CPB Netherlands Bureau for Economic Policy Analysis

CPB Discussion Paper | 217

Childcare subsidies and labour supply Evidence from a large Dutch reform

Leon J. H. Bettendorf
Egbert L. W. Jongen
Paul Muller

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

Leon J.H. Bettendorf* Egbert L.W. Jongen† Paul Muller‡

Abstract

Over the period 2005-2009 the Dutch government increased childcare subsidies substantially, reducing the average effective parental fee by 50%, 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 moderately sized impact on labour supply. Furthermore, the effects are an upper bound since there was also an increase in an earned income tax credit for the same treatment group over the same period. The joint reform increased the maternal employment rate by 2.3%-points (3.0%). Average hours worked by mothers increased by 1.1 hours per week (6.2%). Decomposing the hours effect we find that most of the increase in hours is due to the intensive margin response. A number of robustness checks confirm our results.

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, Katja Baum, Pieter Gautier, Tarjei Havnes, Bas van der Klaauw, Kevin Staub, Dinand Webbink, participants of the 2012 IZA Workshop 'Recent Advances in Labor Supply Modeling' in Dublin, participants of the 2012 IIPF congress in Dresden and seminar participants at CPB Netherlands Bureau for Economic Policy Analysis, the Ministry of Social Affairs and Employment, Tilburg University and VU University Amsterdam. Remaining errors are our own. We are grateful to the 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 Scandinavia, where public spending on childcare is high and participation rates of mothers are high as well. Indeed, several 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). A fruitful way to learn about the causal effect of childcare subsidies on labour supply is to study natural experiments. By comparing the labour supply behaviour of a treatment and control group before and after a policy 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. Over the period 2005-2009 childcare subsidies became much more generous. The average effective parental fee for formal childcare was cut in half, and subsidies were extended to so-called guestparent care (small scale care at the home of the 'guestparent' or at the home of the children). As a result, public spending on childcare skyrocketed, from 1 billion euro in 2004 to 3 billion euro (0.5% of GDP) in 2009 (Ministry of Finance, 2010). Over the same period that childcare subsidies became much more generous, the government also increased targeted earned income tax credits (EITCs) for the same parents.[1,2] Budgettary outlays of these EITCs increased from 0.7 billion euro in 2004 to 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 estimate the joint effect of these reforms using data from the Labour Force Survey of 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 per week. The treatment group consists of parents aged 20 to 50 years old with a youngest child 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 show that

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

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

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 2.3%-points (3.0%). Second, the reform increased the average number of hours worked per week by women in the treatment group by 1.1 hours per week (6.2%). Third, decomposing the hours effect we find that most of the increase in hours comes from the intensive margin, not the extensive margin. A number of robustness checks confirm our results.

There is an extensive literature that considers the relation between parental labour supply and the cost of childcare using structural models and cross-sectional data. An in depth overview is given in Blau and Currie (2006), who report estimated (childcare) price elasticities of female labour participation ranging 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.[3] 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.

A number of papers using natural experiments find rather small labour supply effects. 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 in 2002. Using difference-in-differences they estimate the effect on labour supply to be close to zero. Havnes and Mogstad (2011a) study a staged implementation of subsidized childcare in Norway in the late 1970s. Their difference-in-differences analysis shows that the reform increased the participation rate of mothers by just 1.1%-points. Fitzpatrick (2010) estimates the effect of Pre-kindergarten programs in 1999-2000 on maternal labor supply in the US, and finds no effect on the labour supply of mothers.

[3]For example, unobserved characteristics are likely to influence both the cost of childcare (which depends on income) and the labour supply decision.

However, there are also a number of papers that find bigger labour supply effects, overall or for subgroups. Lefebvre and Merrigan (2008) study a natural experiment in Canada, in the late 1990s, and find an effect on the maternal employment rate of 7.3%-points (13%) and on hours worked of 2.6 hours per week (22%).[4] For the same reform, but using a different dataset, Baker et al. (2008) find an increase in the maternal employment rate of 7.7%-points (14.5%). Nollenberger and Rodriguez-Planas (2012) study a staged implementation of free child care for 3-year olds in Spain in the 1990s. They find an increase in the maternal employment rate of 2.4%-points. This effect may seem modest, but since the maternal employment rate at the start of the policy reform was only 30 percentage points, the effect is large in percentage terms (8.1%). The effect on hours worked is estimated to be 1.0 hours per week (9%). Finally, Gelbach (2002) studies free provision of public schooling in the US and finds an increase in labour participation by single parents of 4%-points. For the same reform, Cascio (2009) also estimates sizeable effects for single mothers, but finds no effect for mothers in couples.

How do we reconcile the different findings? As argued by *e.g.* Havnes and Mogstad (2011a), part of the differences can be attributed to differences in initial labour market conditions, the particular group of women that the study focuses on and the availability of informal childcare. For example, the relatively high initial maternal employment rate in the studies by Lundin et al. (2008), Fitzpatrick (2010) and this paper (see below) can explain the modest effect on the maternal employment rate. Furthermore, the finding that effects are larger for single mothers than for mothers in couples, as in *e.g.* Cascio (2009) and this study, is not surprising given that single mothers appear particularly responsive to changes in financial incentives (Meghir and Phillips, 2010).

However, as pointed out by *e.g.* Cascio (2009) and Havnes and Mogstad (2011a) there are also some potential pitfalls in the empirical strategy that may explain part of the differences. In particular, they suggest that the relatively large effects for the reform in Canada presented in Lefebvre and Merrigan (2008) and Baker et al. (2008) may in part be driven by differential trends in the treatment and control states. In our study we have ten years of pre-reform data. We show that our treatment and control group have the same pre-reform trend, and we find much smaller effects than

[4] All reported percentage points and percentage effects are taken directly from the reported studies.

Lefebvre and Merrigan (2008) and Baker et al. (2008).

We believe our paper is particularly relevant for continental European countries that are currently considering to substantially expand their formal childcare programs. Using a very recent and large natural experiment, starting from an initial maternal employment rate and public spending on childcare arguably quite similar to many other continental European countries,[5] and using a control group that has a similar trend as the treatment group in the 10 years preceding the reform, we show that the impact of childcare subsidies on labour supply is modest.

The structure of the paper is as follows. Section 2 describes the main aspects of the reform we exploit in the empirical analysis. Section 3 discusses our empirical methodology. In Section 4 we present our dataset and some descriptive statistics. Section 5 gives the estimation results for participation and hours worked. Here we also study potentially heterogeneous responses by *e.g.* the level of education, family composition and age of the youngest child. In Section 6 we decompose the effect on hours worked into an extensive margin effect and an intensive margin effect. Here we follow a methodology recently developed by Staub (2010) that allows for a causal interpretation of the decomposition. Section 7 concludes and discusses directions for future research.

2 The natural experiment

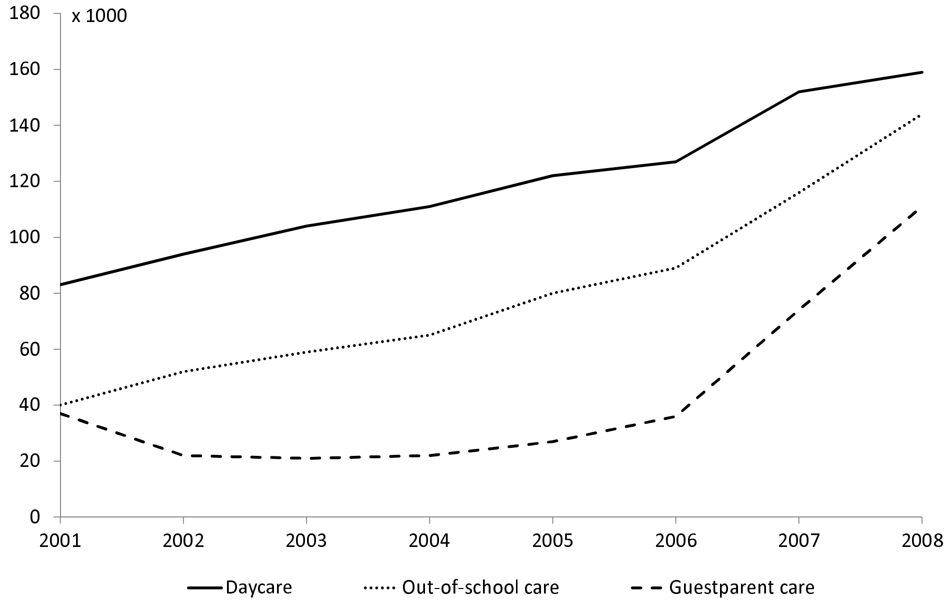
The natural experiment consists of a series of policy changes that took place between 2005 and 2009. The most important changes took place in 2006 and 2007 when subsidies were increased such that on average the parental fee decreased by 50%. Below we give a short historical account of the policy changes, and indicate their significance.

Before the introduction of the *Wet kinderopvang* (Law on childcare) in 2005 centre based childcare was subsidized at different rates. Places subsidized directly by employers and local governments (76% of places[6]) had lower effective parental fees than so-called 'unsubsidized' places (24% of places), the costs of which were partly tax deductible for parents.

[5]See e.g. OECD (2007, Table 3.2, Chart 6.1).

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

Figure 1: Childcare places



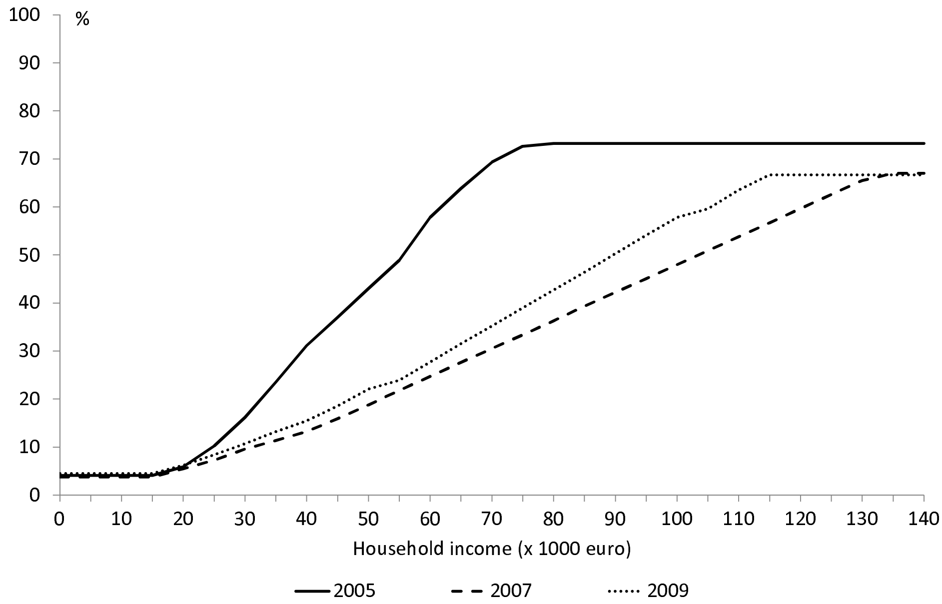
Source: Statistics Netherlands (statline.cbs.nl).

Table 1: Public spending on childcare and EITCs for parents (millions of euro)

Year	2002	2003	2004	2005	2006	2007	2008	2009
Childcare subsidies	725	755	1028	1001	1343	2058	2825	3034
EITCs for parents	410	460	738	830	871	984	971	1290
– *Combinatiekorting*[a]	410	460	479	484	314	324	247	0
– *Inkomensafhankelijke combinatiekorting*[b]	0	0	259	346	557	660	724	1290

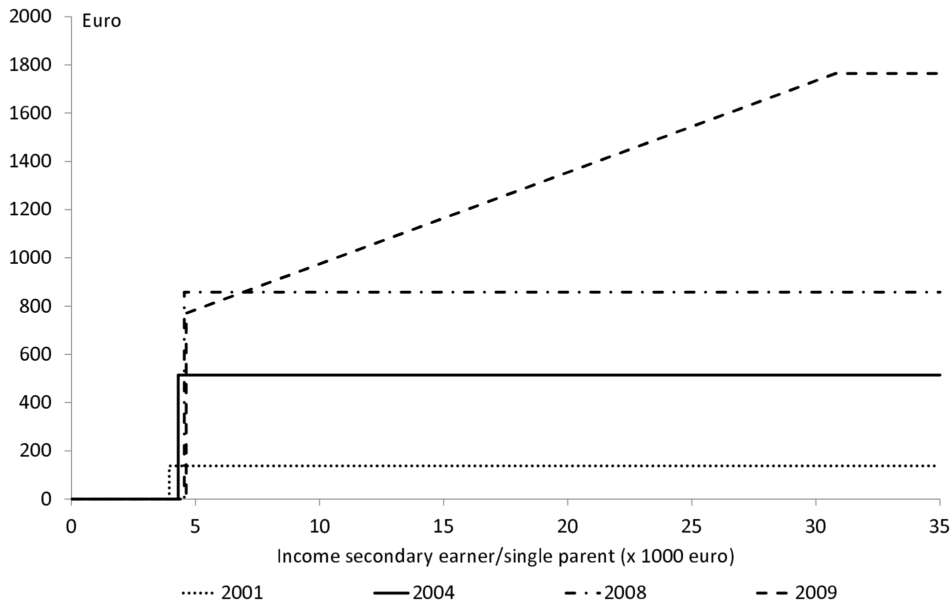
Source: Ministry of Finance (2010) and own calculations (imputation of employers' contribution for childcare up to 2007 with data from the Ministry of Social Affairs and Employment (personal communication) and split of the EITCs for parents in its two components using the MIMOSI model of CPB). [a]The *Combinatiekorting* applies to primary earners, secondary earners and working single parents with a youngest child up to 12 years old. [b]The *Inkomensafhankelijke combinatiekorting* applies to secondary earners and working single parents with a youngest child up to 12 years old.

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.

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 central government. This increased the subsidy somewhat for parents with children going to an 'unsubsidized' place 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 subsidies 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[7], public spending actually fell slightly from 2004 to 2005. Figure 1 shows that the growth rate of the number of childcare places is hardly affected in 2005 relative to the preceding period. Hence, in our empirical analysis we also 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 changes in the parental contribution rate for the 'first child'[8] 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 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.[9,10] 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 data period, in 2009, the

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

[8]The Tax Office defines the first child as the child for which the parents spend the most on formal childcare.

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

[10]Despite the steep increase in the subsidy rate, the average prices of formal childcare places grew more or less in line with the CPI. Over 2005-2009 the average full price for an hour of daycare and out-of-school care grew by 9.6% and 6.0%, respectively (Ministry of Education, Culture and Science, 2009), while the CPI grew by 6.5% (CPB, 2012). Hence, the increase in the subsidy rate was not counteracted by a rise in the full price of childcare.

parental fee was raised again somewhat, but relative to 2005 there was still a large drop for middle and high income households.

Figure 1 shows that the dramatic drop in the contribution rate in 2006 and 2007 spurred the growth in the use of formal childcare in 2006 and beyond. Also, guestparent care took a high flight, as more parents and guestparents became aware of the possibility to claim subsidies. As a result, public spending on childcare went 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).[11] We carefully examined various changes in taxes and subsidies and found that, apart from one, there were no substantial changes in taxes or subsidies targeted at the treatment or control group. The only complication comes from changes in the EITCs for parents with a youngest child up to 12 years old, the *Combinatiekorting* (Combination credit) and the *Inkomensafhankelijke combinatiekorting* (Income dependent combination credit). These EITCs are also targeted exclusively at our treatment group. Figure 3 shows the change in the sum of the *Combinatiekorting* and *Inkomensafhankelijke combinatiekorting* for secondary earners and single parents (mostly women) over the period 2001-2009. Table 1 gives the changes in aggregate 'spending' (revenue losses) on these EITCs. Between 2001 and 2004 these credits increased from 138 to 514 euro, and public expenditures increased from 410 to 738 million euros between 2002 and 2004. Between 2004 and 2008 the individual subsidy increased from 514 euro to 858 euro for secondary earners and single parents, and in 2009 there was another increase for secondary earners and single parents with relatively high earnings. In 2009, the maximum credit was 1,765 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

[11] Another concern might be that what we see is not the result of an increase in childcare demand but the result of a drop in rationing on the formal childcare market. However, the available data on waiting lists suggests that these are rather small, and that the change in waiting lists was much smaller than the change in filled childcare places. For example, the survey data reported in Van Rens and Smit (2011) suggests that the waiting list for daycare (out-of-school care) dropped from 10% (11%) of filled places in 2007 (the first year of the survey) to 7% (6%) of filled places in 2009. The drop in waiting lists is much smaller than the increase in the number of children going to daycare and out-of-school care, which increased by 53% (17%) and 128% (47%) respectively between 2005 and 2009 (2007 and 2009) (Ministry of Education, Culture and Science, 2009).

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 pre-reform 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. This makes it hard to separate the effect of the change in the childcare subsidies from the effect of the change in the EITCs.

3 Methodology

We estimate the effect of the policy change on labour participation using a difference-in-differences 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 childcare costs, 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. As the control group we use parents with a youngest child aged 12 to 17 years old (living at home). These parents are not eligible for childcare subsidies but are otherwise quite similar to 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. This assumption cannot be tested. However, we have 10 years pre-reform data, so we can test whether this assumption holds in the pre-reform period. In Section 4 we estimate the pre-reform trend for both groups and in Section 5 we also estimate placebo treatment effects in periods before the reform as an additional check.

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 a change in behaviour when deciding to pass the new law. Also, if parents anticipated the policy change and adapted their behaviour in advance this too would create a problem for identification of 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 of 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 be no 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. Also, the reform itself should not cause a change in the composition of the groups. This could happen if for example the childcare reform led to a change in fertility rates (since our treatment group is defined by having a young child this would alter the composition of the treatment group). We inspect fertility rates and find no evidence of a change in the trend after 2005 (see the Appendix).

To estimate the treatment effect on participation, we regress participation status on a year fixed effect (α_t), group fixed effects (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 has a youngest child aged up to 12 years old in year s. The common trend of the treatment and control group is captured by the year fixed effects, while the constant difference in participation between the treatment and control group is captured by the group effects. We include different group fixed effects for parents with a youngest child aged 0-3, 4-7 and 8-11 years old. Individual characteristics are included to control for observable 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 to the effect on participation we also 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 that we used in the participation equation. 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 a Heckman selection model (Heckman, 1979) which allows for different effects of the treatment dummies and the covariates on the participation decision and the hours decision. To identify the parameters in the selection model we use a large tax reform in 2001 as an exclusion restriction. Specifically, we include a post tax reform dummy (the period 2001-2009) interacted with education level in the participation equation but not in the hours equation. Bosch and van der Klaauw (2012) show that this reform had a significant effect on the participation rate of women (which was different for different education levels), but had no effect on the hours worked by working women.

4 Data

We use data from the Dutch Labour Force Survey (*Enquete Beroepsbevolking*) of Statistics Netherlands. This is an annual survey which includes approximately 80,000 individuals per year. We have repeated cross-sections for the period 1995-2009. The reform started in 2005, so we have a long data series preceding the policy change 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, education level, native/immigrant, married couple/unmarried couple/single) and household characteristics (number of children, age of the children).[12] 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 mothers with a youngest child aged up to 12 years old. Furthermore we restrict the analyses to mothers aged between 20 and 50 years old. This gives us 202,106 observations for the treatment group (for the full sample period). As a control group we select all mothers with a youngest child aged between 12 and 18 years old. Restricting the control group to mothers aged between 20 and 50 years old we have 61,127 observations in the control group (for the full sample period). We also considered women without children as a potential control group. However, as we will see below, this is not a valid control group since they have a different pre-reform trend in the participation rate and hours worked than the treatment group.

For the DD method to identify the treatment effect it is important that there are no significant changes in the composition of the treatment and control groups before and after the policy change. To investigate whether this is the case we show descriptive statistics of the treatment group and differences between 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 (in the regression we use 5-year category dummies), education (in categories lower, middle and higher educated), a dummy for being single, a dummy for being an immigrant, the size of the household (in the regression we include dummies for families with one, two or three and more children aged below 12) and the age of the youngest child (with separate dummies for 0-3, 4-7 and 8-11 year old).

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 somewhat more likely to be single, and are also somewhat more likely to be lower educated. The share of immigrants is slightly higher in the treatment group, which could be explained by the higher fertility rate of immi-

[12]For each year we restrict our sample to individuals that were interviewed in person. Apart from these, there were three follow-up interviews of the same individual within one year by telephone. Since these are considered less reliable and are basically the same observation (see Statistics Netherlands, 2009), we decided to use only the data from the interviews in person. Unfortunately, we could not make this distinction for 2009, so we have about four times more observations in 2009 than in the other years.

Table 2: Descriptive statistics before and after the reform

	Treatment Group		Differences (Treat-Control)		Normalized differences (Treat-Control)	
	2000-2004	2005-2009	2000-2004	2005-2009	2000-2004	2005-2009
Participation	0.675	0.747	-0.055	-0.025	-0.073	-0.032
Worked Hours	14.53	17.07	-2.739	-1.424	-0.119	-0.059
Age	35.74	36.48	-8.183	-8.211	-0.995	-0.930
Higher educated	0.242	0.298	0.064	0.097	0.03	0.131
Lower educated	0.299	0.226	-0.106	-0.102	-0.135	-0.126
Single	0.098	0.103	-0.050	-0.059	-0.083	-0.093
Immigrant	0.229	0.239	0.038	0.043	0.051	0.054
Household size	3.905	3.890	0.078	0.085	0.050	0.052
Age youngest child	4.405	4.588	-9.922	-9.746	-2.211	-2.058
Observations	60,379	82,151				

Values are means weighted with sample weights. Normalized differences are mean differences divided by the square root of the sum of variances. Source: Labour Force Survey (Statistics Netherlands).

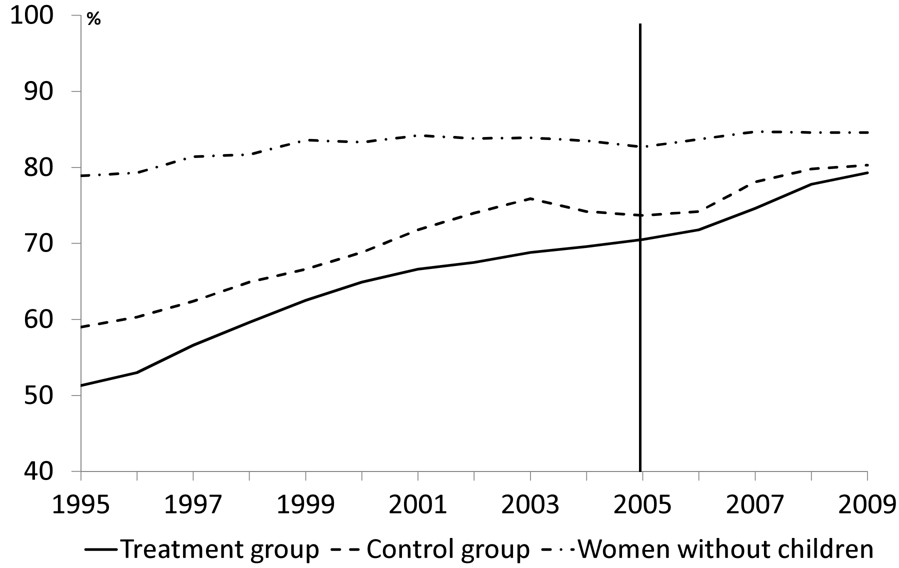
grants. More importantly, the differences in characteristics between the groups are rather stable over time.[13] Therefore we conclude that changes in the composition of unobservable characteristics of the groups are unlikely to bias the estimation results.

Following Imbens and Wooldridge (2009) we also present so-called normalized differences in the last two columns. These are differences scaled by the square root of the sum of variances. With this measure we can determine whether there is sufficient overlap between the treatment and control group with respect to the covariates. Imbens and Wooldridge (2009) argue that if the normalized difference is above one-quarter (in absolute value), linear regression becomes sensitive to the specification. In our case all variables satisfy this condition, with the exception of age and age of the youngest child. The latter is inevitable given our setup since it defines the control and treatment group.

In the DD method we compare the outcomes of the treatment and control group

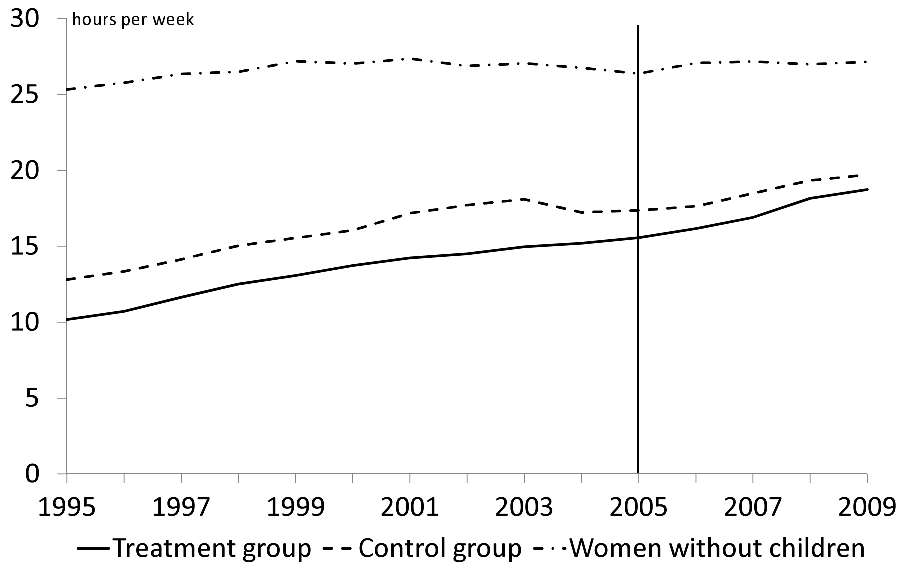
[13] There is some divergence in the share of higher educated, with the share rising faster in the treatment group than the control group. However, inspection of the data shows that this is a gradual process, with no sudden jumps in the shares of lower, middle and higher educated women in the treatment and control group before and after the reform (data available on request).

Figure 4: Participation rate



Source: Labour Force Survey (Statistics Netherlands).

Figure 5: Hours worked per week



Source: Labour Force Survey (Statistics Netherlands).

Table 3: Trendtest participation rate and hours worked per week

	(1) Participation rate	(2) Hours per week
Trend treatment group	0.019***	0.376***
	(0.000)	(0.020)
Trend control group	0.020***	0.465***
	(0.001)	(0.042)
Period	1995-2004	1997-2004
Observations	153,809	121,675
P-value equal trends	0.384	0.053

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

over time. In Figure 4 we plot participation rates of the women in the treatment group (youngest child 0-11), the control group (youngest child 12-17) and for women without children (a potential other control group). The solid vertical line marks the start of the policy reform. 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,[14] whereas women without children clearly have a different pre-reform trend. This suggests that women with an older child are a valid control group for our DD analysis, whereas women without children are not.

As a more formal check on our control group, we perform a pooled regression on participation for the treatment and control groups including all individual characteristics and group dummies 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 and constant remaining differences captured by the group dummies. The estimation

[14]There appears to be somewhat of a spike in the participation rate in the control group around 2003. However, when we add placebo treatment dummies for each individual year in 2000-2004, the individual dummies are not signficantly different from zero, and the treatment effect is virtually unchanged. As an additional check we also looked at the mean values for the controls for the control group. But these show no sudden changes around 2003, see Table A.1 in the Appendix. As a final check we also estimated the treatment coefficients excluding data for 2003-2004, and again found a similar treatment effect as in our main regressions (results available on request).

results are presented in column (1) in Table 3. The estimated trends are very close and we do not reject the hypothesis that they are equal.

In Figure 5 we plot the average number of hours worked per week. Again we see that there is a clear upward trend, both in the treatment group and our preferred control group, whereas the upward trend is absent in the group of women without children. Again, women with an older child seem a valid control group, and women without children do not seem a valid control group. We estimate trends for our treatment and control groups, while controlling for observable characteristics and group fixed effects in the same way as for participation. We find that the trends are the closest when we use data from 1997 onwards. The trend for the control group is still somewhat larger than for the treatment group however, but the difference is not significant at the 5% level (see column (2) in Table 3).

5 Estimation results

5.1 Participation rate

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 consider the results for subgroups.

We estimate equation (1) for the participation rate. We first estimate the model with annual treatment dummies for 2005-2009. We can not reject that the treatment effects per year are equal, see Table A.2 in the Appendix. However, the results suggest a difference between the effects in 2005-2007 and 2008-2009. We therefore decided to estimate and present results with separate treatment effects for 2005-2007 and 2008-2009. These might be considered the short and medium run effects of the reform. Estimates are presented in column (1) in Table 4.[15] In the first three years the effect is 1.5%-points (which is an increase in the participation rate of 2.0%), while in the last two years it increases to 2.3%-points (3.0%).[16] The short

[15]The estimates of the control variables are all significant and in line with expectations, see Table A.2 in the Appendix.

[16]For the two treatment effects we still cannot reject that they are equal. When we estimate one single treatment effect for 2005-2009, we obtain a treatment effect of 0.018 (significant at the 1% level).

Table 4: Effect on participation rate: all women

	(1)	(2)	(3)	(4)	(5)
	OLS	Probit	OLS	OLS	OLS
	95-09	95-09	01-09	95-09	95-09
				No overlap[a]	With placebo
Placebo 00-04					–0.005
					(0.006)
Treatment 05-07	0.015**	0.012*	0.023***	0.025**	0.012*
	(0.006)	(0.006)	(0.007)	(0.011)	(0.007)
Treatment 08-09	0.023***	0.022***	0.032***	0.034***	0.021***
	(0.006)	(0.006)	(0.007)	(0.010)	(0.007)
Observations	263,231	263,231	170,834	249,785	263,231
P-value Placebo=Treat05-07					0.00
P-value Placebo=Treat08-09					0.00

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. [a]We include only parents with a child aged 16-17 in the control group so that individuals in the control group that were previously in the treatment group are excluded.

run estimate is significantly positive at the 5% level, the medium run estimate 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 larger than 1. Still, as a robustness check we present estimates of a probit model in column (2). We find very similar treatment effects (reported values are weighted average marginal effects). As another robustness check we estimate the specification in column (1) using only data from 2001 onwards. The results are given in column (3). Again we find similar, though somewhat larger, treatment effects.

A concern might be that some women that were in the treatment group in the early years are in the control group in the later years. When there is a treatment effect on the participation rate of mothers extending beyond the treatment period (due to for example a career effect), part of the treatment effect may be masked by an effect on the control group. Therefore we also estimate the model using only parents with a child aged 16-17 in the control group so that individuals in the control group that were previously in the treatment group are excluded.[17] This results in the effects reported in column (4). The effect is somewhat larger than our base results, though not significantly different.

The validity of our estimates depends critically on the common trend assumption. To further assess the plausibility of this assumption we also estimate a placebo treatment effect. Specifically, we estimate a treatment effect for some years before 2005. Since no relevant policy change occured in this period (that differs between the treatment and control group) 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 year periods and estimate a placebo treatment effect for 2000-2004. The placebo effect and the two treatment effects are reported in column (5). The placebo treatment effect is not significantly different from zero, while both treatment effects are hardly affected by the inclusion of the placebo dummy. Furthermore, we can reject that the placebo coefficient equals the treatment effect in both cases with very low p-values.

We also estimate the effect for the following subgroups of women: i) lower, middle and higher educated, ii) single women and women in couples, iii) women with a youngest child 0-3 years old (pre primary school), 4-7 years old (first years in primary school), 8-11 years old (last years in primary school) and iv) native and non-native women. We do this by estimating equation (1) for each subsample, thereby allowing differences in all coefficients for covariates between subgroups. Since OLS and probit lead to comparable results, only OLS results are reported for these groups, in Table 5.

With regard to education, we find no effect on lower educated women, while the effect on middle and higher educated women is somewhat higher than the average effect. An explanation for the absence of the effect on lower educated women could be that their subsidy rate did not change much, since lower education is correlated with low income. In columns (4) and (5) we find that the effect on the participation rate of single women is higher than for women in couples. Next, columns (6), (7) and (8) suggest that the effect on the participation rate is somewhat larger for women with a young child than for women whose youngest child is in the last years of primary education. Since older children are less likely to go to childcare, this is in

[17] A trend test reveals that this subgroup of the control group also has a trend similar to the treatment group (results available on request).

Table 5: Effect on participation rate: subgroups of women (OLS)

	(1) Lower educated	(2) Middle educated[a]	(3) Higher educated	(4) Single women	(5) Women in couples
Treat 05-07	0.005	0.021**	0.014	0.030	0.011*
	(0.012)	(0.009)	(0.011)	(0.019)	(0.006)
Treat 08-09	–0.003	0.039***	0.037***	0.047***	0.020***
	(0.012)	(0.008)	(0.010)	(0.018)	(0.006)
Observations	77,604	122,030	63,597	26,453	236,778
	(6) Youngest child 0-3	(7) Youngest child 4-7	(8) Youngest child 8-11	(9) Native	(10) Immigrant[a]
Treat 05-07	0.018***	0.015**	0.012	0.015**	0.036**
	(0.007)	(0.008)	(0.008)	(0.006)	(0.017)
Treat 08-09	0.025***	0.028***	0.015**	0.016***	0.062***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.017)
Observations	155,165	169,193	110,889	222,232	40,999

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. [a]The placebo treatment dummies for 2000–2004 are insignificant for all subgroups except lower educated women, middle educated women, immigrant women and women with a youngest child aged 0-3.

line with expectations. Finally, columns (9) and (10) suggest that the effect on the participation rate of immigrant women is larger than for native women.

However, placebo tests suggest that the control group does not have the same trend as the treatment group for all these subgroups. The placebo treatment dummies for 2000–2004 are insignificant for all subgroups except lower educated women (placebo is negative), middle educated women (placebo is positive), immigrant women (placebo is negative) and women with a youngest child aged 0-3 (placebo is negative).[18] Hence, the treatment effect for these groups may also capture differential trends between the treatment and control group.[19]

[18]Results available on request.

[19]A negative placebo suggests that the treatment effect is underestimated while a positive placebo suggests that the treatment effect is overestimated.

5.2 Hours worked per week

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

As discussed in Section 3 we estimate equation (1) with average hours worked per week as the outcome variable and include all women in this regression, both working women and non-working women. Again we estimate a separate treatment dummy for 2005-2007 and for 2008-2009, the results are reported in column (1) in Table 6. We find positive effects in both periods which are significant at the 1% level. In 2005-2007 the estimated effect is an increase of 0.66 hours per week. In 2008-2009 the effect increases to 1.08 hours per week. Given the average number of hours worked per week for women in 2005-2007 and 2008-2009 (16.2 and 18.4, respectively), these effects are more substantial in percentage terms (6.2% in 2008-2009) than the effects on the participation rate (3.0% in 2008-2009).

Column (2) gives the results for the tobit model. The reported (weigthed average marginal) effects are quite close to the OLS estimates in column (1). However, the tobit model can be restrictive. In column (3) we therefore also present estimates of a Heckman selection model with, as an exclusion restriction, the tax reform in 2001. We report the (weighted average marginal) effect on the entire sample and again find very similar results. Restricting the sample in the linear regression to the years 2001-2009 leads to somewhat larger treatment effects, see column (4). When we restrict the control group to mothers with a youngest child aged 16-17 years old, so that they were never in the treatment group, we find that the estimated effect on hours is again somewhat larger, see column (5). Column (6) presents the effects on hours worked when we add a placebo treatment dummy for 2000-2004 to the OLS specification. The placebo treatment effect is not significantly different from zero, though the coefficient is larger relative to the treatment effect than in the estimates for participation, while both the short and medium run treatment effects are somewhat lower than in column (1). Column (7) presents the effects on hours worked when we add two placebo treatment dummies for 2000-2004 to the participation and hours equation of the Heckman selection model, respectively.[20] In

[20]The reported coefficient is the average marginal effect of switching both placebo dummies from 0 to 1.

Table 6: Effect on hours worked per week: all women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	Tobit	Heckman	OLS	OLS	OLS	Heckman
	97-09	97-09	97-09	01-09	97-09	97-09	97-09
					No overlap[a]	Placebo	Placebo
Placebo 00-04						–0.443*	–0.196
						(0.229)	(0.227)
Treatment 05-07	0.661***	0.751***	0.741***	0.976***	0.708**	0.384	0.613**
	(0.198)	(0.202)	(0.189)	(0.225)	(0.342)	(0.246)	(0.241)
Treatment 08-09	1.075***	1.192***	1.136***	1.375***	1.325***	0.798***	0.999***
	(0.194)	(0.195)	(0.186)	(0.222)	(0.322)	(0.243)	(0.237)
Observations	231,097	231,097	231,097	170,834	217,651	231,097	231,097
P-value Placebo=Treat05-07						0.00	0.00
P-value Placebo=Treat08-09						0.00	0.00

Standard errors in parentheses, * denotes significant at 10% level, ** at 5% level and *** at 1% level. Coefficients in columns (2) and (3) are weighted average marginal effects on the expected outcome for the entire sample. Individual characteristics and year fixed effects are included but not reported. [a]Including only parents with a child aged 16-17 in the control group so that individuals in the control group that were previously in the treatment group are excluded.

this case the placebo is close to zero, while the treatment effects are hardly affected. Equality of the placebo and treatment effects can be rejected with a p-value below 1% in both the OLS and Heckman model.

For subgroups of women we report estimates of the OLS model in Table 7. The pattern is similar to the results for participation. The effect on lower educated women is not significantly different from zero, whereas the effect is larger for higher educated women. The effect for single women is larger than for women in couples. The total effect can be attributed mainly to women with a child younger than 8 years old. Finally, the effect on hours worked is again larger for immigrant women than native women.

For hours worked the placebo treatment dummies for 2000–2004 are insignificant for all subgroups except lower educated women and women with a youngest child 0-3 years old.[21] Hence, for these groups the treatment effect may also capture differential time effects between the treatment and control group.

[21]Results available on request.

Table 7: Effect on hours worked per week: subgroups of women (OLS)

	(1) Lower educated[a]	(2) Middle educated	(3) Higher educated	(4) Single women	(5) Women in couples
Treatment 05-07	0.402	0.767**	0.899**	1.090*	0.552**
	(0.345)	(0.300)	(0.421)	(0.623)	(0.207)
Treatment 08-09	0.340	1.342***	1.617***	1.680***	0.962***
	(0.366)	(0.286)	(0.406)	(0.609)	(0.203)
Observations	65,020	108,152	57,925	23,945	207,152

	(6) Youngest child 0-3[a]	(7) Youngest child 4-7	(8) Youngest child 8-11	(9) Native	(10) Immigrant
Treatment 05-07	1.009***	0.472**	0.309	0.636***	1.308**
	(0.218)	(0.241)	(0.254)	(0.201)	(0.582)
Treatment 08-09	1.418***	1.226***	0.408	0.818***	2.278***
	(0.215)	(0.239)	(0.250)	(0.198)	(0.573)
Observations	135,545	105,170	98,004	193,448	37,649

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. [a]The placebo treatment dummies for 2000–2004 are insignificant for all subgroups except lower educated women and women with a youngest child 0-3 years old.

5.3 Results for men

The effects for men are much less pronounced. We briefly report the main results.[22] We find no effect on their participation rate in any treatment year in any specification. We also check for an effect on subgroups, but find no significant effect for any subgroup as well. However, we do find a negative effect on their hours worked per week of –0.4 to –0.5 hours per week, significant at the 5 or 10% level (depending on the model used). For middle educated men and men with a youngest child 0-3 years old we find a significant negative effect on hours worked of –0.7 and –0.5 hours per week, respectively.[23] The drop in hours worked by men may be the result of the increase in the labour supply of their partners.

[22]Results for men are available on request.

[23]When we include placebo dummies for 2000-2004 for these subgroups the placebo dummies are insignificant.

6 Decomposition

Finally, we decompose the effect on hours worked into an effect on participation (extensive margin effect) and an and an effect on hours worked per worker (intensive margin effect).

When participation status is denoted by Y_i and the number of hours worked by workers by H_i, the expected numbers of hours worked equals

$$E(H_i) = P(Y_i = 1)E(H_i|Y_i = 1)$$

The effect of a binary treatment (T_i) on the expected number of hours worked is $E(H_i|T_i = 1) - E(H_i|T_i = 0)$, assuming that treatment is (conditionally) independent of potential outcomes. This effect can be decomposed into an effect on the probability of participation and an effect on the expected number of hours conditional on participation in the following way (McDonald and Moffitt, 1980)

$$\begin{aligned} E(H_i|T_i = 1) - E(H_i|T_i = 0) =& \Big[P(Y_i = 1|T_i = 1) - P(Y_i = 1|T_i = 0)\Big]E(H_i|Y_i = 1, T_i = 1)+ \\ & P(Y_i = 1|T_i = 0)\Big[E(H_i|Y_i = 1, T_i = 1) - E(H_i|Y_i = 1, T_i = 0)\Big] \end{aligned}$$

Note that this implicitly assumes that the expected number of hours worked conditional on participation and treatment is the same for 'old' participants and 'new' participants. This decomposition may be contaminated by a selection bias, making a causal interpretation impossible (Angrist and Pischke, 2009; Staub, 2010).

An alternative decomposition is proposed by Staub (2010), based on the joint distribution of potential outcomes. This approach explicitly identifies the group of participants (individuals that participate regardless of treatment) and the group of switchers (individuals that switch to participation due to the treatment).[24] The total treatment effect is given by the sum of the treatment effects over both groups

$$\begin{aligned} E(H_i|T_i = 1) - E(H_i|T_i = 0) =& P(i \in S)E(H_i|T_i = 1, i \in S)+ \\ & P(i \in P)\Big[E(H_i|T_i = 1, i \in P) - E(H_i|T_i = 0, i \in P)\Big] \end{aligned}$$

where S and P denote the sets of switchers and participants. We redefine each term to simplify the interpretation

$$TE = \pi^s ATE^s + \pi^p ATE^p$$

[24] Individuals that never participate are irrelevant since they are unaffected by treatment. Furthermore, we assume that the treatment effect on participation is monotone, so we rule out that individuals stop working due to the policy change.

Table 8: Decomposition of treatment effect hours worked per week

	2005-2007 (TE=0.74)			
		π	ATE	π·ATE (% of TE)
Conventional decomposition	Extensive margin	0.014	22.7	0.32 (42%)
	Intensive margin	0.689	0.63	0.43 (58%)
Causal decomposition	Extensive margin	0.014	11.9	0.17 (23%)
	Intensive margin	0.689	0.84	0.58 (77%)
	2008-2009 (TE=1.14)			
		π	ATE	π·ATE (% of TE)
Conventional decomposition	Extensive margin	0.020	23.0	0.46 (40%)
	Intensive margin	0.689	0.99	0.68 (60%)
Causal decomposition	Extensive margin	0.020	12.2	0.24 (22%)
	Intensive margin	0.689	1.30	0.89 (78%)

Results for the Heckman model.

where the supercript denotes the group; π^j (with $j = s, p$) is the probability of belonging to group j and ATE^j is the average treatment effect on group j. The first term is interpreted as the extensive margin effect and the second term as the intensive margin effect. Given the assumptions of the Heckman selection model, all terms on the RHS of both the conventional and the Staub decomposition can be derived (see the Appendix).

We compute the decomposition for the Heckman model. We calculate all terms for each individual in the decomposition and then average over the entire sample using sample weights. The upper part of Table 8 gives the decomposition of the treatment effect for 2005-2007 (0.74), the lower part gives the decomposition for the treatment effect for 2008-2009 (1.14).

Going from left to right, first note that the participation probabilities (π) are (by definition) the same for both decompositions. The average treatment effect on participation (π^s) equals 1.4%-points in the short run and 2.0%-points in the medium run. However, moving to the second column, the conventional decomposition assumes that a new participant works on average about 23 hours per week, which is similar to old participants. This ignores the possibility that new participants work less (or more) hours. In contrast, according to the Staub decomposition a switcher only works about 12 hours per week.[25] Furthermore, the Staub decomposition suggests that the effect on hours worked by old participants is larger than in the conventional decomposition. Using the conventional decomposition the intensive margin effect accounts for 60% of the total hours effect in 2008-2009, while the Staub decomposition suggests that the intensive margin effect accounts for 78% of the total hours effect in 2008-2009.

7 Conclusion

Many countries seek to increase formal labour participation of mothers. Policymakers often point to Scandinavia, where public spending on childcare is high and labour participation rates of mothers is high as well. However, our analysis of a large recent reform in the Netherlands, which cut the parental fee for formal childcare in half, suggests that such a correlation cannot necessarily be interpreted as a causal relation. We conclude that the large policy reform in the Netherlands increased participation of women with young children by a modest 2.3%-point (3.0%). The hours worked effect is larger though, an increase of 1.1 hours per week (6.2%), of which a large share is due to the intensive margin effect. This seems to be partly counteracted by a decrease in average hours worked by men of –0.4 to –0.5 hours per week, though the effect for men is only significant for some specifications. Recall that all these effects should be interpreted as an upper bound, as the government increased an EITC for parents with young children over the same period.

Our findings are quantitatively in between the findings of recent studies for Sweden (Lundin et al., 2008) and Norway (Havnes and Mogstad, 2011a) that find very small effects, and studies for Canada (Lefebvre and Merrigan, 2008) and Spain (Nollenberger and Rodriguez-Planas, 2012) that find rather large effects. We believe that our results are particularly relevant for many other continental European countries that face quite similar starting conditions in terms of maternal employment rates and public spending on childcare, and are considering to spend more.

In this paper we use the Dutch reform to study the relation between childcare subsidies and labour participation. However, the reform could also be used to inves-

[25] The results for the Heckman model suggest that individuals that are less likely to work are also likely to work less hours when working.

tigate a number of other relevant questions. Indeed, Baker et al. (2008) 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 the 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 needs to link the labour participation data to the childcare data, and to link these data to a tax-benefit calculator to determine the effects on government receipts and expenditures.[26] We do not have the data to do this exercise. However, the expansion seems to have been rather costly. Between 2005 and 2009 expenditures on childcare subsidies increased by 2 billion euro and spending on the EITC for secondary earners and single parents increased by 1 billion euro, 3 billion euro in total. This seems a rather large amount given that the increase in participation of mothers due to the reform was about 30 thousand women.[27] Even controlling for a trend growth in these expenditures, and subtracting the abolishment of the 'general' EITC (*Combinatiekorting*) for working parents (480 million euro in 2004, of which the major part went to working fathers), the additional public expenditure per additional working woman seems rather large.

Third, what is the effect of expanding formal childcare on children and families? There are a number of papers that use the same reforms used in the analysis of labour participation to study 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 suggest that this might be an important element to consider in the Dutch reform. Vermeer

[26]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 substantial substitution of informal by formal care.

[27]The net effect on fulltime equivalents is about 30 thousand fulltime equivalents, assuming +1.1 hours per week for women in the treatment group and –0.4 hours per week for men in the treatment group.

et al. (2005) and Kruif et al. (2009) use a large number of internationally comparable indicators for the quality of daycare,[28] 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 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 on labour supply 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.

[28] 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.

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.

Bosch, N. and van der Klaauw, B. (2012). Analyzing female labor supply - evidence from a Dutch tax reform. *Labour Economics*, 19:271–280.

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.

CPB (2012). *Macro-Economic Outlook*. CPB Netherlands Bureau for Economic Policy Analysis, The Hague.

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.

McDonald, J. and Moffitt, R. (1980). The uses of tobit analysis. *Review of Economics and Statistics*, 62(2):318–321.

Meghir, C. and Phillips, D. (2010). Labour supply and taxes. In Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., and Poterba, J., editors, *Dimensions of Tax Design: The Mirrlees Review*. Oxford University Press.

Ministry of Education, Culture and Science (2009). Kerncijfers 2005-2009. Ministry of Education, Culture and Science, The Hague.

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

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

Nollenberger, N. and Rodriguez-Planas, N. (2012). Child care, maternal employment and persistence: a natural experiment from Spain. mimeo, Universitat Autonoma de Barcelona.

OECD (2007). *Babies and Bosses, Reconciling Work and Family Life: A Synthesis of Findings for OECD Countries*. OECD, Paris.

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.

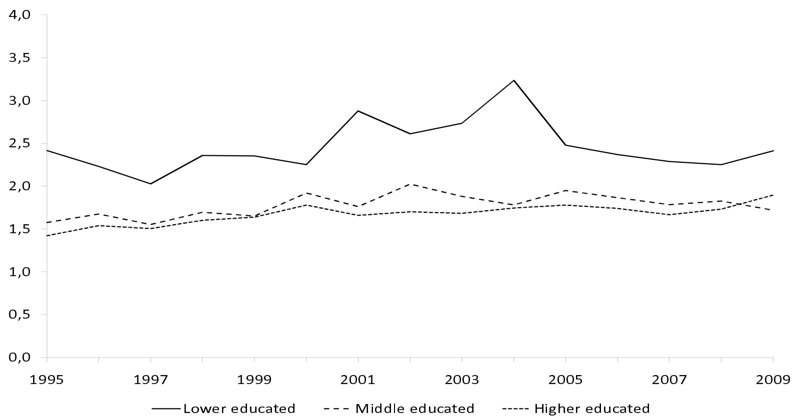
Staub, K. (2010). A causal interpretation of extensive and intensive margin effects in generalized tobit models. University of Zurich, Socioeconomic Institute, Working Paper 1012.

Van Rens, C. and Smit, F. (2011). Wachtlijsten en wachttijden kinderdagverblijven en buitenschoolse opvang - 6e meting. ITS, Radboud University, Nijmegen.

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.

Appendix

Figure A.1: Fertility by level of education



Source: own calculations using the Labour Force Survey (Statistics Netherlands).

Table A.1: Mean values for the control group: 2000-2005

	2000	2001	2002	2003	2004	2005
Participation	0.688	0.718	0.740	0.759	0.742	0.737
Worked hours	16.05	17.17	17.71	18.11	17.23	17.37
Age	43.70	43.61	43.99	44.07	44.23	44.41
Lower educated	0.443	0.418	0.408	0.386	0.375	0.350
Higher educated	0.164	0.165	0.177	0.184	0.199	0.187
Single	0.127	0.147	0.159	0.144	0.160	0.161
Immigrant	0.195	0.199	0.197	0.184	0.180	0.209
Household size	3.819	3.821	3.821	3.840	3.832	3.840
Age youngest child	14.29	14.36	14.37	14.32	14.29	14.28

Values are means weighted with sample weights. Source: Labour Force Survey (Statistics Netherlands).

Table A.2: Effect on participation rate of women

	(1) OLS	(2) OLS
Treatment 2005	0.011	
	(0.009)	
Treatment 2006	0.020**	
	(0.010)	
Treatment 2007	0.013	
	(0.010)	
Treatment 2008	0.021**	
	(0.009)	
Treatment 2009	0.026***	
	(0.005)	
Treatment 05-07		0.015**
		(0.006)
Treatment 08-09		0.023***
		(0.006)
Child 0-3	–0.090***	–0.090***
	(0.005)	(0.005)
Child 4-7	–0.069***	–0.069***
	(0.004)	(0.004)
Child 8-12	–0.027***	–0.027***
	(0.004)	(0.004)
1 child under 18	0.140***	0.140***
	(0.003)	(0.003)
2 children under 18	0.093***	0.093***
	(0.003)	(0.003)
Lower educated	–0.179***	–0.179***
	(0.003)	(0.003)
Higher educated	0.110***	0.110***
	(0.002)	(0.002)
Immigrant	–0.153***	–0.153***
	(0.003)	(0.003)
Single	–0.093***	–0.093***
	(0.004)	(0.004)
Age 20-25	0.107***	0.107***
	(0.009)	(0.009)
Age 25-30	0.008	0.008
	(0.006)	(0.006)
Age 30-35	0.066***	0.066***
	(0.005)	(0.005)
Age 35-40	0.075***	0.075***
	(0.004)	(0.004)
Age 40-45	0.056***	0.056***
	(0.003)	(0.003)
Observations	263231	263231
P-value test equal treatment effects	0.573	0.233

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

Derivation of the decomposition used in Section 6

The Heckman model consists of two equations, explaining the participation status Y_i and the number of hours H_i:

$$Y_i = \mathbf{1}(Z_i\alpha + T_i\alpha_T + V_i) \tag{A.1}$$

$$H_i = \begin{cases} X_i\beta + T_i\beta_T + U_i & \text{if } Y_i = 1 \\ 0 & \text{if } Y_i = 0 \end{cases} \tag{A.2}$$

where treatment is denoted by T_i, and Z_i and X_i are the covariates in the probit and hours equation, respectively. The error terms are assumed to follow a bivariate normal distribution with $\mathrm{E}(V_i) = \mathrm{E}(U_i) = 0$, $\mathrm{Var}(U) = \sigma^2$, $\mathrm{Var}(V) = 1$ and $\mathrm{Cov}(U, V) = \rho\sigma$.

We first derive each term of the RHS of the conventional decomposition:

$$P(Y_i = 1|T_i = 1) - P(Y_i = 1|T_i = 0) = \Phi(Z_i\alpha + \alpha_T) - \Phi(Z_i\alpha) \tag{A.3}$$

$$E(H_i|Y_i = 1, T_i = 1) = X_i\beta + \beta_T + \sigma\rho\frac{\phi(Z_i\alpha + \alpha_T)}{\Phi(Z_i\alpha + \alpha_T)} \tag{A.4}$$

$$P(Y_i = 1|T_i = 0) = \Phi(Z_i\alpha) \tag{A.5}$$

$$E(H_i|Y_i = 1, T_i = 1) - E(H_i|Y_i = 1, T_i = 0) = \beta_T + \sigma\rho\Big[\frac{\phi(Z_i\alpha + \alpha_T)}{\Phi(Z_i\alpha + \alpha_T)} - \frac{\phi(Z_i\alpha)}{\Phi(Z_i\alpha)}\Big] \tag{A.6}$$

where ϕ is the standard normal density function and Φ the standard normal cumulative distribution function. Equation (A.3) defines the change in participation due to treatment. Equation (A.4) is the average number of hours conditional on participation and treatment. As noted, this term is the average number of hours for both 'old' and 'new' participants, such that a causal interpreation is not possible. Next, (A.5) is the probability of participation without treatment and (A.6) is the change in hours conditional on participation due to treatment. This term again includes the expected number of hours conditional on participation and treatment, thereby suffering from the same problem as (A.4).

We also apply the alternative decomposition suggested by Staub (2010), which has a causal interpretation. An individual i is defined as a switcher if $(Y_i|T_i = 0) = 0$ and $(Y_i|T_i = 1) = 1$. This implies that $V_i \in S = [-Z_i\alpha - \alpha_T; -Z_i\alpha >$. Similarly, we have for a participant (someone who participates regardless of treatment), $V_i \in$

$P = [-Z_i\alpha, \infty >$. Using these definitions we can derive:

$$\pi^s = \Phi(Z_i\alpha + \alpha_T) - \Phi(Z_i\alpha) \tag{A.7}$$

$$ATE^s = X_i\beta + \beta_T + \sigma\rho\Big(\frac{\phi(Z_i\alpha + \alpha_T) - \phi(Z_i\alpha)}{\Phi(Z_i\alpha + \alpha_T) - \Phi(Z_i\alpha)}\Big) \tag{A.8}$$

$$\pi^p = \Phi(Z_i\alpha) \tag{A.9}$$

$$ATE^p = \beta_T \tag{A.10}$$

Note that both probabilities (A.7) and (A.9) are equal to those in the conventional decomposition. Both treatment effects are different since they have been conditioned on belonging to the relevant group (switchers or participants). The last term of (A.8) is the expectation of the doubly-truncated normal distribution $E(U_i|V_i \in S)$, using that $U_i = \sigma\rho V_i + \xi$ with $\xi \sim N(0, \sigma^2(1-\rho^2))$ and independent of U_i.

Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T (070) 3383 380

September 2012 | ISBN 978-90-5833-563-0