Chiappori 1997).

where $\tilde{l}^A$ and $\tilde{l}^B$ denote the Marshallian demand for leisure of both household members.

The effect of a change in y^A on each adult's leisure consumption can then be described by the following

$$\frac{\partial l^A}{\partial y^A} = \frac{\partial \tilde{l}^A}{\partial \rho^A}\frac{\partial \rho^A}{\partial y^A} \tag{14}$$

$$\frac{\partial l^B}{\partial y^A} = \frac{\partial \tilde{l}^B}{\partial \rho^B}\frac{\partial \rho^B}{\partial y^A} \tag{15}$$

Hence, we can see that the income effect in this collective framework has two components: the standard income effect, embodied in $\frac{\partial \tilde{l}^A}{\partial \rho^A}$ and $\frac{\partial \tilde{l}^B}{\partial \rho^B}$, and what can be thought of as a collective effect, embodied by $\frac{\partial \rho^A}{\partial y^A}$ and $\frac{\partial \rho^B}{\partial y^A}$. If leisure is a normal good, then we know from the standard labor supply model that the first term can be signed as positive. Moreover, within this framework, this effect can be either magnified or mitigated by the collective effect. Contrary to the case without home production, there is a subtlety in this collective effect. It not only embodies the effect of the change in y^A on the decision structure of the household (described by the Pareto weight, λ), but also the effect of y^A on both the level and the *implicit* price of the domestic good Q.[10] Given that the Oportunidades program induces an increase in the wife's non-labor income, the results obtained from the empirical analysis undertaken in the rest of this paper could help us sign the collective effect within this framework. Furthermore, disentangling how much of this effect is coming from a change in the household decision-making structure and how much is coming from the impact on the allocation of home production hours could potentially become the focus of further theoretical work.

5 Empirical Strategy

5.1 Difference-in-Differences

Given the quasi-experimental design of the Oportunidades' evaluation data, exploiting its panel nature that allows for the availability of baseline and follow-up data for the same households and individuals, as well the classification of control and intervention zones, I implement a difference-in-differences strategy. I then take the following regression equation

$$y_{i,t} = \beta_0 + \beta_1 Treat_i + \beta_2 W2_t + \beta_3 W3_t + \beta_4 (Treat_i \times W2_t) + \beta_5 (Treat_i \times W3_t) + \boldsymbol{\gamma}\mathbf{X}_i + \epsilon_{i,t}$$

where y_{it} corresponds to the amount of weekly hours individual i devotes to the different time use categories defined in subsection 3.4 at time t. $\mathbf{X}_i$ contains a rich set of controls for household-level sociodemographic and economic characteristics, dwelling characteristics and baseline employment and education. Among the household characteristics, I include the household size, an indicator for whether there is a child in school in individual i's household, and whether individual i's household has experienced an income shock as defined in Subsection 3.3 and reported in the baseline survey. Controls for dwelling characteristics include a set of indicator variables regarding inadequate housing conditions such as lack of electricity, lack of running water in the house, dirt floors, walls made of anything but concrete, lack of a proper sewage system, in addition the continuous variable of the number of rooms in the

[10] Under the assumption of constant returns to scale, Chiappori (1997) shows that the price of the domestic good becomes a function of wages alone and, therefore, the ratio of home production hours by both spouses is determined only by the ratio of their wages. Hence, this assumption on the home production function shuts down the implicit domestic good price as a channel through which y^A affects the intra-household allocation of leisure.

dwelling. At the individual level, controls include age, age squared, total years of education, indicators of lagged employment in 2001, 2000 and 1999 and household headship, I also include i's spouse's age, education and lagged employment indicators for 2001, 2000 and 1999.

Furthermore, $Treat_i$ indicates whether individual i is part of a household in the treatment group or in the control group. For the objective of this paper, it would be helpful to consider two definitions of treatment. One of these could be based on whether a household has received any transfer from the program, thereby being incorporated in the program. The second one could be based on a simpler definition that depends on whether a household resides in a control or intervention city block. That is, $Treat_i$ could be defined as either

$$Treat_i = \begin{cases} 1 & \text{cla\_soc} = \text{poor } \& \text{ transfer03} = 1 \\ 0 & \text{cla\_soc} = \text{poor } \& \text{ transfer03} = 0 \end{cases}$$

or

$$Treat_i = \begin{cases} 1 & \text{zona} = \text{intervention } \& \text{ cla\_soc} = \text{poor} \\ 0 & \text{zona} = \text{control } \& \text{ cla\_soc} = \text{poor} \end{cases}$$

For the first definition, I focus on whether an eligible (poor) household has been a recipient of the transfer. The socioeconomic dataset of 2002 contains a variable called *incorp* that captures the program incorporation status of each household as of 2002. However, Angelucci, Attanasio and Shaw (2005), suggest the use of official administrative data on transfers made to participant households to construct an own indicator of program incorporation. For this matter, I used this dataset to create the variable *transfer03* to indicate whether the household received any transfer in 2003. The rationale behind this is that the data for the ENCELURB 2002 was collected during the fall of 2002 and by 2003 the incorporated households should have been receiving some amount of transfers from the program during any of the bimesters of that year. It is worth mentioning that while there are some differences in the distribution of households across treatment and control groups under both definitions, these differences are not significant. On the other hand, for the second definition of treatment, I use the variable $zona$ provided in each of the socioeconomic datasets, which is an indicator for whether a household resides in an intervention or control city block according to the program records.

Moreover, $W2_t$ indicates whether $t = 2003$ and $W3_t$ indicates whether $t = 2004$. Hence, β_4 constitutes the one-year difference-in-differences estimate reported in the next section and β_5 constitutes the two-year difference-in-differences estimate reported in the next section as well.

It is worth mentioning that I make one further restriction in the regressions ran to obtain the results presented in the next section. For the purpose of the analysis, I only keep those observations for which leisure is non-missing for both comparison years. This ensures that I am keeping observations with non-missing values on all time use categories by the way the leisure measure was constructed while keeping the same sample size for all time-use categories. That is, I make sure that the same individuals are being compared in every time-use category so that the results are easier to interpret.

5.2 Matching Difference-in-Differences

5.2.1 Description of the Estimator for the Evaluation of *Oportunidades*

Building upon the work of Heckman, Ichimura and Todd (1997) and Rosenbaum and Rubin (1983) and exploiting the longitudinal nature of the ENCELURB dataset, I intend to implement the following MDID estimator for longitudinal data described in Blundell and Costa Dias (2009)

$$\hat{\alpha}^{MDID} = \frac{1}{N_1} \sum_{i \in T} \left\{ [y_{i,t_1} - y_{i,t_0}] - \sum_{j \in C} \tilde{\omega}_{ij} [y_{j,t_1} - y_{j,t_0}] \right\} \tag{16}$$

where N_1 denotes the number of treated households in the common support region.

That is, to estimate the effect of *Oportunidades* on intra-household time allocation in poor urban households, we compare the difference in outcomes across waves of every treated household, $y_{i,t_1} - y_{i,t_0}$, to an average of the difference in outcomes across time of *observably similar* control households, $y_{j,t_1} - y_{j,t_0}$. The inclusion of control households into the group of observably similar households for a given treated household is determined by the constructed weight, $\tilde{\omega}_{ij}$, which is obtained in the first stage of the implementation of this estimator. This allows us to address the significant observable differences between treatment and control households mentioned in subsection 3.3 while accounting for the presence of unobserved determinants of participation that remain fixed over time.

Throughout this section, I discuss the implementation of the MDID estimator using the two definitions of treatment used in the previous subsection. The first treatment definition to be discussed pertains the one based on transfer receipt and the MDID strategy is applied to recover the average treatment effect (ATT) parameter. The second definition of treatment pertains the assignment of a household to an intervention or control city block and the MDID strategy is applied to recover the average intention to treat (AIT) parameter.

5.2.2 Identifying Assumptions

Heckman, Ichimura and Todd (1998), mention that we can identify ATT if the following conditions hold:

M1) Conditional independence assumption: Selection into treatment is based on observables *only*, also known in the literature as unconfoundedness. That is,

$$(Y_0, Y_1) \perp\!\!\!\perp D | X \tag{17}$$

M2) Common support: In order for the matching estimator's fundamental identification condition, $\mathbb{E}[Y_0|D = 1, X] = \mathbb{E}[Y_0|D = 0, X]$ to hold, both sides of such equality must be simultaneously well-defined for all X. For this to happen, we typically assume that $P(X) \in (0, 1)$. This is key to guarantee that all participants have a counterpart among non-participants. However, Heckman, Ichimura and Todd (1997) suggest that it is better to condition directly on the support common to both participant and nonparticipant groups by estimating the region of common support $S = Supp(X|D = 1) \cap Supp(X|D = 0)$.

These two assumptions together are known as the strong ignorability condition: we can claim that potential outcomes and treatment assignment are statistically independent by conditioning on a set of variables defined over the support common to participants and nonparticipants.

The identification of ATT, however, can rely on weaker versions of the previous two conditions:

M1') Potential outcome of non-participation is statistically independent of treatment assignment conditional on X:

$$Y_0 \perp\!\!\!\perp D | X \tag{18}$$

M2′) Common support: $P(X) < 1$.

That is, participant's distribution of Y_0 given X can be identified using data on nonparticipants over the common support.

Furthermore, following Blundell and Costa Dias (2000), the conditional independence assumption from M1′) can be restated within a difference-in-differences framework as

$$Y_{0,t_1} - Y_{0,t_0} \perp\!\!\!\perp D|X \tag{19}$$

Rosenbaum and Rubin (1983), suggest the use of balancing scores (functions of the relevant covariates) to address the curse of dimensionality faced by covariate matching estimators. They further claim that when the strong ignorability condition holds,

$$Y_0 \perp\!\!\!\perp D|P(X) \tag{20}$$

which, given 19 implies

$$Y_{0,t_1} - Y_{0,t_0} \perp\!\!\!\perp D|P(X) \tag{21}$$

where $P(X) = Pr(D = 1|X)$. This then implies that conditioning on $P(X)$ balances the distribution of Y_0 and, therefore, $Y_{0,t_1} - Y_{0,t_0}$ with respect to D.

Moreover, another implication of the strong ignorability condition is that $D \perp\!\!\!\perp X|P(X)$. This is known as the balancing property of propensity scores (Lee, 2013). That is, individuals with the same propensity score should be observably identical in terms of their vector of covariates, X, and their assignment to treatment can be deemed as random. Alternatively, upon conditioning on $P(X)$, further conditioning on X should not give us any new information about the treatment assignment. This is analogous to the way in which the randomization behind experimental data balances both the observed and unobserved characteristics of treated and control households. More importantly, this is a testable property that can be used to assess the matching quality of the estimator.

5.2.3 Implementation

First Stage: Constructing $\tilde{\omega}_{ij}$

We have that the weight $\tilde{\omega}_{ij}$ is a function of a distance metric made in terms of either the covariates or the propensity score (which is itself a function of said covariates). At this point, it is pertinent to ask the following question: should the matching be made on the covariates or on an estimated propensity score (given that I am not provided a propensity score)? According to Heckman, Ichimura and Todd (1998), neither estimator is necessarily more efficient than the other one. However, as mentioned by the authors, the application of the matching estimator is itself a "data-hungry statistical procedure". That is, if one were to use a nonparametric method as the matching strategy, matching on a high-dimensional X could slow down the rate of convergence of such estimator (i.e. the curse of dimensionality). Conditioning on the propensity score might be a viable way to circumvent such curse, but it requires the estimation of said propensity score. Hence, the use of a nonparametric method to estimate it would lead to the same issue provided that we have a high-dimensional X. A viable solution would be to first estimate the propensity score parametrically, through either a logit or a probit, which usually yield similar results in the case of a binary treatment variable and then use the estimated propensity scores to implement a nonparametric matching method.

Moreover, another crucial issue to deal with at this stage is the choice of conditioning variables since this influences the quality of the matches obtained in the matching stage of the estimator. Heckman, Ichimura and Todd (1998) mention that it is possible to dissect X into two sets of variables, say, T and Z, which need not be mutually exclusive. We can think of the variables in T as those that determine outcomes and those variables in Z as those that determine program participation. A typical recommendation found in the literature is to focus on the intersection of T and Z (Caliendo and Kopeinig, 2005). That is, to choose variables that affect program participation and outcomes simultaneously.

Transfer-based Treatment Definition
Keeping the latter in mind, for my choice of covariates, I build upon the work of Angelucci and Attanasio (2013) and Behrman et al (2011). From the former, I focus on the subset of covariates pertaining household composition, dwelling characteristics, financial indicators (whether the household has some previous loans, savings or even a bank account), and different types of shocks experienced by the household (loss of job by some household member, death of some household member and business loss). From the latter, taking into consideration that the authors in this paper adopt a similar treatment definition as mine, I focus on variables such as participation in other social programs (milk subsidy, breakfast subsidy, tortilla subsidy and procampo), educational attainment of the mother and father, and an index of poverty incidence in the state in which the household resides. The reasoning for including such state index is that it could have affected the households' knowledge about the program. Such knowledge could be considered along two dimensions: the first dimension being the existence of the program and the second dimension being the steps to be taken in order to be incorporated into the program. This stems from the fact that the program was more heavily publicized in poorer states, where the poverty index taken into account for this was precisely the one I currently have, which is provided by the CONEVAL using both census data and household-level income and expenditure data from the INEGI's ENIGH dataset. Table 9 presents the marginal effects at the mean of the covariates included in the probit model used to estimate the propensity scores. It can be noted in this table that I have also included higher powers of some covariates. This resulted from several attempts to balance the treatment and control groups in terms of the covariates, with the given variable selection yielding the best results when testing the aforementioned balancing property of the propensity score.

In determining which variables to include in the model and which to exclude, I used the two criteria suggested by Heckman, Ichimura, Smith and Todd (1998): statistical significance and the model's hit/miss rate. In terms of statistical significance, though some covariates might not be individually significant, they are jointly significant and are, therefore, included in the model. In terms of the the hit/miss rate, my current model correctly predicts program participation 71.5% of the times. This hit rate is quite close to the one Behrman, Gallardo-Garcia, Parker, Todd, and Velez-Grajales (2011) achieve, which is around 73%.

Table 9: Probit Estimates: Marginal Effects at the Mean

	$\Pr(D = 1 \mid X)$	
HH Poverty Index	0.0465	(0.59)
(HH Poverty Index)^2	-0.0618***	(-3.49)
Household size	0.00914	(0.28)
Number of kids, 0-5	0.0963*	(2.56)
Number of kids, 6-12	0.0132	(0.23)
Number of kids, 13-15	-0.0182	(-0.32)
Number of kids, 16-20	-0.0375	(-0.68)
Want more education for children	0.107	(1.35)
(Number of kids in school)^2	-0.0160***	(-3.38)
Number of kids in school, 6-12	0.169***	(3.40)
Number of kids in school, 13-15	0.234***	(4.16)
Number of kids in school, 16-20	0.115	(1.60)
Female head	0.150	(1.65)
Number of rooms	-0.0830***	(-4.42)
Floors made of dirt	0.159***	(5.27)
Walls made of weak material	0.168***	(6.26)
Refrigerator ownership	-0.177***	(-5.67)
Gas stove ownership	-0.143***	(-3.56)
Truck ownership	-0.334***	(-5.51)
Have previous loans	0.112***	(3.84)
Have a bank account	-0.000783	(-0.01)
Have previous savings	0.0200	(0.33)
Death of a family member	0.0628	(1.78)
Job loss	0.0589	(1.83)
Business loss	-0.321***	(-3.73)
Local incidence of poverty	0.0242***	(4.03)
(Local incidence of poverty)^2	-0.000139	(-1.93)
Tortilla subsidy	0.252***	(5.83)
Milk subsidy	-0.0175	(-0.44)
Procampo	-0.0865	(-0.58)
Breakfast subsidy	0.00746	(0.19)
Employed in 2001, mother	-0.0483	(-1.26)
Employed in 2000, mother	0.137**	(3.02)
Employed in 1999, mother	-0.00000797	(-0.00)
Employed in 2001, father	-0.0931	(-1.08)
Employed in 2000, father	-0.0568	(-0.64)
Employed in 1999, father	0.0345	(0.44)
Completed primary school, mother	-0.0154	(-0.05)
Completed secondary school, mother	-0.0697	(-0.22)
Completed high school or above, mother	0.0640	(0.20)
Completed primary school, father	-0.156	(-0.66)
Completed secondary school, father	-0.195	(-0.82)
Completed high school or above, mother	-0.256	(-1.06)
Age of mother	-0.00681*	(-2.29)
Age of father	0.00456*	(1.98)
N	2177	

t-statistics in parentheses

Zone-Based Treatment Definition
For the selection of covariates to be used in this new probit, I carry some of the covariates used in the one for the first definition of treatment I used. Most of these variables are related to household composition in terms of the number of kids by age groups, the number of kids within these age groups that are attending school at baseline, household size, previous loans, previous savings, completed years of education of both parents, the household's poverty index and its square, and different transitory income shocks. Following Angelucci and Attanasio, I created additional variables pertaining the household's income during 1999, 2000 and 2001[11], employment status of the decision makers during 1999, 2000, 2001 and 2002, health checkups of parents and children, and state GDP growth rates for 2000, 2001 and 2002.

There are some differences, however, with some of the variables described above. While the ones I used, broadly capture similar information to theirs, I was unable to use exactly the same ones, and my guess is that this is mainly due to data limitations. Angelucci and Attanasio (2013) use dummy variables for household size as well as for the education level of the household head and spouse. I tried using similar dummy variables but ended up having some of them being dropped by multicollinearity once one of these dummies was dropped by the program for perfectly predicting either success or failure. This might be attributed to differences in the size in our estimation samples. The original sample of Angelucci and Attanasio consisted of all 9,945 eligible households, while mine consisted of 3,125 eligible two-parent households. It is then possible that they had more variation in terms of these indicators in their sample. Hence, I used the continuous counterparts of these variables in order to avoid the issues I was obtaining with the use of the dummy ones.

The results for the estimation of the propensity score can be found in Table 10, where we can see the marginal effects at the mean of all the covariates included in the model used. Following Angelucci and Attanasio, I have also included the F-statistic for the joint significance of the household labor income variables. As done in the previous implementation of this estimator, I also calculated the hit rate of the model, which is of approximately 75% (using a cut-off of 50.5%).

[11]These income variables have been adjusted for inflation using the consumer price index with 2002 base year. Hence, all these variables are reported in terms of pesos measured at 2002 values.

Table 10: Probit Estimates: Marginal Effects at the Mean

	$\Pr(D=1\|X)$	
HH Poverty Index	0.0212	(0.25)
(HH Poverty Index)^2	-0.0386	(-1.85)
Number of kids 0-5	0.0107	(0.24)
Number of kids 6-12	-0.0654	(-0.91)
Number of kids 13-15	0.0334	(0.45)
Number of kids 16-20	-0.0459	(-0.63)
Number of kids in school 0-5	-0.0128	(-0.31)
Number of kids in school 6-12	0.0485	(0.87)
Number of kids in school 13-15	-0.000466	(-0.01)
Number of kids in school 16-20	-0.00576	(-0.07)
Completed years of education, mother	-0.00184	(-0.22)
Completed years of education, father	-0.00795	(-0.92)
Household labor income, 2001	0.000000645	(0.68)
Household labor income, 2000	-0.00000157	(-1.44)
Household labor income, 1999	-0.00000103	(-1.20)
Mother or father was employed in 2001	-0.0540	(-0.56)
Mother or father was employed in 2000	-0.00119	(-0.01)
Mother or father was employed in 1999	0.333***	(4.27)
Previous loans	0.0976**	(2.85)
Previous savings	0.0255	(0.39)
Any death	0.0743	(1.81)
Any jobloss	0.0986**	(2.60)
Employed in 2002, mother	0.140***	(3.92)
Employed in 2002, father	-0.113	(-0.87)
Self-employed, mother	-0.0272	(-0.50)
Self-employed father	-0.0808*	(-2.13)
Household size	0.00812	(0.19)
State GDP growth 2000	-0.127***	(-13.49)
State GDP growth 2001	0.00214	(0.17)
State GDP growth 2002	0.104***	(11.75)
Went to doctor, mother	0.0970*	(2.20)
Went to doctor, father	-0.0442	(-0.72)
Went to doctor, child	0.0118	(0.36)
Income joint significance	6.34*	
N	1489	

t-statistics in parentheses

F-statistic displayed for the income joint significance

Transfer-based Treatment Definition Given the probit model described in the previous subsection, we can estimate the propensity score for each observation used in the model. This yields the following distribution of propensity scores across the group of participant households and the group of non-participant households.

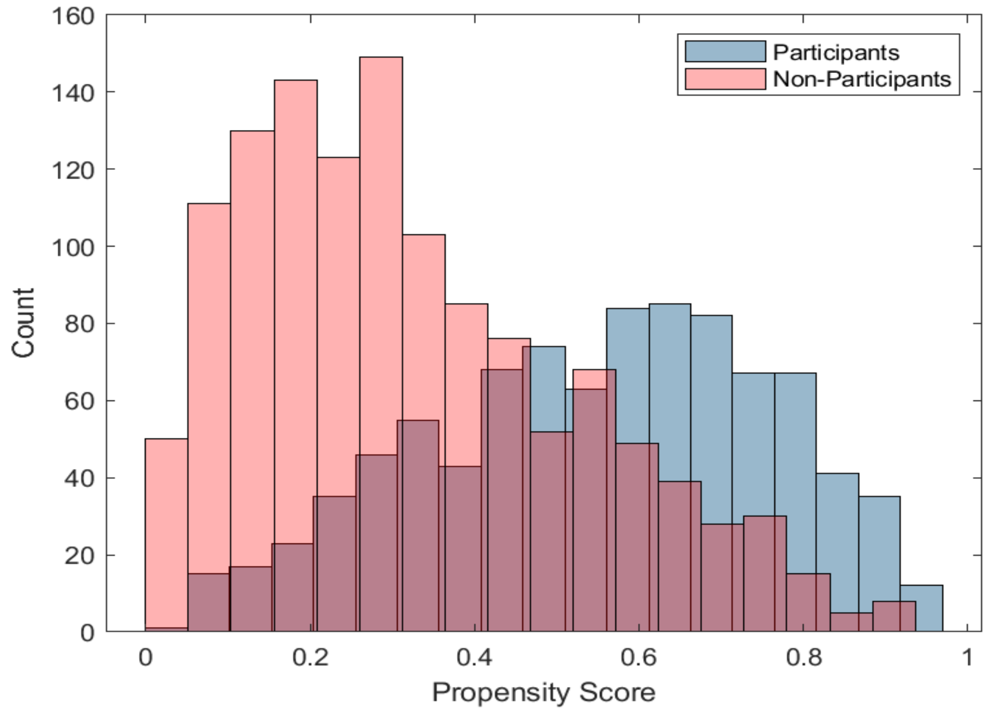

Figure 2: Propensity Scores Histograms, Participant and Non-Participant Households

The most natural approach to adopt in order to impose the common support condition is to simply take the range of propensity scores for which there is some positive amount of observations corresponding to these values in both comparison groups. However, such a loose definition can make it very difficult to balance the two groups in terms of the covariates, especially when using the blocking procedure implemented in the DW algorithm. Even though this was the definition implemented for the results obtained late December, I tried to obtain better results in terms of balancing the sample by tightening such definition more. In order to do so, after applying the minima-maxima definition of the common support, I trimmed the top and bottom 10% of the resulting propensity score distribution.

Once the region of common support has been defined, it is possible to test the balancing condition of the propensity score by implementing the Dehejia and Wahba (2002) algorithm. It is worth mentioning that while there is still some overall imbalance in terms of covariates, when implementing the DW blocking algorithm, such imbalances seem to be less severe. Using 5 blocks, we can see in the table below that most of the covariates for which there is some significant difference across comparison groups are not significant in determining program participation as specified in the probit model used (with the only exception of refrigerator ownership). Moreover, we can also see that the number of covariates for which there are significant differences across groups is not substantial and there is none for the 5th block. I might need to continue increasing the amount of blocks to check for the balancing condition but given the results below, it might be suggestive of the propensity scores allowing for some balancing across groups. This has been the best I have been able to achieve in terms of balancing the sample. However, I might need to explore other balancing tests such as bias reduction or regression tests as mentioned in the literature.

Table 11: Testing the Balancing Condition on the Region of Common Support

	Covariates	Difference
Overall	Household Poverty Index	0.075*
	Number of rooms	-0.079*
	Floors made of dirt	0.105***
	Employed 2000, Mother	0.039**
	Walls made of weak materials	0.165***
	Previous loans	0.050**
	Job loss	0.038**
	Local poverty incidence	5.954***
	Tortilla subsidy	0.026**
Block 1	Business loss	-0.009*
Block 2	Previous savings	-0.030**
	Any death	0.067*
	Female household head	-0.011*
Block 3	Previous savings	0.035*
	Bank account	0.019*
Block 4	Household size	0.319*
	Number of kids, 16-20	0.231*
	Number of kids in school, 16-20	0.208*
	Employed in 2000, mother	0.073*
	Employed in 1999, mother	0.063*
	Age of father	1.782*
Block 5	Household poverty index	0.188*
	Number of kids, 0-5	-0.218*
	Number of kids, 16-20	0.279*
	Number of kids in school, 16-20	0.264*
	Previous loans	-0.124*

Zone-based Treatment Definition

Given the probit model described for this treatment definition, we can estimate the propensity score for each observation used in the model. This yields the following distribution of propensity scores across the group of households in intervention zones and the group of households in control zones. The region of common support is determined in the same way as in the estimation with the transfer-based treatment definition.

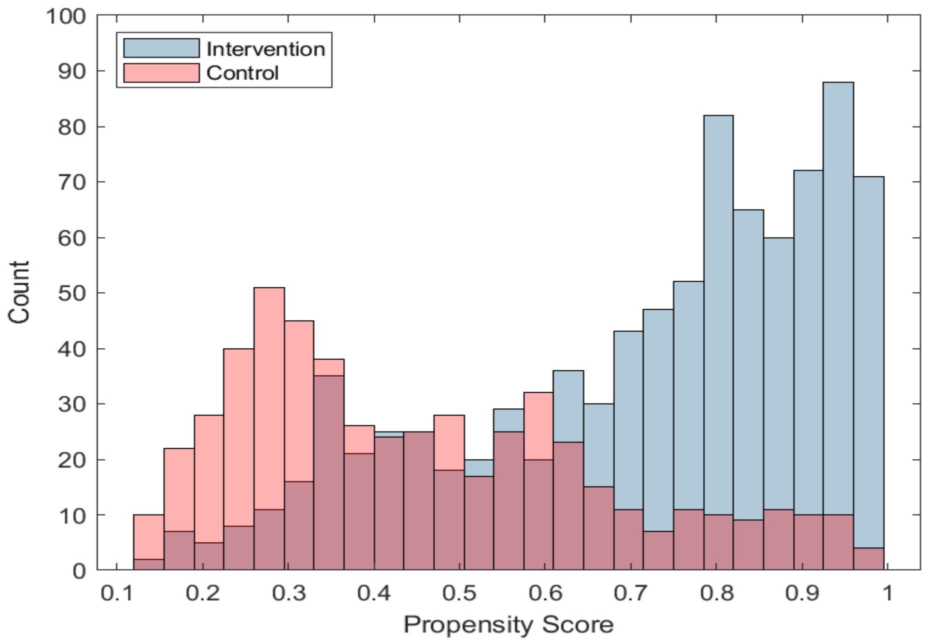

Figure 3: Propensity Scores Histograms, Intervention and Control Households

Once the region of common support is defined, we can then proceed to test the balancing property of the propensity score again. As before, I implement the DW algorithm involving dividing the common support region into equally-spaced blocks and checking whether there are significant differences between intervention and control households within each block. I divided the common support region into 5 equally-spaced blocks and then test whether the difference in the mean values of each of the covariates used is significantly different between the two groups. Before checking whether such differences are significant within each block, I first check this on the entire region of common support. Table 12 shows the results obtained from the implementation of this test.

Table 12: Testing the Balancing Condition on the Region of Common Support

	Covariates	Difference
Overall	Previous savings	-0.024*
	GDP growth rate, 2000	-0.851***
	GDP growth rate, 2001	-0.371**
	GDP growth rate, 2002	0.700***
Block 1	Number of kids, 6-12	0.414*
	Number of kids in school, 13-15	-0.120*
	Completed years of education, mother	0.616**
	Household labor income, 2001	-6,081.853***
	Household labor income, 2000	-4,731.698*
	Any jobloss	-0.088**
	Household size	0.411*
	GDP growth rate, 2002	0.430*
Block 2	Number of kids, 13-15	0.318**
	Number of kids in school, 13-15	0.259**
	Either parent employed in 1999	-0.069*
	Previous savings	-0.066*
	Any death	-0.100**
	Employed in 2002, mother	0.159*
	Went to doctor, child	-0.198***
	Household labor income, 1999	-6,638.915**
Block 3	Household poverty index	0.467***
	Number of kids, 6-12	0.494**
	Number of kids, 16-20	-0.105*
	Number of kids in school, 6-12	0.430**
	Household labor income, 2001	-5,856.899**
	GDP growth rate, 2002	0.525**
Block 4	Household poverty index	0.278**
	Number of kids in school, 0-5	0.076*
	Number of kids in school, 6-12	0.257*
	Household labor income, 2001	-9,240.369***
	Household labor income, 2000	-5,826.076**
	Household labor income, 1999	-6,219.628**
	GDP growth rate, 2000	-0.599***
	GDP growth rate, 2001	-0.531**
Block 5	Number of kids, 13-15	0.152**
	Number of kids in school, 13-15	0.174***
	Self-employed mother	0.084*

Second Stage: Matching
Throughout this stage, it is helpful to keep in mind that the motivation for this estimator is to construct a counterfactual outcome (in my case, across the different time-use categories at hand) for each treated household using the data from *observably* similar control households. For the purpose at hand, the observable similarity between treated and control households is based on their estimated propensity scores and the weights received by each control household is dependent upon the matching strategy implemented. Hence, given the propensity scores estimated in the first stage, I implement both the nearest-neighbor and kernel-based matching strategies.

The nearest neighbor strategy follows the rule

$$C(P_i) = \min_j ||P_i - P_j||, \text{ j} \in I_0$$

where I_0 denotes the control group.

By obtaining the indices of the nearest neighbors of each treated household, I am able to match each treated household's $y_{i,t_1} - y_{i,t_0}$ to that of its nearest neighbor's in the control group. That is, for each individual treated household, I was able to observe both its $y_{i,t_1} - y_{i,t_0}$ and its counterfactual one constructed from its nearest neighbors' $y_{i,t_1} - y_{i,t_0}$. In the case of the one nearest neighbor estimator, I would simply take the outcome of the one matched control household as such counterfactual. In the case of the k-th nearest neighbor estimator (where $k > 1$), I simply took the average of the outcomes of the k matched control households as the counterfactual. Hence, throughout this matching procedure, I assign the weight of $1/k$ to all such matched households irrespective of how dissimilar their propensity scores are to that of the corresponding treated household. A major drawback that became quite evident as k increased is that many of those neighbors matched to treated households were actually bad matches in the sense that their propensity scores were not quite "near" to the treated household's one. Hence, given this, the highest value for k which I decided to try was 10, and even this could be deemed too high given that for some treated households, even the second neighbor was already a significantly lower-quality match.

The drawback noticed in the implementation of the nearest-neighbor matching algorithm naturally leads to the implementation of a kernel-based one in which the weight assigned to each control household is dependent upon the distance between the household's propensity score and that of the corresponding treated household. That is, we have that

$$\tilde{\omega}_{ij} = \frac{K\left(\frac{P_j - P_i}{h}\right)}{\sum_{k \in C} K\left(\frac{P_k - P_i}{h}\right)}$$

I then proceeded to use the Epanechnikov kernel to create the weights, where the kernel used is defined by

$$K(u) = \begin{cases} \frac{3}{4}(1-u^2) & |u| < 1 \\ 0 & \text{otherwise} \end{cases}$$

where the bandwidth, h, is computed following Silverman's rule of thumb for bandwidth selection: $h = 2.345\sigma N^{-0.2}$.

Similarly, I implement the Gaussian kernel, where the kernel is defined by

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

applying again the Silverman rule of thumb for bandwidth selection, $h = 1.059\sigma N^{-0.2}$ in order to obtain results that are comparable to the ones obtained from using the Epanechnikov kernel.

Third Stage: Difference-in-Differences Regression
Once I have obtained a matched sample in which each treated household has been matched to its constructed counterfactual in the previous stage, I run the following regression

$$y_{i,t} = \beta_0 + \beta_1 Treat_i + \beta_2 W2_t + \beta_3 W3_t + \beta_4 (Treat_i \times W2_t) + \beta_5 (Treat_i \times W3_t) + \epsilon_{i,t}$$

The notation is the same as the one defined in the difference-in-differences subsection. However, it is important to mention that from the regression ran in this stage, I only take the coefficients β_4 and β_5. The standard errors are computed in the next stage to account for the propensity score being estimated at an earlier stage of the implementation of the estimator.

Fourth Stage: Bootstrapping
As aforementioned, in order to obtain the standard error for the estimator, I cannot simply take the standard errors from the previous stage and bootstrapping is required. Hence, I have bootstrapped the standard errors to obtain the results presented in the next section. For the bootstrap, I perform 100 repetitions. In each repetition, each of the three stages described above is performed. The definition of the common support with the 10% trimming applied so far has been automated so that it is correctly imposed in each repetition and does not induce unnecessary additional variation in the distribution of the propensity scores in each re-sample.[12] Another important issue to address throughout this stage is that in some repetitions, the bootstrapped sample might have no variation for some dummy variables used in the estimation of the propensity score, which ends up causing some originally significant variables to be dropped from the probit, therefore affecting the estimation at this stage. In order to tackle this and ensure that there is sufficient variation for the relevant dummy variables in each bootstrapped sample, re-sampling is performed with stratification using these variables. This allows for the probit model to be properly estimated in each repetition, therefore allowing for the propensity scores to be estimated in each bootstrapped sample in the same way as in the original one.

[12] The problem with bootstrapping the results for this estimator is that estimation is not carried out in one single program, but in two. For instance, stages 1 and 3 are performed in Stata but stage 2, the matching one, is performed in Matlab. The reasoning for using Matlab for the matching stage is that it allows for better control of the matching process. There are several matching commands in Stata, but these either do not allow for the propensity score to be computed outside of the command (such as teffects nnmatch or psmatch) or their matching process becomes a blackbox as some of my treated households were erroneously being matched with other treated households when constructing the counterfactual outcomes (psmatch2). Hence, matching in Matlab is more straightforward as it becomes clear that treated households are being matched only with control households and the counterfactuals are then adequately constructed. A further advantage of using Matlab is that the non-parametric matching methods are easier to implement as it allows for the bandwidth to be computed in the program, thereby automating the bandwidth selection in these methods.

6 Results

In this section, I present the results obtained from the implementation of the empirical strategies described in Section 5. The row Diff-in-Diff shows the estimates obtained from implementing a difference-in-differences strategy and the row Matching Diff-in-Diff shows the estimates obtained from implementing a matching difference-in-differences strategy. The results from the alternative matching strategies implemented for the matching difference-in-differences estimator can be found in the appendix.

6.1 Transfer-Based Treatment Definition

Table 13: One-Year Results, Both Definitions of Leisure

		Leisure, Definition 1	Leisure, Definition 2
Women	Diff-in-Diff	1.763*	7.984***
		(1.068)	(2.554)
	Observations	1,712	1,712
	Matching Diff-in-Diff	-0.0863	4.757***
		(0.806)	(1.660)
	Observations	1,204	1,204
Men	Diff-in-Diff	-1.633	-0.426
		(1.524)	(1.650)
	Observations	1,810	1,810
	Matching Diff-in-Diff	-0.799	0.241
		(1.001)	(1.141)
	Observations	1,284	1,284

Clustered standard errors at the city block level in parentheses for Diff-in-Diff
Bootstrapped standard errors in parentheses (100 repetitions) for Matching Diff-in-Diff
Matching estimates obtained using an Epanechnikov kernel-based matching strategy.

Following a similar procedure as the one presented in Behrman et al (2012), I present one-year and two-year DID and MDID estimates. The coefficient of interest is found in the row Diff-in-diff while the columns correspond to the different time use outcome variables defined in Subsection 3.4. The tables in this subsection correspond to the results obtained from running the regression specified in Section 5 using the definition of treatment based on transfer receipt. Tables 13 and 14 provide a comparison of the one-year and two-year effects of the transfer receipt on both definitions of leisure defined in Subsection 3.4. The one-year results show that when using a DID approach, there is a significant increase in women's weekly consumption of leisure when analyzing both definitions on leisure, but both the magnitude and statistical significance of such increase is higher for the relatively richer definition of leisure. The one-year estimates obtained from the implementation of the MDID estimator suggest no impact on women's leisure under the first definition, while it suggests a highly significant increase of around 4.8 weekly hours under the second leisure definition. The two-year estimates from implementing a DID strategy and its matching counterpart suggest no impact on women's leisure under the first definition,

Table 14: Two-Year Results, Both Definitions of Leisure

		Leisure, Definition 1	Leisure, Definition 2
Women	Diff-in-Diff	0.298	8.442***
		(1.117)	(2.586)
	Observations	1,722	1,722
	Matching Diff-in-Diff	-0.642	6.111***
		(0.859)	(1.732)
	Observations	1,208	1,208
Men	Diff-in-Diff	-0.194	0.0902
		(1.604)	(1.727)
	Observations	1,832	1,832
	Matching Diff-in-Diff	-1.028	-1.108
		(1.146)	(1.177)
	Observations	1,302	1,302

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Clustered standard errors at the city block level in parentheses for Diff-in-Diff
Bootstrapped standard errors in parentheses (100 repetitions) for Matching Diff-in-Diff
Matching estimates obtained using an Epanechnikov kernel-based matching strategy.

Table 15: One-Year Results, Components of Leisure Definition 2

		Market Work	Child Care	Total Home Production	Core Home Production	Procurement of Goods
Women	Diff-in-Diff	-1.763*	-1.722	-4.499***	-4.587***	0.0883
		(1.068)	(1.505)	(1.545)	(1.485)	(0.204)
	Observations	1,712	1,712	1,712	1,712	1,712
	Matching Diff-in-Diff	0.0863	-0.941	-3.902***	-4.114***	0.212
		(0.806)	(0.836)	(1.119)	(1.085)	(0.128)
	Observations	1,204	1,204	1,204	1,204	1,204
Men	Diff-in-Diff	1.633	-0.318	-0.889**	-0.794**	-0.0947
		(1.524)	(0.393)	(0.380)	(0.345)	(0.0932)
	Observations	1,810	1,810	1,810	1,810	1,810
	Matching Diff-in-Diff	0.799	-0.479	-0.560*	-0.508*	-0.0524
		(1.001)	(0.317)	(0.311)	(0.303)	(0.057)
	Observations	1,284	1,284	1,284	1,284	1,284

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Clustered standard errors at the city block level in parentheses for Diff-in-Diff
Bootstrapped standard errors in parentheses (100 repetitions) for Matching Diff-in-Diff
Matching estimates obtained using an Epanechnikov kernel-based matching strategy.

Table 16: Two-Year Results, Components of Leisure Definition 2

		Market Work	Child Care	Total Home Production	Core Home Production	Procurement of Goods
Women	Diff-in-Diff	-0.298	-2.607*	-5.537***	-5.647***	0.109
		(1.117)	(1.443)	(1.567)	(1.525)	(0.221)
	Observations	1,722	1,722	1,722	1,722	1,722
	Matching Diff-in-Diff	0.642	-1.688**	-5.065***	-5.016***	-0.0495
		(0.859)	(0.840)	(1.177)	(1.127)	(0.141)
	Observations	1,208	1,208	1,208	1,208	1,208
Men	Diff-in-Diff	0.194	0.0148	-0.299	-0.302	0.00352
		(1.604)	(0.431)	(0.396)	(0.356)	(0.102)
	Observations	1,832	1,832	1,832	1,832	1,832
	Matching Diff-in-Diff	1.028	0.562**	-0.481	-0.390	-0.0919
		(1.146)	(0.268)	(0.313)	(0.274)	(0.081)
	Observations	1,302	1,302	1,302	1,302	1,302

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Clustered standard errors at the city block level in parentheses for Diff-in-Diff
Bootstrapped standard errors in parentheses (100 repetitions) for Matching Diff-in-Diff
Matching estimates obtained using an Epanechnikov kernel-based matching strategy.

but a highly significant one under the second definition (around 8 weekly hours using a DID approach and 6 hours using a MDID approach). In the case of men, neither of the empirical strategies on any of the two leisure definitions used suggest any one-year or two-year impact.

Tables 15 and 16 show the effects of the transfer on all the components used for the second leisure definition. Using a DID strategy, the one-year results show a significant decrease in the amount of weekly hours women devote to market work (of approximately 1.8 hours), and to total home production (stemming from a significant decrease of approximately 4.6 hours in the amount of weekly hours devoted to core home production). These changes then lead to a significant increase of approximately 8 weekly hours devoted to leisure. For men, the results suggest a significant decrease in weekly hours devoted to total home production (again driven by a significant decrease in weekly hours devoted to core home production). For the two-year results, there seems to be no significant impact in the time allocation for men. For women, on the other hand, we can see that there is a significant decrease in the amount of weekly hours devoted to child care and total home production, leading to a significant increase of approximately 8 weekly hours devoted to leisure. These results are quite different to what Parker and Skoufias (2000) find in rural households (with the only exception of the reduction in the amount of home production hours for women) and to what Skoufias and di Maro (2006) find in rural households as well. Once the MDID estimator is applied for women, we can see that the one-year results on home production are quite robust, though the significant impact on market work disappears. In the case of the two-year results, the effects on women's child care and home production are quite robust upon the implementation of the MDID estimator. The same argument can be made for men, the results obtained for the one-year and two-year effects under the DID strategy are quite robust upon the implementation of the MDID estimator, with only one exception: the significant positive two-year effect on child care observed upon the implementation of the MDID estimator.

6.2 Zone-Based Treatment Definition

Table 17: One-Year Results, Both Definitions of Leisure

		Leisure, Definition 1	Leisure, Definition 2
Women	Diff-in-Diff	2.486**	7.133**
		(1.126)	(2.967)
	Observations	1,743	1,743
	Matching Diff-in-Diff	-1.997***	2.808
		(0.516)	(1.866)
	Observations	1,260	1,260
Men	Diff-in-Diff	0.602	2.352
		(1.658)	(1.813)
	Observations	1,843	1,843
	Matching Diff-in-Diff	-0.0807	0.185
		(0.866)	(1.086)
	Observations	1,354	1,354

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Clustered standard errors at the city block level in parentheses for Diff-in-Diff

Bootstrapped standard errors in parentheses (100 repetitions) for Matching Diff-in-Diff

Matching estimates obtained using an Epanechnikov kernel-based matching strategy.

The tables in this subsection correspond to the results obtained from running the regression specified in Section 5 using the definition of treatment based on the zone in which the household resides. Tables 17 and 18 provide a comparison of the one-year and two-year effects of the household's city block assignment to intervention zones on both definitions of leisure defined in Subsection 3.4. For women, the one-year DID results are quite similar to what was obtained using the transfer-based treatment definition. However, once the MDID estimator is applied, we can see that there is a switch in the sign of the program's impact on women's leisure hours under the first definition used and the effect on the leisure hours under the second definition used disappears once the MDID estimator is implemented. Still focusing on women, the two-year DID results follow a similar pattern for the two definitions of leisure used. However, upon the implementation of the MDID estimator, we can still observe the switch in signs for the first leisure definition, but under the second leisure definition, we can see that the DID result obtained is quite robust (and actually stronger) after matching. In the case of men, as in the results obtained using the transfer-based treatment definition, neither of the empirical strategies on any of the two leisure definitions used suggest any one-year or two-year impact.

Tables 15 and 16 show the effects of the transfer on all the components used for the second leisure definition. Using a DID strategy, the one-year results for women are quite similar to the ones obtained using the other definition of treatment. For men, this time, the results suggest a significant decrease in total

Table 18: Two-Year Results, Both Definitions of Leisure

		Leisure, Definition 1	Leisure, Definition 2
Women	Diff-in-Diff	2.096*	10.63***
		(1.211)	(2.912)
	Observations	1,753	1,753
	Matching Diff-in-Diff	-1.241***	12.76***
		(0.470)	(1.760)
	Observations	1,278	1,278
Men	Diff-in-Diff	0.468	1.494
		(1.581)	(1.729)
	Observations	1,867	1,867
	Matching Diff-in-Diff	-1.481	-1.169
		(0.994)	(1.073)
	Observations	1,372	1,372

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Clustered standard errors at the city block level in parentheses for Diff-in-Diff
Bootstrapped standard errors in parentheses (100 repetitions) for Matching Diff-in-Diff
Matching estimates obtained using an Epanechnikov kernel-based matching strategy.

Table 19: One-Year Results, Components of Leisure Definition 2

		Market Work	Child Care	Total Home Production	Core Home Production	Procurement of Goods
Women	Diff-in-Diff	-2.486**	-1.325	-3.323**	-3.250**	-0.0726
		(1.126)	(1.642)	(1.622)	(1.563)	(0.225)
	Observations	1,743	1,743	1,743	1,743	1,743
	Matching Diff-in-Diff	1.997***	-0.0102	-4.795***	-4.682***	-0.112
		(0.516)	(0.995)	(0.992)	(0.998)	(0.113)
	Observations	1,260	1,260	1,260	1,260	1,260
Men	Diff-in-Diff	-0.602	-0.293	-1.457***	-1.192***	-0.265***
		(1.658)	(0.423)	(0.408)	(0.371)	(0.102)
	Observations	1,843	1,843	1,843	1,843	1,843
	Matching Diff-in-Diff	0.0807	0.710***	-0.976***	-0.702**	-0.274***
		(0.866)	(0.261)	(0.308)	(0.285)	(0.060)
	Observations	1,354	1,354	1,354	1,354	1,354

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Clustered standard errors at the city block level in parentheses for Diff-in-Diff
Bootstrapped standard errors in parentheses (100 repetitions) for Matching Diff-in-Diff
Matching estimates obtained using an Epanechnikov kernel-based matching strategy.

Table 20: Two-Year Results, Components of Leisure Definition 2

		Market Work	Child Care	Total Home Production	Core Home Production	Procurement of Goods
Women	Diff-in-Diff	-2.096*	-2.825*	-5.704***	-5.826***	0.122
		(1.211)	(1.669)	(1.630)	(1.570)	(0.237)
	Observations	1,753	1,753	1,753	1,753	1,753
	Matching Diff-in-Diff	1.241***	-4.633***	-9.370***	-9.095***	-0.276***
		(0.470)	(0.936)	(1.007)	(0.979)	(0.105)
	Observations	1,278	1,278	1,278	1,278	1,278
Men	Diff-in-Diff	-0.468	-0.215	-0.811**	-0.663*	-0.148
		(1.581)	(0.484)	(0.401)	(0.366)	(0.110)
	Observations	1,867	1,867	1,867	1,867	1,867
	Matching Diff-in-Diff	1.481	0.944***	-1.256***	-1.090***	-0.166***
		(0.994)	(0.214)	(0.272)	(0.269)	(0.033)
	Observations	1,372	1,372	1,372	1,372	1,372

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Clustered standard errors at the city block level in parentheses for Diff-in-Diff
Bootstrapped standard errors in parentheses (100 repetitions) for Matching Diff-in-Diff
Matching estimates obtained using an Epanechnikov kernel-based matching strategy.

home production weekly hours stemming from a significant reduction in the amount of weekly hours devoted to both core home production (of approximately 1.5 hours) and the procurement of goods (of approximately 0.3). For the two-year results, for women, a significant reduction in the amount of weekly hours devoted to market work (of approximately 2.8 hours), child care (of approximately 2.8 hours), and total home production (of approximately 5.7 hours) lead to a significant increase in the amount of weekly hours devoted to leisure of approximately 10.6. On the other hand, for men, using this definition, the results suggest that there is a significant decrease in the amount of weekly hours devoted to total home production stemming from a significant decrease in the amount of weekly hours devoted to core home production. Again, this contrasts the results obtained by Parker and Skoufias (2000) and Skoufias and di Maro (2006) document for rural households. Once the MDID estimator is applied, we can see that the one-year effects on women's home production hours is quite robust. We can also see the switch in the sign of the effect of the program on women's market work hours (which explains the results obtained for leisure definition 1) as a result of the MDID estimator. The results for men seem to be quite robust to the ones obtained under the DID approach, with the exception of child care hours. This resembles what was observed when using the transfer-based treatment definition: once I implement the MDID estimator, the results suggest that there is a significant positive impact on the amount of weekly hours men devote to child care. However, for this zone-based treatment definition, this result holds for both the one-year and two-year impact. On the other hand, mirroring the results for women (though smaller in magniture), the negative impact on total home production hours remains quite robust to the empirical strategy adopted.

Before making any conclusions regarding the MDID results obtained for this treatment definition, it is worth noting that these are still very preliminary in nature. Further robustness checks need to be made, especially for the estimation of the probit that generates the propensity score used in the matching process of the estimator. These robustness checks are precisely the next steps being addressed at this stage of the paper. Hence, these MDID results are subject to change in further versions of this draft.

7 Discussion

The results presented in Section 6 contrast, to some extent, those reported by Parker and Skoufias (2000) and Skoufias and di Maro (2006) while having the advantage of a relatively more ideal reference period of a week rather than a day as well as the availability of time use data across the three waves rather than just one wave. Furthermore, these results are suggestive of a further rejection of the unitary model as they suggest a significant impact of the Oportunidades program on the intra-household time allocation in urban households. Additionally, they highlight the importance of accounting for home production hours when defining leisure. This is evident in the contrasting results obtained when focusing on Leisure Measure 1 against Leisure Measure 2. While Leisure Measure 1 was mostly unaffected by the receipt of the cash transfer for both men and women, women's Leisure Measure 2 seemed to be significantly affected by it. This further confirms the importance of having sufficiently disaggregated time use data as recommended by Apps and Rees (1997), Chiappori (1997), Blundell, Chiappori and Meghir (2005) and Doss (2013). In the absence of data on home production hours, focusing on Leisure Measure 1 would have mislead us to the conclusion that the program did not affect the intra-household allocation of time. More importantly, having such a disaggregated leisure measure allows us to understand where the impact of the program is coming from and how the overall composition of time allocation within the household changes as a result of the program. Particularly, it allows us to assess whether an observed increase in leisure hours induced by the program is coming from a decrease in hours devoted to market work, which tends to be a concern of policymakers when crafting policies that place cash in the hands of beneficiaries. In this case, we might observe an increase in leisure not necessarily coming from a decrease in market work hours, but from a decrease in household production hours or even child care hours or a combination of these three time-use categories.

Besides the empirical implications of the results, there are also potentially interesting theoretical implications. Given that one of the objectives of the program is to improve women's position in the household, an important question is then whether or not women's welfare has been improved by the program based on the results obtained so far. Duflo (2012) mentions that a dimension of female empowerment could be characterized by women's ability to free their time away from domestic production so that she can use this time for market activities or just pure leisure. This dimension takes into consideration that the main source of household inequality between men and women stems from women's gender role as a care-giver, which sets expectations on how women should spend their time. However, finding a theoretical link between changes in women's allocation of hours across the different time-use categories and their welfare could help provide some theoretical support for this argument. Analyzing this link would require developing and estimating a theoretical framework that allows for the distribution of home production hours between men and women within the household to be affected by policies that alter the balance of power in the household and that does not reduce such allocation to a function of only spouses' wages and/or technological considerations. This could become the focus of further work on household responses to targeted benefits.

References

[1] Aguiar, M. & Hurst, E. (2007). Measuring Trends in Leisure: The Allocation of Time Over Five Decades. *Quarterly Journal of Economics*, 122 (3): 969–1006.

[2] Aguiar, M., Hurst, E., & Karabarbounis,L. (2012). Recent Developments in the Economics of Time Use. *Annual Review of Economics*, 4 (1): 373–397.

[3] Angelucci, M. & Attanasio, O. (2013). The Demand for Food of Poor Urban Mexican Households: Understanding Policy Impacts Using Structural Models. *American Economic Journal*, 5(1): 146-178.

[4] Angelucci, M., Attanasio, O. & Di Maro, V. (2012). The Impact of Oportunidades on Consumption, Savings and Transfers. *Fiscal Studies*, 33(3): 305-334.

[5] Angelucci, M., Attanasio, O. & Shaw, J. (2005). El Efecto de Oportunidades sobre el nivel y la composicion del consumo en areas urbanas. *Reporte de Evaluacion Externa: Oportunidades, 2005*

[6] Apps, P.F & Rees, R. (1997). Collective Labor Supply and Household Production. *Journal of Political Economy*, 105(1): 178-190.

[7] Attanasio, O. & Lechene, V. (2002). Tests of Income Pooling in Household Decisions. *Review of Economic Dynamics*, 5(4): 720-748.

[8] Attanasio, O. & Lechene, V. (2014). Efficient Responses to Targeted Cash Transfers. *Journal of Political Economy*, 122(1): 178-222.

[9] Behrman, J.R &. Parker, S.W (2011). *The Impact of the PROGRESA/Oportunidades Conditional Cash Transfers on Health and Related Outcomes for the Aging in Mexico* (PARC Working Paper Series 34).

[10] Behrman, J.R, Gallardo-Garcia, J., Parker, S.W., Todd, P.E. & Velez-Grajales, V. (2011). *Are Conditional Cash Transfers Effective in Urban Areas? Evidence from Mexico*. (Penn Institute for Economic Research Working Paper 11-024).

[11] Behrman, J.R, Gallardo-Garcia, J., Parker, S.W., Todd, P.E. & Velez-Grajales, V. (2012). Are Conditional Cash Transfers Effective in Urban Areas? Evidence from Mexico. *Educ Econ*, 20(3): 233-59.

[12] Bergolo, M. & Galvan, E. (2018). Intra-household Behavioral Responses to Conditional Cash Transfer Programs. Evidence from a Regression Discontinuity Design. *World Development*, 103(1): 100-118.

[13] Blundell, R., Chiappori, P.A. & Meghir, C. (2005). Collective Labor Supply with Children. *Journal of Political Economy*, 113(6): 1277-1305.

[14] Caliendo, M. & Kopeinig, S. (2005). *Some Practical Guidance for the Implementation of Propensity Score Matching*. (IZA Discussion Series Paper No. 1588).

[15] Cherchye, L., De Rock, B. & Vermeulen, F. (2012). Married with Children: A Collective Labor Supply Model with Detailed Time Use and Intrahousehold Expenditure Information. *American Economic Review*, 102(7): 3377-3405.

[16] Chiappori, P.A. (1997). Introducing Household Production in Collective Models of Labor Supply. *Journal of Political Economy*, 105(1): 191-209.

[17] Doss, C. (2013). Intrahousehold Bargaining and Resource Allocation in Developing Countries. *The World Bank Research Observer*, Washington, D.C.: The World Bank.

[18] Dubois, P. & Janvry, A. & Sadoulet, E. (2011). *Effects on School Enrollment and Performance of a Conditional Cash Transfer Program in Mexico* (Working Paper).

[19] Duflo, E. (2012). Women Empowerment and Economic Development. *Journal of Economic Literature*, 50(4): 1051-1079.

[20] Duflo, E. (2003). Grandmothers and Granddaughters: Old Age Pensions and Intrahousehold Allocations in South Africa. *World Bank Economic Review*, 17(1): 1-25.

[21] Heckman, J.J., Ichimura, H. & Todd, P. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies*, 64(4) : 605-654.

[22] Heckman, J.J., Ichimura, H., Smith, J. & Todd, P. (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica*, 66(5) : 1017-1098.

[23] Heckman, J.J., Ichimura, H. & Todd, P. (1998). Matching as an Econometric Evaluation Estimator. *Review of Economic Studies*, 65 : 261-294.

[24] Lee, W.S. (2011). Propensity score matching and variations on the balancing test. *Empirical Economics*, 44(1): 47-80.

[25] Levy, S. (2006). *Progress Against Poverty: Sustaining Mexico's Progresa-Oportunidades Program*. Washington, DC: Brookings Institution Press.

[26] Lise, J. &. Yamada, K. (2017). *Household Sharing and Commitment: Evidence from Panel Data on Individual Expenditures and Time Use* (October 2017 Working Paper).

[27] Mexican Federal Government. *Quantitative Evaluation: Questionnaires and Databases*. Mexico: Prospera External Evaluation Department. Retrieved from `https://evaluacion.prospera.gob.mx/en/eval_cuant/p_bases_cuanti.php` .Prospera

[28] Mexican Federal Government. *Nota Metodologica General Urbana*. Mexico: Prospera External Evaluation Department. Retrieved from `https://evaluacion.prospera.gob.mx/es/wersd53465sdg1/eval_cuant/nota_metodologica_general_urbana_2005_11_15.pdf`.SEDESOL.

[29] Mexican Federal Government. *Mapas de Pobreza por Ingreso 2000, Nacional*. Mexico: Consejo Nacional de Evaluacion de la Politica de Desarrollo Social. Retrieved from `https://www.coneval.org.mx/Medicion/EDP/MP/Paginas/Mapas-2000.aspx`.CONEVAL.

[30] Parker, S.W. & Skoufias, E. (2000). *The Impact of Progresa on Work, Leisure and Time Allocation*. (Final Report). Washington, DC: International Food Research Institute.

[31] Parker, S. & Todd, P. (2017). Conditional Cash Transfers: The Case of Progresa/Oportunidades. *Journal of Economic Literature*, 55(3), 866-915.

[32] Skoufias, E. (2005). *Progresa and Its Impact on the Welfare of Rural Households in Mexico*. (Research Report 139). Washington, DC: International Food Research Institute.

[33] Skoufias, E & di Maro, V. (2006). *Conditional Cash Transfers, Adult Work Incentives, and Poverty*. (Policy Research Working Paper 3973). Washington, DC: World Bank.

[34] Stampini. M, & Tornarolli, L. (2012). *The Growth of Conditional Cash Transfers in Latin America and the Caribbean: Did They Go Too Far?* (IDB Policy Brief n. 185). Washington, D.C.: Inter-American Development Bank.

[35] Todd, P. &. Wolpin, K. (2006). Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility. *American Economic Review*, 96(5): 1384-1417.

[36] Todd, P. &. Wolpin, K. (2008). *Ex-Ante Evaluation of Social Programs* (March 2008 Working Paper).

8 Appendix: Alternative Matching Strategies

Table 21: One-Year ResultsComponents of Leisure Definition 2, Alternative Matching Strategies (Transfer-based Treatment Definition)

		Leisure	Market Work	Child Care	Total Home Production	Core Home Production	Procurement of Goods
Women	Diff-in-Diff	7.984***	-1.763*	-1.722	-4.499***	-4.587***	0.0883
		(2.554)	(1.068)	(1.505)	(1.545)	(1.485)	(0.204)
	Observations	1,712	1,712	1,712	1,712	1,712	1,712
	Matching Diff-in-Diff (1-Nearest Neighbor)	5.847*	0.495	-1.186	-5.156***	-5.307***	0.151
		(3.246)	(1.479)	(1.811)	(1.802)	(1.730)	(0.255)
	Observations	1,204	1,204	1,204	1,204	1,204	1,204
	Matching Diff-in-Diff (2-Nearest Neighbors)	5.600**	1.052	-1.586	-5.066***	-5.209***	0.143
		(2.729)	(1.227)	(1.519)	(1.645)	(1.572)	(0.220)
	Observations	1,204	1,204	1,204	1,204	1,204	1,204
	Matching Diff-in-Diff (3-Nearest Neighbors)	4.437*	1.612	-1.312	-4.738***	-4.974***	0.236
		(2.400)	(1.043)	(1.331)	(1.489)	(1.434)	(0.196)
	Observations	1,204	1,204	1,204	1,204	1,204	1,204
	Matching Diff-in-Diff (5-Nearest Neighbors)	5.177**	0.537	-1.088	-4.626***	-4.881***	0.255
		(2.101)	(0.919)	(1.128)	(1.384)	(1.352)	(0.177)
	Observations	1,204	1,204	1,204	1,204	1,204	1,204
	Matching Diff-in-Diff (10-Nearest Neighbors)	4.654**	0.409	-1.210	-3.853***	-4.142***	0.288*
		(2.012)	(0.872)	(1.012)	(1.335)	(1.313)	(0.142)
	Observations	1,204	1,204	1,204	1,204	1,204	1,204
	Matching Diff-in-Diff (Kernel: Gaussian)	4.792***	0.0690	-0.967	-3.894***	-4.109***	0.215
		(1.663)	(0.803)	(0.836)	(1.131)	(1.096)	(0.127)
	Observations	1,204	1,204	1,204	1,204	1,204	1,204
Men	Diff-in-Diff	-0.426	1.633	-0.318	-0.889**	-0.794**	-0.0947
		(1.650)	(1.524)	(0.393)	(0.380)	(0.345)	(0.0932)
	Observations	1,810	1,810	1,810	1,810	1,810	1,810
	Matching Diff-in-Diff (1-Nearest Neighbor)	0.838	-0.453	0.0389	-0.424	-0.343	-0.0810
		(2.441)	(2.132)	(0.775)	(0.643)	(0.571)	(0.160)
	Observations	1,284	1,284	1,284	1,284	1,284	1,284
	Matching Diff-in-Diff (2-Nearest Neighbors)	-0.0654	0.873	-0.0241	-0.783	-0.649	-0.135
		(2.038)	(1.714)	(0.653)	(0.508)	(0.454)	(0.127)
	Observations	1,284	1,284	1,284	1,284	1,284	1,284
	Matching Diff-in-Diff (3-Nearest Neighbors)	0.674	0.232	-0.305	-0.601	-0.498	-0.103
		(1.808)	(1.492)	(0.586)	(0.454)	(0.417)	(0.105)
	Observations	1,284	1,284	1,284	1,284	1,284	1,284
	Matching Diff-in-Diff (5-Nearest Neighbors)	0.945	0.354	-0.586	-0.713*	-0.569	-0.144
		(1.654)	(1.397)	(0.508)	(0.425)	(0.398)	(0.088)
	Observations	1,284	1,284	1,284	1,284	1,284	1,284
	Matching Diff-in-Diff (10-Nearest Neighbors)	0.0642	0.782	-0.483	-0.363	-0.289	-0.0743
		(1.513)	(1.304)	(0.429)	(0.359)	(0.341)	(0.074)
	Observations	1,284	1,284	1,284	1,284	1,284	1,284
	Matching Diff-in-Diff (Kernel: Gaussian)	0.258	0.787	-0.482	-0.563*	-0.509*	-0.0541
		(1.144)	(1.006)	(0.318)	(0.309)	(0.301)	(0.058)
	Observations	1,284	1,284	1,284	1,284	1,284	1,284

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors in parentheses (100 repetitions)

Table 22: Two-Year Results: Components of Leisure Definition 2, Alternative Matching Strategies (Transfer-based Treatment Definition)

		Leisure	Market Work	Child Care	Total Home Production	Core Home Production	Procurement of Goods
Women	Diff-in-Diff	8.442***	-0.298	-2.607*	-5.537***	-5.647***	0.109
		(2.586)	(1.117)	(1.443)	(1.567)	(1.525)	(0.221)
	Observations	1,722	1,722	1,722	1,722	1,722	1,722
	Matching Diff-in-Diff (1-Nearest Neighbor)	5.780*	2.710*	-2.772	-5.719***	-5.571***	-0.147
		(3.077)	(1.601)	(1.705)	(1.842)	(1.797)	(0.264)
	Observations	1,208	1,208	1,208	1,208	1,208	1,208
	Matching Diff-in-Diff (2-Nearest Neighbors)	5.183*	2.224	-1.709	-5.698***	-5.742***	0.0439
		(2.662)	(1.405)	(1.445)	(1.769)	(1.724)	(0.217)
	Observations	1,208	1,208	1,208	1,208	1,208	1,208
	Matching Diff-in-Diff (3-Nearest Neighbors)	5.078**	1.990*	-1.498	-5.570***	-5.526***	-0.0436
		(2.385)	(1.203)	(1.364)	(1.655)	(1.612)	(0.195)
	Observations	1,208	1,208	1,208	1,208	1,208	1,208
	Matching Diff-in-Diff (5-Nearest Neighbors)	6.009***	0.775	-1.533	-5.250***	-5.212***	-0.0384
		(2.001)	(1.089)	(1.139)	(1.438)	(1.400)	(0.174)
	Observations	1,208	1,208	1,208	1,208	1,208	1,208
	Matching Diff-in-Diff (10-Nearest Neighbors)	5.296***	0.685	-1.477	-4.503***	-4.452***	-0.0513
		(2.033)	(0.988)	(1.103)	(1.387)	(1.342)	(0.169)
	Observations	1,208	1,208	1,208	1,208	1,208	1,208
	Matching Diff-in-Diff (Kernel: Gaussian)	6.161***	0.609	-1.713**	-5.057***	-5.010***	-0.0473
		(1.729)	(0.856)	(0.847)	(1.189)	(1.138)	(0.141)
	Observations	1,208	1,208	1,208	1,208	1,208	1,208
Men	Diff-in-Diff	0.0902	0.194	0.0148	-0.299	-0.302	0.00352
		(1.727)	(1.604)	(0.431)	(0.396)	(0.356)	(0.102)
	Observations	1,832	1,832	1,832	1,832	1,832	1,832
	Matching Diff-in-Diff (1-Nearest Neighbor)	-0.982	0.593	0.86	-0.472	-0.482	0.0108
		(2.428)	(2.298)	(0.620)	(0.540)	(0.500)	(0.156)
	Observations	1,302	1,302	1,302	1,302	1,302	1,302
	Matching Diff-in-Diff (2-Nearest Neighbors)	-1.640	1.818	0.634	-0.812**	-0.708*	-0.104
		(1.849)	(1.843)	(0.477)	(0.408)	(0.366)	(0.129)
	Observations	1,302	1,302	1,302	1,302	1,302	1,302
	Matching Diff-in-Diff (3-Nearest Neighbors)	-1.282	1.271	0.615	-0.604	-0.518	-0.0855
		(1.629)	(1.619)	(0.420)	(0.395)	(0.348)	(0.119)
	Observations	1,302	1,302	1,302	1,302	1,302	1,302
	Matching Diff-in-Diff (5-Nearest Neighbors)	-0.997	0.888	0.718*	-0.609	-0.520	-0.0885
		(1.574)	(1.583)	(0.378)	(0.392)	(0.352)	(0.105)
	Observations	1,302	1,302	1,302	1,302	1,302	1,302
	Matching Diff-in-Diff (10-Nearest Neighbors)	-1.566	1.160	0.738**	-0.332	-0.251	-0.0816
		(1.410)	(1.370)	(0.349)	(0.352)	(0.308)	(0.094)
	Observations	1,302	1,302	1,302	1,302	1,302	1,302
	Matching Diff-in-Diff (Kernel: Gaussian)	-1.077	1.011	0.552**	-0.486	-0.392	-0.0936
		(1.179)	(1.148)	(0.271)	(0.312)	(0.272)	(0.081)
	Observations	1,302	1,302	1,302	1,302	1,302	1,302

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors in parentheses (100 repetitions)

Table 23: One-Year ResultsComponents of Leisure Definition 2, Alternative Matching Strategies (Zone-based Treatment Definition)

		Leisure	Market Work	Child Care	Total Home Production	Core Home Production	Procurement of Goods
Women	Diff-in-Diff	7.133**	-2.486**	-1.325	-3.323**	-3.250**	-0.0726
		(2.967)	(1.126)	(1.642)	(1.622)	(1.563)	(0.225)
	Observations	1,743	1,743	1,743	1,743	1,743	1,743
	Matching Diff-in-Diff (1-Nearest Neighbor)	1.787	2.951*	0.00476	-4.743**	-4.667**	-0.0762
		(3.519)	(1.524)	(1.930)	(2.035)	(2.025)	(0.264)
	Observations	1,260	1,260	1,260	1,260	1,260	1,260
	Matching Diff-in-Diff (2-Nearest Neighbors)	1.114	2.611**	0.799	-4.525***	-4.379**	-0.146
		(2.738)	(1.038)	(1.501)	(1.737)	(1.759)	(0.215)
	Observations	1,260	1,260	1,260	1,260	1,260	1,260
	Matching Diff-in-Diff (3-Nearest Neighbors)	-0.251	2.819***	1.264	-3.831***	-3.687***	-0.144
		(2.399)	(0.983)	(1.255)	(1.420)	(1.427)	(0.184)
	Observations	1,260	1,260	1,260	1,260	1,260	1,260
	Matching Diff-in-Diff (5-Nearest Neighbors)	0.285	2.754***	1.129	-4.168***	-3.943***	-0.225
		(2.153)	(0.777)	(1.188)	(1.226)	(1.233)	(0.155)
	Observations	1,260	1,260	1,260	1,260	1,260	1,260
	Matching Diff-in-Diff (10-Nearest Neighbors)	1.400	3.055***	0.624	-5.078***	-4.980***	-0.0979
		(1.920)	(0.601)	(1.036)	(1.110)	(1.127)	(0.133)
	Observations	1,260	1,260	1,260	1,260	1,260	1,260
	Matching Diff-in-Diff (Kernel: Gaussian)	2.849	2.007***	-0.0289	-4.828***	-4.723***	-0.105
		(1.855)	(0.519)	(0.990)	(0.990)	(0.995)	(0.112)
	Observations	1,260	1,260	1,260	1,260	1,260	1,260
Men	Diff-in-Diff	2.352	-0.602	-0.293	-1.457***	-1.192***	-0.265***
		(1.813)	(1.658)	(0.423)	(0.408)	(0.371)	(0.102)
	Observations	1,843	1,843	1,843	1,843	1,843	1,843
	Matching Diff-in-Diff (1-Nearest Neighbor)	2.335	-1.555	0.356	-1.136*	-0.904	-0.232
		(2.484)	(2.301)	(0.603)	(0.657)	(0.618)	(0.123)
	Observations	1,354	1,354	1,354	1,354	1,354	1,354
	Matching Diff-in-Diff (2-Nearest Neighbors)	1.099	-0.419	0.643	-1.323***	-1.104**	-0.219**
		(1.891)	(1.773)	(0.460)	(0.503)	(0.478)	(0.092)
	Observations	1,354	1,354	1,354	1,354	1,354	1,354
	Matching Diff-in-Diff (3-Nearest Neighbors)	0.602	-0.390	0.690*	-0.902*	-0.681	-0.220**
		(1.650)	(1.653)	(0.405)	(0.493)	(0.461)	(0.087)
	Observations	1,354	1,354	1,354	1,354	1,354	1,354
	Matching Diff-in-Diff (5-Nearest Neighbors)	-0.115	0.295	0.701**	-0.881**	-0.606	-0.274***
		(1.371)	(1.235)	(0.354)	(0.417)	(0.388)	(0.074)
	Observations	1,354	1,354	1,354	1,354	1,354	1,354
	Matching Diff-in-Diff (10-Nearest Neighbors)	0.0666	0.251	0.696**	-1.013***	-0.749**	-0.265***
		(1.174)	(0.956)	(0.327)	(0.360)	(0.331)	(0.072)
	Observations	1,354	1,354	1,354	1,354	1,354	1,354
	Matching Diff-in-Diff (Kernel: Gaussian)	0.170	0.0848	0.718***	-0.972***	-0.701**	-0.272***
		(1.086)	(0.868)	(0.262)	(0.308)	(0.285)	(0.060)
	Observations	1,354	1,354	1,354	1,354	1,354	1,354

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors in parentheses (100 repetitions)

Table 24: Two-Year Results: Components of Leisure Definition 2, Alternative Matching Strategies (Zone-based Treatment Definition)

		Leisure	Market Work	Child Care	Total Home Production	Core Home Production	Procurement of Goods
Women	Diff-in-Diff	10.63***	-2.096*	-2.825*	-5.704***	-5.826***	0.122
		(2.912)	(1.211)	(1.669)	(1.630)	(1.570)	(0.237)
	Observations	1,753	1,753	1,753	1,753	1,753	1,753
	Matching Diff-in-Diff (1-Nearest Neighbor)	14.58***	1.679	-4.992**	-11.26***	-10.69***	-0.571**
		(3.571)	(1.597)	(2.034)	(2.043)	(1.912)	(0.303)
	Observations	1,278	1,278	1,278	1,278	1,278	1,278
	Matching Diff-in-Diff (2-Nearest Neighbors)	12.72***	1.860*	-4.151***	-10.42***	-9.772***	-0.653***
		(2.704)	(1.102)	(1.556)	(1.685)	(1.605)	(0.223)
	Observations	1,278	1,278	1,278	1,278	1,278	1,278
	Matching Diff-in-Diff (3-Nearest Neighbors)	10.26***	2.082**	-3.258***	-9.080***	-8.671***	-0.409**
		(2.319)	(0.879)	(1.243)	(1.476)	(1.432)	(0.189)
	Observations	1,278	1,278	1,278	1,278	1,278	1,278
	Matching Diff-in-Diff (5-Nearest Neighbors)	11.11***	1.822***	-4.210***	-8.726***	-8.378***	-0.347**
		(1.875)	(0.699)	(1.101)	(1.157)	(1.131)	(0.138)
	Observations	1,278	1,278	1,278	1,278	1,278	1,278
	Matching Diff-in-Diff (10-Nearest Neighbors)	12.04***	1.756***	-4.535***	-9.261***	-8.943***	-0.318***
		(1.656)	(0.535)	(0.977)	(1.042)	(1.021)	(0.120)
	Observations	1,278	1,278	1,278	1,278	1,278	1,278
	Matching Diff-in-Diff (Kernel: Gaussian)	12.73***	1.247***	-4.603***	-9.369***	-9.097***	-0.273***
		(1.751)	(0.470)	(0.931)	(1.008)	(0.979)	(0.104)
	Observations	1,278	1,278	1,278	1,278	1,278	1,278
Men	Diff-in-Diff	1.494	-0.468	-0.215	-0.811**	-0.663*	-0.148
		(1.729)	(1.581)	(0.484)	(0.401)	(0.366)	(0.110)
	Observations	1,867	1,867	1,867	1,867	1,867	1,867
	Matching Diff-in-Diff (1-Nearest Neighbor)	1.347	0.131	1.318***	-2.796***	-2.716***	-0.0802
		(2.324)	(2.180)	(0.621)	(0.718)	(0.699)	(0.118)
	Observations	1,372	1,372	1,372	1,372	1,372	1,372
	Matching Diff-in-Diff (2-Nearest Neighbors)	-0.665	1.357	1.208***	-1.901***	-1.808***	-0.0926
		(1.804)	(1.648)	(0.451)	(0.534)	(0.530)	(0.089)
	Observations	1,372	1,372	1,372	1,372	1,372	1,372
	Matching Diff-in-Diff (3-Nearest Neighbors)	-0.228	0.438	1.191***	-1.402***	-1.324***	-0.0782**
		(1.618)	(1.451)	(0.433)	(0.451)	(0.435)	(0.077)
	Observations	1,372	1,372	1,372	1,372	1,372	1,372
	Matching Diff-in-Diff (5-Nearest Neighbors)	-1.354	1.391	1.078***	-1.115***	-0.987***	-0.127*
		(1.304)	(1.162)	(0.319)	(0.377)	(0.374)	(0.066)
	Observations	1,372	1,372	1,372	1,372	1,372	1,372
	Matching Diff-in-Diff (10-Nearest Neighbors)	-1.329	1.381	1.073***	-1.126***	-1.000***	-0.126**
		(1.202)	(1.072)	(0.287)	(0.344)	(0.347)	(0.053)
	Observations	1,372	1,372	1,372	1,372	1,372	1,372
	Matching Diff-in-Diff (Kernel: Gaussian)	-1.228	1.550	0.943***	-1.265***	-1.099***	-0.165***
		(1.062)	(0.988)	(0.214)	(0.270)	(0.267)	(0.034)
	Observations	1,372	1,372	1,372	1,372	1,372	1,372

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors in parentheses (100 repetitions)