CESifo Economic Studies, 2016, 624–649
doi: 10.1093/cesifo/ifw007
Advance Access Publication Date: 31 March 2016
Original article

Unconditional Child Benefits, Mothers' Labor Supply, and Family Well-Being: Evidence from a Policy Reform

Timo Hener*

*Ifo Institute – Leibniz Institute for Economic Research at the University of Munich, Poschingerstr. 5, D-81679 Munich, Germany. e-mail: hener@ifo.de

Received 22 September 2015; Revised 28 January 2016; Accepted 18 February 2016

Abstract

In many parts of the developed world, governments devote a significant share of public funds to unconditional family cash transfers in an attempt to promote the economic well-being of households. But how successful are such policies? Germany has one of the world's most generous child benefit systems, which was subject to a major reform in the mid-1990s. This article exploits the reform using a difference-in-differences approach. The main result suggests that child benefits lead to a substantial reduction of mothers' labor supply at the intensive margin. The result implies that the policy in question is less effective at improving family finances and, consequently, expensive for the taxpayer because increases in benefit receipt are accompanied by negative labor supply responses. However, suggestive evidence seems to support that parents improve their time investment in children.

JEL classification: J18, J22, J13

Key words: child benefits, maternal labor supply, family well-being

1. Introduction

Germany, the fourth largest economy in the world by nominal GDP, dedicates an astounding EUR 200 billion each year to family-related policies. Perhaps even more staggering is the fact that 20% of this amount is spent on a single measure: child benefits. With this, public spending on child benefits exceeds Germany's entire military spending by a factor of 1.3 in 2010. Why do family-related measures play such a prominent role in government spending? On the one hand, there are concerns in many parts of the world that pay-as-you-go pension schemes will become impossible to finance once fertility levels are no longer sufficient to guarantee the long-term replacement of the population. On the other hand, many families with children are susceptible to poverty, which is likely to be an impediment to human capital formation that has long-term effects, severely limiting childhood development. Understanding what types of policies are effective in supporting families is therefore crucial for long-term economic growth and fiscal sustainability.

The question of 'whether' and 'how' monetary incentives influence family decision-making has been the subject of considerable political and scientific debate. In practice, many developed countries use cash transfer programs to prevent child poverty, promote fertility, and improve the economic well-being of families. However, there are substantial differences across countries with respect to program 'conditionalities'. The most well-studied interventions, the Earned Income Tax Credit (EITC) in the USA and the Working Families' Tax Credit (WFTC) in the UK, impose work requirements on the recipients of family cash transfers. A vast literature has shown their effectiveness in encouraging labor supply of lone mothers and fertility. In contrast, Germany mainly relies on unconditional and not means-tested child benefits in supporting families with children. Parents receive these cash transfers at least until their children reach majority. So far, the extent to which Germany's generous child benefit system affects family outcomes is not well understood. The aim of this article is to fill that gap.

This article exploits a major reform of the German child benefit system in 1996 which substantially increased the level of child benefit payments to families. I employ a difference-in-differences (DD) strategy in which couples with children constitute the treatment group, and childless couples the control group. Using data from the Socio-Economic Panel (SOEP) from 1992 to 1998, I study the extent to which the reform affected different dimensions of employment (for example, full-time employment, part-time employment, hours worked), household income, time-use, and life satisfaction.

The main message that emerges from the analysis is that Germany's child benefit system acts as labor supply disincentive for married mothers, who decreased their weekly working hours by 2.3 in response to the 1996 reform. This reduction is explained by a shift away from full-time employment to part-time employment or marginal employment. Indeed, married mothers' full-time employment decreased by 5.5 percentage points, while part-time employment and marginal employment increased by 3.6 and 2.1 percentage points, respectively. As no significant effect on non-employment can be found, the decreased labor market attachment was entirely materialized at the intensive margin. The results are robust to a number of specification and sample selection tests. I exploit that the reform yielded different treatment intensities along the income distribution by estimating heterogeneous treatment effects (TEs) using splits by potential income. I find that the labor market responses were strongest for groups with potential incomes that supposedly experienced the highest treatment intensities. I also find that labor supply disincentives are particularly pronounced in families with more than one child. There are two possible explanations for this. First, more children imply higher child benefits and thus give rise to stronger income effects. Second, families with more children might have finished their fertility plans and become more sensitive to financial incentives. Consistent with the literature, I do not find an effect on fathers' labor supply at the extensive and intensive margins. Moreover, examining household income, I do not find significant increases in families' financial situations since the increase in benefit receipt was almost entirely offset by decreases in employment. As the government tax revenues decrease due to less intensive employment, a 1 EUR child benefit increase costs more than 1 EUR from a public finance point of view. This has already been pointed out by Hoynes and Schanzenbach (2012). Despite the behavioral effects I uncover and changes in the sources of income, I do not find significant effects on mothers' life satisfaction. Descriptive evidence, however, suggests that mothers spent some more time on child care in the home. This may be regarded as direct investment into children from the time resources freed up by child benefits substituting for earnings.

The remainder of this article is organized as follows. The next section gives an overview of the literature on the effects of cash-based family policy instruments. Section 3 describes the child benefit system in Germany and its reform in 1996. Section 4 introduces the estimation strategy, while Section 5 presents the main results and Section 6 the robustness checks. Section 7 concludes.

2. Related literature

There is extensive research on the effects of monetary family benefits in countries other than Germany, particularly on conditional family benefits. For example, the labor market effects of the EITC in the USA and the WFTC in the UK have received a great deal of attention. The distinctive feature of both programs is that eligibility is explicitly tied to labor market attachment or labor earnings. Moreover, the effects of monetary benefits on fertility behavior have been analyzed, and a few studies in this field consider marriage stability, well-being, and health.

The EITC is intended to mitigate low-income families' financial constraints and to make work pay. Under this program's phase-in range of incomes, the credit is increased with income, which results in positive work incentives at the extensive and intensive margin of employment. In the program's constant and phase-out range of incomes, the EITC results in negative work incentives at the intensive margin due to, first, an income effect and then, further up the income distribution, a negative substitution effect. Eissa and Liebman (1996) exploit a 1986 reform of EITC in a DD model using Current Population Survey (CPS) data. They analyze the labor market effects for single mothers, using single women without children as a control group, and find a significant increase in the probability of employment and no effect at the intensive margin on hours worked. Meyer and Rosenbaum (2001) analyze the massive increase in single mothers' labor supply between 1984 and 1996. Considering a wide range of policy changes, including the EITC, they conclude that more generous EITC subsidies and tax changes were responsible for most of the labor supply increase, whereas other welfare program contractions, specifically the means-tested Aid to Families with Dependent Children (AFDC) and the Food Stamps program, had little effect. EITC effects on married couples are analyzed by Eissa and Hoynes (2004). Exploiting an increase of tax credits in 1993 using a DD strategy with childless married couples as the control group, they find negative effects on mothers' labor supply and no effect on fathers' labor supply. Negative effects on labor supply occur because two-adult households usually have sufficient income to fall into the phase-out range of EITC, which implies negative work incentives for the second earner. Hotz and Scholz (2006) constrain their sample to couples at the lowest end of the income distribution and exploit the differences in tax credit due to the number of children. They find that the EITC results in positive work incentives and increases household labor supply.

Explicit means-tested programs for the poorest families in the USA, AFDC, and Food Stamps, which are analyzed by Hoynes (1996) and Hoynes and Schanzenbach (2012), consistently imply negative effects on mothers' labor supply.

The WFTC in the UK implies positive work incentives, as its receipt is restricted to families in which a parent works at least 16 h a week. Negative work incentives are introduced in the income-determined phase-out region. Similar to results for the EITC, Francesconi and van der Klaauw (2007) find, in a DD analysis, that single mothers increase their labor supply; the effect is strongest at the extensive margin of employment. Gregg et al. (2009) show that single mothers' mental health and life satisfaction are increased by the WFTC. However, evidence for couples is mixed. Francesconi et al. (2009) find some positive effects on mothers' labor supply when the fathers' labor market attachment is sufficiently weak, such that the WFTC's eligibility constraint implies a positive work incentive for second earners. Furthermore, they find an increased probability of partnership dissolution due to the improved situation of single motherhood. Blundell et al. (2005), using larger Labour Force Survey (LFS) data, confirm the positive effect on single parents and find only small labor market effects for couples.

Much less is known about the effects of unconditional monetary family policies. In the European context, González (2013) analyzes the introduction of child benefits in Spain. The one-time payment of EUR 2500 was introduced in 2007, with eligibility based on the child's date of birth. This setup allows the author to use a regression discontinuity design to identify causal effects on a variety of outcomes. She finds that eligible mothers return to the labor market later after childbirth, with employment rates decreasing by 4–6 percentage points when the child is 1 year old. Furthermore, she finds positive fertility effects, as birth rates increase and abortions decrease after announcement of the reform.

To the best of my knowledge, there is only a single study on the effects of the German child benefits on family outcomes. In an interesting and closely related study, Tamm (2010) looks at married women with a working partner and analyzes the labor market responses to the 1996 child benefit reform. The author uses two waves of the large German Microcensus with base year 1995 and treatment year 1997 and exploits hours worked as a continuous variable and as a binary indicator of employment. He thoroughly investigates heterogeneity with respect to the number of children, partners' earnings, and educational attainment and checks the pre-trend by including the year 1993. He finds that mothers decreased hours worked by 1 h per week especially with intermediate education, but did not change formal employment. While this study can confirm the main result from Tamm (2010) on working hours using different data, it also adds new evidence by explicitly considering working arrangements like the widespread part-time employment, and considers a vast range of family outcomes including income, well-being, and time use.

The literature also extends to other domains of family outcomes. Several related studies have discovered that monetary family policies can influence fertility. Cohen et al. (2013) exploit the fact that Israel's child benefits vary by the number of children and over time. They find a significant increase in fertility from an increase in child benefits for higher-order children. Brewer et al. (2010) find the WFTC to have a positive effect on fertility, mainly driven by first births. Interestingly, Baughman and Dickert-Conlin (2009) find that the EITC has no effect on, or even slightly reduces, higher-order births. Milligan (2005) uses variation in a fixed child benefit transfer for Canadian families and finds a positive effect on fertility.

Dahl and Lochner (2012) extend this line of research by investigating how monetary family policies affect child outcomes. Using variation in the EITC, they find that increasing credits raises test scores in math and reading. Milligan and Stabile (2009, 2011) analyze the effects of Canadian child benefits on child outcomes. They find significant improvements in test scores and physical and mental health, with boys showing the strongest effects in education and physical health and girls showing the strongest effects in mental health.

An impressive literature has emerged around the related topic of parental leave policies and its effect on families. Concerning maternity leave in Germany, Schönberg and Ludsteck

(2014) apply a DD strategy to a reform and show that expansions led to short-run reductions in maternal employment with smaller long-run effects. Using a similar strategy, Dustmann and Schönberg (2012) find no effect on long-term child outcomes from increases in maternity leave. Lalive and Zweimüller (2009) exploit a parental leave reform in Austria by comparing mothers who gave birth just before and after the reform. They show that treated mothers returned to work much later and had substantially lower earnings in the short run; however, the effects become small in the long run. Furthermore, treated mothers decrease spacing of births and increase fertility.

3. Reform of child benefits

Ever since the 1950s, the German government has provided child benefits to families. The size of the benefit depends on the number of children. These benefits are not means-tested, easily applied for, and parents receive them every month until the child enters the labor market. The child benefits system is the most important and generous family policy program in Germany. The biggest reform to these benefits was triggered by a Supreme Court ruling handed down at the beginning of the 1990s.[1] The Supreme Court ruled as unconstitutional the tax authority's practice of child tax allowances. The judges criticized that the tax allowances did not cover basic living expenses of children. In essence, the ruling meant that the child allowances had to be vastly increased. To address equality of the law reform, the government decided to generously increase child benefits at the same time and reduce the tax deduction by the full amount of child benefits, starting in 1996.

The child benefits reform of 1996 was a combination of changes in tax allowances and child benefits. Until 1995 parents received a small amount of child benefits of EUR 432 and a tax allowance of EUR 2098 per year and per child, as shown in Table 1. From 1996 onward, parents were granted whichever option was most beneficial for them: either the increased child benefits of EUR 1224 or the increased tax allowance of EUR 3203.[2] As a consequence, every family with one child received at least EUR 1224. Only if due to the progressive tax code the deduction from the EUR 3203 child tax allowance exceeds EUR 1224, which is the case at an average tax rate above 38%, the parents would receive higher overall benefits. A simulation of the overall reform effects for a family with one respectively two children over pre-government income is depicted in Figure 1. Over a wide range of incomes, child benefits in 1996 are constant in both family types. The amount of child benefits for a family with two children is exactly twice that for families with one child. The tax deduction becomes more generous only at family incomes of around EUR 75,000 and higher, whereas the mean income of couples in our sample was about EUR 35,000 in 1994. Only in the area around the EUR 75,000 cutoff, the overall financial benefit from child tax allowance plus child benefits was slightly larger before the 1996 reform. In all other areas of the income distribution the reform of 1996 yields an increase in benefits. The financial effect of the reform is illustrated by the vertical distance between the lines for 1996 and 1995. The increase in overall benefits decreases with income up to the threshold around EUR 75,000. Thus, middle-income earners had an unambiguous increase in family transfers. Families with two children and low income received an additional child benefit before

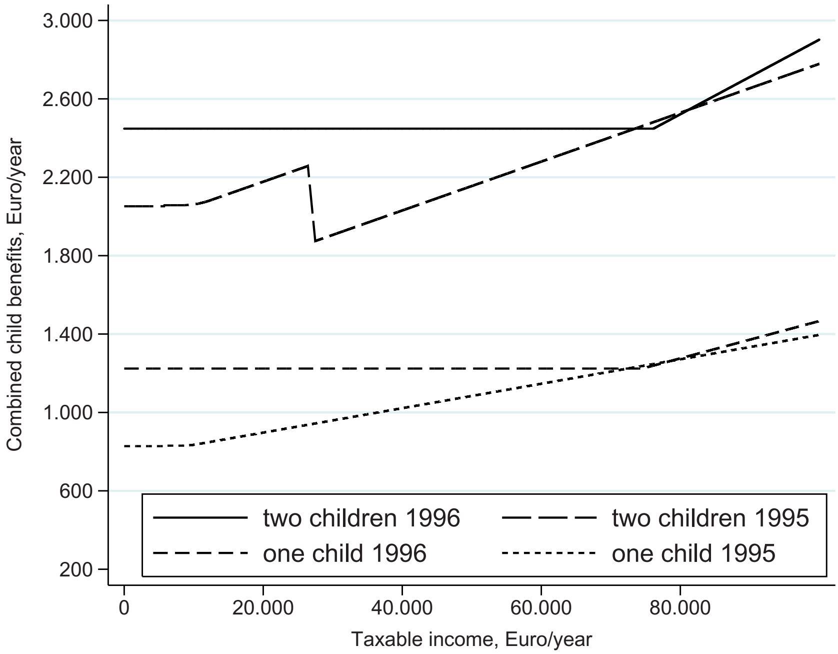
1 See, e.g., BVerfG, decree from 25/9/19922 BvL 5/91.

2 Further increases in 1997 amplified the reform effect to some extent.

Table 1. Child benefits by child order

Year	Yearly child benefits in EUR			Child tax
	1st Child	2nd Child	3rd Child	Allowance
1992–1993	432	432	864	2098
1994-1995	432	432	432	
1996	1224	1224	1836	3203
1997–1998	1344	1344	1836	3534

Notes: Columns show monthly child benefit payments per child by child order and values exchange rate adjusted values in EUR in parentheses.



Note: own simulations

Figure 1. Financial effect of child benefits and child allowance.

the reform, which yields the spike in the graph for 1995. The additional transfer was abolished with the reform. Furthermore, the German social security system implies that low-income families did not benefit from the reform at all, in contrast to what seems to be the case in Figure 1.[3] Families with very low incomes or out of the labor force receive social assistance, before and after 1996. Child benefits are subtracted from social assistance payments, such that low-income families did not experience an increase in child benefits. As social assistance includes housing benefits which depend on the location and individual

3 Before the reform, families with low incomes also received an additional EUR 33 in child benefits, which explains the flat part of the curve at the very left.

situation of a family, there is no unambiguous income threshold. A family with two children needs EUR 10,000–18,000 in post-government income to exceed social assistance payments.[4] In sum, the reform of 1996 had on average a positive effect on family finance. However, there was considerable heterogeneity. Low-income families on social assistance did not benefit, middle income families had the largest financial gain, and very high-income families had little or no increase in payments.

The maximum overall financial gain from the reform for families with two children of about EUR 570 is materialized at incomes of around EUR 27,500, which is close to the 25th percentile of the sample income distribution in 1994. The reform effect at this margin of the distribution implies a 2% increase of pre-government income. The average financial effect of the reform is also substantial as most families are found in the part of the income distribution, where the reform effect is positive.[5] Theoretically, the reform of child benefits increases non-earned income and induces an income effect on parental labor supply decisions. Thus, I expect an increase in demand for leisure and a decrease in mothers' labor supply. The substitution that is induced by the flattening of the child benefits curve is arguably small compared to the increase in child benefits. However, the effect could be heterogenous over the income distribution. At the lower end of the income distribution, the reform effect is ambiguous. The income effect is not present for recipients of social assistance and, if anything, the increased return to employment could imply an incentive to leave social assistance by entering the labor market. At the upper end, the effect should tend to be smaller than for the average earners.

4. Estimation strategy and data

I use SOEP data from 1992 to 1998.[6] The baseline sample is composed of women who have a partner and are 25–55 years old. To be included in the sample, each individual must have completed a personal interview and live in a private household.

A simple regression analysis of a child benefit effect on labor market outcomes is plagued by the usual omitted variable bias, even when making use of a reform as I do here. The bias may originate from time trends or unobserved time-variant factors that correlate with the child benefit reform and the outcome variables over time. To alleviate endogeneity problems in estimation, I use a control group that was unaffected by the reform to generate a counterfactual for the posttreatment period. In effect, this allows identifying an estimation equation that has a missing data problem: an individual can be observed treated or untreated only at a single point in time. Formally, I estimate a DD TE that is described by:

$$\tau = [[y_{t=1}|d = 1) - (y_{t=0}|d = 1)] - [(y_{t=1}|d = 0) - (y_{t=0}|d = 0)] \tag{1}$$

where d denotes the treatment group status and t defines the pre- and posttreatment period.

4 The basic monthly assistance was about EUR 256, which is multiplied by the respective household weight. A family with two children could have a household weight of 3.2, consisting of 1.8 for the two parents, 0.5 for a child aged 6 or younger and 0.9 for a child aged 14 or older. The basic assistance payment does not cover housing assistance, which is granted additionally and varies between families.

5 More than 90% of the couples in the sample had a pre-government income of less than EUR 67,000 before the reform.

6 I use SOEP v26, which can be found under DOI: 10.5684/soep.v26.

I add control variables to the estimation and apply the DD estimation equation:

$$Y = \beta_0 + \beta_1 D + \beta_2 T + \tau(D \times T) + X'\gamma + \epsilon \quad (2)$$

Y is the outcome variable, D is the treatment group indicator, T is the pre- and posttreatment indicator, and τ represents the TE as the coefficient of the interaction $(D \times T)$. X is the additional vector of control variables. ϵ is the error term. I assume that errors cluster by couples and compute cluster-robust standard errors. By restricting the age range of included couples in the sample, I can use the panel data as repeated cross-sections as in each wave new respondents enter the sample and others drop out.[7] A DD model yields causal estimates under the assumption that no time-variant unobserved heterogeneity is correlated with the treatment status and the outcome. All time-constant differences between groups and effects common to both groups are differenced out.

Like Eissa and Hoynes (2004),[8] I use ever childless couples[9] as a control group and impose the same sample restrictions. The treatment group is composed of couples in which the mother has biological children aged 4–18 years.[10] This yields a sample of 11,824 couples of which 2177 belong to the control group. I exclude couples with younger children so as to (i) avoid selection into treatment and (ii) exclude confounding reforms. In 1996, parents were granted the right to a child care slot for 3-year-olds, which raised participation in that particular age group. Four-year-olds are much less likely to be affected by the reform,as child care participation was already high for this group. In robustness checks, I narrow the age range to test for sensitivity in this respect and use different specifications of treatment and control groups.

The pretreatment period is defined as 1992–1994 and the posttreatment period as 1996–1998. I exclude 1995 to avoid early announcement effects. The law was passed on 11 October 1995, but the public was by that date already fully aware of the changes to come. The committee on finance informed parliament on 31 May 31 1995[11] about the planned law changes, and the later enacted law varied very little from this preview. During 1995, political debate over the new law was in full steam and the public could hardly have failed to understand that a major increase in child benefits was coming in 1996. Thus, 1995 should not be included in the pretreatment period.

The other major law change in the 1990s that was exclusive to families was enacted in 1992. The reform increased maternity leave from 1.5 to 3 years and had strong short-term effects on maternal employment (Schönberg and Ludsteck 2014). Thus, 1992 constitutes the beginning of the estimation sample. Changes to custody law may have effects on marriage, divorce, and fertility, as Halla (2013) shows, and thus could violate the common trend assumption, as it affects only couples with children. A major reform to custody law became effective on 1 July 1998,[12] and allowed non-married parents to have joint custody

7 The clustered standard errors therefore account for the fact of repeated observation of couples and the serial correlation of their estimation errors.

8 See also Eissa and Liebman (1996) and Francesconi and van der Klaauw (2007), who use childless single women as a control group for single mothers.

9 Couples who did not have children while in the SOEP survey or before.

10 Couples with younger children are dropped from the sample.

11 See Drucksache 13/1558, Deutscher Bundestag, 13. Wahlperiode.

12 See Bundesgesetzblatt 1997 Teil I Nr.84, p. 2942ff.

of their children. The period overlaps only little with the period of observation, as most interviews are made in the first half of the year and 1998 is the last treatment period. Early announcement of the law should not have an impact in the period of observations, as the supreme court already in 1991 ruled that joint custody must be made possible for non-married parents.[13]

Table 2 sets out means of the dependent and control variables by treatment group status in the pre-reform periods 1992–1994 along with t-test statistics for differences in means. The current employment status variable in the data set is the basis of the employment categories. Full-time and part-time employment are regular employment contracts but with different hours worked. Women and in particular mothers often have part-time jobs. This may partly be owed to the tax system which uses income splitting for married couples. In such a system, the average of the spouses' incomes determines the tax rate. As income taxes are progressive, secondary earners face higher marginal tax rates and may therefore work less. Marginal employment includes jobs with very few hours and irregular part-time jobs. Non-employed are all unemployed persons and persons out of the labor market except those in education. Work intensity differs between the treatment and control group: 38.2% of mothers are full-time employed, 25.4% work part-time, and 3.7% are only marginally employed; in the control group, 62.5% are full-time employed, 13.1% work part-time, and 2.0% are only marginally employed. The fraction of non-employed is higher in the treatment group, 31.8%, than in the control group, 21.4%. Consequently, hours worked per week are lower in the treatment group whether including zero hours or excluding them. The treatment group women average 22 h of work per week including non-working, and 33 h among working women; the control group averages 29 and 37 h, respectively.

Yearly earnings of mothers[14] in the treatment group before taxes and transfers average EUR 12,300 in 2009 prices,[15] women in the control group make EUR 20,664. The post-government household income including rental values of owned homes is weighted by the OECD-modified scale that assigns weights of 1 to the household head, 0.5 to the second adult, and 0.3 to children. This income measure includes transfers and taxes and, thus, also the financial effect of the child benefit reform. The treatment group's equivalence income is EUR 17,699, the control group averages EUR 25,417. The well-being score is the average of the overall life satisfaction question with possible answers from 0 (worst) to 10 (best). The treatment group fares slightly worse on average than the control group. The time-use categories 'cultural events' and 'volunteer work' take on the value of 1 if the respondent claims to engage in them every week or month and is 0 if the answer is less often or never. Cultural events, including concerts, theatre, and talks, are less prevalent in the treatment group, with 8.5% attending regularly, while it is 15.3% in the control group; 9.8% in the treatment group regularly engage in volunteering work, and 9.0% of the control group do the same. Looking at the control variables reveals that families are a bit more often in the middle age categories and less likely to rent a flat or house. Women in both groups are equally qualified; 18.1% of the treatment group and 18.0% of the control group attained higher education degrees defined by the International Standard Classification of Education 1997 coding. Male partners in the control group have more often attained higher

13 See Bundesverfassungsgericht Beschl. v. 07.05.1991, Az.: 1 BvL 32/88.

14 Earnings include all wages and salaries from jobs, also covering self-employment, bonuses like holiday and Christmas bonuses, over-time pay, and profit-sharing.

15 All income-related variables are corrected for the 2009 consumer price index.

Table 2. Descriptive statistics in pre-reform period

Variables	Treatment group	Control group	T-val difference
Dependent variables			
Full-time	0.382	0.625	−14.47
Part-time	0.254	0.131	8.50
Marginal	0.037	0.020	2.80
Non-employed	0.318	0.214	6.61
Hours worked incl. 0s	22.123	29.064	−10.51
Hours worked excl. 0s	33.130	37.484	−8.68
Pre-gov female earnings	12,300	20,664	−18.50
Post-gov weighted hh income	17,699	25,417	−23.98
Well-being	6.764	7.039	−4.39
Cultural events	0.085	0.153	−5.40
Volunteer work	0.098	0.090	0.61
Control variables			
Age 26–30	0.112	0.216	−9.03
Age 31–35	0.240	0.156	5.83
Age 36–40	0.292	0.161	8.64
Age 41–45	0.222	0.158	4.56
Age 46–50	0.090	0.114	−2.40
Age 50–55	0.038	0.145	−13.55
Partner's age 26–30	0.056	0.135	−9.08
Partner's age 31–35	0.165	0.128	2.90
Partner's age 36–40	0.242	0.115	8.94
Partner's age 41–45	0.241	0.137	7.27
Partner's age 46–50	0.141	0.153	−0.96
Partner's age 50+	0.125	0.243	−9.76
Renting	0.600	0.670	−4.16
ISCED higher education	0.181	0.180	0.04
Partner ISCED higher education	0.178	0.228	−3.57
Non-migrant	0.762	0.836	−5.17
Partner non-migrant	0.756	0.815	−3.90

Notes: Depicted are means in the sample by treatment group status. T-values are from regressions of the respective variable on constant and treatment group indicator.

qualifications, 22.8%, than the treatment group partners, 18.8%. In the regression analysis, indicator variables for all education categories are used. In the treatment group, 23.8% of mothers and 24.4% of fathers have a migration background; 16.4% of women and 18.5% of men in the control group have a migration background.

5. Baseline results

I start by estimating mothers' labor market responses to the child benefit reform, as all other effects are determined by this first-order reaction. I distinguish between full-time, part-time, marginal, and non-employment, and look at the hours worked response including zeros and at the intensive margin. Effects on labor market outcomes are shown in Table 3. Estimation on employment measures in Panel A, Column (1) reveals a decrease in

Table 3. Child benefit reform effect on mothers' labor market participation

Panel A	(1)	(2)	(3)	(4)	(5)	(6)
Dep var:	Full-time		Part-time		Marginal	
DD TE	−0.0633***	−0.0552**	0.0325*	0.0355*	0.0184**	0.0212**
	(0.0241)	(0.0250)	(0.0178)	(0.0194)	(0.0082)	(0.0088)
N	11,824	11,057	11,824	11,057	11,824	11,057
R^2	0.0496	0.1765	0.0172	0.0694	0.0042	0.0358
Panel B						
Dep var:	Non-employed		Hours worked			
			incl. 0s		excl.0s	
DD TE	0.0124	−0.0015	−2.7385***	−2.0120**	−3.0273***	−2.3126***
	(0.0221)	(0.0219)	(0.9597)	(0.9535)	(0.6485)	(0.6785)
N	11,824	11,057	11,342	10,618	7,802	7,300
R^2	0.0088	0.0850	0.0310	0.1769	0.0401	0.2520
Controls	No	Yes	No	Yes	No	Yes

Notes: Treatment effects from difference-in-differences estimations (DD TE) are shown in columns from separate regressions. The treatment group is composed of women with partners and children, the control group is composed of childless women with partners. Pre-reform periods are 1992–1994 and post-reform periods are 1996–1998. The reported treatment effects are coefficient estimates of the interaction between the treatment group indicator and the post-reform period indicator. Control variables are age dummies in 5-year groups for both spouses, an indicator of renting, dummies for ISCED education levels for both spouses, dummies for the migration status of both spouses, dummies for the federal state of residence and dummies for the month of the interview.

Cluster-robust standard errors with clusters at the household level in parenthesis.

*Significant at 10%; **significant at 5%; ***significant at 1% level.

full-time work of 6.3 percentage points. Adding control variables yields a TE of 5.5 percentage points. The first estimate is statistically significant at the 1% level, the second at the 5% level. A proportion of full-time work arrangements seems to be substituted by part-time employment. The latter increases by 3.3 percentage points without controls and by 3.6 percentage points with controls, both of which are only just statistically significant (at the 10% level). Mothers' marginal employment also increases, by 1.8 percentage points without controls and by 2.1 percentage points with controls, with both estimates being significant at the 5% level. There is no significant effect on non-employment with small point estimates. The increased child benefits seem to drive mothers out of full-time employment and into less-intensive employment arrangements.

The shift toward part-time work is also reflected in hours worked, as the results in Panel B of Table 3 show. Hours worked, including non-working individuals, decrease by 2.7 h per week without controls, or by 2.0 h per week with controls, where both estimates are statistically significant at conventional levels. Considering only positive hours worked in the dependent variable yields similar results. Hours worked per week then decrease by 3.0 without controls, or by 2.3 with controls, and both estimates are highly significant. The slightly larger reduction in hours worked among working mothers suggests that the main employment effect manifests at the intensive margin. This is consistent with the results from the employment measures. In general, the hours of work results are consistent with a mother's desire for a reduction in the time spent at work.

The child benefit reform as a combination of benefits and allowances entails heterogeneity in the treatment intensity that can be exploited to test for differences in the reform effects and the plausibility of the baseline results. The treatment varies over the income distribution with the largest effects at middle incomes. To exploit the treatment heterogeneity over the income distribution, I split the treatment group by income quintiles. As the contemporaneous income is endogenous in the estimation equation for employment outcomes, I use a simple variant of a Mincer earnings equation (Mincer 1974) to estimate the potential income of couples in each year.[16] The yearly distributions of potential income are then used to define the quintile identity for each couple. Recall that very low-income families could show ambiguous TEs due to the social assistance benefit system. Furthermore, high-income earners could benefit little in some range of the income distribution. If one compares the thresholds of the observed income quintiles in the last pre-reform observation with the distribution of financial benefits of the reform implied by Figure 1, it is most likely that almost all families benefitted. In fact, the observed incomes of the 20th, 40th, 60th, and 80th percentile are EUR 24,279, EUR 33,318, EUR 41,773, and EUR 53,569. Thus, the lowest quintile may exhibit lower but still positive treatments on average. The strongest treatments should be expected in the second quintile. Also, the highest quintile should still exhibit substantial treatments as most families are below the range that implies zero treatment and some fewer very high-income earners could be receiving large benefits through the child allowance. The DD estimation equation is tailored as to allow for heterogeneous TEs:

$$Y = \beta_0 + \sum_{i=1}^{5} \beta_{1i} D_i + \beta_2 T + \sum_{i=1}^{5} \tau_i (D_i \times T) + X'\gamma + \epsilon \tag{3}$$

where D_1, D_2,..., D_5 are indicator variables for the treatment group in the five income quintiles. Coefficients τ_1, τ_2,..., τ_5 indicate the quintile-specific TEs.

Table 4 shows the results of the quintile DD estimation. I find considerable heterogeneity in the TEs on full-time employment. The lowest quintile shows basically no effect of the reform, while the largest TEs are found for the second quintile. In the specification with control variables in Column (2), full-time employment decreases by 7.8 percentage points and the estimate is statistically significant at the 5% level. The reductions for the third and fifth quintile are of almost similar size and the same statistical significance level, and there is no significant reduction in the fourth quintile. While the overall picture seems to fit the predictions, the negative effect in the highest quintile may come as a surprise. The treatment intensity should be lower in this group, which suggests a smaller TE in the DD estimation. A possible explanation is that there is additional TE heterogeneity over income, which could come about if high-income families are more flexible in their work arrangement and

16 The income regression by year is the log of pre-government household income regressed on the female's and her partner's years of schooling, the female's and her partner's potential labor market experience, and its square defined as the difference between age and years of schooling, and indicator variables for the number of children. The regressions are performed on the sample as used in the baseline estimation with female age being restricted to 25 through 55. The linear prediction of this estimation is used to specify income quintiles.

Table 4. Heterogeneous treatment effects on mothers' labor market outcomes

Panel A	(1)	(2)	(3)	(4)	(5)	(6)
Dep var:	Full-time		Part-time		Marginal	
1st quintile	−0.0543	−0.0039	0.0103	0.0119	0.0076	0.0077
	(0.0336)	(0.0330)	(0.0286)	(0.0290)	(0.0131)	(0.0131)
2nd quintile	−0.1108***	−0.0776**	0.0436	0.0331	0.0419***	0.0338**
	(0.0336)	(0.0326)	(0.0289)	(0.0290)	(0.0138)	(0.0139)
3rd quintile	−0.0950***	−0.0715**	0.0449	0.0437	0.0307**	0.0295**
	(0.0343)	(0.0331)	(0.0289)	(0.0288)	(0.0130)	(0.0129)
4th quintile	−0.0485	−0.0499	−0.0107	0.0025	0.0186	0.0226*
	(0.0356)	(0.0334)	(0.0285)	(0.0280)	(0.0121)	(0.0122)
5th quintile	−0.1002***	−0.0723**	0.0926***	0.0868***	0.0093	0.0089
	(0.0345)	(0.0320)	(0.0299)	(0.0291)	(0.0135)	(0.0135)
N	10,946	10,933	10,946	10,933	10,946	10,933
R^2	0.0574	0.1792	0.0173	0.0704	0.0051	0.0371

Panel B						
Dep var:	Non-employed		Hours worked incl. 0s		Hours worked excl.0s	
1st quintile	0.0435	−0.0121	−3.5716***	−0.8401	−2.5678**	−0.9173
	(0.0329)	(0.0318)	(1.3561)	(1.2956)	(1.1241)	(1.0801)
2nd quintile	0.0245	0.0090	−4.3693***	−3.1308**	−4.6934***	−3.1100***
	(0.0328)	(0.0317)	(1.3475)	(1.2792)	(1.0786)	(1.0297)
3rd quintile	0.0155	−0.0068	−2.7981**	−1.8654	−2.8332***	−2.1806**
	(0.0313)	(0.0305)	(1.3344)	(1.2740)	(1.0227)	(0.9492)
4th quintile	0.0455	0.0287	−3.0850**	−2.7800**	−1.7644*	−2.3569***
	(0.0317)	(0.0308)	(1.3786)	(1.2877)	(0.9961)	(0.8962)
5th quintile	−0.0023	−0.0252	−3.2204**	−1.9084	−4.0621***	−2.8273***
	(0.0313)	(0.0299)	(1.3634)	(1.2533)	(0.9970)	(0.8920)
N	10,946	10,933	10,510	10,497	7,266	7,260
R^2	0.0241	0.0827	0.0510	0.1781	0.0474	0.2543
Controls	no	yes	no	yes	no	yes

Notes: Treatment effects from potential income quintile specific difference-in-differences estimations are shown in columns from separate regressions. The treatment group is composed of women with partners and children split by potential income quintile, and the control group is composed of childless women with partners. Pre-reform periods are 1992–1994 and post-reform periods are 1996–1998. The reported treatment effects are coefficient estimates of the interaction between the treatment group indicator and the post-reform period indicator. Control variables are age dummies in 5-year groups for both spouses, an indicator of renting, dummies for ISCED education levels for both spouses, dummies for the migration status of both spouses, dummies for the federal state of residence and dummies for the month of the interview.

Cluster-robust standard errors with clusters at the household level in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1% level.

less dependent on income of the—presumably second—earner. Another possibility is that the treatment intensity is indeed strong for very high income earners through the child tax allowance. Even stronger differences in the TEs become visible when looking at the alternative work arrangements. In the fifth quintile, part-time employment increases by 8.7 percentage points in the specification with control variables. The estimate is statistically highly significant. Thus, the decrease in full-time employment equivalently increases part-time employment. In the second and third quintile, the part-time estimates are half as large and not statistically significant. Marginal employment, however, increases by 3.4 percentage points and 3.0 percentage points, respectively, and the estimates are statistically significant at the 5% level. There is also a small increase in the fourth quintile, significant at the 10% level. In the fifth quintile, the estimates are close to zero as in the first quintile. None of the estimates for non-employment are statistically significant. Hours worked decrease according to the heterogeneity in work arrangements. Including all couples in Column (4) of Panel B, the estimation with controls shows negative effects in the second and fourth quintile. In Column (6), with zeros excluded, all quintiles except the first show negative effects on the intensive margin of work of −2.2 to −3.1 h per week. Despite the highest quintile substituting full-time work with mainly part-time work, while the lower quintiles to the same extent increase marginal work arrangements, the decrease in hours is similar. A possible explanation is that the differences in hours worked between work arrangements of women in high-income quintiles was initially higher because full-time employment hours are higher in that quintile.

To this point, the negative effect on female labor supply potentially counteracts the financial stimulus of child benefits. Therefore, in Table 5, I explore the response of individual earnings and household income to the reform. Columns (1) and (2) of Panel A show that the reform has no significant effect on pre-government female earnings, with negative point estimates. Splitting the effect by potential income quintile in Panel B reveals considerable heterogeneity. Results in Column (2) show negative point estimates for quintile 1–4, with the second quintile showing a highly significant estimate of EUR −2,596 and the third quintile a just significant estimate of EUR −1,588 in pre-government earnings. That the strongest effects are found in these two quintiles corresponds to the differences in treatment intensity and the results on labor supply. The fifth quintile shows a significant increase of EUR 2650 in earnings despite the fact that employment intensity was reduced. Further investigation of the issue reveals that excluding the top 1% earnings from the regression turns the coefficient for the fifth quintile small and positive without statistical significance. More importantly, policy makers are interested in the overall welfare effect of the reform. Thus, Columns (3) and (4) of Table 5 show the effect on the equivalent household income after taxes and transfers and, thus, including the direct child benefits income increase of the reform. In Panel A, the TEs are positive but small and statistically insignificant. Tests for treatment heterogeneity in Panel B do not reveal any significant effects on household resources either. The reform that should increase disposable income of families, thus, is entirely consumed by negative labor supply effects of mothers and corresponding lower earnings.

If child benefits decrease employment intensity but do not significantly improve a family's financial situation, are they used to substitute work time for other activities and lead to more favorable family outcomes? The closest measure of utility in surveys is overall life satisfaction. Columns (1) and (2) in Panel A of Table 6 show that the reform did not significantly increase life satisfaction. Similarly, Panel B shows no significant effects in any of the

Table 5. Child benefit reform effect on income

	(1)	(2)	(3)	(4)
Dep var:	Female pre-gov earnings		Post-gov weighted hh income	
Panel A				
DD TE	−656.1149	−668.7967	542.5939	90.4717
	(731.5327)	(721.6584)	(542.8881)	(485.4036)
N	11,824	11,057	11,824	11,057
R^2	0.0557	0.1843	0.0827	0.2950
Panel B				
1st quintile	−2,624.5946***	−1,137.2678	−942.4028	−248.7549
	(871.5384)	(846.9323)	(593.6563)	(537.7209)
2nd quintile	−3,165.2225***	−2,595.7319***	−255.3646	−401.0264
	(905.4001)	(871.9170)	(601.7923)	(541.2213)
3rd quintile	−2,365.4263***	−1,587.9608*	16.4361	224.0966
	(906.1513)	(862.0659)	(596.9436)	(526.5106)
4th quintile	−1,058.4716	−1,326.4459	−497.4615	−295.4141
	(1,015.2689)	(960.7723)	(794.0601)	(740.1856)
5th quintile	2,384.3423*	2,650.4246**	736.9547	454.3305
	(1295.5095)	(1250.4228)	(749.9361)	(658.7542)
N	10,946	10,933	10,946	10,933
R^2	0.1001	0.1872	0.1600	0.3252
Controls	No	Yes	No	Yes

Notes: In Panel A, treatment effects from difference-in-differences estimations (DD TE) are shown in columns from separate regressions. The treatment group is composed of women with partners and children, the control group is composed of childless women with partners. In Panel B, treatment effects from potential income quintile specific difference-in-differences estimations are shown in columns from separate regressions. The treatment group is composed of women with partners and children split by potential income quintile. Pre-reform periods are 1992–1994 and post-reform periods are 1996–1998. The reported treatment effects are coefficient estimates of the interaction between the treatment group indicator and the post-reform period indicator. Control variables are age dummies in 5-year groups for both spouses, an indicator of renting, dummies for ISCED education levels for both spouses, dummies for the migration status of both spouses, dummies for the federal state of residence and dummies for the month of the interview.

Cluster-robust standard errors with clusters at the household level in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1% level.

quintiles. Looking at time-use shows that some of the freed-up daytime is used for supposedly meaningful activities. Columns (3) and (4) show that the propensity of visiting cultural events is unaffected. Though, the propensity to engage in volunteering work, shown in Columns (5) and (6), increases significantly by 3.4 percentage points and 3.2 percentage points, respectively, in Panel A. The quintile-specific effects in Panel B are revealing heterogeneity that fits well to the earlier results. The first quintile shows no reaction and the second quintile shows the largest effect of 6.9 percentage points significant at the 1% level. The just significant effects for the third and fourth quintile of 4.0 percentage points are somewhat smaller and nothing is found in the fifth quintile.

Even though some of the time may be used to increase welcome activities, family policy is foremost concerned with child well-being. If the labor supply reaction to the reform is

Table 6. Child benefit reform effect on mothers' well-being and time use

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var:	Life satisfaction		Cultural events		Volunteer work	
Panel A						
DD TE	0.0378	0.0209	−0.0088	−0.0045	0.0343**	0.0316**
	(0.0839)	(0.0863)	(0.0176)	(0.0171)	(0.0147)	(0.0153)
N	11,792	11,032	9874	9210	9848	9185
R^2	0.0036	0.0853	0.0091	0.0900	0.0028	0.0664
Panel B						
1st quintile	0.0175	0.0573	−0.0141	−0.0008	0.0059	0.0148
	(0.1271)	(0.1226)	(0.0196)	(0.0194)	(0.0181)	(0.0185)
2nd quintile	−0.1334	−0.1636	0.0108	0.0145	0.0793***	0.0692***
	(0.1263)	(0.1220)	(0.0209)	(0.0208)	(0.0207)	(0.0203)
3rd quintile	0.1133	0.1144	−0.0173	−0.0111	0.0455**	0.0404*
	(0.1235)	(0.1176)	(0.0209)	(0.0204)	(0.0213)	(0.0211)
4th quintile	−0.0511	0.0087	−0.0134	−0.0177	0.0421*	0.0400*
	(0.1222)	(0.1177)	(0.0238)	(0.0231)	(0.0244)	(0.0240)
5th quintile	0.1073	0.0453	0.0032	−0.0003	0.0177	0.0093
	(0.1152)	(0.1091)	(0.0269)	(0.0263)	(0.0259)	(0.0260)
N	10,921	10,908	9113	9103	9088	9078
R^2	0.0051	0.0863	0.0387	0.0915	0.0201	0.0688
Controls	No	Yes	No	Yes	No	Yes

Notes: In Panel A, treatment effects from difference-in-differences estimations (DD TE) are shown in columns from separate regressions. The treatment group is composed of women with partners and children, and the control group is composed of childless women with partners. In Panel B, treatment effects from potential income quintile specific difference-in-differences estimations are shown in columns from separate regressions. The treatment group is composed of women with partners and children split by potential income quintile. Pre-reform periods are 1992–1994 and post-reform periods are 1996–1998. The reported treatment effects are coefficient estimates of the interaction between the treatment group indicator and the post-reform period indicator. Control variables are age dummies in 5-year groups for both spouses, an indicator of renting, dummies for ISCED education levels for both spouses, dummies for the migration status of both spouses, dummies for the federal state of residence, and dummies for the month of the interview.
Cluster-robust standard errors with clusters at the household level in parenthesis.
*Significant at 10%; **significant at 5%; ***significant at 1% level.

translated into more time investment in children, it could be welfare improving nonetheless. As the time used for child care has no plausible correspondence in the control group, simple mean over time graphs by potential income quintile are used to investigate this matter. Figure 2 shows the means of the number of hours used caring for the child on a typical weekday. After in most groups the means seem stable in the pre-reform periods, some increases can be observed post-reform that are compatible with early announcement effects in 1995. Especially in the second quintile, the one with the largest treatment intensity, child care time increases by more than half an hour. Also, the third quintile seems to experience an increase in child care time. To a lesser extent, the first and fifth quintile also show some increase, but it seems rather short-lived or not timed well with the reform, while the fourth quintile reports stable hours over the whole period of observation. Exploiting the differences in treatment intensity, Figure 3 allows a closer look at the distribution of child care hours. The graph shows the differences in the density of child care time use for treated

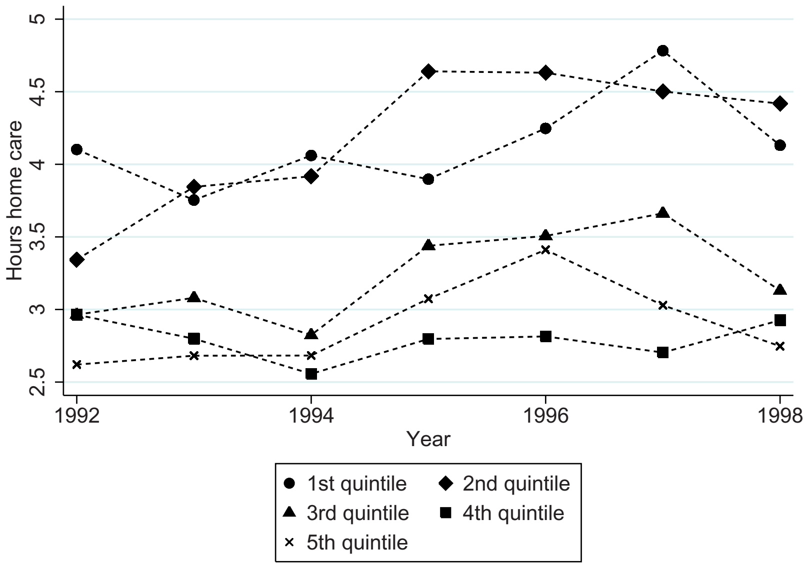


Figure 2. Child care time use by income quartiles.

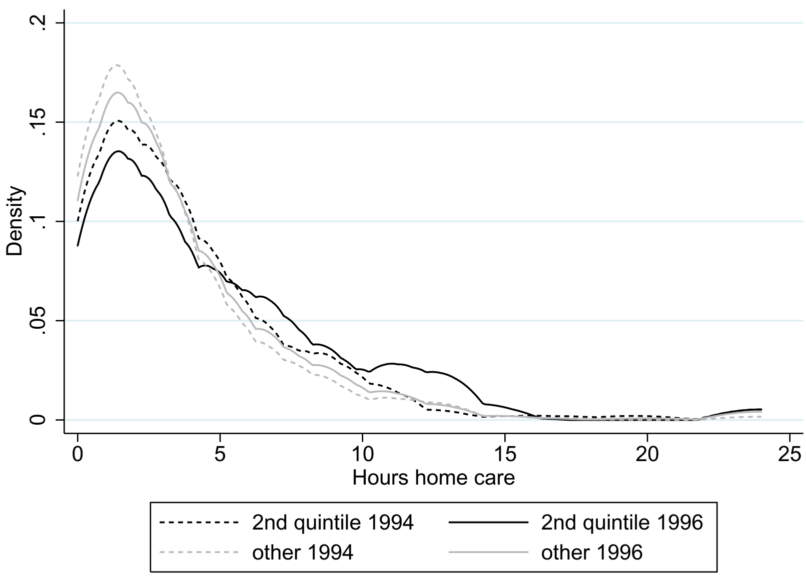


Figure 3. Density of child care time use by treatment intensity. *Notes:* Density plot of the hours of home care with epanechnikov kernel and half-bandwidth of 1. The sample is composed of the treatment group in years 1994 and 1996. The dashed lines depict the densities in 1994 and the solid lines depict the densities in 1996. The black lines represent densities for the second income quintile, the one with the greatest treatment intensity, and the gray lines represent the densities of the other income quintiles.

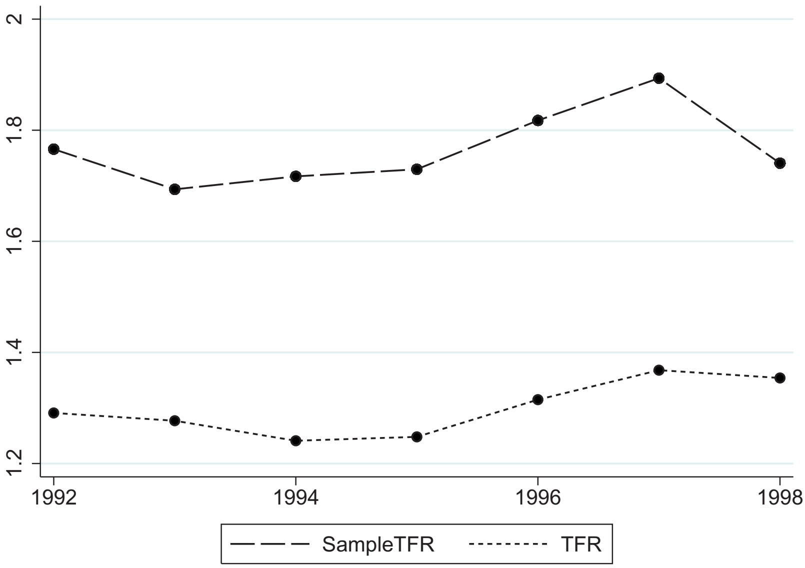


Figure 4. Total fertility rate. *Notes:* TFR denotes the total fertility rate from the Federal Statistical Office and the sample TFR is the TFR computed in the SOEP sample.

mothers between 1994 and 1996 within the second quintile of potential income and within the remaining quintiles. It is evident that low hours of child care decrease in both groups after the reform. However, mothers in the second quintile also less often care up to 5 h and markedly increase the density around 7 h and 11–14 h. The descriptive evidence is of course not a formal test,[17] although the overall picture seems compatible with a substitution from hours worked toward time invested in children.

Testing for effects of the reform on fertility is similarly impaired by the absence of a plausible control group. As the reform affected all people, none would be excluded when having a child, whether it is the first- or a higher-order birth. Moreover, the survey might be too small to recover fertility trends correctly. First, Figure 4 shows the total fertility rate (TFR) from the Statistical Office and the TFR from the SOEP data sample[18] over the years of observation. The latter yields a higher value, as only couples are included in the sample, but both lines follow the same trend. It is apparent that the TFR increases in 1996. The post-reform means of the official TFR figure are always higher than the pre-reform means. The increase is moderate in 1996, which is consistent with a normal fertility lag, as the reform could have taken effect some time during 1995. The increase from 1995 to 1997 is 0.1 children per woman, indicating a substantial increase. Further indication of fertility effects comes from Figure 5 that extracts the TFR of the second potential quintile as the most affected group.[19] The graph shows little variation in the other four quintiles, but fertility of

17 An unreported regression analysis using the second quintile as the treatment group yields small positive but insignificant estimates on the mean hours in child care.

18 Approximated by the individual birth probability multiplied by 30, the number of age-years in the TFR statistics. The sample restriction is now having a partner and the woman being under age 50.

19 Using all five quintiles separately would yield very volatile time lines due to the low number of births in subsamples.

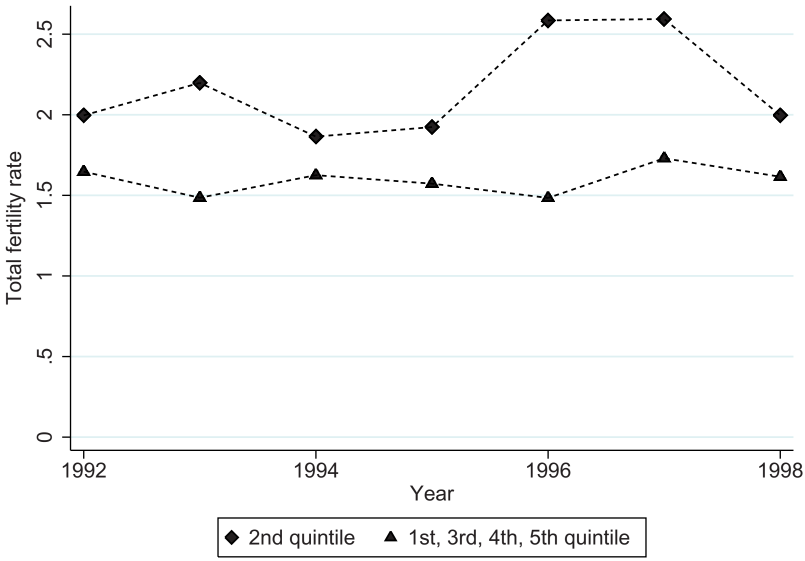


Figure 5. TFR by income quintiles.

the second quintile clearly increases in 1996 and 1997 before it returns to the pre-reform level in 1998. This descriptive evidence is consistent with international evidence of positive fertility effects from monetary benefits (see, for example, Cohen et al. 2013; Milligan 2005).

6. Tests of robustness

To check the robustness of the main results, I explore some possible confounding factors and check for robustness to misspecification of the main results on full-time and part-time employment. To account for state-specific employment trends that may be correlated with state-specific trends in family formation, I include state-specific time trends in the baseline DD model. Results in Columns (1) and (2) of Table 7 show results for full-time and part-time employment using this specification. The negative effect on full-time employment is 5.2 percentage points without controls and 5.3 percentage points with controls, with statistical significance at the 5 % level. Thus, the baseline result on full-time employment is confirmed and robust to state time trends. The estimate on part-time employment with controls is 3.5 percentage points, significant at the 10% level. This is very similar to the baseline result. Only without controls the estimate becomes insignificant. The unreported estimates on the other labor market responses are all very robust to the inclusion of state time trends.

To further test robustness, I allow trends specific to the treatment and the control group prior to treatment and a trend shift posttreatment. Following Francesconi and van der Klaauw (2007) and Francesconi et al. (2009), I apply the extended DD estimation equation:

$$Y = \alpha_1 + \alpha_2 d + (\alpha_{31} + \alpha_{32} d)t + [\alpha_{41} + \alpha_{42}(t - s)]I(t \geq s) + \beta d I(t \geq s) + X'\gamma + \epsilon \quad (4)$$

where t is a time trend, I(.) is an indicator function that takes on the value of unity for $(t \geq s)$, where s is the reform year, and zero otherwise. All other regressors are defined as in

Table 7. Robustness checks of baseline results

Robustness:	(1) State trends	(2)	(3) Group trends	(4)	(5) SE Clustering	(6)
Panel A						
Dep var:	Full-time employment					
DD TE	−0.0518**	−0.0529**	−0.0709*	−0.0206	−0.0633*	−0.0552**
	(0.0248)	(0.0250)	(0.0404)	(0.0414)	(0.0324)	(0.0240)
Pre-treatment interaction			0.0017	−0.0086		
			(0.0094)	(0.0098)		
N	11,824	11,057	11,824	11,057	11,824	11,057
R^2	0.1394	0.1777	0.0500	0.1767	0.0496	0.1765
Panel B						
Dep var:	Part-time employment					
DD TE	0.0284	0.0346*	0.0637**	0.0472	0.0325**	0.0355**
	(0.0183)	(0.0196)	(0.0301)	(0.0320)	(0.0124)	(0.0166)
Pre-treatment interaction			−0.0076	−0.0029		
			(0.0074)	(0.0079)		
N	11,824	11,057	11,824	11,057	11,824	11,057
R^2	0.0463	0.0714	0.0173	0.0694	0.0172	0.0694
Controls	No	Yes	No	Yes	No	Yes
State interactions	Yes	Yes	No	No	No	No
Trend interactions	No	No	Yes	Yes	No	No

Notes: In columns (1) and (2) regressions control for interactions of federal state dummies and the treatment group. In columns (3) and (4), regressions use the amended trends DD model. In columns (5) and (6), regressions use state-level clustering for calculation of standard errors. Treatment effects from difference-in-differences estimations (DD TE) are shown in columns from separate regressions. The treatment group is composed of women with partners and children, and the control group is composed of childless women with partners. Pre-reform periods are 1992–1994 and post-reform periods are 1996–1998. The reported treatment effects are coefficient estimates of the interaction between the treatment group indicator and the post-reform period indicator. Control variables are age dummies in 5-year groups for both spouses, an indicator of renting, dummies for ISCED education levels for both spouses, dummies for the migration status of both spouses, dummies for the federal state of residence, and dummies for the month of the interview.
Cluster-robust standard errors with clusters at the household level in parenthesis.
*Significant at 10%; **significant at 5%; ***significant at 1% level.

the baseline DD model. In addition to the standard DD, I thus impose a general time trend estimated by α_{31}, a treatment-specific time trend deviation estimated by α_{32} and a general posttreatment trend estimated by α_{42}; α_{41} is the posttreatment shift parameter and β is the TE estimate. The estimation therefore allows group-specific time trends prior to treatment and a general trend shift posttreatment. The estimation also facilitates a test for the equality of pretreatment trends, similar to a placebo test, by the coefficient on α_{32}.

Columns (3) and (4) report results of the TE from the extended DD estimation. The effect on full-time employment is 7.1 percentage points without controls, just barely statistically significant. Including controls drops the coefficient to 2.1 percentage points and it loses significance. However, the interaction effect on the pretreatment trend, depicted in the row below, is statistically insignificant in both specifications and even changes sign. Thus, the common trend assumption appears to be valid for full-time employment. In Columns (3) and (4) of Panel B, we see the effects on part-time employment derived from the extended DD model. Part-time work increases by 6.4 percentage points when estimated without control variables and the estimate is statistically significant at the 5% level. With control variables, the effect loses significance and shows a point estimate of 4.7 percentage points. The pretreatment trends are statistically insignificantly different between treatment and control group, suggesting that the common trend assumption is valid for part-time employment as well. Note, however, that trend specifications based on few waves are very sensitive to fluctuations. Results based on the extended trend specification DD should be interpreted with caution as interpolation of small errors in trends is exacerbated over time.[20]

Lastly, I explore the sensitivity of standard errors to clustering. Hypothesis testing might over-reject the null hypothesis if errors are correlated. As repeated observations are certainly correlated, I cluster on couples in the baseline results. However, correlation might be induced between individuals as well at a higher level of aggregation. Thus, it might be advisable to cluster at the highest level of aggregation at which I would expect correlation, here, states (Moulton 1990; Pepper 2002; Bertrand et al. 2004). Results in Columns (5) and (6) report baseline results with clustering on 16 states instead of individuals. The results continue to be statistically significant. The standard errors for one of the estimates increase in Column (5) of Panel A and decrease for three of the estimates in Column (6) of Panel A and both in Panel B compared to the baseline. There appears to be no tendency toward over-rejection with individual clustering. This is unsurprising as, in this exercise, the treatment is not correlated with higher aggregation units, as it would be with state-specific treatments.

I exploit some additional variation in the treatment intensity to test for heterogeneity in the TE. Child benefits are paid per child and thus the treatment intensity is the larger the more children a family has. And a bigger treatment should yield larger TEs. Moreover, families with different numbers of children might react differently depending on whether or not they have reached their desired family size. In Table 8, I show results on female labor market responses by number of children. For full-time employment in Columns (1) and (2) of Panel A, I find small, negative, and insignificant estimates for mothers with one child. The TE grows to 7.5 percentage points without controls and 7.0 percentage points with controls for families with two children and becomes highly statistically significant at the 1% level. For families with three children, the effect is similar—around 7 percentage points and statistically significant at the 5% level. TEs on part-time employment are small, positive, and statistically insignificant for families with one child. Estimates of 4.4 percentage points or 4.5 percentage points become statistically significant for families with two children. For larger families, the effects decrease again and become insignificant. Overall, the prediction that larger child benefit increases should yield larger shifts from full-time to part-time employment is validated. There is no significant effect on non-working in any of the family sizes, as shown in Columns (5) and (6), although families with three children show point estimates of almost 5 percentage points. Marginal employment also increases for families with one child, by 2.1 percentage points or 3.0 percentage points, significant at

20 The hours worked results in fact become small and insignificant, when using the full trend specification with all controls. A likely reason is the trend's sensitivity to small deviations in the last pre-reform period.

Table 8. Reform effect by number of children on mothers' labor market outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Dep var:	Full-time		Part-time		Marginal	
1 child	−0.0363	−0.0190	0.0161	0.0184	0.0219**	0.0296***
	(0.0301)	(0.0311)	(0.0248)	(0.0263)	(0.0110)	(0.0114)
2 children	−0.0746***	−0.0697***	0.0435**	0.0453**	0.0259***	0.0275***
	(0.0264)	(0.0270)	(0.0204)	(0.0217)	(0.0095)	(0.0101)
3 children+	−0.0695**	−0.0724**	0.0254	0.0254	−0.0055	−0.0009
	(0.0316)	(0.0321)	(0.0278)	(0.0289)	(0.0135)	(0.0141)
N	11,076	10,349	11,076	10,349	11,076	10,349
R^2	0.0525	0.1763	0.0207	0.0718	0.0062	0.0383
Panel B						
Dep var:	Non-employed		Hours worked			
			incl. 0s		excl.0s	
1 child	−0.0022	−0.0293	−1.9628*	−0.8248	−2.4782***	−2.2741***
	(0.0269)	(0.0271)	(1.1631)	(1.1696)	(0.8346)	(0.8317)
2 children	0.0064	−0.0019	−3.0775***	−2.5001**	−3.6819***	−2.7097***
	(0.0245)	(0.0245)	(1.0452)	(1.0367)	(0.7240)	(0.7346)
3 children+	0.0489	0.0458	−2.9137**	−2.8077**	−1.6170	−1.3379
	(0.0323)	(0.0326)	(1.3176)	(1.3280)	(1.0759)	(1.0996)
N	11,076	10,349	10,623	9937	7539	7052
R^2	0.0147	0.0776	0.0378	0.1754	0.0444	0.2555
Controls	No	Yes	No	Yes	No	Yes

Notes: Treatment effects from difference-in-differences estimations (DD TE) are shown in columns from separate regressions. The treatment group is composed of women with partners and children and split by the number of children, the control group is composed of childless women with partners. Pre-reform periods are 1992–1994 and post-reform periods are 1996–1998. The reported treatment effects are coefficient estimates of the interaction between the treatment group indicator and the post-reform period indicator. Control variables are age dummies in 5-year groups for both spouses, an indicator of renting, dummies for ISCED education levels for both spouses, dummies for the migration status of both spouses, dummies for the federal state of residence, and dummies for the month of the interview.

Cluster-robust standard errors with clusters at the household level in parenthesis.

*Significant at 10%; **significant at 5%; ***significant at 1% level.

conventional levels. Effects are of similar magnitude and precision for families with two children. Mothers of three children show no change in marginal employment. Thus, the weakest attachment to the labor market is not sensitive to treatment size.

In Panel B of Table 8, I show the family size heterogeneity on hours worked outcomes. Mothers of one child show only marginal decreases in hours worked including zeros and significant decreases between 2.5 and 2.7 h at the intensive margin. However, for both hours worked variants, there is a significant decrease for mothers of two children. In both categories, effect sizes are also larger for mothers of two children than for mothers of one child. Families with three children decrease their hours worked mainly at the extensive margin, hours worked including zeros, but only insignificantly at the intensive margin. This result is consistent with the effects on employment categories. As full-time work is decreased,

Table 9. Child benefit reform effect on fathers' labor market outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Dep var	Full-time		Part-time		Marginal	
DD TE	0.0079	0.0195	−0.0015	−0.0013	0.0003	−0.0021
	(0.0223)	(0.0222)	(0.0069)	(0.0072)	(0.0050)	(0.0050)
N	11,344	11,057	11,344	11,057	11,344	11,057
R^2	0.0107	0.0861	0.0040	0.0185	0.0022	0.0098
Panel B						
Dep var	Non-employed		Hours worked incl. 0s		excl.0s	
DD TE	−0.0099	−0.0200	−0.4932	0.2929	−1.0698*	−0.6808
	(0.0202)	(0.0198)	(1.0402)	(1.0189)	(0.5676)	(0.5853)
N	11,344	11,057	10,813	10,544	9,683	9,450
R^2	0.0062	0.0916	0.0088	0.1121	0.0044	0.0840
Controls	No	Yes	No	Yes	No	Yes

Notes: Treatment effects from difference-in-differences estimations (DD TE) are shown in columns from separate regressions. The treatment group is composed of men with partners and children, the control group is composed of childless men with partners. Pre-reform periods are 1992–1994 and post-reform periods are 1996–1998. The reported treatment effects are coefficient estimates of the interaction between the treatment group indicator and the post-reform period indicator. Control variables are age dummies in 5-year groups for both spouses, an indicator of renting, dummies for ISCED education levels for both spouses, dummies for the migration status of both spouses, dummies for the federal state of residence, and dummies for the month of the interview.
Cluster-robust standard errors with clusters at the household level in parenthesis.
*Significant at 10%; **significant at 5%; ***significant at 1% level.

it may even be substituted by non-employment, which would not show up at the intensive margin of hours worked.

Family policy studies typically investigate female labor supply, as it is generally lower than that of males and more responsive to policies. However, in principle, fathers' labor supply could similarly respond to a child benefit reform, as the income effect should work through household-level finances. Also, males could respond due to changes in the intra-household time allocations when females reduce labor supply. Therefore, I show results for male partners in Table 9. None of the employment status outcomes shows a significant response to the child benefit reform and all estimates are small. There is some weak indication of a reduction in hours worked excluding zeros. However, the result is only just significant and vanishes when including control variables.

A possible concern with the estimation is that the child care reform mentioned earlier might not have only increased the participation of 3-year-olds but also that of 4-year-olds. As a robustness check, I therefore use families in the treatment group whose youngest child is at least 7 years old and thus of school age. The results confirm the baseline estimates, with just significant negative effects on full-time employment and significant positive effects on marginal employment of mothers. Effect sizes are comparable to the baseline as well.

Result tables are available in the Supplementary Appendix. In a sample of single women and single mothers, I find a similar pattern of results; however, the precision of the estimates is considerably lower and, thus, are not included here.

7. Conclusion

Whether and how monetary family benefits affect the lives of families depends largely on the conditionality of the benefit. Conditional benefits requiring parents to work to be eligible have been shown to be effective in keeping single mothers in the labor market and to have positive side effects like higher fertility (Eissa and Liebman 1996; Meyer and Rosenbaum 2001; Francesconi and van der Klaauw 2007; Gregg et al. 2009; Blundell et al. 2005; Brewer et al. 2010). Unconditional monetary benefits have been studied less and, nonetheless, play an important role in many countries. In contrast to conditional benefits, the work incentives of unconditional benefits are typically negative and thus suggest a very different conclusion.

In this article, I showed that child benefits, even if unconditional, have significant behavioral effects. Mothers tend to work less, which is mainly due to reductions in work intensity. Increased propensities of part-time and marginal work arrangements substitute full-time contracts and yield fewer hours worked. The reduction in employment is of concern as it counteracts the principal motive behind family cash transfers. That is, child benefits do not appear to increase family income as the induced employment reductions reduce earned family income.

This renders child benefits an expensive policy as the tax revenue lost due to employment reduction elevates the costs of providing the benefit. Moreover, it calls into question whether unconditional family benefits are an advisable policy tool. Especially, as an alternative policy, subsidizing child care has favorable labor market effects on mothers (Bauernschuster and Schlotter 2015) and, thus, is much less of a budgetary burden.

However, it is not all bad news. Although mothers do not appear to be happier after the child benefit reform, there is suggestive evidence that they spend more time with their children. If child benefits free up time otherwise allocated to work, and that time is invested in children, unconditional financial benefits could indeed be a valuable policy tool. Future research thus could consider more directly the effects of financial family policy on time investment in children.

Supplementary Material

Supplementary Material is available at *Cesifo* online.

Acknowledgements

I thank Helmut Rainer, Stefan Bauernschuster, Natalia Danzer, Christian Holzner, and Janina Reinkowski for their work on an earlier related government report for the Federal Ministry of Family, Senior Citizens, Women and Youth, and the Federal Ministry of Finance.

References

Bauernschuster, S. and M. Schlotter (2015), "Public Child Care and Mothers' Labor Supply—Evidence from Two Quasi-Experiments," *Journal of Public Economics* 123, 1–16.

Baughman, R. and S. Dickert-Conlin (2009), “The Earned Income Tax Credit and Fertility”, *Journal of Population Economics* 22, 537–563.

Bertrand, M., E. Duflo, and S. Mullainathan (2004), “How Much Should We Trust Differences-In-Differences Estimates?”, *The Quarterly Journal of Economics* 119, 249–275.

Blundell, R., M. Brewer, and A. Shephard (2005), “Evaluating the Labour Market Impact of Working Families’ Tax Credit using Difference-in-Differences,” HMRC Working Papers 4, HM Revenue and Customs, London, UK.

Brewer, M., A. Ratcliffe, and S. Smith (2010), “Does Welfare Reform Affect Fertility? Evidence from the UK”, *Journal of Population Economics* 25, 1–22.

Cohen, A., R. Dehejia, and D. Romanov (2013), “Financial Incentives and Fertility”, *Review of Economics and Statistics* 95, 1–20.

Dahl, G. B. and L. Lochner (2012), “The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit”, *The American Economic Review* 102, 1927–1956.

Dustmann, C. and U. Schönberg (2012), “Expansions in Maternity Leave Coverage and Children’s Long-Term Outcomes”, *American Economic Journal: Applied Economics* 4, 190–224.

Eissa, N. and H. W. Hoynes (2004), “Taxes and the Labor Market Participation of Married Couples: The Earned Income Tax Credit”, *Journal of Public Economics* 88, 1931–1958.

Eissa, N. and J. B. Liebman (1996), “Labor Supply Response to the Earned Income Tax Credit”, *The Quarterly Journal of Economics* 111, 605–637.

Francesconi, M., H. Rainer, and W. van der Klaauw (2009), “The Effects of In-Work Benefit Reform in Britain on Couples: Theory and Evidence”, *Economic Journal* 119, F66–F100.

Francesconi, M. and W. van der Klaauw (2007), “The Socioeconomic Consequences of In-Work Benefit Reform for British Lone Mothers”, *Journal of Human Resources* 42, 1–31.

González, L. (2013), “The Effect of a Universal Child Benefit on Conceptions, Abortions, and Early Maternal Labor Supply”, *American Economic Journal: Economic Policy* 5, 160–188.

Gregg, P., S. Harkness, and S. Smith (2009), “Welfare Reform and Lone Parents in the UK”, *Economic Journal* 119, F38–F65.

Halla, M. (2013), “The Effect of Joint Custody on Family Outcomes”, *Journal of the European Economic Association* 11, 278–315.

Hotz, V. J. and J. K. Scholz (2006), “Examining the Effect of the Earned Income Tax Credit on the Labor Market Participation of Families on Welfare,” Working Paper 11968, National Bureau of Economic Research, Cambridge, MA.

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

Hoynes, H. W. and D. W. Schanzenbach (2012), “Work Incentives and the Food Stamp Program”, *Journal of Public Economics* 96, 151–162.

Lalive, R. and J. Zweimüller (2009), “How does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments”, *The Quarterly Journal of Economics* 124, 1363–1402.

Meyer, B. D. and D. T. Rosenbaum (2001), “Welfare, The Earned Income Tax Credit, and the Labor Supply of Single Mothers”, *The Quarterly Journal of Economics* 116, 1063–1114.

Milligan, K. (2005), “Subsidizing the Stork: New Evidence on Tax Incentives and Fertility”, *Review of Economics and Statistics* 87, 539–555.

Milligan, K. and M. Stabile (2009), “Child Benefits, Maternal Employment, and Children’s Health: Evidence from Canadian Child Benefit Expansions”, *American Economic Review* 99, 128–132.

Milligan, K. and M. Stabile (2011), “Do Child Tax Benefits Affect the Well-being of Children? Evidence from Canadian Child Benefit Expansions”, *American Economic Journal: Economic Policy* 3, 175–205.

Mincer, J. (1974), *Schooling, Experience, and Earnings*, Columbia University Press, New York, NY.

Moulton, B. R. (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *The Review of Economics and Statistics* 72, 334–338.

Pepper, J. V. (2002), "Robust Inferences from Random Clustered Samples: An Application using Data from the Panel Study of Income Dynamics", *Economics Letters* 75, 341–345.

Schönberg, U. and J. Ludsteck (2014), "Expansions in Maternity Leave Coverage and Mothers Labor Market Outcomes after Childbirth", *Journal of Labor Economics* 32, 469–505.

Tamm, M. (2010), "Child Benefit Reform and Labor Market Participation," *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)* 230, 313–327.