Child Care Provision: Semiparametric Evidence from a Randomized Experiment in Mexico.

Pierre Dubois* and Marta Rubio-Codina†

December 2010‡

Abstract

We estimate semiparametrically the impact of the Mexican conditional cash transfer program *Oportunidades* on the time mothers and older sisters spend taking care of children aged under 3, using the randomization of program placement and the methodology in Lewbel (2000). Results support the existence of substitution effects: mothers in treatment households are more likely to substitute for their older daughters' time to child care. As a result, daughters devote more time to schooling and less taking care of their younger siblings. Overall, total household time allocated to child care increases. These findings indicate that *Oportunidades* not only fosters human capital accumulation through keeping teenage girls in school but also through more and arguably better (mother provided) child care.

Keywords: Child Care Provision, Substitution Effects, Semiparametric Estimation, Conditional Cash Transfer Programs.

JEL Codes: D10, J13, J22, I00.

*Toulouse School of Economics (GREMAQ, INRA, IDEI): pierre.dubois@tse-fr.eu. 21 allée de Brienne, F-31000 Toulouse.

†Institute for Fiscal Studies: marta_r@ifs.org.uk. IFS, 7 Ridgmount Street, London, WC1E 7AE.

‡We would like to thank Márton Csillag, Paul Gertler, Sylvie Lambert, Arthur Lewbel, Thierry Magnac, seminar participants at IPEA, and participants at the NEUDC, ESEM, LAMES, RES and IATUR meetings for useful comments and suggestions on earlier versions of this paper. All errors are our own.

1 Introduction

In recent years, the economic case for public investment in Early Child Development (ECD) has become increasingly forceful. Carneiro and Heckman (2003) argue that returns to investments at early ages are higher than returns to investments later on in life mainly because of dynamic complementarities –namely, early childhood learning fosters and facilitates later learning. Both in developed and developing countries, poor ECD outcomes –often linked to poor family environments– are associated with inadequate school readiness and poor school performance (Currie 2001); a lower earning capacity (Currie and Thomas 2001); higher criminality (Carneiro and Heckman 2003; Garces et al. 2002); and lower levels of social integration.

While the relevance of stimulation and the home environment on ECD is well established empirically (see Walker et al 2007 for a review), caregivers may fail to provide adequate care and stimulation if they lack sufficient time, energy, knowledge and money. Because these resources are often scarce in impoverished rural environments, there is large scope for Conditional Cash Transfer (CCT) programs –such as the *Oportunidades* program in Mexico– to improve the circumstances in which children from beneficiary families begin their lives.[1] Even if CCT programs are not specifically designed as ECD interventions per se, the monetary incentives and additional benefits they provide –in the form of nutritional supplements, health checkups, and educational talks– are likely to affect child rearing practices within a household.

In this paper, we investigate whether *Oportunidades* affects child care provision through altering the allocation of time given over to child care amongst household members. We

[1] Gertler and Fernald (2004) provide a description of the high prevalence of low ECD outcomes amongst *Oportunidades* eligible children.

exploit time use data on the randomized *Oportunidades* evaluation sample to semiparametrically identify the impact of the program on participation and on the extent of participation in child care activities for mothers and sisters of under 3 year old children. We focus the analysis on mothers and their older daughters as they are found to be the two main child care providers in the household. Moreover and given the program design, the older daughter is the household member likely to contribute the most to the total transfer amount received by the household, conditional on her attending school. This strengthens the case for greater economic incentives to enhance substitution effects in the allocation of time devoted to child care between mothers and their older daughters. OLS, probit and semiparametric estimates support the existence of such a substitution effect. We find a 14% increase in mother provided child care in treatment households with teenagers 12 to 17 and children less than 3 years old. In turn, older daughters –ages 12 to 17– reduce their child care participation by 40%, and increase their participation in school activities by 9%. Overall, total household time to child care increases, which implies net increases in child care quantity.

The contribution of this study is twofold. Methodologically, we exploit the experimental nature of the *Oportunidades* evaluation data to obtain a semiparametric estimate of treatment on time allocation. We apply the Lewbel (2000) estimator for qualitative response models to binary and ordered data, and argue that randomization of treatment offers a unique setting in which to implement this recently developed estimator.[2] In terms of findings, the analysis provides evidence that *Oportunidades* increases human capital accumulation both by keeping teenage girls in school and through more and arguably "better" –mother provided– child care. Hence, linking benefits to school attendance can

[2] There are relatively few empirical applications of the Lewbel (2000) estimator. These include the works of Cogneau and Maurin (2001) and Goux and Maurin (2005).

simultaneously solve two related inefficiencies. First, an inefficiency in the levels of school attendance and domestic work provided by older daughters. Second, an inefficiency in the levels of child care provided in the household.[3]

The interest of economists in child care arrangements initially revolved around the responsiveness of female labor supply and child care demand to job related child care subsidizing policies (Heckman 1974; Michalopoulos et al. 1992; Averett et al. 1997). Since Blau and Robins (1988), a number of studies have addressed family labor supply, fertility and child care provision decisions within an intrahousehold time allocation framework (Mueller 1984; Tiefenthaler 1997). Following the expansion of CCT interventions worldwide, an increasingly extensive literature has developed around the impacts of these programs on child health and nutrition –see Lagarde et al. (2007) for a review– and more recently, on ECD (Gertler and Fernald 2004; Fernald et al. 2008; Paxson and Schady 2010; Macours et al. 2008). This paper contributes to both literatures by shedding some light on one of the mechanisms –namely, changes in household time allocation– through which CCT programs can affect child care provision, and in turn, ECD.

The remainder of the paper is organized as follows. The next section describes the rural *Oportunidades* program. In Section 3, we discuss the potential mechanisms at play behind the observed effects. Section 4 describes the data and the sample of analysis. In section 5, we present the Lewbel (2000) semiparametric estimator and discuss identification. Results are presented in section 6 and section 7 concludes.

[3] Of course this is under the assumption that the pre-program levels of schooling and child care provision were indeed inefficient, or in other words households were constraint in their choices.

2 The Rural *Oportunidades* Program

The Mexican government established the *Oportunidades* CCT program in 1997.[4] The primary objective was to break the inter-generational transmission of poverty by alleviating current poverty while investing in the human capital of the next generation. To this end, the program provides financial incentives (cash) to parents to invest in the human capital of their children, encouraging investments in health, nutrition and education. The program is one of the largest CCT interventions in the world, with approximately 3 billion US dollars distributed to some 5 million beneficiary households in 2008.[5]

When *Oportunidades* began rolling out in 1997, program eligibility was determined in two stages (Skoufias et al. 2001). First, underserved communities were identified based on the proportion of households living in very poor conditions as defined using data from the 1995 census (*Conteo de Población y Vivienda*). Second, low-income households within those communities were chosen by means of a proxy means test. The score was constructed using basic socio-economic data collected on all households in the communities through the *Encuesta de Características Socioeconómicas de los Hogares* (ENCASEH). This process designated 52% of the households in selected communities as elegible for benefits.[6] All eligible households living in treatment communities were offered *Oportunidades* and over 90% enrolled. Once enrolled, households received benefits for a three-year period conditional on meeting the program requirements with the possibility of being recertified.

[4] *Oportunidades* was originally known as *Progresa*. The name was changed during the Fox administration in 2000. The data used here are from the rural *Progresa* evaluation. However we refer to the program under its current name, *Oportunidades*, throughout the text.

[5] http://www.oportunidades.gob.mx/informacion_general/main.html

[6] Subsequently, the Government decided that a subset of households had been unduly excluded and expanded the eligibility criteria to include a set of slightly wealthier households in a process called "densification" (Hoddinott and Skoufias 2004).

Monetary benefits represent on average over 20% of total household income. They are given bimonthly to the female head of the household in two forms. The first is a nutritional grant given to *all* families to spend on more and better nutrition. It is complemented with nutritional supplements and immunization directed to 0 to 2 year olds, and to pregnant and lactating women; and regular health checkups for all household members. The second is an educational grant given to *each* child younger than 18 and enrolled in school between the third grade of primary school and the third grade (last) of secondary school. The scholarship rises substantially after graduation from primary school and is higher for girls than boys during secondary school. It is received conditional on children attending a minimum of 85% of school days and on not repeating a grade more than twice. The female head of the household must also attend monthly "pláticas" or educational talks on health related topics.

Skoufias (2005) discusses the program at length and provides a review of its impacts.

3 *Oportunidades* and Child Care Provision

According to traditional household models, family utility is maximized when household members allocate their time to the production of those commodities in which they have a comparative advantage (Becker 1973). Women's believed comparative advantage in home time would thus explain part of the gender gap in market work participation and female specialization in household activities, including the care of children. Even if the traditional division of labor between genders is less and less patent in western economies, it continues to be firmly established in rural Mexico (INEGI 2002; Parker and Skoufias 2000). As an illustration, Figure A plots participation rates in child care by age and sex in the *Oportunidades* evaluation sample. Child care is clearly a female activity: while participation rates oscillate between 40 and 60% for adult women; men participation rates are around 8%. A peak is observed for both men and women in their twenties. Note also that female participation increases sharply from the age of 12.

These patterns in the allocation of time devoted to child care can be framed into different types of household models. We could use the "separate spheres" bargaining model of the household developed by Lundberg and Pollak (1993), or household collective models (Chiappori 1988; Browning, Chiappori and Lechene 2010), or even a standard unitary household model. Each specific set up would allow modelling the possible effects of the CCT program on household time allocation, using interior or corner solutions to account for the fact that child care provision is a female (wife/mother) activity. The objective of this paper is not to infer which model is more appropriate but rather to focus on the allocation of time to child care by the two main caregivers in the household –namely the mother and her 12 to 17 year old daughter– and how it changes as a result of the intervention.[7] To this end, we take the unitary household model approach as it provides the simplest framework for interpretation of results, without loss of generality.

We consider that the mother allocates her time between child care and leisure, and her first daughter's time between child care, schooling, and leisure. She chooses the optimal levels to maximize her utility function –which is a function of total time to child care, her and her daughter's leisure, and her daughter's schooling– subject to a budget constraint, and to her and her daughter's time constraints.[8] In this stylized household, the *Oportunidades* intervention amounts to:

[7] We assume that market child care services are unavailable in these disadvantaged rural communities.

[8] Such solution would coincide with the noncooperative equilibrium allocations reached by the mother under a "separate spheres" bargaining model. This would be the household final equilibrium if, for example, transaction costs were very high. We indeed observe in the data that household members outside the "sphere" (males) do not alter their (initially very low) contribution of time to child care.

(i) an increase in the mother's nonlabor income given the nutritional grant.

(ii) the provision of a minimum level of maternal care, given the required attendance to the "pláticas", preventive health visits and the nutritional supplements.

(iii) a reduction in the price of schooling given the educational grant that the 12 to 17 year old daughter receives conditional on attendance. This implies that time in child care is more expensive relative to time in school for daughters living in treatment households.

The budget constraint in treatment households thus integrates the change in the price of schooling and the unconditional nutritional grant. Assuming interior solutions, each of the intervention components described above result in:

(i) an ambiguous effect on total maternal child care. If child care is assumed a normal good, maternal child care provision increases with income controlled by the mother through an income effect. However, because child care requires maternal time as an input factor, increases in income might increase leisure, and reduce child care time.

(ii) direct increases in the total quantity of child care.

(iii.a) increases in the daughter's time to school through the own-substitution effect because of the reduction in the price of schooling.

(iii.b) reductions in the daughter's time allocated to child care through the own-substitution effect, assuming the daughter's child care and schooling times are substitutes.

(iii.c) increases in maternal child care time through the cross-substitution effect, given her time and her first daughter's are substitute inputs in the production of child care.

In all treatment households, maternal time allocated to child care will be affected by the nutritional grant (income effect) and compliance with the program requirements (preventive health visits and attendance at "pláticas"). However, cross-substitution effects in child care time (effect $iii.c$) will only arise amongst those mothers whose daughters are eligible to receive the educational grant. In the empirical exercise, we will exploit both the random allocation of benefits and heterogeneity in mother's offspring to disentangle the cross-substitution effect from the composite of the other two effects (income effect and compliance with program requirements). This composite effect will be a residual in our empirical specification, whose components are unidentifiable.

4 Experimental Design and Data

We benefit from the fact that the Mexican Government was committed to a rigorous evaluation of the impact of *Oportunidades* using a controlled-randomized evaluation design. Given budgetary and logistical constraints, the Government could not enrol all eligible families in the country simultaneously and decided to phase in the enrolment of entire communities over time instead. As part of this process, the Government randomly chose 320 treatment and 186 control communities in seven states for a total of 506 experimental communities.[9] Eligible households in treatment communities began receiving benefits in April of 1998; whereas eligible households in control communities were not incorporated until November of 1999. In order to minimize anticipation effects, households in control communities were not informed that *Oportunidades* would provide benefits to them until

[9] Behrman and Todd (1999) test for statistical differences in the distributions of a large set of observable characteristics between treatment and control households and assess the validity of the randomized allocation of treatment.

two months before incorporation.[10]

The data used in this paper comes from the Oportunidades rural evaluation surveys, the Encuestas de Evaluación de los Hogares Rurales (ENCEL), and the ENCASEH baseline data. The ENCEL interviewed all households in the 506 evaluation communities every six months between 1998 and 2000, and again in November 2003. The May 1999 survey collected additional data on time use for all household members older than eight.[11] By then, treatment households had enjoyed benefits for over a year, while no control household had yet received transfers. This allows us to obtain an estimate of the average treatment effect on time devoted to child care.[12]

We construct a dataset of mothers older than 18 years of age living in eligible households in May 1999. We then match each mother to the characteristics and time allocation of her older daughter –younger than 18 and still living in the household. We use information on time allocation to construct our main dependent variables. Specifically, the question on time devoted to child care reads: "how many minutes did household member i devote yesterday to the care of small children, the elderly or the sick?". To narrow the scope of the question to the care of young children, we restrict the sample of analysis to mothers of children younger than 3 living in households where there are no elderly or sick

[10] Attanasio et al. (2005) find no evidence of anticipation effects among control households.

[11] Parker and Skoufias (2000) and Rubio-Codina (2010) provide further details on these data. Rubio-Codina (2010) estimates the impacts of the program on intra-household time allocation but does not specifically focus on time to child care nor does it disentangle the different roles of mothers and daughters to this and other activities.

[12] Time use data was also collected as part of the November 2000 ENCEL round. As all eligible households were already receiving benefits by then, we use this cross-section to test for differential effects resulting from different lengths of exposure to the program benefits. We find increases in maternal time to child care for mothers that have been in the program for longer (i.e. households in the original treatment group). Results are available upon request.

members that might require care. Given the data available, this implies dropping out of the sample households with: (i) elders older than 65 that did not engage in any paid or unpaid work activity during the week before the interview; and (ii) members older than 6 that reported being unable to perform regular daily activities during the month prior to the interview.[13]

The final sample consists of 4,036 mothers with children younger than 3 (see Table 1). This represents about one third (34%) of all eligible households, as classified according to the original classification scheme.[14] Approximately 26% of these women are also mothers of a 12 to 17 teenage girl. Note that the proportion of mothers with different offspring compositions is very similar in treatment and control households and remains similar to the original (randomized) distribution –60% treatment and 40% control. This suggests that the potential for sample selection and sample composition biases is negligible.

Panel I in Table 2 shows summary statistics of the dependent variables for treatments and controls in May 1999. In the last column, we report a test of the equality of raw means. Conditioning on having a teenager aged 12 to 17, mothers in the treatment group have higher participation rates in child care than mothers in the control group although

[13] Although the program has improved self-reported health status for children and adults (Gertler and Boyce 2001) as well as children's nutritional status (Rivera et al. 2004; Behrman and Hoddinott 2005), we argue that excluding these households does not bias our estimates because: (i) they only represent 6.6% and 1.6% of the households in the estimation sample, respectively; (ii) these proportions are balanced across treatment and control groups; and (iii) parametric estimates are robust to keeping these households in the estimation.

[14] We exclude "densified" households from the analysis so as to not attribute treatment effects to households that were not treated until later because of the administrative delays in the disbursement of transfer payments. Because households were categorized as "densified" in both treatment and control communities and under the same criteria, excluding them does not compromise the internal consistency of our estimates.

the difference is not statistically significant. For first daughters aged 12 to 17, there is a 6.8 points significant reduction in child care participation and a 7.9 points significant increase in school participation given treatment. There are no significant differences in the amount of time devoted to any of the activities considered conditional on participation.

In Panel II we report summary statistics on maternal and household characteristics in May 1999; and on baseline household and community characteristics for the sample of mothers under analysis. Mothers of children younger than 3 are around 30 and have 3.5 years of education on average. Less than 2% are the head of the household and between 6% (control) and 9% (treatment) work for a wage. 41% of the mothers in the sample are indigenous. On average, they have between 1 and 2 children younger than 3, 3 children younger than 7, and 1 child ages 8 to 17. The test of equality of means shows no significant differences for any of the variables reported, as is expected from random assignment. More importantly, there is not sufficient evidence to reject the hypothesis of (statistically) equal offspring composition between mothers in treatment and control groups during the intervention years. The only exception is the number of sons ages 12 to 17, which is significantly larger for treatment mothers.[15]

5 Semiparametric Estimation and Identification

Our objective is to identify empirically whether there have been changes in time allocated to child care in the household as a result of the intervention, and understand the underlying mechanisms at play. For this purpose, we next exploit the exogenous variation introduced

[15] We cannot test the exogeneity of treatment by comparing baseline time allocation patterns between treatments and controls due to lack of data. Randomization should however guarantee that they were not statistically different.

by the random assignment of households to treatment and control communities, as well as heterogeneity in household composition.

5.1 Empirical Specification

We specify the overall program impact on child care provision as:

$$\tilde{y}_i = \alpha_0 v_i + \alpha_1 T_i + \sum_{r=1}^{R} \beta_r x_{ri} + \varepsilon_i \qquad \forall i = 1, .., N \tag{1}$$

where $\tilde{y}_i$ is the number of hours individual i allocates to child care, T_i is a binary variable equal to 1 if i lives in an original treatment community and 0 otherwise; v_i is the total transfer amount the household *has potentially received* since taking up the program; and the x_{ri} are R individual, household and community characteristics (listed below). It is important to recall at this point that while all eligible household in treated villages receive benefits ($T_i = 1$), they do not all receive equivalent transfer amounts (because the program rules are such that cash benefits differ notably according to the education levels of children). Moreover, because of administrative (i.e. random) implementation delays, households may have delayed take-up of the program within treatment villages. Thus, α_1 identifies the average effect of treatment conditional on total accumulated transfers received v_i. In other words, it identifies the counterfactual effect of being treated by the conditional cash transfer program given some transfers v_i received by the household. Notably, the total average effect of the program on a given individual also depends on the amount of transfers received by the household; which, as noted earlier, is a function of characteristics x_{ri} by design. Finally, note that in our specification, interaction terms between v_i and T_i are already taken into account since $v_i T_i = v_i$, by construction. This will prove important in the semiparametric method as it relies on an exogeneity and partial independence assumption that would not be satisfied if some omitted interaction term appears in the error term ε_i.

Since the true number of hours is not directly observed, $\tilde{y}_i$ is a latent variable. For simplicity, we model the observed variable y_i in two different ways: (*i*) as a participation dummy that equals 1 if i spends a positive number of hours on child care ($\tilde{y}_i > 0$) and 0 otherwise ($\tilde{y}_i \leqslant 0$); and (*ii*) as an interval indicator that can attain up to k different values depending on the extent of time i devoted to child care (up to one hour, between one and two hours, etc.).[16]

In a second specification, we interact the household treatment status with dummies controlling for the mothers' offspring composition to capture heterogeneous responses across mothers living in different household environments:

$$\tilde{y}_i = \alpha_0 v_i + \alpha_1 T_i + \alpha_2 S_i^j + \alpha_3 T_i S_i^j + \textstyle\sum_{r=1}^{R} \beta_r x_{ri} + \varepsilon_i \qquad \forall i = 1, .., N \tag{2}$$

where S_i^j equals 1 if the situation j is true. We consider two possible situations: (*i*) mother i has offspring aged 12 to 17, and (*ii*) mother i has daughters aged 12 to 17. As noted in section 3, cross-substitution effects in maternal time allocated to child care will only arise amongst those mothers who live in treatment households ($T_i = 1$) and have school aged children eligible to receive the educational grant ($S_i^j = 1$). We argue that the coefficient on the interaction of treatment and offspring composition, α_3, identifies this cross-substitution effect. Given the child care participation patterns shown in Figure A –i.e. females older than 12 are the main child care providers in these communities– we expect any cross-substitution effect to take place between mothers and their 12 to 17 year

[16]Even if time devoted to child care is a continuous variable as is measured in minutes, the empirical distribution shows a discrete support on specific values, mainly hours and half hours. For this reason, we categorize total hours into a limited number of intervals defined sensitively to the underlying thresholds in the empirical distribution of hours and guaranteeing enough power in each cell. This allows our method to be unaffected by households answering with approximations and rounding to the closest half hour.

old daughters.[17] In this specification, the coefficient on the treatment dummy, α_1, is the remaining effect of the program on the mother's time to child care. It is composed by a combination of the impacts coming from the nutritional grant (income effect) and from compliance with the program requirements (preventive health visits and attendance to the "pláticas") that affect all mothers in treatment households –regardless of their offspring composition.[18]

In the event of cross-substitution effects between mothers and first daughters, one should also expect reductions in the older daughter's participation and in the time she spends caring for her younger siblings. To test this, we estimate equation (1) on the (extent of) child care participation of first daughters –aged 12 to 17– of those mothers in the sample described above.[19] We additionally estimate equation (1) for first daughters' school participation and for their extent of schooling and leisure time in order to obtain a broader picture of their time allocation. Finally, we also estimate equations (1) and (2) on the extent of leisure time for mothers.

5.2 Semiparametric Identification

Parametric identification of the parameters in equations (1) and (2) is possible using a standard probit model –when y_i is dichotomous– and an ordered probit model –when y_i is

[17] Moreover, and as noted before, because girls enrolled in secondary school receive the largest transfer amounts, teenage girls are the household members that most significantly contribute to total household transfers (conditional on secondary school enrolment).

[18] These effects could be confounded with other factors –such as maternal education– also correlated with household composition. However, the random allocation of benefits and the fact that the sub-sample of treatment and control mothers is balanced (see Table 2) dismisses such concern.

[19] Recall that the number of daughters aged 12 to 17 is balanced between mothers in the treatment and control groups (see Table 2). Note also that over 70% of the mothers in the estimation sample have only one daughter 12 to 17 years old.

polychotomous– under the assumption that the error term of the latent variable follows a normal distribution and the normalization of a parameter ($\alpha_0 = 1$, for example). However, parametric identification relies too heavily on the chosen distribution of the error term. In our setting, additional problems arise given that the error term is likely heteroscedastic for two reasons: (i) we are testing for heterogeneous treatment effects across households with different demographic compositions (i.e. α_1 and α_3 are different); and (ii) Oportunidades sampled a large number of randomized communities (clusters), each consisting of relatively few poor correlated households. As such, maximum likelihood will only be valid as the number of observations in the cluster tends to infinity with the cluster unit fixed.

We have tested –and generally rejected– normality using Conditional Moments and other standard tests. Results are available upon request. Consequently, and in order to avoid the problems of specifying a parametric model, we estimate these discrete choice models semiparametrically. Semiparametric estimation has the advantage of not imposing any particular distribution (normal, logistic, etc.) on the latent variable errors and allows them to suffer from conditional heteroscedasticity of unknown form. Khan and Tamer (2009) provide an alternative distribution free estimator of such single index models. However, we opt for the method of Lewbel (2000) because of the suitability of our data for such purpose. Indeed, we can exploit the randomization of treatment in the data to justify the partial independence assumption (Assumption A2 below), which is at the basis of the method. Moreover, this assumption, has the advantage of guaranteeing that the error term has a conditional mean which does not depend on interaction effects between partially independent regressors and other regressors, thus avoiding some form of mispecification error in the single index model.

Let us denote y, v and ε the column vectors of y_i and v_i and ε_i, respectively. X is the

$n*(3+R)$ matrix of right hand side variables except v_i also called the "special regressor"; and β the column vector $[\alpha_1, \alpha_2, \alpha_3, \beta_1, ..., \beta_R]$. $I_{\{.\}}$ is an indicator function that equals one if $\cdot$ is true and zero otherwise. Lewbel (2000) considers the binary choice model,

$$y = I_{\{v+X\beta+\varepsilon>0\}} \tag{3}$$

and the following assumptions, whose empirical validity will be discussed in the next subsection:

A.1: Continuity: the conditional distribution v given X is continuous.

A.2: Partial Independence: the conditional distribution of ε is independent of v given X, $F_\varepsilon(\varepsilon|v, X) = F_\varepsilon(\varepsilon|X)$.

A.3: Large Support: the conditional distribution of v given X has support $[v_L, v_H]$ that contains zero: $v_L \leq 0 \leq v_H$. The support of $-X'\beta - \varepsilon$ is a subset of $[v_L, v_H]$.

A.4: Uncorrelated errors: $E(\varepsilon X) = 0$, as in linear models.

Under assumptions A.1 to A.4, Lewbel (2000) shows that β can be estimated (with root N consistency) by an ordinary least squares regression of y^* on X, where

$$y^* = \frac{y - I_{\{v>0\}}}{f(v|X)} \tag{4}$$

and $f(v|X)$ denotes the conditional probability density function of v given X. If the distribution of v is unknown, a nonparametric first stage is needed to estimate it. Alternatively, an ordered data estimator can be used under more stringent conditions. For simplicity and precision in the estimation, we will assume that v follows a normal distribution as parametrized in Lewbel (2006).[20]

[20] A description of the implementation of the Lewbel estimator is available upon request.

Lewbel (2000) shows that the proposed methodology extends to ordered response models with K choices defined as:

$$y = \sum_{k=0}^{K-1} k I_{\{\alpha_k < v + X\beta + \varepsilon \leq \alpha_{k+1}\}} \tag{5}$$

where $\alpha_0 = -\infty$ and $\alpha_K = +\infty$. In this case, the transformation of the dependent variable is written as:[21]

$$y^* = \frac{\frac{y}{K-1} - I_{\{v>0\}}}{f(v|X)} \tag{6}$$

Assumptions A.1 to A.4 imply that the conditional probability of success, $pr(y = 1|v, X)$, increases monotonically and varies from 0 to 1 over the support $[v_L, v_H]$ of v. As this is admittedly very restrictive in empirical applications, Magnac and Maurin (2007) propose an alternative assumption to A.3: a symmetry condition on the tails of the errors ε. Let $y_{v_L} = X'\beta + v_L + \varepsilon$ be the propensity of success for individuals with the smallest v, v_L; and $y_{v_H} = -(X'\beta + v_H + \varepsilon)$ the propensity of failure for individuals with the largest v, v_H. Then, the symmetry condition can be expressed as:

A.5: $E(X' y_{v_L} I_{\{y_{v_L}>0\}}) = E(X' y_{v_H} I_{\{y_{v_H}>0\}})$

A.5 requires that the propensity of success y_{v_L}(or $pr(y = 1|v_L, X)$) and the propensity of failure y_{v_H} (or $pr(y = 0|v_H, X)$) are identically distributed. If so, the Lewbel (2000) estimator is unbiased. If symmetry of the tails is not satisfied, it is always possible to choose conditional distributions for y_{v_L} and y_{v_H} –by trimming outliers in the distribution

[21] Lewbel (2000) also proposes an extension to censored data which consists in applying the ordered data estimator repeatedly. More precisely, it involves: (i) defining a continuum of values for α_k (thresholds), (ii) obtaining a $\hat{\beta}_k$ for each threshold defined, and (iii) efficiently combining the $\hat{\beta}_k$'s. Estimating as many $\hat{\beta}_k'$'s as values takes the dependent variable is an unnecessary computational burden. For simplicity and given the empirical distribution of the dependent variable, we chose to categorize total hours into a limited number of intervals, as discussed earlier.

of $v-$ in such a way that symmetry is more likely satisfied (Magnac and Maurin 2007). Moreover, Khan and Tamer (2010) have shown that the rate of convergence of the Lewbel (2000) estimator can be slower than the parametric one and that numerical instability can happen depending on tail distributions. We will thus devote some particular attention to the support condition in order to apply this method and have tested the stability and robustness of our estimator to variations in the trimming parameters.

5.3 The Special Regressor

Our chosen special regressor v_i is the *potential* transfer amount that the household *should have* accumulated at the time of analysis (May 1999) since it first received benefits. To construct v_i, we take household's composition and children's school enrolment in 1997 (baseline). We compute the accumulated transfer amount by applying the program benefit allocation rules since first take up, assuming no school drop out and no grade repetition. Note that v_i and T_i differ since different treatment households are eligible for different transfer amounts depending on the gender and age of their school-aged children and on whether they are enrolled in school. Since v_i in May 1999 is predicted projecting forward baseline household composition and school enrolment, the required assumption for identification is that *pre-program* household composition and school enrolment are exogenous to program allocation. This is guaranteed by the randomization and validated by Behrman and Todd (1999) for the evaluation subsample. As seen in Table 2, our subsample of analysis is also balanced in terms of observable baseline and current characteristics.

We next argue that the total amount of transfers potentially received at t is a valid special regressor as it satisfies the assumptions required for identification: continuity, monotonicity, partial independence and large support.

First, the amount of accumulated potential transfers is a continuous variable (A.1) and includes zero in the support (A.3). It is indeed positive for treatment households and zero for control households.

Second, randomization of treatment at the community level guarantees that the treatment dummy is independent of observables and unobservables, $(\varepsilon_i, X_i) \perp T_i$. In turn, this implies that $\varepsilon_i \perp v_i | X_i$ (A.2) since the accumulated potential transfer is exogenous by construction. As defined above, the cash transfers are $v_i = T_i^* f^*(D_{i97})$ where $f^*(.)$ is known, given by the design of the program and a function of a subset of baseline observables $D_{i97} \subset X_{i97}$ and T_i^* represents the length of time the household has been receiving benefits by May 1999. We can also rewrite v_i as $v_i = T_i f(D_i) u_i$ where T_i is the randomized treatment status and $f(.)$ is a known function of a subset of May 1999 observables D_i (household demographics, school attendance and grade attended) and $D_i \subset X_i$. Thus, by construction, $u_i = \frac{T_i^* f^*(D_{i97})}{T_i f(D_i)}$ is an exogenous variable composed by two random elements: (i) the administrative difficulties that delayed the reception of benefits amongst beneficiaries; and (ii) any departure in household demographics and children's school attendance in May 1999 from the situation predicted using baseline information. Note that $u_i \neq 0$ implies that v_i is a nondeterministic function of some of the other regressors in X_i, as is required for identification.

As noted, a testable implication of assumptions A.1 to A.4 and A.5 is the monotonicity of the conditional probability of success $pr(y_i = 1 | v_i, X_i)$ over the support of v_i. In this setting, monotonicity amounts to assuming that: (i) the mother's child care time and the first daughter's schooling time are nondecreasing functions of the cumulative potential transfers (v_i); and (ii) the first daughter's child care time is a nonincreasing function of v_i. We have discussed the theoretical validity of these assumptions in section 3 and "tested" its empirical validity (results available upon request).

6 Results and Discussion

We semiparametrically estimate the effect of *Oportunidades* on participation in child care, and on the extent of participation in child care and leisure for mothers of children under 3 and for the older daughter –aged 12 to 17– of these mothers. For the first daughter, we also estimate the program impact on participation and on the extent of participation in school. For comparison purposes, we first report OLS estimates (Models A), then probit –or ordered probit, depending on the dependent variable– results (Models B), and finally the semiparametric (Lewbel) estimates (Models C). As our interest concerns the application of the Lewbel method, we will focus the discussion on the semiparametric results.[22] In all sets of estimates, we trim observations with values of the special regressor v_i in the top 7% of its distribution to conform to the symmetry condition (A.5). We also trim extremely low values of the conditional probability density function of v_i as they imply, by construction, outlier observations of the transformed dependent variable y^*. An analysis of the sensitivity of the Lewbel estimator to trimming is available upon request.

All regressions include the following explanatory variables: maternal age, age squared, years of education, ethnicity, head of the household status and whether she is a paid worker; first daughter's education (in years); baseline and contemporary household demographic composition; baseline assets (dirt floor, electricity and farm size); and community

[22] Note that the OLS, probit (or ordered probit) and Lewbel estimates are not comparable in magnitude for various reasons. First, OLS estimates are likely biased given the binary and ordered nature of the dependent variables. Second, probit and ordered probit impose a normalization to one of the variance of the errors for identification. Third, OLS, probit and ordered probit do not impose the normalization to one of the coefficient on the special regressor as in the Lewbel method.

characteristics (male agricultural wage in the community, distance to large urban center, distance to secondary school, presence of pre-school and presence of junior high school imparted via TV, or "telesecundaria" in the community). Estimates are robust to the exclusion of these controls.

6.1 Maternal Time

Table 3 shows estimates of the impact of the program in May 1999. Panels I and II present results for mothers' participation in child care (binary outcome) and for their extent of participation (ordered polychotomous outcome). For mothers, the extent of participation in child care variable takes six different values: $k \in \{0, 1, ..., 5\} = \{$no time devoted to child care, up to one hour, between one and two hours, two and four hours, four to seven, seven to thirteen$\}$.[23] Estimation results follow very similar patterns for both types of outcomes.

Models 1A to 1C in Panel I show no significant effect of treatment on maternal participation in child care when all mothers are pooled together neither parametrically nor semiparametrically. However, when treatment is interacted with whether the mother has offspring aged 12 to 17, the effect becomes positive and significant. Moreover, the coefficient on having 12 to 17 year old children alone is negative and significant. We interpret these findings as indicative of: (i) a cross-substitution effect between mothers and their older children in child care provision; and (ii) *Oportunidades* attenuates the cross-substitution effect. The semiparametric coefficient on having children 12 to 17 interacted with treatment is 0.44 (Model 2C, Panel I) and the mean marginal effect is 7.15

[23] Results are robust to a redefinition of these categories which were defined based on the empirical distribution of hours and to ensure enough power in each cell.

percentage points.[24] Given an initial participation rate of 51.4%, this results in a 13.9 percentage increase in participation in child care for mothers with children 12 to 17. We further interact treatment with a dummy equal to 1 if the mother has a daughter –as opposed to a child– aged 12 to 17 and find that the coefficient on this interaction is also positive and significant (Model 3C, Panel I).

Similar effects are observed on the extent of participation in child care. Mothers of daughters 12 to 17 are more likely to increase the amount of time taking care of their younger children given treatment (Model 2C, Panel II). As before, the coefficient on treatment alone is negative and non statistically significant (Model 1C, Panel II). The analysis on the extent of maternal leisure in Panel III shows opposite effects: mothers with children younger than 3 and teenagers 12 to 17 enjoy less leisure given treatment. This reduction is partly explained by increases in time to child care, to other domestic activities and to compliance with the program requirements.

The qualitative evidence in Adato et al. (2000) endorses these findings. During the summer of 1999, the authors conducted focus groups with 230 beneficiary and non-beneficiary women to learn about their perceptions on the program. The authors report: "Another reason that women's time burden increases is because of the need to do work that was previously done by children who are now attending school, particularly secundaria.

[24] For dichotomous outcomes, we compute the marginal effects on the estimated coefficients as: $M_{ij} = \frac{\triangle[1-\hat{G}(-v_i - x_i\hat{\beta})]}{\triangle x_j} = \frac{\triangle[1-\hat{G}(z_i\hat{\beta})]}{\triangle x_j} = \hat{G}(\hat{z}_i^0) - \hat{G}(\hat{z}_i^1)$, where M_{ij} is the effect of switching the j^{th} binary variable, x_j, from 0 to 1 on the probability that y_i equals one; and $\hat{z}_i^1$ is the value of the index $-v_i - x_i\hat{\beta}$ when the j-th binary variable is set to 1, and similarly for $\hat{z}_i^0$ when the value of j is set to zero. $\hat{G}(.)$ is the estimated cumulative distribution function of the probability of y_i given z_i. We nonparametrically estimate $\hat{G}(.)$ running the kernel regression of y_i on z_i. Results reported in the text were computed using a Gaussian Kernel and 500 equally spaced points in the range of z_i and are robust to larger numbers of points.

However, their mothers see this as worthwhile in order for their children to study. (...) Although some women said that the father also does some of this work, more often it was the mother." (Adato et al. 2000, p. xiii).

6.2 First Daughter's Time

Next, we turn our attention to the allocation of time by first daughters. Panels I and II in Table 4 present the estimated program impact on first daughters' participation in child care and on their extent of participation. The latter variable takes four different values: $k \in \{0, 1, 2, 3\} = \{$no time devoted to child care, up to one hour, between one and two hours, more than two hours$\}$.

Semiparametric estimates evidence a reduction of 7.1 percentage points in the first daughter's probability of participating in child care (estimated coefficient of -0.032, Model C in Panel I) or a 40.1% decrease. Significant reductions are also observed in terms of the extent of participation (Model C in Panel II). A concern is that the observed reductions in teenage child care participation are in fact driven by an increase in school attendance of other siblings in primary school age. Unfortunately, this hypothesis is not testable as almost all teenage girls with siblings younger than 3 also have siblings in primary school age (6 to 11). It is well-known, however, that the program had little effect on primary school enrolment (Schultz 2004), which makes this an unlikely driver of the results found.

About 65% of the first daughters in the sample that engage in child care activities do not attend school. 3% of them report that they are not enrolled because they have to help in the house.[25] Moreover, conditional on school enrolment, another 3% report having to take care of their siblings as one of the reasons why they miss school. Other

[25] Note, however, that a majority of teenage girls report not going to school because they do not have enough money (51%), they do not like it (18%), or the school it too far (8%).

more frequent reasons are teacher absenteeism, illness, or care of the sick.

These figures are somewhat indicative that child care and schooling are substitute activities. Not surprisingly, we semiparametrically identify a significant increase in participation in schooling activities (attendance and homework) for first daughters aged 12 to 17 (Model C in Panel III). The estimated coefficient of 0.08 translates into a marginal effect of 3.9 percentage points or a 9.3% increase in schooling, which is consistent with the positive treatment impact on female secondary school enrolment reported in Schultz (2004) and Parker and Skoufias (2000). The increase in the extent of time allocated to schooling activities is, however, not significant. Moreover, we observe a significant reduction in the extent of leisure for first daughters (Model C in Panel V). This might suggest that these girls reduce (or stop) their contribution to child care and instead, take up schooling full time, as a result of the reduction in the price of schooling. Another possibility is that the intervention affects the allocation of teenage girls' time to other activities.[26]

6.3 Quantity and Quality of Care

In this subsection, we estimate equations (1) and (2) on total household hours devoted to child care. We define total household hours to child care as the sum of hours each household member reports spending in child care and categorize the variable as k ={0,1,...,6} ={no time allocated to child care, up to one hour, one to two hours, two to three hours, three to five, five to seven, more than seven}.

We should expect no effect on total household hours if mothers and their 12 to 17 year

[26] As a robustness check, we re-estimate all previous regressions on the restricted sample of mums and older daughters whose time information refers to weekdays. The estimated coefficients (available upon request) are very similar –albeit less precisely estimated– to the ones reported here.

old daughters are the only two household members taking care of the very young and if mothers fully substitute for their daughters' time. Conversely, if mothers –and possibly other household members– increase time devoted to child care by more than the amount previously devoted by teenage girls, α_3 in equation (2) should be positive and significant. A negative and significant α_3 would imply a reduction in total household hours to child care in treatment households. Table 5 shows a significant increase in the extent of total household hours to child care in treatment households with children under 3 and teenagers ages 12 to 17. This implies that *Oportunidades* fosters net increases in total child care provided within the household.

The next natural question is whether mothers alone are more than compensating for care time previously provided by their older daughter or whether other household members are also contributing. Table 6 shows results from estimating (1) and (2) using the share of child care hours provided by household member m over total household hours as the dependent variable. We consider the following "sets" of household members: mothers of 0 to 3 year old; other adult women living in the household; adult men; brothers aged 8 to 11 and 12 to 17; and sisters aged 8 to 11 and 12 to 17 of under 3 year old children.

Results are consistent with the premise that the mother is the household member providing a larger share of child care in substitution of the older daughter and overall. The coefficient on treatment interacted with girls (daughters) 12 to 17 on the share of mother's time is positive and almost significant at 10%. The coefficient on treatment for the share of time devoted to child care for daughters aged 12 to 17 continues to be negative and significant. Surprisingly, daughters aged 8 to 11 –who are the sisters of the 12 to 17 year old daughter– also seem to increase their share of child care provision in treatment households. Because of the low participation rates for this subpopulation (around 3%), we are inclined to think that this effect is driven by a few outliers.

Hence, and as long as we believe that the mother is more productive in the provision of child care, these findings are suggestive that the program entails gains not only in the quantity but also in the quality of the care given to the very young. The biology and psychology literature have repeatedly acknowledged the mother as the best child nurturer. Frequent breast-feeding and mother warmth are widely recognized as key care practices (UNICEF 2001). Variations in the quality of maternal care are also proven to produce lasting changes in stress reactivity, anxiety, and memory function in the offspring (Grantham-McGregor et al. 2007). The works of Case and Paxson (2001) and Case et al. (2001) are examples in the economic literature of the important role the biological mother –as opposed to the stepmother– plays in the adequate investment in the child's health and education. Moreover, beneficiary mothers increase their knowledge on parenting both through the interaction with medical staff at the health centers and by attending the "pláticas" –the educational talks covering best health, hygiene and nutrition practices.

On the other hand, one could argue that environmental factors (the education of the caregiver, for example) matter more than biological attachment. In the current context, this would imply that older daughters are better caregivers as they are, on average, more educated than their mothers. However, when we interact the treatment dummy with years of education of the mother we find stronger substitution effects amongst more educated mothers (results available upon request). This result confirms that the increase in quantity of care is, on average, likely to go hand in hand with increases in quality.[27]

[27] We are currently using more recent rounds of the *Oportunidades* evaluation data to investigate whether increases in maternal time to child care do improve child cognitive and non-cognitive development. Because we have to rely on a somewhat different identification strategy, we consider this issue to be beyond the scope of this paper.

7 Conclusion

This paper provides robust semiparametric evidence that does not rely on *ad hoc* parametric models of the effect of the *Oportunidades* program on child care provision.

After showing that child care provision is a female activity in rural Mexico, we have argued that the intervention might lead to increases in the quantity of care provided within the household, using the set up provided by a unitary model. The nutritional supplements, health checkups and "pláticas" result in direct increases in the quantity of child care mothers provide and young children receive. Moreover, the conditional-on-attendance education grants result in a reallocation of time to "better paying" activities given the change in the relative shadow values of household members' time it entails. As a consequence, increases in maternal care in substitution for care previously provided by her older daughter were expected. These increases can arguably result in more and better care if the mother is assumed to be a better caregiver.

We have applied the Lewbel (2000) semiparametric method and provided evidence in support of such a mother-daughter substitution hypothesis. While mothers with 0 to 3 and 12 to 17 year old children are significantly less likely to participate in child care, this behavior is reversed given treatment. In addition, older daughters devote their freed up time to schooling. We also observe an increase in total household time given over to child care in these households. These findings suggest that –by linking benefits to school attendance– the *Oportunidades* program fosters human capital accumulation both through keeping teenage girls in school, and through more and "better" child care. Increased maternal care is likely to lead to better development in the early ages and increased school readiness. Note that it would have been unfeasible to increase the levels of child care provided by directly conditioning the reception of benefits to maternal time allocated to child care, as it is not possible to monitor how much time mothers spend with their children.

Interestingly, we have found no significant effect of treatment on child care provision amongst mothers with no children older than 12 or on the pooled sample of mothers. This suggests two things: first, the program mainly alters the household allocation of time devoted to child care through the reduction in the price of schooling (educational grant) and the resulting mother-daughter cross-substitution effect. Second, the educational talks and preventive care do not influence maternal child care provision as substantially as desired. Gertler and Fernald (2004) point at the inadequate development of the "pláticas" as an explanation of the inexistent program impacts on cognitive development. An alternative explanation could be that mothers in control communities also attend these talks, which would confound the estimated treatment effect. However, it seems unlikely that households in control communities know about them, let alone travel to treatment communities to attend. In any event, and given the important role of ECD in long run individual and societal welfare, CCT programs could re-consider introducing more intense parental and community training activities oriented to promote child stimulation and early education.

References

[1] Adato, Michelle, Bénédicte de la Brière, Dubravka Mindek and Agnes Quisumbing (2000), "The Impact of Progresa on Women's Status and Intrahousehold Relations", IFPRI Report, Washington D.C.

[2] Attanasio, Orazio, Costas Meghir and Ana Santiago (2005), "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate

Progresa", IFS Working Paper EWP10/14, London.

[3] Averett, Susan L., Elizabeth H. Peters and Donald M. Waldman (1997), "Tax Credits, Labor Supply, and Child Care", *Review of Economics and Statistics*, 79(1):125-135.

[4] Becker, Gary S. (1973), "A Theory of Marriage: Part I", *Journal of Political Economy,* 81(4): 813-846.

[5] Behrman, Jere R. and Petra E. Todd (1999), "Randomness in the experimental samples of PROGRESA (education, health, and nutrition program)", IFPRI Report, Washington D.C.

[6] Behrman, Jere R. and John Hoddinott (2005), "Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition", *Oxford Bulletin of Economics and Statistics*, 67(4): 547-569

[7] Blau, David M. and Philip K. Robins (1988), "Child-Care Costs and Family Labor Supply", *Review of Economics and Statistics*, 70(3):374-381.

[8] Browning Martin, Pierre A. Chiappori, Valérie Lechene (2010), "Distributional effects in Household Models: Separate Spheres and Income Pooling", *Economic Journal*, 120(6): 786-799.

[9] Carneiro, Pedro and James J. Heckman (2003), "Human Capital Policy", in *Inequality in America: What Role for Human Capital Policies*, James Heckman and Alan Krueger, eds. MIT Press. Cambridge, MA.

[10] Case, Anne, I-Fen Lin and Sara McLanahan (2001) "Educational Attainment in Blended Families”, *Evolution and Human Behavior*, 22(4): 269-289.

[11] Case, Anne and Christina Paxson (2001), “Mothers and Others: Who Invests in Children’s Health?”, *Journal of Health Economics*, 20(3):301-328.

[12] Chiappori, Pierre A. (1988), “Rational household labor supply”, *Econometrica*, 56(1): 63-89

[13] Cogneau, Denis and Eric Maurin (2001), “Parental income and school attendance in a low-income country: a semi-parametric analysis”, DIAL Working Paper, 2001/16.

[14] Currie, Janet (2001), “Early Childhood Education Programs”, *The Journal of Economic Perspectives*, 15(2):213-238.

[15] Currie, Janet and Duncan Thomas (2001), “Early Test Scores, Socioeconomic Status and Future Outcomes”, *Research in Labor Economics*, 20:103-132.

[16] Fernald, Lia C., Paul J. Gertler and Lynette M. Neufeld (2008) “Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico’s *Oportunidades*", *Lancet*, 371:828-37.

[17] Garces, Eliana, Duncan Thomas and Janet Currie (2002), “Longer-Term Effects of Head Start”, *The American Economic Review*, 92(4):999-1012.

[18] Gertler, Paul J. and Simone Boyce (2001), “An Experiment in Incentive-Based Welfare: The Impact of PROGRESA on Health in Mexico”, mimeo UC Berkeley.

[19] Gertler, Paul J. and Lia C. Fernald (2004), “The Medium Term Impact of *Oportunidades* on Child Development in Rural Areas”, mimeo UC Berkeley.

[20] Goux, Dominique and Eric Maurin (2005), “The effect of overcrowded housing on children’s performance at school”, *Journal of Public Economics*, 89(5-6):797-819.

[21] Grantham-McGregor, Sally, Yin Bu Cheung, Santiago Cueto, Paul Glewwe, Linda Richter, Barbara Strupp and the International Child Development Steering Group (2007), “Developmental Potential in the First 5 Years for Children in Developing Countries”, *Lancet*, 369:60-70.

[22] Heckman, James J. (1974) “Effects of Child-Care Programs on Women’s Work Effort”, *Journal of Political Economy*, 82(2):S136-S163.

[23] Hoddinott, John and Emmanuel Skoufias (2004), “The Impact of OPORTUNIDADES on Consumption”. *Economic Development and Cultural Change*, 53(1):37-61.

[24] INEGI (2002), *Uso del Tiempo y Aportaciones en los Hogares Mexicanos*, Instituto Nacional de Estadística (INEGI), Aguascalientes.

[25] Khan Shakeeb and Elie Tamer (2009), “Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities”, *Journal of Econometrics*, 152 (2):104-119.

[26] Khan Shakeeb and Elie Tamer (2010), “Irregular Identification, Support Conditions, and Inverse Weight Estimation”, forthcoming *Econometrica.*

[27] Lagarde, Mylene, Andy Haines and Natasha Palmer (2007), “Conditional Cash Transfers for Improving Uptake of Health Interventions in Low- and Middle-Income Countries: A Systematic Review”, *Journal of the American Medical Association*, 298(16):1900-1910.

[28] Lewbel, Arthur (2000), "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables", *Journal of Econometrics*, 97(1): 145-177.

[29] Lewbel, Arthur (2006), "Simple Endogenous Binary Choice and Selection Panel Model Estimators", Boston College Working Papers in Economics No. 613.

[30] Lundberg, Shelly and Robert A. Pollak (1993), "Separate Spheres Bargaining and the Marriage Market", *Journal of Political Economy*, 101(6):988-1010.

[31] Macours, Karen, Norbert Schady and Renos Vakis (2008), "Cash Transfers, Behavioral Changes, and the Cognitive Development of Young Children: Evidence from a Randomized Experiment", Policy Research Working Paper Series 4759, The World Bank.

[32] Magnac, Thierry and Eric Maurin (2007), "Identification and Information in Monotone Binary Models", *Journal of Econometrics*, 127(1):76-104.

[33] Michalopoulos, Charles, Philip K. Robins and Irwin Garfinkel (1992), "A Structural Model of Labor Supply and Child Care Demand", *Journal of Human Resources*, 27(1):166-203.

[34] Mueller, Eva (1984), "The value and allocation of time in rural Botswana", *Journal of Development Economics*, 15(1-3):329-360.

[35] Parker, Susan and Emmanuel Skoufias (2000), "The Impact of Progresa on Work, Leisure, and Time Allocation", Final Report, IFPRI, Washington, D.C.

[36] Paxson, Christina and Norbert Schady (2010), "Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador", *Economic Development and Cultural Change*, 59(1):187-229.

[37] Rivera, Juan A., Daniela Sotes-Alvarez, Jean-Pierre Ilabricht, Teresa Shamah and Salvador Villalpando (2004), "Impact of the Mexican Program for Education, Health, and Nutrition (Progresa) on Rates of Growth and Anemia in Infants and Young Children", *Journal of the American Medical Association*, 291(21):2563-2570.

[38] Rubio-Codina, Marta (2010), "Intra-household Time Allocation in Rural Mexico: Evidence from a Randomized Experiment", *Research in Labor Economics*, 31: 219-257

[39] Schultz, Paul (2004), "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program", *Journal of Development Economics*, 74(1): 199-250.

[40] Skoufias, Emmanuel (2005) "PROGRESA and Its Impacts on the Welfare of Rural Households in Mexico", IFPRI Report 139, Washington D.C.

[41] Skoufias, Emmanuel, Benjamin Davis and Sergio de la Vega (2001), "Targeting the Poor in Mexico: An Evaluation of the Selection of Households into PROGRESA", *World Development*, 29(10):1768-1784.

[42] Tiefenthaler, Jill (1997), "Fertility and Family Time Allocation in the Philippines", *Population and Development Review*, 23(2):377-397.

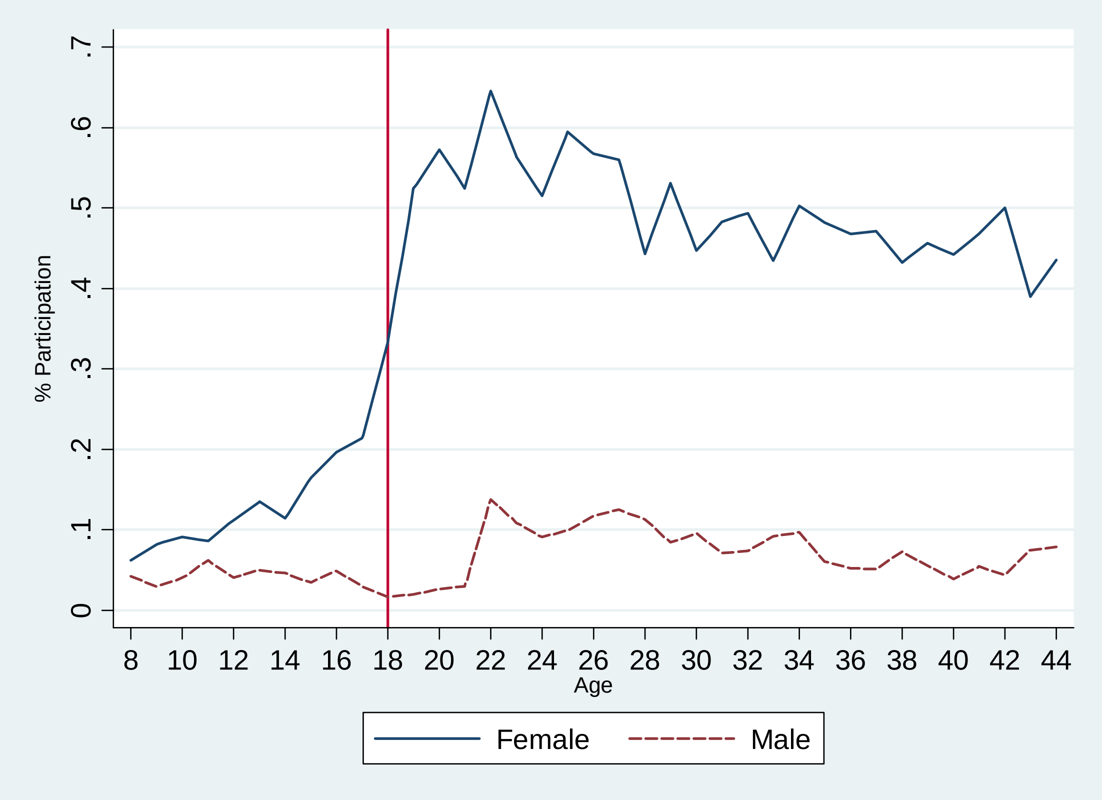
[43] UNICEF (2001), *The State of World's Children*: UNICEF, New York.

[44] Walker, Susan P., Theodore D. Wachs, Julie Meeks Gardner, Betsy Lozoff, Gail A. Wasserman, Ernesto Pollitt, Julie A. Carter and the International Child Development

Steering Group (2007), "Child Development: Risk Factors for Adverse Outcomes in Developing Countries", *Lancet*, 369:145-157.

FIGURES and TABLES

Figure A: Child Care Participation Rates by Sex and Age (May 1999)



Note: Households with children under 3 years old, no elder older than 65 not working and no sick individuals. Households deemed eligible for benefits according to the original classification scheme.

Table 1: Sample of Mothers with Kids Younger than 3 Years Old

	All	Treatment	Control
Mothers of Kids Younger than 3	4036	2571 (63.7%)	1465 (36%)
Mothers of Kids Younger than 3 & Teenagers 12 to 17	1509	973 (64.5%)	536 (35.5%)
Mothers of Kids Younger than 3 & Teenage Girls 12 to 17	979	636 (64.9%)	343 (35.1%)

Notes: Sample of mothers 18 to 44 living in households with no elder older than 65 not working and no sick individuals in May 1999. Households deemed eligible for benefits according to the original classification scheme.

Table 2: Comparison of Raw Means between Treatment and Control Groups

	Treatment Group			Control Group			
I. Dependent Variables (May 1999)	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**t-stat**
Participation in Child Care (All Mothers) =1	2571	50.80	0.500	1465	50.99	0.500	-0.075
Participation in Child Care (Mothers of 12 to 17 Teenagers) =1	973	47.28	0.500	536	42.35	0.495	1.421
Participation in Child Care (Mothers of 12 to 17 Teenage Girls) =1	636	49.37	0.500	343	44.90	0.498	1.107
Participation in Child Care (First Daughter) =1	626	14.22	0.350	337	21.07	0.408	-2.127*
Participation in School (First Daughter) =1	626	45.69	0.499	336	37.80	0.486	1.826+
Hours in Child Care (All Mothers)†	1306	3.76	2.862	747	3.74	2.854	0.088
Hours in Child Care (Mothers of 12 to 17 Teenagers)†	460	3.41	2.681	227	3.36	2.681	0.202
Hours in Child Care (Mothers of 12 to 17 Teenage Girls)†	314	3.32	2.703	154	3.39	2.631	-0.247
Hours in Child Care (First Daughter)†	89	2.31	2.247	71	2.06	1.497	0.846
Hours in School (First Daughter)†	286	6.15	1.586	127	6.00	1.471	0.969
Total Household (Hh) Hours in Child Care (All Hhs)†	1353	4.04	3.086	783	4.03	3.028	0.047
Total Hh Hours in Child Care (Hhs with 12 to 17 Teenagers)†	500	4.05	3.276	248	4.16	3.183	-0.420
Total Hh Hours in Child Care (Hhs with 12 to 17 Teenage Girls)†	345	4.03	3.435	170	4.38	3.241	-1.082
II. Covariates							
Current Maternal Characteristics (May 1999)							
Age	2571	29.66	6.418	1465	29.57	6.451	0.412
Years of Education	2534	3.55	2.720	1447	3.51	2.781	0.224
Hh Head =1	2537	1.89	0.136	1447	1.94	0.138	-0.094
Indigenous =1	2534	40.84	0.492	1445	43.04	0.495	-0.374
Working for a Wage =1	2534	8.72	0.282	1446	5.95	0.237	1.601
Current Household Characteristics (May 1999)							
Number of Kids 0 to 3	2571	1.25	0.465	1465	1.25	0.476	-0.382
Number of Kids 4 to 7	2571	1.05	0.798	1465	1.05	0.814	0.266
Number of Sons 8 to 11	2571	0.42	0.624	1465	0.38	0.605	1.500
Number of Sons 12 to 17	2571	0.37	0.697	1465	0.32	0.615	2.390*
Number of Daughters 8 to 11	2571	0.40	0.604	1465	0.40	0.614	-0.233
Number of Daughters 12 to 17	2571	0.33	0.644	1465	0.32	0.631	0.588
Years of Education Daughters 12 to 17	2571	1.38	2.589	1465	1.29	2.512	0.918
Baseline Household Characteristics (October1997)							
Number of Kids 0 to 3	2571	1.38	0.732	1465	1.41	0.740	-1.116
Number of Kids 4 to 7	2571	1.09	0.851	1465	1.07	0.847	0.571
Number of Teenagers 8 to 17	2571	0.75	1.046	1465	0.72	1.002	0.837
Number of Adults 18 to 54	2571	2.26	0.779	1465	2.26	0.827	-0.284
Number of Adults Over 55	2571	0.14	0.415	1465	0.13	0.381	0.657
Electricity =1	2568	61.06	0.488	1464	62.09	0.485	-0.219
Dirtfloor =1	2563	71.91	0.450	1460	76.23	0.426	-1.503
Animal and Land (more than 3 ha) Ownership =1	2561	29.48	0.456	1463	31.99	0.467	-0.854
Baseline Community Characteristics (October1997)							
Pre-school =1	2514	91.45	0.280	1417	91.11	0.285	0.123
Junior High School Imparted via TV =1	2571	21.78	0.413	1465	25.26	0.435	-0.608
Health Center =1	2571	76.66	0.423	1465	82.87	0.377	-1.568
Distance to Closest Secondary School (Km)	2571	2.45	2.171	1465	2.67	2.782	-0.562
Minimum Distance to Large Urban Centre (Km)	2571	105.46	44.046	1465	101.52	47.534	0.716
Monthly Community Agricultural Male Wage (pesos)	2506	6.27	0.313	1395	6.29	0.304	-0.572

Notes: +Significant at 10%; *Significant at 5%. Sample of mothers 18 to 44 living in households with no elder older than 65 not working and no sick individuals in May 1999. Households deemed eligible for benefits according to the original classification scheme.
†Conditional on being positive.

Table 3: Maternal Participation in Child Care and Maternal Extent of Participation in Child Care and Leisure

		OLS	PROBIT	LEWBEL	OLS	PROBIT	LEWBEL	OLS	PROBIT	LEWBEL
		Mod 1A	**Mod 1B**	**Mod 1C**	**Mod 2A**	**Mod 2B**	**Mod 2C**	**Mod 3A**	**Mod 3B**	**Mod 3C**
I. Child Care Participation =1	Special Regressor (v)	0.345	5.843	-	-5.314	-8.472	-	-2.533	-1.879	-
(Mean Dep Var = 0.51)		(8.853)	(22.617)	-	(9.951)	(25.386)	-	(9.190)	(23.580)	-
	Treatment (T) =1	0.014	0.020	0.007	-0.006	-0.033	-0.054	0.007	0.001	-0.003
		(0.030)	(0.077)	(0.005)	(0.030)	(0.078)	(0.053)	(0.030)	(0.077)	(0.008)
	Children 12 to 17 =1				-0.166**	-0.428**	-0.600*			
					(0.045)	(0.120)	(0.276)			
	T*Children 12 to 17 =1				0.095*	0.246*	0.435**			
					(0.037)	(0.096)	(0.128)			
	Daughters 12 to 17 =1							-0.101	-0.259	-0.097
								(0.071)	(0.182)	(0.085)
	T*Daughters 12 to 17 =1							0.062	0.164	0.257*
								(0.043)	(0.109)	(0.111)
II. Extent of Child Care Hours	Special Regressor (v)	6.700	11.828	-	-21.370	-6.995	-	1.093	6.041	-
(Mean Dep Var = 1.94)		(32.380)	(22.201)	-	(36.543)	(25.143)	-	(34.413)	(23.769)	-
	Treatment (T) =1	0.037	-0.000	-0.001	-0.001	-0.030	-0.059	0.030	-0.008	-0.007
		(0.107)	(0.071)	(0.002)	(0.107)	(0.070)	(0.052)	(0.109)	(0.071)	(0.007)
	Children 12 to 17 =1				-0.344*	-0.287**	-0.369*			
					(0.144)	(0.105)	(0.149)			
	T*Children 12 to 17 =1				0.297*	0.217*	0.306**			
					(0.131)	(0.090)	(0.093)			
	Daughters 12 to 17 =1							-0.180	-0.145	-0.065
								(0.232)	(0.160)	(0.056)
	T*Daughters 12 to 17 =1							0.094	0.098	0.142*
								(0.138)	(0.096)	(0.061)
III. Extent of Leisure	Special Regressor (v)	13.184	-2.596	-	-21.415	-23.310	-	3.206	-7.842	-
(Mean Dep Var = 7.54)		(30.137)	(19.251)	-	(35.890)	(22.717)	-	(32.887)	(20.842)	-
	Treatment (T) =1	0.001	-0.035	-0.004	0.031	-0.016	0.056	0.014	-0.028	0.004
		(0.104)	(0.064)	(0.003)	(0.105)	(0.064)	(0.052)	(0.105)	(0.064)	(0.007)
	Children 12 to 17 =1				0.250+	0.164+	0.493*			
					(0.139)	(0.086)	(0.218)			
	T*Children 12 to 17 =1				-0.316*	-0.191*	-0.457**			
					(0.126)	(0.079)	(0.118)			
	Daughters 12 to 17 =1							0.059	0.041	0.100
								(0.217)	(0.135)	(0.067)
	T*Daughters 12 to 17 =1							-0.162	-0.086	-0.233**
								(0.141)	(0.088)	(0.078)

Notes: (N = 3710). +significant at 10%, *significant at 5%, **significant at 1%. SE in parentheses and clustered at the community level. Observations in the top 7% of Fv and with f(v|x) > 1*10-5 have been trimmed. All regressions include the following covariates: maternal age, age squared, years of education, ethnicity, head of the household status and whether she is a paid worker; the first daughter's years of education; baseline and contemporary household demographic composition; baseline household assets (dirt floor, electricity and farm size); male agricultural wage in the community at baseline, distance to large urban center, distance to secondary school, presence of pre-school and presence of junior high school imparted via TV, or "telesecundaria" in the community.

Table 4: First Daughters (Ages 12 to 17) Time Use - Child Care, Schooling and Leisure

		OLS	PROBIT	LEWBEL
		Model A	**Model B**	**Model C**
I. Child Care Participation =1	Special Regressor (v)	-6.994	-40.453	-
(Mean Dep Var = 0.18)		(12.140)	(53.705)	-
	Treatment (T) =1	-0.085	-0.379+	-0.032**
		(0.052)	(0.220)	(0.011)
II. Extent of Child Care Hours	Special Regressor (v)	-4.755	-38.338	-
(Mean Dep Var = 0.40)		(26.676)	(53.429)	-
	Treatment (T) =1	-0.132	-0.356+	-0.032**
		(0.112)	(0.215)	(0.011)
III. School Participation =1	Special Regressor (v)	35.239+	100.092*	-
(Mean Dep Var = 0.42)		(18.103)	(50.348)	-
	Treatment (T) =1	-0.066	-0.185	0.008*
		(0.077)	(0.214)	(0.004)
IV. Extent of Hours in School	Special Regressor (v)	87.903+	86.352*	-
(Mean Dep Var = 2.51)		(46.025)	(43.990)	-
	Treatment (T) =1	-0.137	-0.146	0.002
		(0.194)	(0.188)	(0.002)
V. Extent of Leisure	Special Regressor (v)	17.391	10.248	-
(Mean Dep Var = 9.04)		(62.764)	(37.904)	-
	Treatment (T) =1	-0.180	-0.116	-0.013*
		(0.239)	(0.143)	(0.006)

Notes: (N = 758) +significant at 10%, *significant at 5%, **significant at 1%. SE in parentheses and clustered at the community level. Observations in the top 7% of Fv and with f(v|x) > 1*10-4 have been trimmed. All regressions include covariates (see Table 3).

Table 5: Total Household Hours in Child Care

	OLS	PROBIT	LEWBEL	OLS	PROBIT	LEWBEL	OLS	PROBIT	LEWBEL
Extent Household Hours in Child Care	**Model 1A**	**Model 1B**	**Model 1C**	**Model 2A**	**Model 2B**	**Model 2C**	**Model 3A**	**Model 3B**	**Model 3C**
Special Regressor (v)	16.364	17.024	-	-19.451	-3.856	-	13.278	11.820	-
	(39.153)	(21.341)	-	(44.630)	(24.416)	-	(41.990)	(23.157)	-
Treatment (T) =1	-0.003	-0.029	-0.001	-0.039	-0.056	-0.059	-0.004	-0.035	-0.007
	(0.129)	(0.070)	(0.002)	(0.129)	(0.069)	(0.052)	(0.130)	(0.070)	(0.007)
Children 12 to 17 =1				-0.290	-0.243*	-0.342*			
				(0.189)	(0.110)	(0.138)			
T*Children 12 to 17 =1				0.340*	0.220*	0.299**			
				(0.165)	(0.090)	(0.091)			
Daughter 12 to 17 =1							-0.084	-0.107	-0.061
							(0.297)	(0.161)	(0.054)
T*Daughter 12 to 17 =1							0.041	0.080	0.148*
							(0.177)	(0.097)	(0.062)

Notes: (N=3710) Mean Dep. Var. = 2.16 hours. *significant at 5%, **significant at 1%. SE in parentheses and clustered at the community level. Observations in the top 7% of Fv and with f(v|x) > 1*10-5 have been trimmed. All regressions include covariates (see Table 3).

Table 6: Share of Child Care Hours Provided by Different Household Members

	OLS												
	Share Mums	**Share Mums**	**Share Other Women**	**Share Other Women**	**Share Men**	**Share Men**	**Share Sons 8 to 11**	**Share Sons 8 to 11**	**Share Sons 12 to 17**	**Share Sons 12 to 17**	**Share Daughters 8 to 11**	**Share Daughters 8 to 11**	**Share Daughters 12 to 17**
Treatment (T) =1	0.025	0.009	0.012	-0.013	-0.006	-0.008	-0.002	-0.007	0.008	0.005	-0.010	-0.026	-0.031*
	(0.021)	(0.024)	(0.024)	(0.033)	(0.006)	(0.008)	(0.005)	(0.008)	(0.007)	(0.011)	(0.011)	(0.016)	(0.015)
Girls 12 to 17 in the Household =1		-0.072*		-0.051		0.002		-0.011		0.006		-0.048**	
		(0.033)		(0.046)		(0.010)		(0.011)		(0.015)		(0.018)	
T*Girls 12 to 17 in the Household =1		0.051		0.058		0.006		0.011		0.005		0.041*	
		(0.032)		(0.045)		(0.009)		(0.011)		(0.015)		(0.018)	
Observations	4036	4036	476	476	4100	4100	1382	1382	1047	1047	1381	1381	988
Mean Share Child Care Hours	0.44	0.44	0.10	0.10	0.03	0.03	0.02	0.02	0.02	0.02	0.04	0.04	0.08

Notes: *significant at 5%, **significant at 1%. SE in parentheses clustered at the community level. All regressions include covariates (see Table 3). Share Household Member m =Total Hours to Child Care of Household Member m / Total Household Hours to Child Care.

1 Appendix for Referee's Inspection Only

2 Testing Normality Using a Conditional Moment Test

Conditional Moment (CM) tests are based on *moment functions* of the form:

$$m(z, \theta) = w(x, \theta) r(z, \theta)$$

where r is a restriction function, z are the exogenous variables, x is a subvector of z, and $w(x, \theta)$ is a weighting function. If the model is correctly specified, the conditional score vector with respect to the vector of parameters θ (the gradient of the likelihood) provides a vector of functions that satisfy the CM restrictions implied by the economic model $E[r(z, \theta_o|x)] = 0$. Newey (1985) derives a CM specification test where the test-statistic is the r-squared resulting from regressing a vector of ones on the moment restrictions and on the scores of the likelihood function. This statistic is distributed as a $\chi^2(r)$ with r restrictions. Newey (1985) provides an application of this test in the binary probit setting. Combined with Rudd's (1984) parametrization of nonnormality, the moment function writes:

$$m(z, \theta) = \lambda(x, \theta)[y - \Phi(x\theta)][(x\theta)^2, (x\theta)^3]'$$

where $\lambda(.) = \frac{\phi(.)}{\Phi(.)(1-\Phi(.))}$. Rudd (1984) assumes that conditional on x, θ and $\gamma = (\gamma_0, \gamma_1, \gamma_2)$, $\varepsilon = y^* - x\theta$ has distribution function $\Phi = (\gamma_0 + \eta + \gamma_1 \eta^2, \gamma_2 \eta^3)$ where γ_1 and γ_2 satisfy $1 + 2\gamma_1 \eta, 3\gamma_2 \eta^2$ for all η (RESET-type tests). This CM test is like a Lagrange Multiplier test for the hypothesis $\gamma_1 = \gamma_2 = 0$.

Pagan and Vella (1989) propose a "*regression analogue*" procedure to compute the

Newey (1985) statistic and its variance. The authors prove that if: (i) the x's are independently distributed random variables, and (ii) the sample first order conditions evaluated at the Maximum Likelihood estimator are zero; then, the variance of the Newey's (1985) statistic is the estimated covariance matrix of the errors in each of the equations with $m(z, \theta)$ as a dependent variable and unity and $r(z, \theta)$ as regressors. As such, the nonnormality CM test above can be computed by regressing the CM restriction on a constant and the scores of the likelihood. If the coefficient on the constant is significantly different from zero, then the null hypothesis (normality) is rejected.

We implement Pagan and Vella's (1989) "*regression analogue*" to test whether the errors in the mothers' and the first daughters' child care participation and hours equations satisfy the normality assumption required by the probit and tobit models. Table A1 presents results. RESET-type tests reject normality for the parametrization of nonnormality as the cube of the residuals for mothers' participation (at the 10 percent significance level) and child care hours (at the 5 percent significance level). Similarly, when tested jointly, the RESET-tests almost reject normality at the 10 percent for participation and at the 5 percent for hours. When parametrized as the square of the residuals, normality is also rejected (at the 10 percent significance level) for the error distribution of the first daughters' schooling hours equation.

Newey (1985) notes that CM tests may fail to reject the null due to their low power against some form of mispecification, even if such mispecification results in inconsistent estimates. Moreover, Monte Carlo simulations have shown that RESET-type tests have low power compared with alternative tests based in the comparison of parametric versus semiparametric estimates or choice probabilities. Given this, we have implemented some additional normality tests known to perform well in relatively small samples – namely a joint test of skewness and kurtosis and the Shapiro-Wilk and Shapiro-Francia tests. These tests reject normality in both cross-sections. Finally, plots of the non-parametric residuals for the tobit and probit specifications of the main equations, and plots of the residual quantiles on the normal distribution quantiles also reject normality. Note that these tests are simultaneously testing the assumption that the error term has variance one, as required for identification of probit and tobit models.

3 Monotonicity of the Special Regressor

Figure A1 shows conditional (on a certain set of X_i's) kernel estimates and its pointwise confidence intervals of the following kernel regressions: (i) maternal participation in child care, (ii) first daughters' participation in child care, and (iii) first daughters' participation in school; on the household stock of potential transfers (the special regressor v_i). We have followed Algorithm 4.2.1 in Härdle (1989) to compute the confidence bands. Recall that in the current setting, the monotonicity assumption implies that: (a) mothers' participation in child care activities and first daughters' school participation are nondecreasing in v_i, and (b) first daughters' participation in child care is nonincreasing in v_i. None of the Kernel estimates rejects monotonicity.

4 Testing the Symmetry Condition

Tables A2 and A3 show the sensitivity of the Lewbel methodology to trimming outliers of the distribution of the special regressor, F_{v_i}; and of its estimated conditional density at different points, $\hat{f}(v_i|x_i)$. By construction, extremely low observations of $\hat{f}(v_i|x_i)$ imply outlier observations of the transformed dependent variable y_i^*. For both tables and for each dependent variable, each column presents estimation results on a different subsample, as indicated. For each block of results, the first column shows results on all observations. In the second column, observations with $\hat{f}(v_i|x_i) \leq 10^{-5}$ are trimmed. Note that dropping these observations reduces standard errors substantially. In the third column, we additionally trim the top 2 % of F_{v_i}. Finally, in the fourth and fifth columns we trim the top 5 and 10 % of F_{v_i}. We included estimates resulting from trimming F_{v_i} at an intermediate level (7 %) in the main tables of results (Tables 3 and 4). As estimates in these tables show, this two-step trimming process is crucial to improve the precision of the semiparametric estimates.

5 Implementation of the Lewbel (2000) Estimator

Following Lewbel (2006) and for ease in the implementation of the estimator, we parametrize the density of v_i as $v_i = x_i'\gamma + \sigma\eta_i$. Now, the unobserved error term η_i is assumed to have a normal density function, f_η, with mean zero and variance one. This implies that identification is also based on η_i being orthogonal to ε_i and x_i. Given the parameters $\lambda = (\beta, \gamma, \sigma)$, the conditional density function becomes: $f(v_i|x_i, \lambda) = \frac{1}{\sigma} f_\eta \left(\frac{v_i - x_i'\gamma}{\sigma} \right)$.

For each cross-section, we apply the following steps to obtain consistent parameter estimates:

1a. Construct v_i, this is to say, the demeaned potential transfer amount accumulated over time by the household.

1b. Replace v_i with $-v_i$ for the following dependent variables: first daughter's child care participation and hours, and mothers's and first daughter's leisure hours; such that y_i is nondecreasing in v_i conditional on x_i, for all y_i.

2. Linearly regress v_i on x_i across observations i in the estimation subsample (as defined above) and obtain $\hat{\gamma}$. As noted, we trim the top 7 % of the distribution of the special regressor v_i (the stock of potential transfers accumulated at t).

3. Compute $\widehat{\sigma}^2$ as the sample average of $(v_i - x_i'\hat{\gamma})^2$.

4. Construct $\hat{f}(v_i|x_i, \hat{\lambda}) = \frac{1}{\sigma} f_\eta \left(\frac{v_i - x_i'\gamma}{\sigma} \right)$.

5. Trim very small values of the conditional density distribution $\hat{f}(v_i|x_i, \hat{\lambda})$, as they result in extreme observations of the transformed dependent variable y_i^*. As shown earlier, this reduces standard errors substantially.

6. Construct $y_i^* = y^*(v_i, x_i, \hat{\gamma}, \hat{\sigma}) = \frac{y_i - I_{\{v_i > 0\}}}{\hat{f}(v_i|x_i, \hat{\lambda})}$ for dichotomous outcomes and $y^* = y^*(v_i, x_i, \hat{\gamma}, \hat{\sigma}) = \frac{\frac{y_i}{K-1} - I_{\{v_i > 0\}}}{\hat{f}(v_i|x_i, \hat{\lambda})}$ for polychotomous outcomes.

7. Linearly regress y_i^* on x_i to obtain $\hat{\beta}$.

8. Bootstrap the previous linear regression to obtain parameter standard errors and confidence intervals.

9. Compute the marginal effects on the estimated coefficients for dichotomous outcomes as follows: $M_{ij} = \frac{\triangle[1 - \hat{G}(-v_i - x_i\hat{\beta})]}{\triangle x_j} = \frac{\triangle[1 - \hat{G}(z_i\hat{\beta})]}{\triangle x_j} = \hat{G}(\hat{z}_i^0) - \hat{G}(\hat{z}_i^1)$, where M_{ij} is the effect of switching the j^{th} binary variable, x_j, from 0 to 1 on the probability that y_i equals one; and $\hat{z}_i^1$ is the value of the index $-v_i - x_i\hat{\beta}$ when the j-th binary variable is set to 1, and similarly for $\hat{z}_i^0$ when the value of j is set to zero. $\hat{G}(.)$ is the estimated cumulative distribution function of the probability of y_i given z_i. Nonparametrically estimate $\hat{G}(.)$ running the kernel regression of y_i on z_i. Results reported in the text were computed using a Gaussian Kernel and 500 equally spaced points in the range of z_i, and are robust to larger number of points. $\hat{G}(\hat{z}_i^0)$

and $\hat{G}(\hat{z}_i^1)$ are predicted by taking the average of the $\hat{G}(\hat{z}_{i-1})$ and $\hat{G}(\hat{z}_{i+1})$, for all $\hat{z}_{i-1} > \hat{z}_i^0 \geqslant \hat{z}_{i+1}$ or $\hat{z}_{i-1} > \hat{z}_i^1 \geqslant \hat{z}_{i+1}$. Lastly, the observation-by-observation mean marginal effects are computed by taking the mean over the sample of marginal effects: $\bar{M}_j = \frac{1}{N}\sum_{i=1}^{I} M_{ij}$.

References

[1] Härdle, Wolfgan (1989) *Applied Nonparametric Regression*, Econometric Society Monographs. Cambridge University Press.

[2] Newey, Whitney K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica*, 53(5): 1047-1070.

[3] Pagan, Adrian and Frank Vella (1989), "Diagnostic Tests for Models Based on Individual Data: A Survey", *Journal of Applied Econometrics*, 4(Dec): S29-S59.

[4] Ruud, Paul A. (1984), "Tests of Specification in Econometrics", *Econometric Reviews*, 3(2): 211-242.

Referee's Appendix: Figures and Tables

Figure A1: Monotonicity Tests

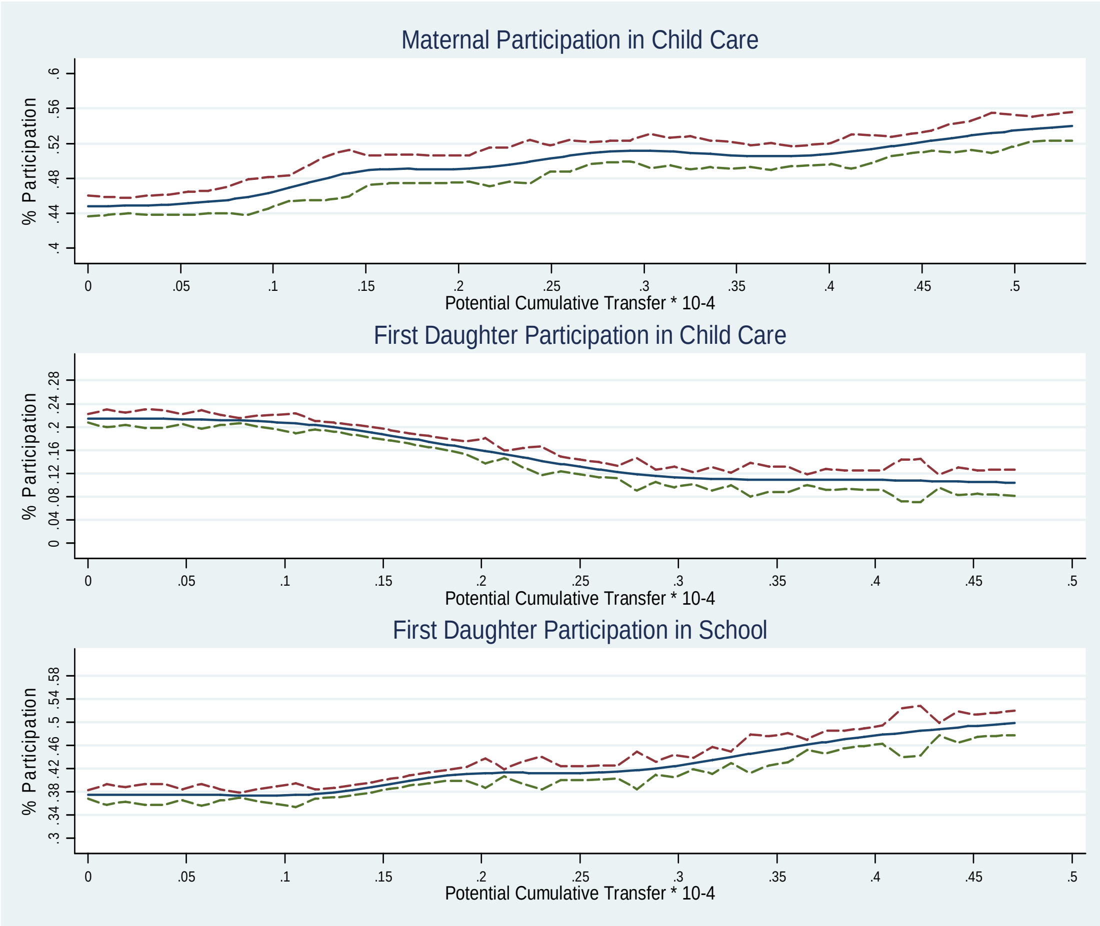


Table A1: "Regression Analogue" Test for Normality (Pagan and Vella, 1989)

	Model	Moment Restriction	T-Stat	Joint Test: Chi-Squared (2)
Mother's Participation in Child Care	Probit	E(Pred^2*GR)	-1.18	4.56
		E(Pred^3*GR)	-1.70	(0.1025)
Mother's Hours in Child Care	Tobit	E(Pred^2*GR)	-0.44	5.71
		E(Pred^3*GR)	-2.20	(0.0575)
Mothers' Leisure Hours	Tobit	E(Pred^2*GR)	-1.39	4.29
		E(Pred^3*GR)	-1.31	(0.1171)
First Daughter's Participation in Child Care	Probit	E(Pred^2*GR)	-0.54	0.29
		E(Pred^3*GR)	0.51	(0.8651)
First Daughter's Hours in Child Care	Tobit	E(Pred^2*GR)	-0.04	0.01
		E(Pred^3*GR)	0.01	(0.9948)
First Daughter's Participation in School	Probit	E(Pred^2*GR)	-0.76	0.58
		E(Pred^3*GR)	0.54	(0.7501)
First Daughter's Hours in School	Tobit	E(Pred^2*GR)	-1.82	3.33
		E(Pred^3*GR)	1.07	(0.1893)
First Daughter's Leisure Hours	Tobit	E(Pred^2*GR)	-0.01	0.57
		E(Pred^3*GR)	-0.03	(0.7530)

Notes: Pred are liniar predictions and GR are generalized residuals.

Table A2: Analysis of the Sensitivity of Results to Trimming Outliers of Fv and f(v|x) Maternal Participation in Child Care and Extent of Participation in Child Care and Leisure

		All	All & $f(v\|x) > 1*10^{-5}$	98% Fv & $f(v\|x) > 1*10^{-5}$	95% Fv & $f(v\|x) > 1*10^{-5}$	90% Fv & $f(v\|x) > 1*10^{-5}$
		Mod 2B - 1	**Mod 2B - 2**	**Mod 2B - 3**	**Mod 2B - 4**	**Mod 2B - 5**
I. Child Care Participation =1	Treatment (T) =1	-0.100	-0.029	-0.049	-0.027	-0.005
		(0.346)	(0.050)	(0.030)	(0.028)	(0.012)
	Children 12 to 17 =1	-19.347	-0.430+	-0.475*	-0.464*	-0.392*
		(17.068)	(0.251)	(0.202)	(0.223)	(0.187)
	T*Children 12 to 17 =1	4.141	0.361**	0.410**	0.381**	0.384**
		(2.882)	(0.138)	(0.150)	(0.139)	(0.143)
	Observations	4036	4031	3910	3789	3593
		Mod 3B - 1	**Mod 3B - 2**	**Mod 3B - 3**	**Mod 3B - 4**	**Mod 3B - 5**
	Treatment (T) =1	-0.017	-0.002	-0.007	0.002	0.011
		(0.196)	(0.018)	(0.008)	(0.009)	(0.011)
	Daughter 12 to 17 =1	0.919	-0.131	-0.075	-0.111	-0.089
		(3.471)	(0.149)	(0.089)	(0.108)	(0.118)
	T*Daughter 12 to 17 =1	3.633	0.272*	0.197*	0.281*	0.442*
		(2.672)	(0.129)	(0.091)	(0.138)	(0.202)
	Observations	4036	4033	3911	3790	3595
II. Extent of Child Care Hours		**Mod 2B - 1**	**Mod 2B - 2**	**Mod 2B - 3**	**Mod 2B - 4**	**Mod 2B - 5**
	Treatment (T) =1	-0.168	-0.057	-0.089*	-0.046	-0.012
		(0.167)	(0.050)	(0.040)	(0.029)	(0.010)
	Children 12 to 17 =1	-7.471	-0.323	-0.318*	-0.337*	-0.283*
		(6.783)	(0.236)	(0.136)	(0.164)	(0.117)
	T*Children 12 to 17 =1	1.647	0.300*	0.296**	0.301**	0.262**
		(1.094)	(0.122)	(0.081)	(0.097)	(0.090)
	Observations	4036	4031	3910	3789	3593
		Mod 3B - 1	**Mod 3B - 2**	**Mod 3B - 3**	**Mod 3B - 4**	**Mod 3B - 5**
	Treatment (T) =1	-0.048	-0.027	-0.030*	-0.007	0.000
		(0.077)	(0.023)	(0.014)	(0.008)	(0.006)
	Daughter 12 to 17 =1	0.653	-0.125	-0.074	-0.090	-0.066
		(1.253)	(0.124)	(0.076)	(0.084)	(0.065)
	T*Daughter 12 to 17 =1	1.349	0.192*	0.140*	0.159*	0.232*
		(1.027)	(0.097)	(0.061)	(0.077)	(0.097)
	Observations	4036	4033	3911	3790	3595
III. Extent of Leisure		**Mod 2B - 1**	**Mod 2B - 2**	**Mod 2B - 3**	**Mod 2B - 4**	**Mod 2B - 5**
	Treatment (T) =1	0.226	0.029	0.066*	0.032	0.008
		(0.305)	(0.037)	(0.034)	(0.023)	(0.009)
	Children 12 to 17 =1	15.467	0.360+	0.473**	0.409*	0.328*
		(14.142)	(0.206)	(0.172)	(0.171)	(0.130)
	T*Children 12 to 17 =1	-3.143	-0.333**	-0.394**	-0.416**	-0.351**
		(2.267)	(0.114)	(0.103)	(0.111)	(0.102)
	Observations	4036	4031	3910	3789	3593
		Mod 3B - 1	**Mod 3B - 2**	**Mod 3B - 3**	**Mod 3B - 4**	**Mod 3B - 5**
	Treatment (T) =1	0.087	0.022	0.021+	0.004	-0.007
		(0.152)	(0.024)	(0.011)	(0.009)	(0.008)
	Daughter 12 to 17 =1	-1.573	0.138	0.118	0.140	0.087
		(2.579)	(0.127)	(0.095)	(0.100)	(0.087)
	T*Daughter 12 to 17 =1	-2.749	-0.246*	-0.249*	-0.281**	-0.386**
		(2.125)	(0.105)	(0.100)	(0.107)	(0.148)
	Observations	4036	4033	3911	3790	3595

Notes: +significant at 10%, *significant at 5%, **significant at 1%. SE in parentheses. Outliers have been trimmed at different points of the accumulated potential transfer distribution (the special regressor v) for different specifications (columns) as indicated. Fv is the empirical distribution of v; f(v|x) denotes the conditional probability density function of v given x as estimated from the data (predicted). Model 2B and Model 3B labels relate to the labels in Table 3. All regressions include covariates (see Table 3).

Table A3: Analysis of the Sensitivity of Results to Trimming Outliers of Fv and f(v|x) First Daughters (Ages 12 to 17) Time Use - Child Care, Schooling and Leisure

		All	All & $f(v\|x) > 1*10^{-4}$	98% Fv & $f(v\|x) > 1*10^{-4}$	95% Fv & $f(v\|x) > 1*10^{-4}$	90% Fv & $f(v\|x) > 1*10^{-4}$
I. Child Care Participation =1		**Mod B - 1**	**Mod B - 2**	**Mod B - 3**	**Mod B - 4**	**Mod B - 5**
	Treatment (T) =1	-0.028*	-0.028*	-0.026**	-0.031**	-0.030**
		(0.011)	(0.011)	(0.009)	(0.011)	(0.009)
	Obervations	963	963	872	807	704
II. Extent of Child Care Hours	Treatment (T) =1	0.011	0.011	0.007	0.008+	0.009*
		(0.008)	(0.008)	(0.005)	(0.004)	(0.004)
	Obervations	962	962	871	807	702
III. School Participation =1	Treatment (T) =1	-0.029**	-0.029**	-0.026**	-0.031**	-0.030**
		(0.011)	(0.011)	(0.009)	(0.011)	(0.009)
	Obervations	963	963	872	807	704
IV. Extent of Hours in School	Treatment (T) =1	0.002	0.002	0.002	0.002	0.004
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	Obervations	961	961	870	807	702
V. Extent of Leisure	Treatment (T) =1	-0.010+	-0.010+	-0.009*	-0.012*	-0.013*
		(0.005)	(0.005)	(0.004)	(0.005)	(0.006)
	Obervations	973	973	880	815	710

Notes: +significant at 10%, *significant at 5%, **significant at 1%. SE in parentheses. Outliers have been trimmed at different points of the accumulated potential transfer distribution (the special regressor v) for different specifications (columns) -as indicated. Fv is the empirical distribution of v; f(v|x) denotes the conditional probability density function of v given x as estimated from the data (predicted). Model B label relates to the labels in Table 4. All regressions include covariates (see Table 3).