Childcare subsidies and labour supply: evidence from a large Dutch reform

Leon Bettendorf* Egbert Jongen† Paul Muller‡

January 31, 2012

Abstract

Over the period 2005-2009 the Dutch government cut the parental fee for formal childcare in half and extended subsidies to so-called guestparent care. We estimate the labour supply effect of this reform with a difference-in-differences strategy, using parents with older children as a control group. We find that the reform had a modest impact on labour supply. Furthermore, the effects are an upper bound since there was also an increase in an EITC for the same treatment group over the same period. The joint reform increased the participation rate of young mothers by 2.5%-points. We find no effect on the participation rate of young fathers. Average hours worked by young mothers increased by 1.1 hours per week. Average hours worked by young fathers decreased by 0.4 hours per week. A number of robustness checks suggest that these results are robust.

JEL codes: C21, H40, J13, J22

Keywords: Childcare subsidies, labour participation, hours worked, difference-in-differences

*CPB Netherlands Bureau for Economic Policy Analysis. E-mail: L.J.H.Bettendorf@cpb.nl.

†CPB Netherlands Bureau for Economic Policy Analysis. E-mail: E.L.W.Jongen@cpb.nl.

‡VU University Amsterdam and Tinbergen Institute. Corresponding author. VU University Amsterdam, Faculty of Economics and Business Administration, De Boelelaan 1105, 1081 HV Amsterdam. Tel.: +31-20-5986582. E-mail: p.muller@vu.nl. We have benefitted from comments and suggestions by Alberto Abadie, Henk-Wim de Boer, Cindy Elsnerus, Michiel van Goor, Pierre Koning and colleagues at VU University Amsterdam and CPB Netherlands Bureau for Economic Policy Analysis. Remaining errors are our own. We are grateful to the Dutch Ministry of Social Affairs and Employment for funding this research.

1 Introduction

Many countries seek to increase the labour participation of young mothers. Policymakers often point to the Scandinavian countries, where high public spending on childcare subsidies goes hand in hand with high participation rates of mothers. Indeed, several other countries have recently adopted part of the Scandinavian model by providing generous childcare subsidies to young parents (*e.g.* Canada, the Netherlands and the US) or are considering to do so (*e.g.* Germany). Will this bring labour participation up to Scandinavian levels? Unfortunately, correlation is not causation. Other factors may be driving the correlation. A fruitful way to learn about the causal effect of childcare subsidies on labour supply is to study so-called natural experiments. By comparing the labour supply behaviour of a treatment group before and after a policy reform with the labour supply behaviour of a control group before and after the reform we can isolate the effect of the reform.

In this paper we study the causal effect of childcare subsidies on labour supply by means of a large natural experiment in the Netherlands. Specifically, over the period 2005-2009 childcare subsidies became much more generous, with the most significant changes in 2006 and 2007. The average parental fee for formal childcare for children up to 12 years old was cut in half, and subsidies were extended to so-called guestparent care. The effects we find provide an upper bound of the effects of these reforms on labour supply as there was also an (less pronounced) increase in the earned income tax credit (EITC) targeted at the same parents. We estimate the joint effect of these reforms using microdata from the Labour Force Survey from Statistics Netherlands for the period 1995-2009, employing a difference-in-differences (DD) strategy. We estimate the effect on the participation rate and hours worked.

Since 2005, childcare subsidies have become much more generous in the Netherlands. First, with the introduction of the *Wet kinderopvang* (Law on childcare) in 2005, subsidies were given to parents rather than childcare institutions. Effectively, this reduced the parental fee for parents that could not find a subsidized childcare place prior to 2005. In 2006 and 2007, the government then substantially increased the subsidy rate and started paying the employers' contribution for childcare (which reduced the parental fee for parents that did not get a (voluntary) contribution from their employer for childcare costs). The cumulative effect of these changes was to reduce the average parental fee from 37% to 18% of the full price. Furthermore, with the introduction of the *Wet kinderopvang* coverage of the subsidy was extended to so-called guestparent care. This is small scale care at the home of the 'guestparent' or at the home of the children. As a result, it became interesting for many parent-guestparent combinations to formalize the informal care and receive the generous subsidy from the government. Due to the drop in the parental fee for centre based care, and the extension to guestparent care, public spending on childcare skyrocketed from 1 billion euro in 2004 to 3 billion euro (.5% of GDP) in 2009 (Ministry of Finance, 2010). Over the same period that childcare subsidies became much more generous, the government also increased the targeted EITC for the same parents.[1,2] Budgettary outlays of this EITC rose from close to 750 million euro in 2004 to close to 975 million euro in 2008 and 1.3 billion euro in 2009 (Ministry of Finance, 2010). Since both policies target the same treatment group, we can only determine the labour supply effect of the joint reform.

We employ a DD strategy to identify the effect of the joint reform on the labour supply of the treatment group. The treatment group consists of parents aged 20 to 50 years old with a youngest child aged up to 12 years old. As a control group we use parents aged 20 to 50 years old with a youngest child 12 to 17 years old. We convincingly show that this is a valid control group because the trends in participation and average hours worked of both groups are very similar before the reform. Unfortunately, we do not have linked individual data on the use of childcare and labour supply. Hence, we estimate an intention-to-treat effect.

Our main findings are as follows. First, we find that the reform increased the participation rate of women in the treatment group by a modest 2.5%-points (3.3%). We find no effect on the participation rate of men. Second, the reform increased the average number of hours worked per week by women in the treatment group by 1.1 hours per week (6.6%). The average number of hours worked per week by men in the treatment group dropped by 0.4 hours. This may be the result of an income effect due to higher childcare subsidies or a higher net wage for their partners. Third, we do not find an effect for lower educated women. This can be explained by the

[1]The *Combinatiekorting* (Combination credit) and the *Aanvullende combinatiekorting* (Additional combination credit). The names refer to the combination of labour participation and care for children.

[2]Furthermore, both the reform in childcare subsidies and the increase in the targeted EITC were mostly targeted middle and high income earners.

fact that for low incomes the subsidies hardly changed in the reform. Fourth, we find larger effects for single mothers than for mothers in couples. These results are robust to a number of sensitivity checks.

There is an extensive literature that considers the relation between parental labour supply and the cost of childcare using cross-sectional data. An in depth overview is given in Blau and Currie (2006). They show that estimated (childcare) price elasticities of female labour participation range from 0.06 to -3.60. They argue that only a small part of this variation is due to differences in the composition of the sample or different data sources. Most of the variation seems to be due to identification problems related to the endogeneity of the explanatory variables. Unobserved characteristics are likely to influence both the cost of childcare (which *e.g.* depends on income) and the labour supply decision. To solve this problem, exogenous variation in the cost of childcare is needed. Therefore, the focus has shifted to quasi-experimental methods that use policy changes as exogenous variation in prices. As a result, there is a small but growing body of literature that studies the impact of changes in childcare costs resulting from policy reforms on labour supply.

Lundin et al. (2008) study the effect of changes in the price of childcare on female labour supply in Sweden, using regional variation in the effect of a policy reform that put a cap on prices for parents. The resulting average drop in the hourly childcare price for parents was 55%.[3] Using difference-in-differences they estimate the effect on labour supply to be close to zero. They argue that they find an effect close to zero because the childcare system was already well-developed and highly subsidized before the reform.

Havnes and Mogstad (2011a) explore the effect of the introduction of universally accesible and subsidized childcare in Norway in the late 1970s. They apply a difference-in-differences approach using time and regional variation in the staged implementation of the policy to identify the effect of the price decrease on labour supply. They find that the reform increased the participation rate of mothers by just 1.1%-point (Havnes and Mogstad, 2011a, Table 2). Most of the increase in the use of formal childcare (94%) comes from substitution of informal childcare.

Gelbach (2002) estimates the effect of free provision of public schooling for young children on maternal labour supply in the US. He finds that public school enrollment increased participation of single mothers by 4%-points and average working hours by

[3]See Lundin et al. (2008, Table 1).

2.2 hours per week. For the same reform, Cascio (2009) also estimates sizeable effects for single mothers, but finds no effect for mothers in couples. Fitzpatrick (2010) also estimates the effect of universal availability of prekindergarten programs on overall preschool enrollment and maternal labor supply in the US, using a regression-discontinuity design. She finds no effect on the labour supply of mothers. She argues that the difference with Gelbach (2002) is due to the later period she considers, where the participation of mothers was already high to begin with.

Finally, Lefebvre and Merrigan (2008) study a large natural experiment in Canada. They compare labour supply outcomes of mothers in Québec, where the cost of childcare was substantially decreased (to 5$ a day), to outcomes of mothers in other regions in Canada. They find an effect on participation of mothers of 7.3%-points. Hours worked increase by 133 hours per year. Baker et al. (2008) study the same reform, using a different dataset, and find an increase in the employment rate of women in couples of 7.7%-points.

We believe our paper is an interesting extension to this small but growing body of literature. First of all, to the best of our knowledge, this is the first quasi-experimental study of the impact of childcare costs in a continental European country. This makes our results particularly informative for other continental European countries that are considering a substantial increase in childcare subsidies. Furthermore, we consider a very recent reform, which also contributes to the relevance of our results. Finally, we have ten years of prereform data. This allows us to formally test whether or not the treatment and control group share the same prereform trend. This is a crucial assumption in DD analyses. Some authors, *e.g.* Cascio (2009) and Havnes and Mogstad (2011a), have suggested that the large effects of the Canadian studies may have been driven in part by different trends in the treatment and control states. With our data we can show that the treatment and control group have the same prereform trend, and find similar small effects as Lundin et al. (2008), Cascio (2009) and Havnes and Mogstad (2011a).

The structure of the paper is as follows. Section 2 describes the main aspects of the reform we exploit in the empirical analysis. Our empirical methodology, including robustness checks, is discussed in Section 3. In Section 4 we present our dataset and some descriptive statistics. Section 5 presents the estimation results for participation and hours worked. Section 6 concludes and discusses directions for future research.

2 The natural experiment

What we define as the policy reform is actually a series of policy changes that took place between 2005 and 2009, where the most important changes took place in 2006 and 2007 when the parental fee was cut in half. Below we give a short historical account of the policy changes, and indicate their significance.

Before the introduction of the *Wet kinderopvang* (Law on childcare) in 2005 there were three types of formal childcare places. In 2004, places subsidized by employers made up 67% of daycare places and 58% of out-of-school care places.[4] Places subsidized (directly) by local and central governments made up 9% of daycare places and 18% of out-of-school care places. The remainder, 24% of daycare places and 25% of out-of-school care places, were places that were not directly subsidized, but part of the formal childcare expenditures was tax deductible.

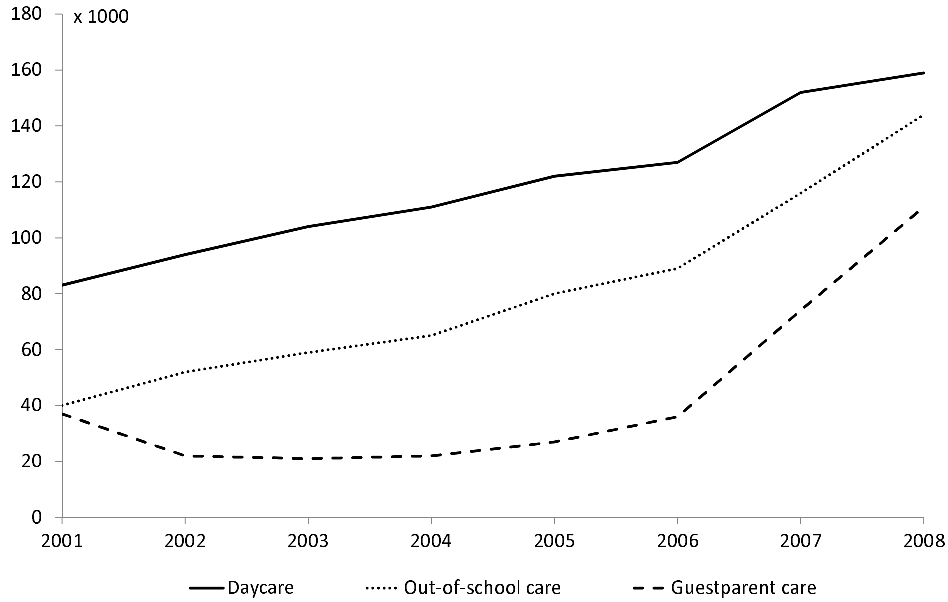
The introduction of the *Wet kinderopvang* in 2005 unified the subsidies for childcare places. From 2005 onwards, all formal places qualified for the same subsidy from the government. This increased the subsidy somewhat for parents with children going to a place that was not directly subsidized before 2005. With the introduction of the *Wet kinderopvang* so-called guestparent care also became eligible for the subsidy. This is small scale care at the home of the guestparent or the children. But the unification of the way places were subsidised and the extension to guestparent care had only a minor effect on public spending on formal childcare, see Table 1. Indeed, presumably because the subsidy was actually reduced somewhat for the highest incomes[5], public spending actually fell slightly from 2004 to 2005. When we look at the number of childcare places, see Figure 1, there is hardly an effect on the growth rate in childcare places in 2005 relative to the preceding period. Hence, in our empirical analysis we do not expect to find large labour supply effects for 2005.

In 2006 and 2007 the subsidy rate was increased drastically. Figure 2 shows the resulting change in the parental contribution rate (for the first child) between 2005 and 2007. First, note that the parental fee depends on the income of the household. In all years, households with the lowest income receive the highest subsidy (up to 96% of the full price). For the lowest income households the subsidy rate hardly changed between 2005 and 2007. For the middle income households the subsidy rate

[4]Source: Statistics Netherlands (statline.cbs.nl).

[5]See Plantenga et al. (2005).

Figure 1: Child care places



Source: Statistics Netherlands (statline.cbs.nl).

Table 1: Public spending on childcare and tax credits for working parents (million of euro)

Year	2002	2003	2004	2005	2006	2007	2008	2009
Childcare subsidies	725	755	1028	1001	1343	2058	2825	3034
Tax credits working parents[a]	415	465	740	830	865	980	950	1250

Source: Ministry of Finance (2010) and own calculations (imputation of employers' contribution for childcare up to 2007 with data from the Ministry of Social Affairs and Employment (personal communication)).

[a] Public spending on *Combinatiekorting* and *Aanvullende combinatiekorting*.

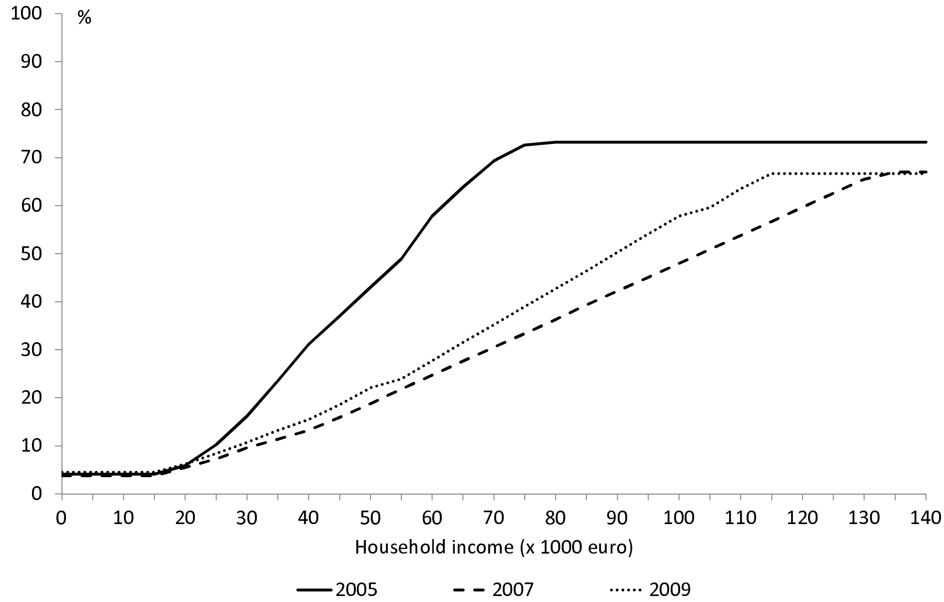
went up by 20 to 40%-points, whereas the increase in the subsidy for the highest income households was somewhat smaller than for middle income households. On average, the parental cost share in the full price dropped from 37% in 2005 to 18% in 2007.[6] Next to the drop in parental fees, from 2007 onwards schools were obliged to act as an intermediary for parents and childcare institutions to arrange out-of-school care. Figure 2 further shows that at the end of our dataperiod, in 2009, the parental fee was raised again somewhat, but relative to 2005 there was still a large drop for most middle and high income households. In the empirical analysis we also estimate the effect of the policy changes on labour participation by level of education. It is important to note that, on average, low educated women have a low household income where there was hardly any change in the parental contibution rate.

Figure 1 shows that the dramatic drop in the contribution rate in 2006 and 2007 spurred the growth in the use formal childcare in 2006 and beyond. Also, guestparent care took a high flight. As a result, public spending on formal childcare skyrocketed, increasing from 1 billion euro in 2005 to 3 billion euro in 2009, see Table 1.

In DD analyses it is crucial to consider other policies that might influence the outcome variables for the treatment or control group (differently). We carefully examined various changes in taxes and subsidies and found that there were no substantial changes in taxes or subsidies targeted at the treatment or control group, apart from one. The only complication comes from changes in the EITCs for working parents with a youngest child up to 12 years old, the *Combinatiekorting* (Combination credit) and the *Aanvullende combinatiekorting* (Additional combination credit). These EITCs are also targeted exclusively at our treatment group. Figure 3 shows the change in the sum of the *Combinatiekorting* and *Aanvullende combinatiekorting* for secondary earners and single parents (mostly women) over the period 2001-2009. Table 1 gives the changes in public expenditures on these credits. Between 2001 and 2004 these credits increased from 138 to 514 euro, and public expenditures increased from 415 to 740 million euros between 2002 and 2004. Between 2004 and 2008 there was an almost similar rise in the individual subsidy from 514 euro to 858 euro, and in 2009 there was another increase for secondary earners and single parents with

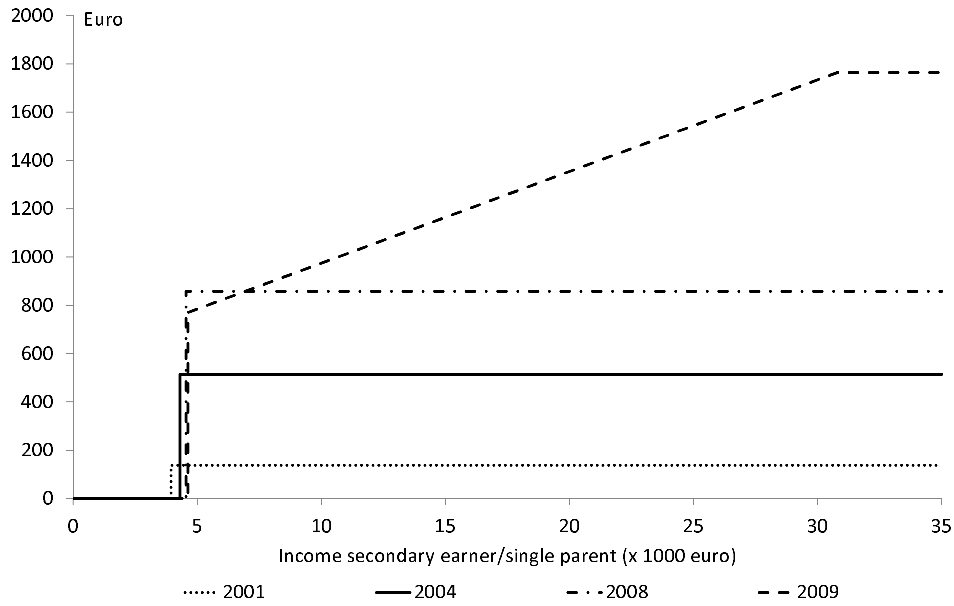
[6] Source: Tax Office data provided by the Ministry of Social Affairs and Employment (personal communication).

Figure 2: Parental contribution rate for the first child



Source: own calculations using publicly available subsidy tables.

Figure 3: EITC secondary earners and single parents



Source: Tax Office.

substantial earnings. In 2009, the maximum credit was 1765 euro, where the maximum was reached at 30,803 euro of gross individual income (in 2009 the minimum wage of a fulltime worker was 16,776 euro). Since these credits target the same group as childcare subsidies, we can only determine the joint effect of the changes in childcare subsidies and these credits. Note however, that over the period 2005-2008 there was a rise in the tax credits quite similar to the period 2001-2004. Hence, part of the rise in the targeted EITCs is captured in the prereform trend. Furthermore, note that the change in the credits for working parents in 2009 was mostly targeted at middle and high income earners, like the change in childcare subsidies.

3 Methodology

We estimate the effect of the policy change on labour participation using a difference-in-differences (DD) approach (see *e.g.* Angrist and Pischke, 2009; Blundell and Costa Dias, 2009; Imbens and Wooldridge, 2009). This method estimates the effect of a reform by comparing the change in outcomes of the treatment group before and after the reform, using the change in outcomes of a valid control group to control for other changes over time. Our treatment group consists of parents influenced by the change in child care cost, which in the Netherlands are parents with a youngest child up to 12 years old.

The control group should not be influenced by the policy change, but should be comparable to the treatment group. We use parents with a youngest child living at home aged between 12 and 18. These parents are not eligible to care subsidies but are otherwise quite similar to parents in the treatment group. The common trend assumption requires that in the absence of the policy reform the change in outcome of the treatment group would have been equal to the change in the outcome of the control group. By definition this assumption cannot be tested. However, we have data on ten years before the reform, so we can formally test whether this assumption holds in the pre-reform period. In Section 5 we estimate the pre-reform trend for both groups and we also estimate placebo treatment effects in periods before the reform.

We also need the policy change to be exogenous for an unbiased estimate of the treatment effect. A problem would arise if, for example, the government was anticipating certain behavior when deciding to pass the new law. Also if parents expected the policy change in advance and adapted their behavior before 2005 this would create a problem for estimating the treatment effect. In our case, both issues are unlikely. First, inspection of the data shows that there is no change in the long-term trend in the years before the reform that could have induced the policy changes from 2005 onwards. Second, the most important policy change is the reduction of the parental fee in 2006-2007. Since this reduction was not included in the law on childcare in 2005 parents were unable to anticipate these changes before 2005. Both assumptions are supported by the outcomes of the placebo tests that we report in Section 5.

Finally, there should not be a change in the composition of the treatment and control group not controlled for by observable characteristics. To assess whether this is the case, we compare characteristics of the treatment and control group before and after the policy change in Section 4 and find that the composition of both groups in terms of observable characteristics is quite stable.

To estimate the treatment effect on participation, we regress participation status on a year fixed effect (α_t), a group fixed effect (γ_g), individual characteristics (X_i) and a set of treatment dummies for each year after the reform (D_{gs}):

$$y_{igt} = \alpha_t + \gamma_g + X_i\beta + \sum_{s=2005}^{2009} \delta_s D_{gs} + \epsilon_{igt}. \tag{1}$$

D_{gs} is a set of dummies equal to one if individual i belongs to the treatment group in year s. The common trend is captured by the year fixed effects, while the constant difference in participation between the treatment and control group is captured by the group effect. Individual characteristics are included to control for possible changes in the composition of the groups over time, although we will show that these changes are small. In equation (1) we allow the treatment effect to be different in each year after the policy change. We impose restrictions on the effects if the estimated coefficients are not significantly different from each other. To correct for potential heteroskedasticity we report robust standard errors (Bertrand et al., 2004). Since we are interested in the population average effect, we weight all regressions using sample weights (Cameron and Trivedi, 2005).

Participation is a discrete variable, so equation (1) is a linear probability model. This model may suffer from the problem that the estimated probabilities could be outside the 0-1 interval. After estimation we check whether this is the case and also estimate a probit model as a robustness check.

Next, we estimate the effect on hours worked per week. We follow Angrist and Pischke (2009) and estimate a linear model with the same sample of individuals as for participation. So we estimate equation (1) with y denoting the number of hours worked, potentially zero. Hours worked has a mass point at zero. Therefore we also estimate a tobit model as a robustness check. Finally, we also estimate the Heckman selection model (Heckman, 1979) which allows for different effects of the covariates on the participation decision and the hours decision. The Heckman model ideally involves an exclusion restriction to identify the parameters. In our case it is difficult to find a variable that influences the participation decision but not the hours decision. So we use the same set of explanatory variables in both equations and identify the parameters from the non-linearity in the correction for selection.

4 Data

We use data from the Dutch Labour Force Survey (*Enquete Beroepsbevolking*) from Statistics Netherlands. The survey is performed annually and includes around 80,000 individuals per year. We have repeated cross-sections for the period 1995-2009. Since the reform started in 2005, we have a long data set preceding the policy change, so we have sufficient information to check the common trend assumption crucial in DD analyses. The survey includes labour supply information (participation and average weekly hours worked), individual characteristics (age, gender, education level, native/immigrant, married couple/unmarried couple/single) and household characteristics (number of children, age of the children).[7] We use sample weights in the regression to estimate effects that are representative for the Dutch working age population.

From this dataset we select our treatment group of parents with a youngest child aged up to 12 years old. Furthermore we restrict the analyses to parents aged

[7]For each year we restrict our sample to individuals that were interviewed in person. Apart from these, there were around three follow-up interviews of the same individuals within one year by telephone. Since these are considered less reliable and most outcomes do not change a lot within a year (Statistics Netherlands, 2009, see), we decide to only use the in-person interviews. Unfortunately, we could not make this distinction for 2009, so we have about four times more observations in 2009 than in the others years.

Table 2: Characteristics women in treatment and control group before and after the policy reform

	Treatment Group		Control Group	
	2000-2004	2005-2009	2000-2004	2005-2009
Participation	0.67	0.75	0.73	0.77
Worked Hours	14.53	17.07	17.27	18.50
Age	35.74	36.48	43.92	44.69
High educated	0.24	0.30	0.18	0.20
Low educated	0.30	0.23	0.41	0.33
Single	0.10	0.10	0.15	0.16
Immigrant	0.23	0.24	0.19	0.20
Household size	3.90	3.89	3.83	3.81
Age youngest child	4.41	4.59	14.33	14.33
Observations	60379	82151	16706	27272

Values are means weighted with sample weights. Source: Labour force survey (Statistics Netherlands).

between 20 and 50 years old. This gives us 202,106 observations for mothers and 179,548 observations for fathers. As a control group we select all parents with a youngest child aged between 12 and 18. Restricting the control group to parents aged between 20 and 50 years old gives 61,127 observations for women and 45,129 observations for men.

As discussed in Section 3 it is important that the composition of the groups remains the same before and after the policy change. To investigate whether this is the case for women we show characteristics of the treatment and control group for 2000-2004 and 2005-2009 in Table 2. The table shows the outcome variables participation and hours worked per week and the explanatory variables age, education (in categories low, middle and high educated), a dummy for being single, a dummy for being an immigrant, the size of the household (in the regression we only include a dummy for large households) and the age of the youngest child (not included in the regression).

Differences in most characteristics are small, though of course there is a substantial difference in the age of the parent and the age of the youngest child. Mothers in the control group are slightly more likely to be single. Mothers in the control group are also somewhat lower educated. The share of immigrants is slightly higher in the treatment group, which could be explained by the higher fertility rate of immigrants. More importantly, the changes in characteristics within each group are small. In both groups only the average age and education level increase slightly. So as far as any changes occur, they are quite similar in control and treatment group and we conclude that changes in composition of the groups are unlikely to harm the estimation results.

4.1 Participation

In the DD method we compare the outcomes of the treatment and control group over time. In Figure 4 we plot participation rates of women and men for the two groups. The solid vertical line marks the start of the policy reform. For women we see that both the treatment and control group exhibit an upward trend before the policy change, while participation is always higher for the control group. Furthermore, the rate of growth is very similar for the two groups. This indicates that the control group is suitable for the DD analysis for women.

For men we do not observe a trend, participation is rather stable. The participation of the treatment and control are almost identical over the years, hence also for men the control group seems valid.

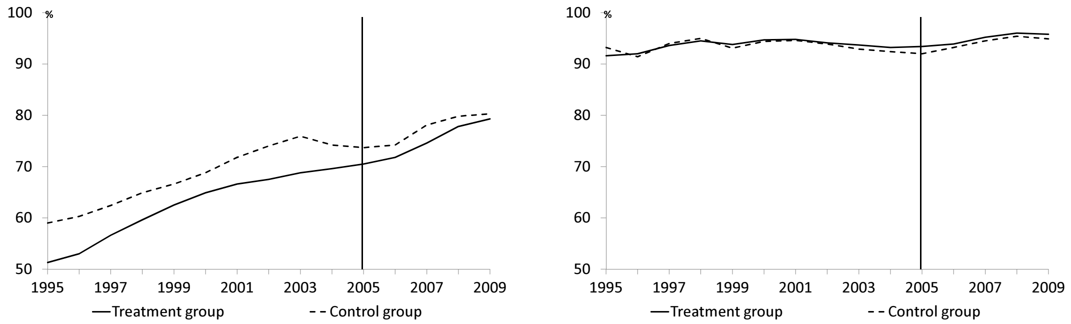
As a more formal check, taking into account changes in the characteristics within the groups, we perform a pooled regression on participation for the treatment and control groups including all individual characteristics and a group fixed effect, for the pre-reform period. We also include a linear trend interacted with the group dummy. In this way we obtain an estimate of the long-term trend for each group, while controlling for the observable characteristics. The estimation results are presented in Table 3. Both for women and men the estimated trends are very close and we do not reject the hypothesis that they are equal.

4.2 Hours worked per week

In Figure 5 we plot the average number of hours worked per week, for women and men. Again we see that there is a clear upward trend for women, both in the treatment and control group. The rate of increase is very similar in both groups before the policy change, such that also for hours worked the control group seems

Figure 4: Labour participation

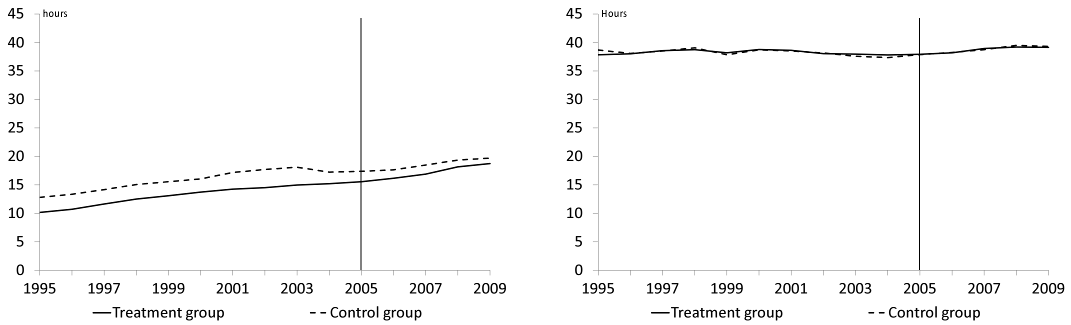
(a) Women (b) Men



Source: Labour Force Survey (Statistics Netherlands)

Figure 5: Hours worked per week

(a) Women (b) Men



Source: Labour Force Survey (Statistics Netherlands)

Table 3: Trendtest participation and hours worked per week

	Participation		Hours worked per week	
	Women	Men	Women	Men
Trend treatment group	0.020***	0.004***	0.390***	-0.029
	(0.000)	(0.000)	(0.020)	(0.022)
Trend control group	0.021***	0.003***	0.446***	-0.085*
	(0.001)	(0.001)	(0.041)	(0.046)
Included years	1995-2004	1995-2004	1997-2004	1997-2004
Observations	153809	133568	121675	104772
P-value test equal trends	0.834	0.180	0.223	0.269

Note: Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Individual characteristics and a group fixed effect are included but not reported.

suitable for the DD analysis. For men the number of hours worked is very stable between 1995 and 2004 and the values for the treatment and control group are almost identical.

We estimate trends for treatment and control group, while controlling for observable characteristics in the same way as for participation. We find that the trends are most similar when data from 1997 onwards is used.[8] In Table 3 the positive trend for women is very similar in the treatment and control group and equality is not rejected. For men we find no significant trend for the treatment group and a small negative trend for the control group that is significant at the 10% level. The difference between the two is not significant. So we conclude that for hours worked the common trend assumption is unlikely to be violated for both women and men using data from 1997 onwards.

5 Estimation results

5.1 Participation rate of women

We first present the estimation results for the effect of the reform on the participation rate of all women in the treatment group, and subsequently we consider the results

[8]Before 1997 the trends of the control and treatment group are slightly different, so we decided to use data from 1997 onwards.

for subgroups.

5.1.1 All women

We estimate equation (1) for participation and show results in Table 4. In column (1) we estimate 5 year-specific treatment effects. We find positive coefficients for each treatment year. In 2005 and 2007, the effect is not significantly different from zero. The estimates of the control variables are all significant and as expected. The treatment group dummy has a negative coefficient. Also lower educated women, immigrants, single women and women in large families have a lower participation rate. High educated women participate more (than middle educated women). The effect of age is non-linear, increasing up to a peak around age 37.

In a formal test we can not reject that the treatment effects per year are equal. We show results with two treatment effects in column (2). We distinguish between the first three years and the last two years, which might be considered the short and medium run effect of the reform. Both estimates are significantly positive at the 1% level. In the first three years the effect is 1.6%-points (which is an increase in the participation rate of 2.3%), while in the last two years it increases to 2.5%-points (3.3%). The effect of the other variables does not change. When estimating two treatment effects, we still cannot reject that they are equal (see the p-value in last line). For completeness, when we estimate one single treatment effect for 2005-2009, we estimate an effect of 0.020 (significant at the 1% level).

As noted in Section 3 the linear probability model may predict values outside the 0-1 interval. We find that only 0.14% of the predicted values are (slightly) larger than 1. Still, as a robustness check we present estimates of a probit model in column (3). We find very similar treatment effects (reported values are average marginal effects). As a final robustness check we estimate the specification in column (2) with data from 2001 onwards in column (4). Again we find similar treatment effects.

The validity of our estimates depends critically on the common trend assumption. To further assess the plausibility of this assumption we estimate placebo treatment effects. Specifically, we estimate a treatment effect for some years before 2005. Since no relevant policy change occured in this period we should not find a significant effect. If we do find an effect this could indicate that the control and treatment group do not have the same pre-reform trend. We divide our dataset in three 5

Table 4: Effect on participation rate of women

	(1) OLS	(2) OLS	(3) Probit	(4) OLS (01-09)
Treatment 2005	0.014			
	(0.009)			
Treatment 2006	0.021**			
	(0.010)			
Treatment 2007	0.014			
	(0.010)			
Treatment 2008	0.023**			
	(0.009)			
Treatment 2009	0.027***			
	(0.005)			
Treatment 05/06/07		0.016***	0.013**	0.023***
		(0.006)	(0.006)	(0.007)
Treatment 08/09		0.025***	0.022***	0.032***
		(0.006)	(0.007)	(0.007)
Group	–0.082***	–0.082***	–0.077***	–0.076***
	(0.004)	(0.004)	(0.004)	(0.005)
Low educated	–0.176***	–0.176***	–0.155***	–0.174***
	(0.003)	(0.003)	(0.002)	(0.003)
High educated	0.105***	0.105***	0.119***	0.087***
	(0.002)	(0.002)	(0.003)	(0.003)
Immigrant	–0.151***	–0.151***	–0.142***	–0.160***
	(0.003)	(0.003)	(0.003)	(0.004)
Single	–0.079***	–0.079***	–0.073***	–0.079***
	(0.004)	(0.004)	(0.003)	(0.004)
Large family	–0.203***	–0.203***	–0.186***	–0.209***
	(0.005)	(0.005)	(0.005)	(0.007)
Age	0.076***	0.076***	0.062***	0.101***
	(0.011)	(0.011)	(0.011)	(0.014)
Age^2 (x1000)	–1.364***	–1.364***	–1.038***	–2.099***
	(0.314)	(0.314)	(0.306)	(0.393)
Age^3(x1000)	0.006**	0.006**	0.004	0.013***
	(0.003)	(0.003)	(0.003)	(0.004)
Observations	263231	263231	263231	170834
P-value test equal treatment effects	0.577	0.235	0.268	0.184

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Year fixed effects are included but not reported. Probit results are average marginal effects.

Table 5: Placebo treatment participation women

	(1)
	OLS
Placebo 00-04	–0.002
	(0.006)
Treatment 05/06/07	0.016**
	(0.007)
Treatment 08/09	0.024***
	(0.007)
P-value test Placebo=Treatment 05/06/07	0.01
P-value test Placebo=Treatment 08/09	0.00
Observations	263231

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Individual characteristics and year fixed effects are included but not reported.

year periods and estimate a placebo treatment effect for 2000-2004. The placebo effect and the two actual treatment effects are reported in Table 5. The placebo treatment effect is not significantly different from zero, while both treatment effects remain almost unchanged compared to the basic specification. Furthermore, we can reject that the placebo coefficient equals the treatment effect in both cases with very low p-values.

5.1.2 Subgroups

We also estimate the effect separately for each level of education, and separately for single women and women in couples. We do this by estimating equation (1) with the particular subsample, thereby allowing differences in all coefficients between subgroups.[9] Since OLS and probit lead to comparable results, only OLS results are reported for the five groups in Table 6. With regard to education, we find no effect

[9]Since the composition of these groups is different, we need to check if the control group is suitable for each subgroup. We estimate trends for each group in the same way as we did for the entire sample. We find that for each subgroup the control group has a comparable pre-reform trend. Results are available on request.

Table 6: Effect on participation rate for subgroups of women

	(1) Low educated	(2) Middle educated	(3) High educated	(4) Single women	(5) Women in couples
Treatment 2005/2006/2007	0.007	0.022**	0.016	0.030	0.013**
	(0.012)	(0.009)	(0.011)	(0.019)	(0.006)
Treatment 2008/2009	–0.003	0.042***	0.038***	0.045**	0.021***
	(0.012)	(0.008)	(0.010)	(0.018)	(0.006)
Observations	77604	122030	63597	26453	236778

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Individual characteristics and year fixed effects are included but not reported.

on low educated women, while the effect on middle and high educated women is somewhat higher than the average effect. An explanation for the absence of the effect on low educated women could be that their subsidy rate did not change much, since low education is correlated with low income. In column (4) and (5) we find that the effect on the participation of single women is higher than for women in couples.

Also for subgroups we perform placebo tests and find that, with the exception of middle educated women, the placebo effect is never significantly different from zero and always significantly different from the treatment effect.[10]

5.2 Participation rate of men

Next, consider the effects of the reform on the participation rate of men. In column (1) in Table 7 we find no effect on participation in any year after the policy change. All other variables have the same sign as for women and all are significant at the 1% level. The highest participation for men is achieved at age 35. In column (2) we estimate only two treatment effects, but both are not significantly different from zero. Also the probit estimation in column (3) and the OLS specification with data from 2001 onwards in column (4) give no significant effect. We also find that the placebo treatment is not significant. We finally check if an effect exists on the

[10]Results available on request.

Table 7: Effect on participation rate of men

	(1) OLS	(2) OLS	(3) Probit	(4) OLS (01-09)
Treatment 2005	0.008			
	(0.008)			
Treatment 2006	0.000			
	(0.007)			
Treatment 2007	0.004			
	(0.007)			
Treatment 2008	0.004			
	(0.006)			
Treatment 2009	0.002			
	(0.004)			
Treatment 05/06/07		0.004	0.003	0.002
		(0.004)	(0.004)	(0.005)
Treatment 08/09		0.003	0.006	0.000
		(0.004)	(0.004)	(0.005)
Group	–0.008***	–0.008***	–0.005***	–0.004
	(0.002)	(0.002)	(0.002)	(0.004)
Low educated	–0.053***	–0.053***	–0.041***	–0.047***
	(0.002)	(0.002)	(0.001)	(0.002)
High educated	0.017***	0.017***	0.024***	0.018***
	(0.001)	(0.001)	(0.002)	(0.001)
Immigrant	–0.134***	–0.134***	–0.091***	–0.119***
	(0.003)	(0.003)	(0.002)	(0.003)
Single	–0.112***	–0.112***	–0.069***	–0.100***
	(0.009)	(0.009)	(0.004)	(0.010)
Large family	–0.057***	–0.057***	–0.035***	–0.061***
	(0.004)	(0.004)	(0.002)	(0.006)
Age	0.068***	0.068***	0.039***	0.050***
	(0.011)	(0.011)	(0.007)	(0.014)
Age^2 (x1000)	–1.547***	–1.547***	–0.849***	–1.160***
	(0.294)	(0.294)	(0.185)	(0.367)
Age^3(x1000)	0.011***	0.011***	0.006***	0.008***
	(0.003)	(0.003)	(0.002)	(0.003)
Observations	224676	224676	224676	143569
P-value test equal treatment effects	0.950	0.832	0.191	0.808

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Probit results are average marginal effects. Year fixed effects are included but not reported.

same set of subgroups as for women, but find no significant estimate for any of the subgroups as well.[11]

5.3 Hours worked per week by women

Next to participation we are also interested in the effect on the average number of hours worked in the treatment group. Again we start with the results for all women, and subsequently consider the effects for subgroups.

5.3.1 All women

As discussed in Section 3 we estimate several specifications of equation (1) with average hours worked per week as the outcome variable. We first include all women in this regression, both working women and non-working women. The results are presented in column (1) in Table 8. We find positive effects in both periods which are significant at the 1% level. In 2005-2007 the estimated effect is an increase of 0.75 hours per week. In 2008-2009 the effect increases to 1.14 hours. Given the average number of worked hours per week for women in 2005-2007 and 2008-2009 (16.2 and 18.4, respectively), these effects are more substantial (6.6% in 2008-2009) than the effects for participation (3.3% in 2008-2009).

Column (2) gives the results for the tobit model, which assumes normally distributed error terms in the latent variable. The reported average (unconditional) marginal effects show that the estimates are very close to the OLS estimates. The tobit model could still be too restrictive. In column (3) we therefore also present estimates of a Heckman selection model. We report the average marginal effect on the entire sample and again find very similar results.

In the last three columns we estimate the same three models, but report the effects on the subsample of working women only. In the case of OLS this implies restricting the sample to workers. For the tobit and Heckman model the estimation does not change, but the average marginal effects are evaluated only for the working sample. The estimated (conditional) effects are somewhat smaller in this case. Depending on the model, the effect in 2005-2007 is between 0.6-0.7 hours, while in 2008-2009 it ranges between 0.9-1.0 hours. The similarity of the results with different models indicates that the results are robust.

[11] Results available on request.

Table 8: Hours worked per week by women

	All women			Working women		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	tobit	Heckman	OLS	tobit	Heckman
Treatment 05/06/07	0.745***	0.818***	0.834***	0.707***	0.602***	0.680***
	(0.198)	(0.200)	(0.182)	(0.177)	(0.147)	(0.148)
Treatment 08/09	1.137***	1.230***	1.217***	0.982***	0.906***	0.991***
	(0.195)	(0.191)	(0.177)	(0.173)	(0.141)	(0.144)
Observations	231097	231097	231097	166201	231097	231097

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Coefficients in columns (2) and (3) are average marginal effects on the observed outcome for the entire sample, coefficients in columns (5) and (6) are average marginal effects on the sample of working women. Individual characteristics and year fixed effects are included but not reported.

Table 9: Placebo treatment hours worked women

	(1)
	OLS (all women)
Placebo 00-04	–0.315
	(0.229)
Treatment 05/06/07	0.548**
	(0.246)
Treatment 08/09	0.940***
	(0.243)
P-value test Placebo=Treatment 05/06/07	0.00
P-value test Placebo=Treatment 08/09	0.00
Observations	231097

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Individual characteristics and year fixed effects are included but not reported.

Table 10: Hours worked per week for subgroups of women (OLS)

	(1) Low educated	(2) Middle educated	(3) High educated	(4) Single women	(5) Women in couples
Treatment 05-07	0.464	0.801^{***}	1.038^{**}	1.128^{*}	0.643^{***}
	(0.345)	(0.301)	(0.421)	(0.623)	(0.207)
Treatment 08-09	0.265	1.422^{***}	1.715^{***}	1.651^{***}	1.042^{***}
	(0.366)	(0.286)	(0.408)	(0.610)	(0.203)
Observations	65020	108152	57925	23945	207152

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Individual characteristics and year fixed effects are included but not reported.

Similar to the analysis of the participation rate, we also estimate a placebo effect on hours worked per week. We add the placebo treatment dummy for 2000-2004 to the unconditional OLS model and report results in Table 9. The placebo treatment effect is not significantly different from zero, while both treatment effects are. Equality of the placebo and the treatment effects can be rejected with a p-value below 0.01.

5.3.2 Subgroups

For subgroups of women we estimate the unconditional OLS model and report estimates in Table 10. The pattern is similar to the results for participation. The effect on low educated women is not significantly different from zero, whereas the effect is larger for high educated women. The effect for single women is larger than for women in couples.

5.4 Hours worked per week by men

We find no impact of the reform on the participation rate of men, but there seems to be a negative effect on their hours worked per week. We first estimate the unconditional effect with the OLS, tobit and Heckman model, see column (1), (2) and (3) in Table 11. In 2005-2007 the unconditional effect is not significantly different from zero (except for the Heckman model), but in 2008-2009 it is significant in all three specifications (although only at the 10% level with OLS and tobit) and ranges from -0.37 to -0.53. We have two potential explanations for this negative effect. First, it may simply be that the income effect of the reform dominates for men. Second, it may be the case that men respond to the increased labour supply of their partners by reducing their hours.

In column (4), (5) and (6) we estimate the conditional effect (the effect on working men). Since participation of men is around 95%, the average marginal effects in the tobit and Heckman model are very similar to the effects on all men. In column (4) we find slightly larger and significant effects in the OLS model if non-working men are dropped from the sample. We conclude that hours worked for men have decreased by 0.4-0.5 hours per week in 2008-2009.

The placebo treatment effect in the unconditional OLS model is estimated in Table 12. The placebo effect is not significantly different from zero. The negative treatment effects are slightly smaller but not significant. As a result we can not reject that the placebo effect is equal to the treatment effect in 2005-2007.

Finally, we estimate the effect on hours worked by men for the five subgroups. In Table 13 we report the coefficients estimated by OLS. We find that the negative effect only appears among middle educated men. For low and high educated men the coefficients are not significantly different from zero. For single men we find no effect, but standard errors are large due to the small sample of single men with young children. For all groups the placebo effect is not significant.

Table 11: Hours worked per week by men

	All men			Working men		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	tobit	Heckman	OLS	tobit	Heckman
Treatment 05/06/07	–0.177	–0.154	–0.350**	–0.348**	–0.157	–0.350**
	(0.239)	(0.249)	(0.163)	(0.163)	(0.255)	(0.163)
Treatment 08/09	–0.396*	–0.374*	–0.526***	–0.525***	–0.383*	–0.526***
	(0.217)	(0.224)	(0.162)	(0.162)	(0.229)	(0.162)
Observations	195880	195880	195880	186125	195880	195880

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Coefficients in columns (2) and (3) are average marginal effects on the observed outcome for the entire sample, coefficients in columns (5) and (6) are average marginal effects on the sample of working women. Individual characteristics and year fixed effects are included but not reported.

Table 12: Placebo treatment hours worked men

	(1)
	OLS (all men)
Placebo 00-04	0.162
	(0.239)
Treatment 05/06/07	–0.077
	(0.281)
Treatment 08/09	–0.297
	(0.263)
P-value test Placebo=Treatment 05/06/07	0.35
P-value test Placebo=Treatment 08/09	0.05
Observations	195880

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Individual characteristics and year fixed effects are included but not reported.

Table 13: Hours worked per week for subgroups of men (OLS)

	(1) Low educated	(2) Middle educated	(3) High educated	(4) Single men	(5) Men in couples
Treatment 05-07	0.289	–0.492	–0.114	–0.098	–0.123
	(0.531)	(0.361)	(0.351)	(1.935)	(0.240)
Treatment 08-09	0.483	–0.778**	–0.467	0.034	–0.398*
	(0.474)	(0.326)	(0.335)	(1.776)	(0.216)
Observations	50443	85023	60414	2973	192907

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Individual characteristics and year fixed effects are included but not reported.

6 Conclusion

Many countries seek to increase formal labour participation of mothers. Policymakers often point to Scandinavia, where high public spending on childcare goes hand in hand with high labour participation rates of mothers. However, our analysis of a large recent reform in the Netherlands, which cut the parental fee for formal childcare in half, suggests that we should be careful in interpreting this correlation as a causal relation. We conclude that the policy reform in the Netherlands increased participation of women with young children by a modest 2.5%-point. Average hours worked of these women increased by 1.1 hours per week, whereas average hours worked of the fathers decreased by 0.4 hours per week. Furthermore, these effects are an upper bound as the government increased an EITC for parents with young children over the same period. Our findings are in line with recent studies using Swedish (Lundin et al., 2008) and Norwegian (Havnes and Mogstad, 2011a) data, which also find small if any effect on the labour participation of parents.

In this paper we have used the Dutch reform to study the relation between childcare subsidies and labour participation. However, the reform could also be used to investigate a number of other relevant questions. Indeed, Baker et al. (2008, p. 711) argue that a full evaluation of publicly financed childcare requires answers to three questions, which we take up below.

First, how does public financing affect the quality and quantity of formal childcare, and to what extent does it lead to substitution of informal childcare? This requires microdata on price and use of formal and informal childcare over time. One of the side effects of the policy reform is that since 2005 we have potentially good microdata on the use of formal childcare, since all subsidies now run via the Tax Office. However, finding reliable informal childcare data remains a challenge.

Second, how do childcare subsidies affect labour participation and what is the net cost to the government? We have answered the first part of this question. For the second part one would to link the labour participation data to the childcare data, and we would need to link these data to a tax-benefit calculator to determine the effects on government receipts and expenditures.[12] However, with expenditures on childcare subsidies and the targeted EITC rising by some 2.5 billion euro, and labour participation rising by about 30 thousand persons[13] and 30 thousand full-time equivalents[14], it is unlikely that this reform paid for itself with additional tax revenues and savings on benefits.

Third, what is the effect of expanding formal childcare on children and families? There are a number of papers that use the same reforms to consider the effects on children and families (see *e.g.* Loeb et al., 2007; Baker et al., 2008; Havnes and Mogstad, 2011b). For the moment no such study exists for the Netherlands. However, a number of recent studies suggests that this might be an important element to consider in the Dutch reform. Vermeer et al. (2005) and Kruif et al. (2009) use a large number of internationally comparable indicators for the quality of daycare,[15] and find a disturbing trend. On a scale from 1 (bad) to 7 (excellent), their sample scored on average 4.8 in 1995, 4.3 in 2001, 3.2 in 2005 and a meager 2.8 in 2008. Furthermore, in 2008, 49% of daycare centres got a rating 'insufficient' and 51% got a rating of 'poor', while none of the 200 daycare centres got a rating of 'good'. Hence, it seems important to study how the policy reform affected children and

[12] Despite the substantial rise in female participation found in Baker et al. (2008) they still calculate the net effect on government finances to be negative, in part due to a substantial substitution of informal by formal care.

[13] 30 thousand persons is 2.5%-points of working women in the treatment group.

[14] The net effect of +1.1 hours per week for women in the treatment group and –0.4 hours per week for men in the treatment group.

[15] Specifically, they use the ITERS-R (Infant/Toddler Environment Rating Scale - Revised) for 0-2.5 year olds, and the ECERS-R (Early Childhood Environment Rating Scale - Revised) for 2.5 to 5 year olds.

families, and how participation in formal childcare affects children and families in general.

We would also be interested in how these effects may differ in the short and long run. In particular, we have used data up to 2009 (more recent microdata is not available). Since the major changes in the parental fee took place in 2006 and 2007, we consider our results 'medium run' effects. It would be interesting to study what happened after 2009. However, we can only look so far. Faced with the dramatic rise in public expenditures on formal childcare, the current government plans to substantially decrease subsidies for formal childcare. Indeed, by 2015 the average parental fee is expected to rise to 34% (Ministry of Social Affairs and Employment, 2011). However, this will provide us with an interesting new natural experiment, to study *e.g.* whether the response of parents to changes in the parental fee is symmetric for decreases and increases.

References

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press, Princeton.

Baker, M., Gruber, J., and Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy*, 116(4):709–745.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Blau, D. and Currie, J. (2006). *Handbook of the Economics of Education*, chapter Preschool, day care, and after school care: who's minding the kids?, pages 1163–1278. Elsevier.

Blundell, R. and Costa Dias, M. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, 44(3):565–640.

Cameron, A. and Trivedi, P. (2005). *Microeconometrics: methods and applications.* Cambridge University Press.

Cascio, E. (2009). Maternal labor supply and the introduction of kindergartens into american public schools. *Journal of Human Resources*, 44(1):140–170.

Fitzpatrick, M. (2010). Preschoolers enrolled and mothers at work? The effects of universal prekindergarten. *Journal of Labor Economics*, 28(1):51–85.

Gelbach, J. (2002). Public schooling for young children and maternal labor supply. *American Economic Review*, 92(1):307–322.

Havnes, T. and Mogstad, M. (2011a). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics*, 95:1455–1465.

Havnes, T. and Mogstad, M. (2011b). No child left behind: subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy*, 3:97–129.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Imbens, G. and Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

Kruif, R. d., Riksen-Walraven, M., Gevers Deynoot-Schaub, M., Helmerhorst, K., Tavecchio, L., and Fukkink, R. (2009). Pedagogische kwaliteit van de opvang voor 0- tot 4-jarigen in Nederlandse kinderdagverblijven in 2008. NCKO, Leiden.

Lefebvre, P. and Merrigan, P. (2008). Child-care policy and the labor supply of mothers with young children: a natural experiment from Canada. *Journal of Labor Economics*, 26(3):519–48.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., and Rumberger, R. (2007). How much is too much? The effects of duration and intensity of child care experiences on children's social and cognitive development. *Economics of Education Review*, 26(1):52–66.

Lundin, D., Mörk, E., and Öckert, B. (2008). How far can reduced childcare prices push female labour supply? *Journal for Labor Economics*, 15:647–659.

Ministry of Finance (2010). Het kind van de regeling, rapport brede heroverwegingen 5. The Hague.

Ministry of Social Affairs and Employment (2011). Vereenvoudiging en beperking kindregelingen. The Hague.

Plantenga, J., Wever, Y., Rijkers, B., and de Haan, P. (2005). Arbeidsmarktparticipatie en de kosten van kinderopvang. *Economisch Statistische Berichten*, 4455:115.

Statistics Netherlands (2009). EBB: methoden en definities. http://www.cbs.nl/NR/rdonlyres/150506F0-38F8-4AAD-B252-F91C952DFDF6/0/2010ebbmethodenendefinities2009a.pdf.

Vermeer, H., IJzendoorn, R., de Kruif, R., Tavecchio, L., Risken-Walraven, M., and van Zeijl, J. (2005). Kwaliteit van de Nederlandse kinderdagverblijven: trends in kwaliteit in de jaren 1995-2005. NCKO, Leiden.