GRAPE Group for Research in APplied Economics

GRAPE Working Paper # 53

Evaluating the 500+ child support program in Poland

Filip Premik

FAME | GRAPE, 2021

Foundation of Admirers and Mavens of Economics
Group for Research in Applied Economics

Evaluating the 500+ child support program in Poland

Filip Premik
FAME | GRAPE
University of Minnesota

Abstract

We investigate immediate effects of a large scale child benefit program introduction on labor supply of the household members in Poland. Due to nonrandom eligibility and universal character of the program standard evaluation estimators are likely to be inconsistent. In order to address this issues we propose a novel approach which combines difference-in-difference (DID) propensity score based methods with covariate balancing propensity score (CBPS) by Imai and Ratkovic (2014). The DID part solves potential problems with non-parallel outcome dynamics in treated and non-treated subpopulations resulting from non-experimental character of the data, whereas CBPS is expected to reduce significantly bias from the systematic differences between treated and untreated subpopulations. We account also for potential heterogeneity among households by estimating a range of local average treatment effects which jointly provide a reliable view on the overall impact. We found that the program has a minor impact on the labor supply in periods following its introduction. There is an evidence for a small encouraging effect on hours worked by treated mothers of children at school age, both sole and married. Additionally, the program may influence the intra-household division of duties among parents of the youngest children as suggested by simultaneous slight decline in participating mothers' probability of working and a small increase in treated fathers' hours worked..

Keywords:

child benefits, labor supply, program evaluation, difference-in-difference estimation, covariate balancing propensity score

JEL Classification

C21, C23, I38, J22

Corresponding author

Filip Premik (filip.premik@gmal.com)

Acknowledgements

The author thanks Marcin Bielecki, Michał Gradzewicz, Marcin Kolasa, Iga Magda, Kamila Sławińska, Ewa Stanisławska, Karol Szafranek, Joanna Tyrowicz, Małgorzata Walerych, Jakub Witkowski, Jan Werner, Dan Yu for long discussions and inspirations. Valuable comments were given by the participants of NBP Summer Workshop 2017, Nordic Econometric Meeting, WSE Panel Data conference. Special thanks go to the participants of STATA Users' Group Meeting for technical insights and comments on development of CBPS package. The financial support of National Science Centre (grant # 2020/37/N/HS4/03637) is gratefully acknowledged.

Published by: FAME | GRAPE
ISSN: 2544-2473

Foundation of Admirers and Mavens of Economics
ul. Koszykowa 59/7
00-660 Warsaw
Poland

W | grape.org.pl
E | grape@grape.org.pl
TT | GRAPE_ORG
FB | GRAPE.ORG
PH | +48 799 012 202

1 Introduction

The effects of unconditional non-equivalent income support instruments on labor supply are at the core of labor economics. The interest in child-care programs and labor supply has been growing since the seminal paper of Heckman (1974). Autonomous income in a form of subsidy pushes the budget constraint upwards, expanding the feasible set of choices between consumption and leisure. Depending on the properties of the utility function, one can expect both a contraction in hours worked and unchanged labor supply with an increased consumption. Income support instruments are frequently under scrutiny for the possible effect of discouraging employment: a utility flow generated by the social transfer might exceed the utility from working even if the earned income was higher than the transfer, because work induces some nonzero disutility (Besley and Coate, 1992; Spencer, 2003). The direction of the effect may also be ambiguous because of nonlinearities in a budget set which appear due to a relationship between wage rate and hours worked (Burtless and Hausman, 1978), a stigma effect (Moffitt, 1983; Hoynes et al., 1996) or tied wage–hours contracts (Averett and Hotchkiss, 1997). Child benefit program is also supposed to limit barriers in second earner access to the labor market constituted by costs of care (Kimmel, 1998) and affect labor supply through this channel.

A analysis of labor supply response to the child benefit program requires some concern about the child-care provision mechanisms (Blau and Robins, 1988; Connelly, 1992; Ribar, 1992, 1995). This proves to be a key issue in the evaluation of non-equivalent child care benefit programs especially when the eligibility relies mainly on the number of children in the family. This literature focuses attention on the distinction between first and second earner. The latter is mainly the one that is responsible for providing the most of non-market child care. The former in turns brings the most of labor income, but at the same time may still provide some child-care. This intra-household allocation may be the case even among the single parent households, as they might have an access to the informal child care at low (or zero cost) provided by other children, friends, relatives living in the household, as long as they contribute to household home production.

Given these theoretical premises, the actual effects of income support program on the labor supply remain an empirical question even for the monetary non-equivalent transfers. The biggest challenge lies in the causal identification in the data: individuals adjust labor supply and consumption for a variety of reasons. Assigning that certain income support instrument has been the cause and labor supply adjustment has been the effect is at the heart of much of modern microeconometrics and labor economics. Empirical investigations of the effects of universal transfers on labor supply are difficult due to lack of an appropriate counter-factual. If eligibility is based on the number of children, ineligible households (i.e. those who do not receive the transfer) are not a good control group – there may be other, unobservable differences between the eligible and ineligible households (Graham and Beller, 1989).

Due to the recent strong development, program evaluation methods are getting more and more attention in studies of the labor supply. Blundell et al. (2004) perform difference-in-difference propensity score matching investigating the effects of a mandatory job search program. Baker et al. (2008) describe the effects of highly subsidized, universally accessible child care introduction in Quebec, which led to a significant rise in maternal labor supply. Luna (2011) investigates the introduction of an unconditional child benefit in Spain on family well-being using regression discontinuity design. The study suggests that mothers eligible for the program stayed longer out of the labor market after giving birth. Koebel and Schirle (2016) in turn exploits difference-in-difference strategy to measure the effects of Canadian Universal Child Care Benefit. They find out that the child care benefit program has a negative effect on legally married woman and increases labor supply of single mothers.

In this paper we contribute to the program evaluation literature by proposing a novel approach to estimate the effects of child support instrument on household labor supply in a situation, when the instrument design invalidate other estimation methods. Our estimator combines an idea of covariate balancing propensity score (CBPS, Imai and Ratkovic (2014)) with difference-in-difference (DID) estimators by Abadie (2005) and Heckman et al. (1997). The DID estimators solve potential problems with non-parallel outcome dynamics in treated and non-treated subpopulations, which might result from nonexperimental character of the data. CBPS is expected to reduce significantly bias resulting from systematic differences between treated and untreated subpopulations.

The new estimators are applied to investigate the immediate effects of introduction of a large scale child benefit program in Poland on labor supply of the household members. The eligibility to the program is not random because it is driven by the number of children within a household. Additionally, the universal character of the program makes it difficult to define appropriate control groups, as among eligible there are various types of households on different steps of the life cycle. The study indicates that the introduction of a large scale child benefit program has a minor impact on women's labor supply and almost no effect on men's labor supply. All of these effects are immediate, i.e. they concern the period directly after the introduction of the program. We found a small encouraging impact on hours worked by treated mothers of children at school age, both sole and married. Additionally, there is a weak evidence for a limited decrease in probability of employment among married mothers of the youngest children (0-5) that participate in the program simultaneous with a slight increase in men's hours worked in the same child age category. It would suggest a program-induced change in the division of household duties between men and women. Although this effect seems barely significant, we expect it to strengthen in the subsequent quarters after introduction of the program. The labor supply for the rest of parents remains unaffected. One of the possible explanations for this result is that the program has been introduced quite recently which is why households might not have had sufficient time to fully adjust their behavior. Another key issue is whether the households perceived the transfers as a permanent or transitory change in family's income.

The paper is organized as follows. In section 2 we discuss some simple theoretical models of labor supply and explain why their predictions for the movement in labor supply as a response to an increase in non-labor income are ex-ante ambiguous. In section 3 we describe the identification strategy that allows us to identify the treatment effects. In section 4 we propose a new estimator to account for potential incomparability between treated and control group. In section 5 we describe the data set. In section 6 we present and discuss the empirical results. Section 7 concludes.

2 Theoretical considerations

Economic theory does not predict the direction of change in labor supply in response to the exogenous non-labor income shock. This section revises different approaches to this comparative statics exercise and illustrates the issue using simple examples.

2.1 Non-convexities in the budget set

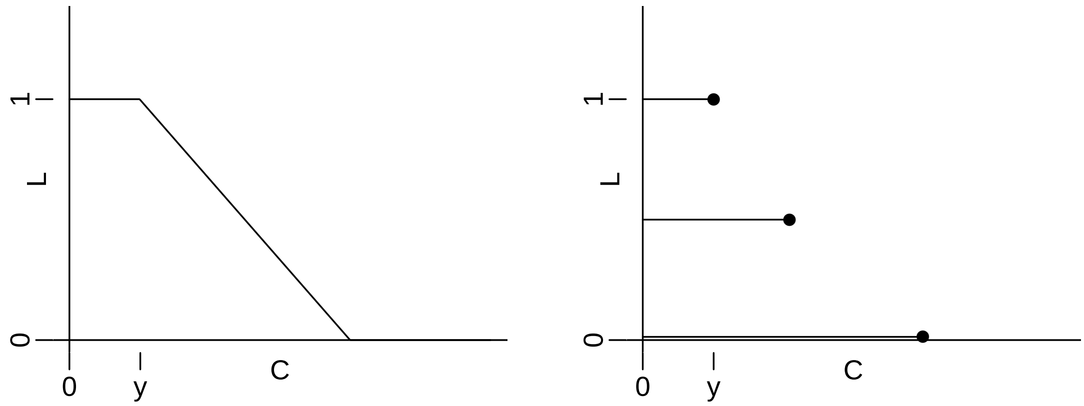
One of the explanations for why labor supply may not necessarily decrease as a response to an exogenous non-labor income shock is related to kinks in the budget set, i.e. points at which the boundary of the budget set in not differentiable (for the reference, see e.g. Heckman (1974), Burtless and Hausman (1978) and Moffitt (2002) for the review). Consider a simple static model of choosing consumption and leisure:

$$\max_{c,\ell} \quad U(c,\ell)$$
$$c = y + w(1-\ell)$$

where c is consumption good, l is leisure (normalized to closed unit interval), p is price of consumption good, w wage rate. Assume that U is increasing in both arguments. Figure 1 presents the budget set of the consumer.

Having obtained a non-labor income y, the consumer may sustain nonnegative consumption even if they spend the whole time endowment on leisure. Then the amount of feasible consumption increases with a decrease in labor. Consider an exogenous change in non-labor income y. Moffitt (1990) show that unless the consumer's choice initially was on the kink of the budget set, the direction of change while considering comparative statics with respect to y is ambiguous.

Moreover, the budget set that consumer faces is likely to be discontinuous. As the data shows, the vast majority of jobs observed are full-time jobs. This indicates that the choice of the amount of leisure feasible to the agent is discrete. Translating this finding into the budget set one obtains a discrete set of possible labor/leisure choices. This partially discretize the budget set as well,



() **kinked budget set** () **discontinuous budget set**

Figure 1: Consumer's budget set.

leading not only to kink-type points as discussed above, but also to discontinuities of the budget set. Therefore, depending on the shape of preferences, labor supply response to a shock in y has an ambiguous direction.

2.2 Preferences shifters

Another explanation for an a-priori ambiguous direction of changes in labor supply as a response to non-labor income shocks are preference shifters. Consider a simplified version of life-cycle consumption-leisure choice model. Suppose an economy contains a positive measure of a-priori identical consumers. Assume finite horizon, no uncertainty or discounting. Suppose that all consumers enjoy perfect foresight.

Focus the attention on consumer i. They solve:

$$\max_{c,\ell} \quad \sum_{t=0}^{T} \frac{c_t^{1-\sigma}}{1-\sigma} - \frac{\chi}{1+\frac{1}{\epsilon}}(1-\ell_t)^{1+\frac{1}{\epsilon}}$$
$$c_t + a_{t+1} = w_t(1-\ell_t) + y$$
$$c_t \geq 0, a_{t+1} \geq -a, a_0 = 0$$

where y is an endowment of exogenous, non-labor income, a_t is the standard Arrow-securities, w_t is wage, all in period t. The parameters of interest are Frisch elasticity of labor ϵ, relative risk aversion coefficient γ and individual disutility from work χ.

One may easily show that the optimal solution to the problem satisfies:

$$c_t \equiv c_i \quad \forall_t$$

$$c - \frac{1}{T}\Big(\sum_t w_t (\frac{w_t}{\chi + i})^{\epsilon}\Big) c^{-\gamma\epsilon} - Ty = 0$$

$$1 - \ell_t = (\frac{w_t}{\chi})^{\epsilon} c^{-\gamma\epsilon}$$

By the implicit function theorem, c is increasing in the endowment y. This result is consistent with economic intuition. However, once one plug it into the equation for within period labor supply, the direction of the effect depends on the risk-aversion parameter and Frish elasticity.

An additional insight comes if one allows χ to vary between consumers. This is especially appealing in the context of child benefit program, as one may expect the opportunity cost of working an additional hour to differ between households that need to provide care services to their children and those who do not have kids. In the simple form, the heterogeneity may depend on the endowment of non-income labor: $\chi_i = \psi(y, \theta_i)$ for some function ψ and household i specific parameter vector θ_i. In this case, childless household have less disutility from working. At the same time, leisure (that implicitly accounts also for time spent on child care services) is less costly once y is relatively high. This heterogeneity enters the equation of labor supply, which in turn implies the effects of increase in the non-labor endowment are not necessarily monotone between households.

2.3 Home production

Household labor supply decisions may be affected also by home production technology. Consider a version of Ghez and Becker (1975) household time use model. In this type of models each type of commodity that enters household's final consumption is produced through a (usually constant returns to scale) technology that combines input of market goods and time spent by household. Let the number of commodities be J. The home-production function of commodity j in period t is given by:

$$c_{jt} = F^j(\xi_{jt} x_{jt}, \nu_{jt} h_{jt}),$$

where x_{jt} and h_{jt} denotes the input of market goods and time, respectively, to produce commodity j, and ξ_{jt} and ν_{jt} denote the potential technology shifters for input of goods and time. Then, the life-cycle utility function of a household with discount factor β and period utility function $U(c_1, ..., c_j)$ is given by:

$$\sum_{t=0}^{T} \beta^t U(c_{1t}, ...c_{Jt}).$$

Each household maximizes its expected life-cycle utility subjects to sequential budget constraint and time constraint:

$$\sum_{j=1}^{J} p_{jt} x_{jt} + a_{t+1} = w_t n_t + (1 + r_t) a_t + y_t$$

$$n_t + \sum_{j=1}^{J} h_{jt} = 1,$$

where p_{jt} and w_t denotes the prices of input goods and labor, a_t and r_t denotes bonds and interest rate, and y_t denotes non-labor income. Household's endowment of time is normalized to 1, and n_t and h_{jt} denote time spent on market production and home production of commodity j respectively. One may note that this general model formulation nests the standard household problem, where the utility is a function of just consumption goods and leisure. That is, households maximize $\sum_{t=0}^{T} \beta^t U(c_{1t}, c_{2t})$ subject to budget constraint and $n_t + \ell_t = 1$, with $c_{1t} = x_t$ and $c_{2t} = \ell_t$.

To simplify the considerations, specify three commodities that the household utility depends on: a home production good, childcare, and leisure. The household allocates its time endowment among labor for market production n_t, time spent on home production h_{1t} and childcare h_{2t} maximizes $\mathbb{E}_0 \sum_{t=0}^{T} \beta^t U(c_{1t}, c_{2t}, \ell_t)$ subject to $p_{1t}x_{1t} + p_{2t}x_{2t} + a_{t+1} = w_t n_t + (1 + r_t)a_t + Y_t$ and $n_t + h_{1t} + h_{2t} + \ell_t = 1$, where $c_{jt} = F^j(\xi_{jt}x_{jt}, \nu_{jt}h_{jt})$ is characterized by home production function F^1 and childcare production function F^2.

The technology shifters ξ_{jt} and ν_{jt} and non-labor income y_t may be introduce uncertainty to the model. Denote the state in period t s_t and the history of states up to period t $s^t = (s_0, s_1, ..., s_t)$. Let $\lambda_t(s^t)\beta^t$ be the multiplier of household budget constraint and $\pi(s^t)$ be the unconditional probability of the realization of s^t. Then, the first order conditions (F.O.C.s) with respect to $x_{jt}(s^t)$, $h_{jt}(s^t)$, $n_t(s^t)$, and a_{t+1} are given by:

$$[x_{jt}(s^t)] \qquad \pi(s^t)\frac{\partial U}{\partial c_{jt}(s^t)}\frac{\partial F^j}{\partial x_{jt}(s^t)} = \lambda_t(s^t)p_{jt}(s^t)$$

$$[h_{jt}(s^t)] \qquad \frac{\partial U}{\partial c_{jt}(s^t)}\frac{\partial F^j}{\partial h_{jt}(s^t)} = \frac{\partial U}{\partial \ell_t(s^t)}$$

$$[n_t(s^t)] \qquad \pi(s^t)\frac{\partial U}{\partial \ell_t(s^t)} = \lambda_t(s^t)w_t(s^t)$$

$$[a_{t+1}(s^t)] \qquad \lambda_t(s^t) = (1 + r_{t+1}(s^{t+1}))\beta\lambda_{t+1}(s^{t+1}).$$

After combining the first three F.O.C.s, the intra-production substitution between input goods x_{jt} and time h_{jt} is characterized by

$$\frac{\partial F^j}{\partial x_{jt}(s^t)} \Big/ \frac{\partial F^j}{\partial h_{jt}(s^t)} = p_{jt}(s^t)/w_t(s^t),$$

and the cross-production substitution is characterized by

$$\frac{\frac{\partial F^1}{\partial x_{1t}(s^t)} \Big/ \frac{\partial F^1}{\partial h_{1t}(s^t)}}{\frac{\partial F^2}{\partial x_{2t}(s^t)} \Big/ \frac{\partial F^2}{\partial h_{2t}(s^t)}} = \frac{p_{1t}(s^t)}{p_{2t}(s^t)}.$$

Therefore, the elasticities of substitution of home production technology F^j and utility function U determine optimal allocation of input goods and time use.

These elasticities are likely to vary with household characteristics such as the number or age of children. If the non-labor income increases, the relative cost of market good inputs decrease relatively to the time input. For example, with a larger budget for current period, depending on their preference and technology, the household may want to allocate more time spent with children. However it may be the case that market care (preschool, extracurricular classes and activities, etc.) if significantly more efficient in term of care production technology, so that the additional income is spread as inputs for other types of home production.

2.4 First and Second earner time allocation

Labor supply decision within the household is likely to be made simultaneously for each of the members [1]. Blundell et al. (2016b) find that, in a household with children, the response to permanent and transitory shocks differ between female and male[2]. In response to a transitory wage shock, the elasticity of husbands' labor supply does not depend on the presence of children, while with young kids in the households, women labor supply elasticity is significantly larger. In response to a permanent increase in wage, men allocate more time in market labor on the cost of both leisure and childcare, whereas women spend less time working in the market but more time working at home with children. However, when the woman receives a permanent wage increase and reduces her time spent in childcare significantly, it is the man that reduces his hours, increases leisure and tends to spend slightly more time providing childcare.

We will account for this issues in the empirical part, considering separate effects for males and females within a household.

3 Identification Strategy

The program *Rodzina 500 Plus* is one of the most significant policy interventions affecting the household sector in Poland in the recent years. Introduced in the second quarter of 2016, this large scale program (cost of app. 2% of GDP per year) provides a monthly non-equivalent benefit of roughly 20% of net average wage to each family with two children, and another 20% of net average wage for each third and next child. Additionally, families below a certain income level are entitled to obtain the benefit also for the first-born child. In order to obtain the benefit, eligible households are supposed to register in local authorities. This kind of simplicity strongly

[1] Mroz (1987) provides some evidence that the income earned by husband is largely exogenous to woman's labor supply, Connelly (1992) in turn provides the conditions under which single parent optimization is consistent with household optimization.

[2] In more general terms first and second earner. Data suggests that the mother is more likely to be the second earner in the presence of children in the household.

stimulates participation among the eligibles. The authorities predict that around 2.7 million families bringing up to 3.7 million children are or will be enrolled in the program. The main goal of the program is to improve financial well-being of families upbringing children and stimulate fertility.

The program was announced in the middle of the first quarter of 2016. The first payment should have been distributed in April, but some households obtained the benefit for the initial months with some delay. The short time before announcement and realization ensure the absence of any anticipation effects that would interfere the causal inference. However, we shall discuss a caveat that results from the introduction of the program in the middle of 2016Q2 and the delay. In 2016Q2 some of the households were interviewed before the program has actually begun. Others might have not obtained the benefit even though they have satisfied all the legal requirements. Therefore, one may question whether they have succeeded to adjust the economic behavior to the program, which actually have not yet affected them. This is why we prefer to perceive 2016Q2 as still a pre-treatment period. The data provide some evidence supporting this claim. We elaborate more on that issues in the following sections.

Furthermore, it was not clear if households perceive the benefit as a long run increase in the non-labor income when it was introduced. If they expected the program to be terminated in the reasonably short horizon or believed that the eligibility requirements might change, then they were likely to treat the increase in non-labor income as transitory. Therefore, the labor supply would not adjust. Moreover, even household perceived the program as permanent, the process of labor supply adjustment may take several periods due labor market frictions, already existing contracts, etc. Hence, movements in aggregate labor supply were likely be fully observed in data with some delay. However, the longer the span covered by the study, the greater the risk that the sample would include households which responded to *Rodzina 500 Plus* with increased fertility. This is why we focus on the immediate effects of the program utilizing data for 2016Q3.

The estimation of the effects of nonequivalent child benefit program on labor supply falls into the category of program evaluations. However, the eligibility to the program is not random as long as we believe that the number of children within a household is not random. Indeed, economic theory predicts that the number of children might be set as a result of household optimization (Rosenzweig and Schultz, 1985). It implies that families that do not obtain the benefit may differ systematically from the treated subpopulation. Additionally, the number of the children is also likely to affect the outcome variable i.e. the labor supply measures (Connelly, 1992; Nakamura and Nakamura, 1992; Black et al., 2013).

To address this issues and consistently estimate the effects of interest we develop a new econometric approach. We combine an idea of covariate balancing propensity score (Imai and Ratkovic, 2014) with difference-in-difference estimator by Abadie (2005) and Heckman et al. (1997). The strategy based on DID exploits the quasi-natural experiment character of the child benefit program – it was introduced swiftly, not allowing for the anticipation effects. The estimators of

Abadie (2005) and Heckman et al. (1997) solve potential problems with non-parallel outcome dynamics in treated and non-treated subpopulations, which might result from non-experimental character of the data and systematic differences between treated and non-treated. However, for the difference-in-difference method to be a reliable estimator of the causal effect, one needs to assure a commonality of trends prior to the treatment (Imbens and Wooldridge, 2009). Given the stark differences between eligible and ineligible households, this assumption is not likely to be satisfied. Hence, we utilize CBPS estimates of propensity score to assure proper adjustment of the control to the treated group. The CBPS estimator employs moment-based approach to force balancing of the conditional distributions of covariates. It is expected to reduce significantly bias resulting from the systematic differences between treated and untreated subpopulations. It should also help to assure commonality of trends prior to the treatment, thus validate the DID approach. The details of our empirical approach are provided in the next sections.

Finally, the universal character of the program makes it difficult to define appropriate control groups. Among the treated there are plenty of households on different stages of the life cycle and composition (Graham and Beller, 1989). It is necessary to ensure comparability of outcomes between households assigned to the treatment and control group. Previous research proposed to utilize local average treatment effect to be more confident about the causal identification (Angrist and Evans, 1998; Jacobsen et al., 1999; Agüero and Marks, 2008; Cristia, 2008; Lefebvre and Merrigan, 2008). Therefore, we pursue the local average treatment strategy in defining the control groups. Several approaches are considered. The first idea consist of comparing households with (at least) two children and the youngest child under the age of 6 to the households with one child under the age of 6. The idea behind this identification scheme relies on the premise that children require the most attention when they are the youngest and this attention cannot be easily replaced through the market for caring service. At this period of child development, the mother in general is more likely to be inactive (the caring leave after the maternity leave) which implies that the least disincentive may be sufficient to turn working mothers into non-working mothers. We justify the choice of a threshold by the fact that even though the school duty is imposed on children from 7, children at the age 6 might go to a so called zero-class, which is frequently run in a proper primary school. Another concern relates to comparing women as second earners in households with two or more children to households with women as single earners eligible for the program. Employment elasticities are larger for single than married mothers (Connelly and Kimmel, 2003; Blundell et al., 2016a). Additionally, both single and married woman obtain kind of a non-labor income: single mothers in a form of benefit, full family mothers in a form of husband's earnings. However, single women are supposed to be more constrained by means of providing child care services. One may also expect that more frequently single women are financially constrained in comparison with the married. Therefore, the benefit may loosen the constraints and in this way enable single women to move to new optimum, which is not necessarily a corner-solution.

4 Estimation

We propose a novel approach in estimation of average treatment effects on the treated that combines the balancing propensity score matching (or weighting) with difference-in-difference. The following paragraphs introduce the notation and describe our approach in detail.

We use the standard program evaluation notation throughout the paper. We observe a household i in two periods indexed by $t \in \{0, 1\}$. Treatment D_{it} is binary and is observed only in $t = 1$, hence $D_{it} \equiv D_{i1} \equiv D_i$. Therefore, $\forall_i \quad D_{i0} = 0$. Potential outcomes are denoted as Y_{it}^D, where $t \in \{0, 1\}$, $D \in \{0, 1\}$. For instance, Y_{i1}^1 is the value of outcome variable in the world a unit i receives the treatment. The outcome variable observed by an econometrician is $Y_{it} = Y_{it}^0(1 - D_{it}) + Y_{it}^1 D_{it}$. We consider a set of pre-treatment covariates $X_i \equiv \{x_{i0}\}$, assuming population conditional independence $Y_{it}^0, Y_{it}^1 \perp\!\!\!\perp D_i | X_i$. We are interested in average treatment effect on the treated (ATT):

$$\tau^{ATT} \equiv \tau = \mathbb{E}[Y_{it}^1 - Y_{it}^0 | D_{it} = 1] = \mathbb{E}[Y_{i1}^1 - Y_{i1}^0 | D_i = 1] \tag{1}$$

or its conditional-on-covariates versions. The conditional probability of treatment assignment is given by $P[D_i = 1|X] = F_\beta(X_i)$, where β is a set of parameters. Let $f_\beta(X_i)$ denote the first derivative of $F_\beta(X_i)$ with respect to the parameters. If nothing else is mentioned, $F_\beta(X_i)$ is the logistic cumulative distribution function.

4.1 Covariate balancing propensity score

Since the influential paper of Rosenbaum and Rubin (1983), the propensity score related methods have been continuously gaining attention in research, especially in a context of causal analysis (see for example Dehejia and Wahba, 1999; Caliendo and Kopeinig, 2008; Austin, 2011). Propensity score as a conditional probability of treatment assignment serves mainly for matching or weighting. The true propensity score is rarely known, so it must be estimated by a researcher. The standard framework is limited to maximum likelihood binomial models. However, as it stems from the likelihood theory, misspecification of the propensity score model leads to inconsistent estimators of binomial model parameters and therefore to inconsistent estimates of propensity score. That in turn might result in potentially heavily biased causal analysis (Zhao, 2004). Usually we do not know the correct functional form of data generating process, which is why researchers tend to experiment with many probably incorrect specifications and choose *the best* one. Due to the already mentioned reasons, the estimation of conditional probability of being eligible to the child benefit program is excessively exposed for those threats. Therefore, instead of estimating many probably misspecified propensity scores and checking their quality by a covariate balance check (Dehejia and Wahba, 2002), we employ the approach of Imai and Ratkovic (2014). They exploit the dual nature of propensity score. It is both conditional probability of treatment assignment and covariate balancing score. They derive a set of moment conditions that identify a propensity score vector that automatically weight the control group such that their weighted distribution matches with that of the treatment group:

$$\mathbb{E}\Big[D\tilde{X} - \frac{F_\beta(X_i)(1-D)\tilde{X}}{1-F_\beta(X_i)}\Big] = 0 \tag{2}$$

One may set $\tilde{X} = X$ to balance the first moments, $\tilde{X} = X^2$ to balance second moments, etc. Restrictions that impose balance on a chosen moment for all covariates are sufficient to assure just-identification of the parameters β from the probability function $F_\beta(X_i)$. In general, imposing restrictions for more than one class of moments leads to an over-identified estimator, expected to be more biased in finite samples but asymptotically more efficient. One may add either first order conditions from binary ML model with $F_\beta(X_i)$ as the cumulative probability function or balancing conditions for (higher order) moments of covariates. Additional, the over-identified version enables the researcher to perform a standard specification test of over-identifying restrictions. In the context of CBPS estimation, the specification test might be perceived as a verification of the reliability of unconfoundedness assumption (Imai and Ratkovic, 2014).

Let start with the just-identified case. The sample moment conditions are:

$$\frac{1}{N}\sum_i \frac{N}{N_1} \cdot \frac{T - F_\beta(X_i)}{1-F_\beta(X_i)} \cdot X_i = 0 \tag{3}$$

The just-identified models are at the core interest in the paper as we have rather a small sample. We consider over-identified version only to perform the specification tests.

In summary, we prefer to use the CBPS method as it is robust for misspecification of the probability model instead of ML estimator. Additionally, it balances the distribution of covariates which is advantageous in the context of the paper due to expected heterogeneity between control and treated groups.

4.2 Abadie's difference-in-difference

Difference-in-difference estimators exploit the time dimension in data to generate means of a natural experiment. Following Abadie (2005), we maintain the assumption of parallel path of outcomes:

$$\mathbb{E}[Y^0_{i1} - Y^0_{i0}|X_i, D_i = 1] = \mathbb{E}[Y^0_{i1} - Y^0_{i0}|X_i, D_i = 0] \tag{4}$$

which states that in the absence of treatment outcomes in both groups would behave in the same way[3] and an overlap assumption $P[D_i = 1|X_i] < 1$ – a well known identification condition for selection on observables. These assumptions are enough to uncover ATT.

[3] Abadie (2005) notices that this assumption identifies only $\mathbb{E}[Y^1_{i1}|X_i, D_i = 0]$. $\mathbb{E}[Y^1_{i1}|X_i, D_i = 1]$ remains unrestricted so it does ATE. However, we focus only on ATT so it is of minor concern in the paper.

Abadie (2005) proposes an estimation procedure that relies on the least squares approximation to $\tau_{X_k} = \mathbb{E}[Y_1^1 - Y_1^0 | D_i, X_k]$:

$$\tau_{X_k} = \underset{\theta \in \Theta}{\arg\min} \quad \mathbb{E}\Big[F_\beta(X_i) \cdot \Big(\frac{D_i - F_\beta(X_i)}{(F_\beta(X_i))(1 - F_\beta(X_i))} \cdot (Y_{i1} - Y_{i0}) - g(X_k; \theta)\Big)^2\Big] \tag{5}$$

where $X_k \subset X$ and $g(X_k; \theta)$ is the approximating function. If $g(X_k; \theta) \equiv \theta$ then $\tau_{X_k} = \tau$. We are interested mainly in the *homogeneous*[4] average effect on the treated. Considering sample analogs, the equation 5 implies the following first order conditions:

$$\frac{1}{N} \sum_i -F_\beta(X_i) \left(\frac{D_i - F_\beta(X_i)}{1 - F_\beta(X_i)} \Delta Y_i - F_\beta(X_i) \tau \right) = 0 \tag{6}$$

Our contribution is to combine moment conditions from equations (3) and (6) in order to obtain a GMM representation for a one-step ATT estimator. It is appealing in many ways, as it combines advantages of both CBPS and Abadie's DID method. First, it is a one-step method so there is no need to adjust the DID estimation for the stochastic nature of estimated previously propensity score. Second, we move an estimation process into the GMM famework. It allows us to obtain the (efficient) covariance matrix immediately. Third, the one-step estimation is justified because the way propensity score is defined produces balancing properties that are automatically satisfied. Fourth, given conditional independence assumption is satisfies it is robust for functional form misspecification, as neither Imai and Ratkovic (2014) nor Abadie (2005) assume any functional form for the true data generating process. In further discussions we refer to this estimator as ACBPS.

In the empirical part we apply just-identified ACBPS as a efficient GMM, where the parameters needed to estimate consistently the moment covariance matrix come from pre-estimated logit and logit-based Abadie's DID. We compare the results with Heckman's difference-in-difference propensity score estimators using both logit and CBPS propensity score estimates. Our method should be advantageous over the conventional estimators in the presence of improperly balanced control group. Additionally, we expect it to be less biased in small samples than ACBPS as it performs matching instead of GMM estimation. Finally, we present also results using standard OLS estimator. However, this method is expected to perform poor due to lack of a control for the imbalance between treatment and control groups. We maintain an assumption of conditional independence, i.e. conditioning on the set of covariates described in the next section, there is no systematic differences in probability of obtaining the benefit between treated and control subsamples.

4.3 Robustness checks

We compare estimates from our estimators with other estimators for ATT to verify the robustness of the results. We investigate:

[4]The quality of any non-constant approximation for the effects of interest is expected to be poor due to the limited number of observations in our sample.

↳ Basic DID (as a baseline, Card and Krueger (1994))

$$\tau = (\overline{y}_{11} - \overline{y}_{01}) - (\overline{y}_{10} - \overline{y}_{00})$$

↳ Regression-based DID controlling for covariates:

$$\Delta Y_i = \beta_0 + \tau D_i + X_i\beta + \varepsilon_i$$

↳ Abadie's estimator in the original two step formulation, with ML logit estimation for the propensity score.

↳ Heckman's estimator in the original formulation with ML logit estimation for the propensity score and kernel matching.

The Stata codes for ACBPS and CBPS estimation are available on request. In the paper we also make use of codes by Houngbedji (2016) and Leuven and Sianesi (2003).

5 Data

We utilize data from Labor Force Survey (LFS) collected by the Central Statistical Office in Poland. It is a rotating panel in which households are interviewed in two consecutive quarters and then re-interviewed in two respective periods a year after. The time dimension allows us to apply the DID strategy. We present the estimates of immediate effects of the program, i.e. the sample covers 2016Q3 as the only post-treatment period for the main results.[5] Finally, we obtain a balanced panel with $T_0 = 2015Q3$ and $T_1 = 2016Q3$.

Table 1: Covariates controlled for in the models' specifications

No.	Type of variables	Specific variables
(1)	household	presence of non-parental household members, small city and village indicators, mother's age at birth of the youngest child
(2)	mother demographic	age, age squared, work experience, education
(3)	mother's job	indicators for public company, working as an employee, tenure contract, full-time job, binary variable for working in the field of education
(4)	father demographic	age, age squared, work experience, education
(5)	father's job	indicators for public company, working as an employee, tenure contract, full-time job, binary variable for working in the field of education

We consider only families with children and drop households in which another child was born between T_0 and T_1. Depending on the definition of control groups we keep households with

[5] As a sensitivity analysis we present the results where both 2016Q2 and 2016Q3 for a given household are considered post-treatment periods.

various characteristics in the sample. We distinguish five main groups of covariates that we control for in order to satisfy the unconfoundedness assumption. They are described in Table 1.

6 Results

This section presents estimates of a range of local treatment effects concerning the impact of introducing non-equivalent child support on labor supply. We consider various definitions of control and treated subpopulation to provide an assessment of the total effect. Wherever the sample size is sufficient for identification, we separate effects on households with respect to the age of the youngest child, as it is a significant determinant of (especially women's) labor supply (Jacobsen et al., 1999). The sets of control variables are described in the previous section.

The main results are presented in Tables 2-7 and are calculated assuming that 2016Q2 is a pre-treatment period. One might compare them with Tables 8-12 in the Appendix A which contain the estimates using both 2016Q2 and 2016Q3 as post-treatment periods. Even though using more data limits problems with small number of observations, especially in the subgroups defined over a certain age of the youngest child in household, the estimated impact is usually even weaker and less precise. It suggests that 2016Q2 should not be perceived as a quarter when the program has already been introduced. This claim is additionally supported by the fact that estimates obtained solely using 2016Q2 as a post-treatment period are not only statistically insignificant but also numerically close to zero in almost all specifications.

The specification test mentioned in the tables is the over-identifying restrictions test for CBPS propensity score. It is valid both for Heckman's and Abadie's approaches. They contain just-identified models, hence the estimates of propensity score probability function parameters are the same. Rejection of the null hypothesis might suggest violation of the conditional independence assumption indicating failure in accounting for observable factors.

6.1 Full families with children at the same age

We begin with comparisons between families in which both mother and father are present. The treated group is defined as a subset of households that have at least two children — thus are eligible for the participation in the child benefit program, whereas families with only one children are classified as a control group. However, some of the control group households may also be eligible. The data gives no opportunity to verify their participation status as household's income is not observed. However, one should not expect a high percentage of potentially treated in the control group, as they contain at least both father and mother so their total income is likely to exceed the eligibility threshold. We separate the effects for parents who worked in the pre-treatment period and those who did not.

First, we analyze paths of labor supply of parents who worked in the pre-treatment period. Table

2 presents the estimated effects for women, Table 3 for men. The average effects on the pooled sample are insignificant for both men and woman. However, they do not pass the specification test, thus one might doubt whether the necessary identification assumptions have been satisfied. However, we do not reject the null hypothesis for models estimated on subsamples with regard to the age of the youngest child. It suggests that households differ significantly with age of the youngest child and so their outcomes should not be compared in a correct causal analysis.

Most local effects for both women and men who worked in T_0 and have obtained a benefit indicate that they worked in T_1 with approximately the same probability as those employed in $T = 0$ who have only one child. This finding is robust to the choice of specification. In the majority of comparison groups, not only are the parameters statistically insignificant but they tend to zero also numerically for all subgroups except for the last one. There appear some exceptionally significant OLS or logit-based DID estimates, but they are likely to be biased because of inability to account for the systematic differences between households. Additionally, other estimates in their comparison group are way smaller and insignificant.

Two comparisons require further comments. First, all estimates consistently suggest a small decrease in labor participation for treated mothers whose youngest child is below 6. The estimates are obtained using a sample of moderate size. Even though they are not significant, the considerable number of observations suggest that this tendency does not necessarily come from the sampling error. It might indicate a weak regularity in quitting the job by participating married females. This regularity is consistent with economic theory, as additional money from non-equivalent transfer shifts the budget line so the allocation with no labor (and full-time maternal care) becomes available. The youngest children need maternal care in the highest degree. However, due to the immediate character of the estimates the effect is weak, as potentially more time is required to effectively quit a job in a response to the benefit program introduction. Moreover, estimates on the sample containing both 2016Q2 and 2016Q3 (Table 8) confirm the presence of slight decline in labor market participation. Second, similar situation takes place in case of married mothers whose youngest child is above 11. However, here the effects are likely to equal to zero. Significant regression estimates emphasize rather the fact that each treated mother from this subpopulation in the sample is employed in both T_1 and T_0 whereas 4% of the untreated dropped out from the employment state. Moreover, only 80 treated observations do not increase the credibility of the result.

Slightly different story can be told about hours worked. According to the estimates, the program affects positively hours worked by mothers of at least two children who are older than 11, i.e. they have finished the primary school or are about to do so. The average impact amounts to approximately 6-8 hours. This result is consistent with an intuition and theoretical predictions. Teenagers do not require full time maternal care. On the contrary, the market services might outperform informal care by means of the quality, as the needs of teenagers are different than those of younger children. Therefore, given a non-equivalent transfer mothers become able to

Table 2: **Effects on mother's who worked in T_0.**

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	0.005	0.003	0.006	0.010	0.004	0.005
	(0.582)	(0.764)	(0.548)	(0.622)	(0.748)	(0.635)
N= 1664, N treated = 820, Specification test for CBPS p-val = .001						
youngest child younger than 6						
τ	-0.021	-0.032	-0.021	-0.032	-0.026	-0.026
	(0.161)	(0.069)	(0.131)	(0.541)	(0.144)	(0.145)
N= 702, N treated = 408, Specification test for CBPS p-val = .872						
youngest child between 6 and 11						
τ	0.027	0.017	0.020	0.057^{*}	0.016	0.011
	(0.074)	(0.278)	(0.187)	(0.029)	(0.410)	(0.573)
N= 575, N treated = 332, Specification test for CBPS p-val = .899						
youngest child older than 11						
τ	0.039^{***}	0.039^{**}	0.036	0.033	0.045^{*}	0.030
	(0.000)	(0.004)	(0.186)	(0.359)	(0.025)	(0.104)
N= 387, N treated = 80, Specification test for CBPS p-val = .904						
Hours worked						
all households						
τ	1.437	1.348	1.128	1.310	0.947	1.350
	(0.054)	(0.101)	(0.225)	(0.122)	(0.303)	(0.127)
N= 1664, N treated = 820, Specification test for CBPS p-val = .001						
youngest child younger than 6						
τ	0.927	0.563	0.228	0.719	0.767	0.861
	(0.425)	(0.679)	(0.857)	(0.561)	(0.587)	(0.549)
N= 702, N treated = 408, Specification test for CBPS p-val = .872						
youngest child between 6 and 11						
τ	0.757	0.085	-0.320	-0.493	-0.228	-0.375
	(0.534)	(0.948)	(0.804)	(0.692)	(0.875)	(0.795)
N= 575, N treated = 332, Specification test for CBPS p-val = .899						
youngest child older than 11						
τ	5.641^{**}	5.371^{**}	7.599	6.152^{**}	7.683^{***}	5.785^{**}
	(0.002)	(0.004)	(0.091)	(0.004)	(0.001)	(0.008)
N= 387, N treated = 80, Specification test for CBPS p-val = .904						

We control for the following set of covariates: (1), (2), (3), (4), (5) and labor status of the father – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

afford additional after-school classes or other form of activities for a child gaining additional time to work. Provided the leisure plays a minor role in total utility, both mother and a child might spend the time more productively. One should emphasize that this effect concerns adjustments of labor supply only at the intensive margin. Furthermore, it is worth mentioning that in the sample 21% and 13% of the untreated and treated mothers respectively decreased the number of working hours, whereas 14% and 25% of the untreated and treated respectively increased the number of hours worked. Therefore, the data background is different than in case of probability of being employed in this comparison. It supports the conclusion about the effect's significance. One should also notice that all but one estimates for hours worked by females with teenager as the youngest child are statistically significant. Only the estimate from the ACBPS model seems insignificant. Nevertheless, it is numerically close to the rest of effects. We interpret this results as a lack of precision. ACBPS is a conventional GMM estimator and therefore might be inefficient in small samples.

Table 3: Effects on father's who worked in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	0.011	0.011	0.011	0.015	0.010	0.013
	(0.059)	(0.084)	(0.251)	(0.445)	(0.198)	(0.087)
N= 2355, N treated = 1255, Specification test for CBPS p-val = 0						
youngest child younger than 6						
τ	0.006	0.011	0.019	-0.078	0.014	0.007
	(0.492)	(0.242)	(0.330)	(0.713)	(0.223)	(0.531)
N= 1214, N treated = 754, Specification test for CBPS p-val = .251						
youngest child between 6 and 11						
τ	0.024*	0.017	0.004	0.049*	0.008	0.010
	(0.040)	(0.122)	(0.593)	(0.040)	(0.556)	(0.503)
N= 704, N treated = 406, Specification test for CBPS p-val = .344						
youngest child older than 12						
τ	0.020**	0.011	0.010	0.010	0.015	0.015
	(0.008)	(0.198)	(0.321)	(0.769)	(0.311)	(0.246)
N= 437, N treated = 95, Specification test for CBPS p-val = .801						
Hours worked						
all households						
τ	1.218	1.408*	1.728*	1.737*	1.690*	1.856*
	(0.056)	(0.038)	(0.041)	(0.019)	(0.035)	(0.017)
N= 2355, N treated = 1255, Specification test for CBPS p-val = 0						
youngest child younger than 6						
τ	1.764	2.097*	3.215*	3.512	2.590*	2.152
	(0.051)	(0.034)	(0.014)	(0.135)	(0.024)	(0.054)
N= 1214, N treated = 754, Specification test for CBPS p-val = .251						
youngest child between 6 and 11						
τ	-0.299	0.020	-0.639	-0.812	-0.312	-0.741
	(0.796)	(0.987)	(0.644)	(0.515)	(0.822)	(0.593)
N= 704, N treated = 406, Specification test for CBPS p-val = .344						
youngest child older than 11						
τ	3.743*	2.612	2.720	2.606	3.167	2.662
	(0.033)	(0.158)	(0.332)	(0.172)	(0.174)	(0.207)
N= 437, N treated = 95, Specification test for CBPS p-val = .801						

We control for the following set of covariates: (1), (2), (3), (4), (5) and labor status of the mother – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The evidence for an impact of the program on the previously employed men's hours worked is also weak. One should not focus on the pooled sample estimates as they do not pass the specification test. However, the results for men who worked in T_0 and whose youngest child is below 6 are consistently positive and barely significant, especially in the specifications using CPBS approach. It would suggest that proper balancing allows us to uncover a weak effect. Even though an increase in labor supply by 2-3 hours weekly seems insignificant economically, it is interesting to look at these estimates taking into account the negative estimates for married woman in the same child age group. Holding all concerns about the size of the effect mentioned before, the child benefit program might cause an intra-household change of the social roles. Consider a household in which the mother quit the job in order to provide more maternal care. The amount of money received as a benefit is typically smaller than earnings lost as a result of job resignation. Therefore, it sounds reasonable that the father increases hours worked to smooth at least partially family's income.

Table 4: Effects on mother's who didn't work in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	0.007	0.027	0.007	0.010	0.030	0.008
	(0.789)	(0.345)	(0.806)	(0.722)	(0.427)	(0.825)
N= 906, N treated = 528, Specification test for CBPS p-val = .001						
youngest child younger than 6						
τ	-0.048	0.038	0.028	0.042	0.025	0.027
	(0.207)	(0.328)	(0.403)	(0.196)	(0.621)	(0.598)
N= 599, N treated = 396, Specification test for CBPS p-val = .355						
youngest child older than 5						
τ	0.018	0.027	0.049	0.050	0.050	0.060
	(0.616)	(0.454)	(0.212)	(0.124)	(0.332)	(0.223)
N= 307, N treated = 132, Specification test for CBPS p-val = .778						
Hours worked						
all households						
τ	0.343	1.217	0.582	0.723	1.309	0.683
	(0.723)	(0.246)	(0.563)	(0.471)	(0.322)	(0.596)
N= 906, N treated = 528, Specification test for CBPS p-val = .001						
youngest child younger than 6						
τ	-1.479	1.455	1.051	1.439	0.991	0.987
	(0.287)	(0.326)	(0.391)	(0.260)	(0.590)	(0.591)
N= 599, N treated = 396, Specification test for CBPS p-val = .355						
youngest child older than 5						
τ	0.835	0.859	1.920	1.800	1.499	1.954
	(0.513)	(0.481)	(0.170)	(0.115)	(0.369)	(0.248)
N= 307, N treated = 132, Specification test for CBPS p-val = .661						

We control for the following set of covariates: (1), (2), (4), (5) and labor status of the father – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tables 4 and 5 presents analogous effects of the program on the subsample of parents who did not work in T_0. The results are identical regardless the specification. No estimate is significant either statistically or economically. However, this comparison scenario suffers more intensely from the lack of data. The subsamples of males and females who did not work in T_0 are substantially smaller than for the employed parents. It is especially visible in case of the fathers, as in general men are more likely to be employed.

Table 5: Effects on father's who did not work in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	0.061	0.008	0.058	-0.042	-0.010	-0.012
	(0.317)	(0.910)	(0.487)	(0.712)	(0.919)	(0.871)
N= 215, N treated = 93, Specification test for CBPS p-val = .508						
Hours worked						
all households						
τ	2.426	0.440	2.398	-1.272	-0.425	-0.105
	(0.343)	(0.881)	(0.512)	(0.788)	(0.918)	(0.974)
N= 215, N treated = 93, Specification test for CBPS p-val = .508						

We control for the following set of covariates: (1), (2), (3), (4) and labor status of the mother – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In conclusion, the program seems to have minor impact on labor status of parents both employed and unemployed in a pre-treatment period. Considering hours worked, the results indicate a moderate increase of 5-6 hours weekly worked by treated mothers in comparison with the non-participating, whose youngest child is a teenager. Moreover, regardless of the estimator applied the qualitative interpretation of this outcome remains the same. However, the small number of observations (80 treated mothers) undermines the credibility of the results. An interesting story can be told about households with both parents and youngest child below 6. There is weak evidence on a small decrease in probability of employment for women and moderate increase in father's hours worked. We expect this effect to become more evident in the subsequent quarters as households will have more time to adjust their behavior. The program *Rodzina 500 Plus* does not differentiate the treated and untreated parents among the rest of comparison groups.

6.2 Single participating mothers and ineligible full families

In the next step we compare single mothers with two or more children with women from the full family which have only one child. Due to the relatively small number of single woman households we are not able to separate the effects with respect to the age of the youngest child. Moreover, estimation on the subsamples of men is infeasible due to an insufficient number of single fathers in the sample. Fortunately, all specifications pass the over-identification test. The rationale for the comparisons between married and single females lies in larger employment elasticities for the latter (Connelly and Kimmel, 2003). They might react strongly for any disincentive to work. However, the data provide no evidence supporting this hypothesis. The program does not differentiate the probability of working in T_1 between sole participating mothers and matched untreated married females regardless whether they worked in T_0 or not.

Considering hours worked, we obtain slightly significant estimates from Heckman's models on the employed in T_0. These estimators utilize matching so are expected to outperform the rest of

Table 6: Effects on lone mothers who worked in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	0.004	0.025	0.085	0.039	0.034	0.031
	(0.870)	(0.267)	(0.261)	(0.110)	(0.083)	(0.114)
N= 903, N treated = 59, Specification test for CBPS p-val = .645						
Hours worked						
all households						
τ	0.906	2.013	3.243	1.759	2.855*	2.692*
	(0.538)	(0.157)	(0.209)	(0.257)	(0.020)	(0.028)
N= 903, N treated = 59, Specification test for CBPS p-val = .645						

We control for the following set of covariates: (1), (2), (3) – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

methods as the number or treated observations is dramatically low. At the same time, ACBPS indicates an effect even stronger numerically, but (not surprisingly because of the GMM nature of the estimate) insignificant. Although one should not claim economic importance of the effect that amounts to 3 hours weekly, the estimate provide an average of all effects for various ages of the female's youngest child. Therefore, we interpret it as a premise that some sole treated mothers work in T_1 significantly longer than their untreated counterpart. The positive sign of the estimate suggests that the program provides incentive to work for lone mothers of rather older children. In such a case the difference between the quality of informal and market care is not dramatic. Additional non-labor income allows those mothers to buy more market care and therefore work longer, whereas married woman might have used informal care provided by other members of a household in both periods. We shall emphasize that the positive effects concern only those who are employed and decide to work longer. There is no connection with changing the labor market status between unemployment and employment.

Table 7: Effects on lone mothers who did not work in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	-0.030	0.058	0.044	0.058	0.064	0.063
	(0.627)	(0.344)	(0.574)	(0.319)	(0.434)	(0.406)
N= 419, N treated = 41, Specification test for CBPS p-val = .256						
Hours worked						
all households						
τ	-0.140	2.503	2.477	2.562	2.658	2.734
	(0.953)	(0.278)	(0.443)	(0.249)	(0.380)	(0.335)
N= 419, N treated = 41, Specification test for CBPS p-val = .256						

We control for the following set of covariates: (1), (2) – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7 Conclusions

In this paper we provide evidence that the introduction of a large scale child benefit program has a minor immediate impact on women's labor supply and almost no effect on men's labor supply. We found a small encouraging effect on treated woman's hours worked on the subpopulation who worked in the pre-treatment period – both single and married. It refers rather to mothers whose youngest child is at school age. Additionally, there are some signals that married mothers of the youngest children who obtain benefits decrease their probability of employment in comparison with the untreated. Simultaneously, the effect for fathers' hours worked in the same comparison group are positive and barely significant. It would suggest a program-induced change in the division of household duties between men and women. We expect this result to become more evident in subsequent quarters. Except for that, all other effects on both employment and hours worked equal zero. Each device, either simple as OLS regression or more sophisticated as CBPS-based DID estimators, mostly produces estimates that are statistically insignificant and economically unimportant. Results are strongly robust to the choice of estimation strategy. The labor supply of the treated follow the same path as the untreated regardless whether they take care of an infant, they worked in the pre-treatment period and are single mothers.

There are some possible explanations for such outcomes. First, as mentioned before we provide estimates of immediate effects of the program introduction. Households might not have enough time to adjust their behavior for the presence of an additional non-labor income. The process of quitting or finding a job is not immediate and may last over a quarter. Second, one should not expect any reaction in labor supply if households do not perceive the benefit as a long term icrease in income. Third, the data utilized in the paper concern the third quarter which contains summer holidays period. It would not be rare if the families postpone any important decisions for subsequent quarters. Fourth, the precision of estimators suffer from the small size of sample as only one wave of the LFS might be utilized in estimation.

References

Abadie, A. (2005): "Semiparametric difference-in-differences estimators," *The Review of Economic Studies*, 72, 1–19.

Agüero, J. M. and M. S. Marks (2008): "Motherhood and female labor force participation: evidence from infertility shocks," *The American Economic Review*, 98, 500–504.

Angrist, J. D. and W. N. Evans (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, 450–477.

Austin, P. C. (2011): "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies," *Multivariate Behavioral Research*, 46, 399–424.

Averett, S. L. and J. L. Hotchkiss (1997): "Female labor supply with a discontinuous, nonconvex budget constraint: incorporation of a part-time/full-time wage differential," *Review of Economics and Statistics*, 79, 461–470.

Baker, M., J. Gruber, and K. Milligan (2008): "Universal child care, maternal labor supply, and family well-being," *Journal of political Economy*, 116, 709–745.

Besley, T. and S. Coate (1992): "Workfare versus welfare: Incentive arguments for work requirements in poverty-alleviation programs," *The American Economic Review*, 82, 249–261.

Black, D. A., N. Kolesnikova, S. G. Sanders, and L. J. Taylor (2013): "Are children "normal"?" *Review of Economics and Statistics*, 95, 21–33.

Blau, D. M. and P. K. Robins (1988): "Child-care costs and family labor supply," *The Review of Economics and Statistics*, 374–381.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016a): "Female labor supply, human capital, and welfare reform," *Econometrica*, 84, 1705–1753.

Blundell, R., M. C. Dias, C. Meghir, and J. Reenen (2004): "Evaluating the employment impact of a mandatory job search program," *Journal of the European economic association*, 2, 569–606.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016b): "Consumption Inequality and Family Labor Supply," *American Economic Review*, 106, 387–435.

Burtless, G. and J. A. Hausman (1978): "The effect of taxation on labor supply: Evaluating the Gary negative income tax experiment," *Journal of political Economy*, 86, 1103–1130.

Caliendo, M. and S. Kopeinig (2008): "Some practical guidance for the implementation of propensity score matching," *Journal of economic surveys*, 22, 31–72.

Card, D. and A. B. Krueger (1994): "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 84, 772–793.

Connelly, R. (1992): "The effect of child care costs on married women's labor force participation," *The review of Economics and Statistics*, 83–90.

Connelly, R. and J. Kimmel (2003): "Marital status and full–time/part–time work status in child care choices," *Applied Economics*, 35, 761–777.

Cristia, J. P. (2008): "The Effect of a First Child on Female Labor Supply Evidence from Women Seeking Fertility Services," *Journal of Human Resources*, 43, 487–510.

Dehejia, R. H. and S. Wahba (1999): "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs," *Journal of the American statistical Association*, 94, 1053–1062.

——— (2002): "Propensity score-matching methods for nonexperimental causal studies," *Review of Economics and statistics*, 84, 151–161.

Ghez, G. and G. S. Becker (1975): *The Allocation of Time and Goods over the Life Cycle*, NBER.

Graham, J. W. and A. H. Beller (1989): "The effect of child support payments on the labor supply of female family heads: An econometric analysis," *Journal of Human Resources*, 664–688.

Heckman, J. J. (1974): "Effects of child-care programs on women's work effort," *Journal of Political Economy*, 82, S136–S163.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997): "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme," *The review of economic studies*, 64, 605–654.

Houngbedji, K. (2016): "ABSDID: Stata module to estimate treatment effect with Abadie semiparametric DID estimator," Statistical Software Components, Boston College Department of Economics.

Hoynes, H. W. et al. (1996): "Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation under AFDC-UP," *Econometrica*, 64, 295–332.

Imai, K. and M. Ratkovic (2014): "Covariate balancing propensity score," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76, 243–263.

Imbens, G. W. and J. M. Wooldridge (2009): "Recent developments in the econometrics of program evaluation," *Journal of economic literature*, 47, 5–86.

Jacobsen, J. P., J. W. Pearce III, and J. L. Rosenbloom (1999): "The effects of childbearing on married women's labor supply and earnings: using twin births as a natural experiment," *Journal of Human Resources*, 449–474.

Kimmel, J. (1998): "Child care costs as a barrier to employment for single and married mothers," *Review of Economics and Statistics*, 80, 287–299.

Koebel, K. and T. Schirle (2016): "The Differential Impact of Universal Child Benefits on the Labour Supply of Married and Single Mothers," *Canadian Public Policy*, 42, 49–64.

Lefebvre, P. and P. Merrigan (2008): "Child-care policy and the labor supply of mothers with young children: A natural experiment from Canada," *Journal of Labor Economics*, 26, 519–548.

Leuven, E. and B. Sianesi (2003): "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing," Statistical Software Components, Boston College Department of Economics.

Luna, L. G. (2011): "The effects of a universal child benefit," Tech. rep.

Moffitt, R. (1983): "An economic model of welfare stigma," *The American Economic Review*, 73, 1023–1035.

——— (1990): "The Econometrics of Kinked Budget Constraints," *The Journal of Economic Perspectives*, 4, 119–139.

Moffitt, R. A. (2002): "Welfare programs and labor supply," *Handbook of public economics*, 4, 2393–2430.

Mroz, T. A. (1987): "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions," *Econometrica: Journal of the Econometric Society*, 765–799.

Nakamura, A. and M. Nakamura (1992): "The econometrics of female labor supply and children," *Econometric Reviews*, 11, 1–71.

Ribar, D. C. (1992): "Child care and the labor supply of married women: Reduced form evidence," *Journal of human resources*, 134–165.

——— (1995): "A structural model of child care and the labor supply of married women," *Journal of Labor Economics*, 13, 558–597.

Rosenbaum, P. R. and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Rosenzweig, M. R. and T. P. Schultz (1985): "The Supply and Demand of Births. And their Life-Cycle Consequences," *American Economic Review*, 75, 992–1015.

Spencer, D. (2003): "Love's labor's lost? the disutility of work and work avoidance in the economic analysis of labor supply," *Review of Social Economy*, 61, 235–250.

Zhao, Z. (2004): "Using matching to estimate treatment effects: data requirements, matching metrics, and Monte Carlo evidence," *Review of economics and statistics*, 86, 91–107.

Appendix

A Estimates using data for both 2016Q2 and 2016Q3

Table 8: Effects on mother's who worked in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	0.005	0.003	0.006	0.010	0.004	0.005
	(0.582)	(0.764)	(0.548)	(0.622)	(0.748)	(0.635)
N= 1664, N treated = 820, Specification test for CBPS p-val = .001						
youngest child younger than 5						
τ	-0.021	-0.032	-0.021	-0.032	-0.026	-0.026
	(0.161)	(0.069)	(0.131)	(0.541)	(0.144)	(0.145)
N= 702, N treated = 408, Specification test for CBPS p-val = .872						
youngest child between 6 and 11						
τ	0.027	0.017	0.020	0.057*	0.016	0.011
	(0.074)	(0.278)	(0.187)	(0.029)	(0.410)	(0.573)
N= 575, N treated = 332, Specification test for CBPS p-val = .899						
youngest child older than 11						
τ	0.039***	0.039**	0.036	0.033	0.045*	0.030
	(0.000)	(0.004)	(0.186)	(0.359)	(0.025)	(0.104)
N= 387, N treated = 80, Specification test for CBPS p-val = .904						
Hours worked						
all households						
τ	1.437	1.348	1.128	1.310	0.947	1.350
	(0.054)	(0.101)	(0.225)	(0.122)	(0.303)	(0.127)
N= 1664, N treated = 820, Specification test for CBPS p-val = .001						
youngest child younger than 5						
τ	0.927	0.563	0.228	0.719	0.767	0.861
	(0.425)	(0.679)	(0.857)	(0.561)	(0.587)	(0.549)
N= 702, N treated = 408, Specification test for CBPS p-val = .872						
youngest child between 6 and 11						
τ	0.757	0.085	-0.320	-0.493	-0.228	-0.375
	(0.534)	(0.948)	(0.804)	(0.692)	(0.875)	(0.795)
N= 575, N treated = 332, Specification test for CBPS p-val = .899						
youngest child older than 11						
τ	5.641**	5.371**	7.599	6.152**	7.683***	5.785**
	(0.002)	(0.004)	(0.091)	(0.004)	(0.001)	(0.008)
N= 387, N treated = 80, Specification test for CBPS p-val = .904						

We control for the following set of covariates: (1), (2), (3), (4), (5) and labor status of the father – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Effects on father's who worked in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	0.011	0.011	0.011	0.015	0.010	0.013
	(0.059)	(0.084)	(0.251)	(0.445)	(0.198)	(0.087)
N= 2355, N treated = 1255, Specification test for CBPS p-val = 0						
youngest child younger than 5						
τ	0.006	0.011	0.019	-0.078	0.014	0.007
	(0.492)	(0.242)	(0.330)	(0.713)	(0.223)	(0.531)
N= 1214, N treated = 754, Specification test for CBPS p-val = .251						
youngest child between 7 and 11						
τ	0.024^{*}	0.017	0.004	0.049^{*}	0.008	0.010
	(0.040)	(0.122)	(0.593)	(0.040)	(0.556)	(0.503)
N= 704, N treated = 406, Specification test for CBPS p-val = .344						
youngest child older than 12						
τ	0.020^{**}	0.011	0.010	0.010	0.015	0.015
	(0.008)	(0.198)	(0.321)	(0.769)	(0.311)	(0.246)
N= 437, N treated = 95, Specification test for CBPS p-val = .801						
Hours worked						
all households						
τ	1.218	1.408^{*}	1.728^{*}	1.737^{*}	1.690^{*}	1.856^{*}
	(0.056)	(0.038)	(0.041)	(0.019)	(0.035)	(0.017)
N= 2355, N treated = 1255, Specification test for CBPS p-val = 0						
youngest child younger than 5						
τ	1.764	2.097^{*}	3.215^{*}	3.512	2.590^{*}	2.152
	(0.051)	(0.034)	(0.014)	(0.135)	(0.024)	(0.054)
N= 1214, N treated = 754, Specification test for CBPS p-val = .251						
youngest child between 6 and 11						
τ	-0.299	0.020	-0.639	-0.812	-0.312	-0.741
	(0.796)	(0.987)	(0.644)	(0.515)	(0.822)	(0.593)
N= 704, N treated = 406, Specification test for CBPS p-val = .344						
youngest child older than 11						
τ	3.743^{*}	2.612	2.720	2.606	3.167	2.662
	(0.033)	(0.158)	(0.332)	(0.172)	(0.174)	(0.207)
N= 437, N treated = 95, Specification test for CBPS p-val = .801						

We control for the following set of covariates: (1), (2), (3), (4), (5) and labor status of the mother – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Effects on mother's who didn't work in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	0.007	0.027	0.007	0.010	0.030	0.008
	(0.789)	(0.345)	(0.806)	(0.722)	(0.427)	(0.825)
N= 906, N treated = 528, Specification test for CBPS p-val = .001						
youngest child younger than 6						
τ	-0.048	0.038	0.028	0.042	0.025	0.027
	(0.207)	(0.328)	(0.403)	(0.196)	(0.621)	(0.598)
N= 599, N treated = 396, Specification test for CBPS p-val = .355						
youngest child older than 5						
τ	0.018	0.027	0.049	0.050	0.050	0.060
	(0.616)	(0.454)	(0.212)	(0.124)	(0.332)	(0.223)
N= 307, N treated = 132, Specification test for CBPS p-val = .778						
Hours worked						
all households						
τ	0.343	1.217	0.582	0.723	1.309	0.683
	(0.723)	(0.246)	(0.563)	(0.471)	(0.322)	(0.596)
N= 906, N treated = 528, Specification test for CBPS p-val = .001						
youngest child younger than 6						
τ	-1.479	1.455	1.051	1.439	0.991	0.987
	(0.287)	(0.326)	(0.391)	(0.260)	(0.590)	(0.591)
N= 599, N treated = 396, Specification test for CBPS p-val = .355						
youngest child older than 5						
τ	0.835	0.859	1.920	1.800	1.499	1.954
	(0.513)	(0.481)	(0.170)	(0.115)	(0.369)	(0.248)
N= 307, N treated = 132, Specification test for CBPS p-val = .661						

We control for the following set of covariates: (1), (2), (4), (5) and labor status of the father – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Effects on father's who did not work in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	0.061	0.008	0.058	-0.042	-0.010	-0.012
	(0.317)	(0.910)	(0.487)	(0.712)	(0.919)	(0.871)
N= 215, N treated = 93, Specification test for CBPS p-val = .508						
Hours worked						
all households						
τ	2.426	0.440	2.398	-1.272	-0.425	-0.105
	(0.343)	(0.881)	(0.512)	(0.788)	(0.918)	(0.974)
N= 215, N treated = 93, Specification test for CBPS p-val = .508						

We control for the following set of covariates: (1), (2), (3), (4) and labor status of the mother – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Effects on lone mothers who worked in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	0.004	0.025	0.085	0.039	0.034	0.031
	(0.870)	(0.267)	(0.261)	(0.110)	(0.083)	(0.114)
N= 903, N treated = 59, Specification test for CBPS p-val = .645						
Hours worked						
all households						
τ	0.906	2.013	3.243	1.759	2.855^{*}	2.692^{*}
	(0.538)	(0.157)	(0.209)	(0.257)	(0.020)	(0.028)
N= 903, N treated = 59, Specification test for CBPS p-val = .645						

We control for the following set of covariates: (1), (2), (3) – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Effects on lone mothers who did not work in T_0

	OLS		Abadie		Heckman	
	raw	with x	cbps	logit	cbps	logit
Probability of working in T_1						
all households						
τ	-0.030	0.058	0.044	0.058	0.064	0.063
	(0.627)	(0.344)	(0.574)	(0.319)	(0.434)	(0.406)
N= 419, N treated = 41, Specification test for CBPS p-val = .256						
Hours worked						
all households						
τ	-0.140	2.503	2.477	2.562	2.658	2.734
	(0.953)	(0.278)	(0.443)	(0.249)	(0.380)	(0.335)
N= 419, N treated = 41, Specification test for CBPS p-val = .256						

We control for the following set of covariates: (1), (2) – see table 1. p-values in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$