Mothers working during preschool years and child skills. Does income compensate?*

Cheti Nicoletti

Department of Economics, University of York; ISER University of Essex; IZA

Kjell G. Salvanes

Norwegian School of Economics, CEPR, CESifo, HCEO and IZA

Emma Tominey

Department of Economics, University of York; HCEO; IZA

January 10, 2022

Abstract

Increasing mothers' labour supply in a child's preschool years may reduce time investments, yielding a negative direct effect on mid-childhood and teenage outcomes. But as mothers' work hours increase, income will rise. Can income compensate for the negative effect of hours? Our mediation analysis exploits exogenous variation in both mothers' hours and family income. Results suggest a negative, insignificant direct effect from increasing mother's hours on child test scores. However the positive mediating effect of income creates a positive total effect on test scores of 26% of a standard deviation for 10-hours increase in mother's weekly hours in preschool years.

Keywords: Child development, test scores, parental investments
JEL codes: J13, J24, I22, I24

*We thank James Heckman, Alessandra Voena and Marco Fransesconi for comments and participants in seminars at EALE/SOLE 2020, The University of Chicago, Workshop on the Economics of Family, Gender and Old-Age London School of Economics, University of Surrey, University of Sheffield, University of Padua, University of Amsterdam, University of Bristol, London School of Economics, University of Bath, DIW Berlin, LUISS Rome and the WOLFE York 2018 workshop.

1 Introduction

Parental resources of time and money have an important role in explaining the transmission of disadvantage across generations. Children born in more disadvantaged families grow up with lower income resources, poorer parental time investments and worse educational achievements than children born in more privileged families (see McLanahan 2004; Guryan, Hurst, and Kearney 2008; Kalil, Ryan, and Corey 2012, Dotti Sani and Treas 2016). An important target of policies aiming to raise parental time investments and household income is the mother's labour supply. For example maternity leave policies allow working mothers to invest more time with newborn children, while in-work benefits offer financial incentives for mothers to work, to reduce poverty in working families with children. We ask in this paper whether increasing mothers' labour supply in a child's preschool years has a positive or negative effect on child outcomes in childhood and teen years.

There are several challenging issues in answering this question. First, there are two channels though which increased maternal labour supply affects children. On the one hand, an increase in mothers' work hours may reduce the time available to spend with her child which may harm her child's human capital development. On the other hand, an increase in working hours will raise household income and therefore the resources available to invest in children. To understand how an increase in mothers' work hours affect their children, it is therefore necessary to consider not only the direct effect of mothers' hours worked on child outcomes through a potential reduction in time investments in children, but also the indirect effect through an increase in income. Secondly, we have to deal with the endogeneity of both work hours and income.

While many papers have estimated the effect of mothers' labour supply decisions on child outcomes, there is a void in our knowledge of the mechanisms through which mothers' work hours affect children's outcomes.[1]

[1] Blau and Grossbergm (1992) and Ermisch and Francesconi (2013) estimate a negative effect of mothers' working hours on child outcomes whilst Bernal and Keane (2011) estimates a negative effect of childcare on outcomes in single parent households. A set of papers analysing the effect of extensions to maternity leave on child outcomes have found positive effects for expansions from no paid leave to 12 weeks (Carneiro, Løken, and Salvanes 2015) or effects just for high educated mothers (Liu and Skans 2010, Danzer and Lavy 2013) but otherwise small or no effects on children of extending the parental leave period beyond the first months (Rasmussen 2010, Baker and Milligan 2010, Dustmann and Schönberg 2012, Baker and Milligan 2015, Dahl, Løken, Mogstad, and Salvanes 2016, Huebener, Kuehnle, and Spiess 2019).

We address this research question by (i) estimating the causal total effect of an increase in mothers' hours worked during pre-school years on children school achievements; (ii) exploring the mechanism behind the effect of mother's labour hours by decomposing the causal total effect into the part explained by a time investment reduction (the direct effect) and the part explained by an increase in income (the mediator effect). We leverage Norwegian population wide administrative data in the analysis, where we focus on first-born children and assess the effect of mothers' work decisions in the years before the child starts school on the child's educational achievements measured by school test scores at age 11 and 15. Household income and mother's work hours in the pre-school period are averaged over the 5 years from when the child is 1 to 5 years old.[2]

Our first contribution is empirical, aiming to better understand the causal mechanisms explaining the total effect of an increase in mothers' work hours in preschool years on first-born children's educational achievements. The focus on mothers work hours as the measure of labour supply is justified by the fact that recent changes in labour supply of mothers have been along the intensive margin (see Nicoletti, Salvanes, and Tominey 2018 and Blau and Kahn 2013). In our sample 88% of mothers work during the pre-school years and much of the variation in the mothers' labour supply comes from how many hours they work rather than from whether they work or not. Our paper is most similar to Agostinelli and Sorrenti (2021) who nonetheless ask a different question to ours, investigating the effect of mothers' hours and family income for children aged 6-18 on their contemporaneous outcomes. Instead we focus on inputs in the pre-school years - a sensitive period for parental investments (Cunha and Heckman 2007). On top of this, Agostinelli and Sorrenti (2021) exploit the Earned Income Tax Credit (EITC) to identify the effect of income and therefore rely on variation from the bottom of the income distribution. Our complementary work uses different instruments whose variation does not rely on the population of low-income households but on variation coming from almost the whole population of parents.

Our second contribution is the novel application of the partially overlapping peer approach to construct instruments for mothers' hours and family income.[3] Whilst the partially

[2]We exclude the first year after child's birth, owing to the fact that parental labour decisions in this period are mainly driven by parental leave entitlement. In our data period in Norway, mothers are entitled to up to 52 weeks of paid leave and fathers are entitled to share a proportion of the leave.

[3]See Bramoullé, Djebbari, and Fortin (2009) and Lee, Liu, and Lin (2010) for the theoretical framework and Nicoletti, Salvanes, and Tominey (2018), De Giorgi, Pellizzari, and Redaelli (2010), De Giorgi,

overlapping peer approach has been used to identify the effect of peers on individual or household behaviour, ours is the first paper to use peers' of peers' behaviour as an instrument to identify the effect of an individual or household behaviour on outcomes.

Our third contribution is to adapt the mediation analysis approach (see e.g. Heckman and Pinto 2015) to decompose the total effect of an increase in work hours into the *causal* direct effect and the *causal* mediator effect through income and to extend this approach to address the endogeneity of *both* household income and work hours.

Using the terminology from the mediation analysis literature, our treatment is mothers' working hours and our mediator is household income. Our main econometric challenge is to address the endogeneity of the treatment and mediator. Studies based on randomized treatments[4] needs only to address the issue of endogeneity of the mediators and they do it by imposing specific assumptions,[5] using instrumental variables for the mediators,[6] or a difference-in-difference method[7]. In our setting, where we do not have a randomized treatment, we will take into account the endogeneity of both the treatment and mediator using an instrumental variable approach.[8]

To consistently estimate the direct and mediator effects, we estimate three equations for child school achievement, household income, and mother's hours respectively. We allow the error in each of the equations to be correlated with each other to reflect the endogeneity issues, which we solve by adopting instruments for both income and hours and considering a three-stage least squares estimation (3SLS).

In the setting of our paper, parents make decisions about their labour supply in the five years before their first child starts school. This is a period where parents face uncertainty about the returns to their decisions and are likely to be influenced by their peers - other parents of a first child. Indeed, analyzing a different question, Nicoletti, Salvanes, and Tominey (2018) find that hours mothers choose to work in their child's pre-school years

Frederiksen, and Pistaferri (2020), Nicoletti and Rabe (2019) for empirical applications.

[4]E.g. Imai, Keele, Tingley, and Yamamoto (2011), Heckman, Pinto, and Savelyev (2013), Heckman and Pinto (2015), Acharya, Blackwell, and Sen (2016) and Aklin and Bayer (2017).

[5]E.g. Heckman and Pinto (2015), Acharya, Blackwell, and Sen (2016), Fagereng, Mogstad, and Rønning (2021), Macmillan and Tominey (2020).

[6]E.g. Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2020) and Aklin and Bayer (2017).

[7]Deuchert, Huber, and Schelker 2019.

[8]Mediation analyses using instruments for both the treatment and mediator has been rarely used (see Powdthavee, Lekfuangfu, and Wooden 2013, Frölich and Huber 2017, and Huber 2020).

are influenced by their family peers. Household income may be affected by peers through peer effects in consumption (De Giorgi, Frederiksen, and Pistaferri 2020 and Lewbel, Norris, Pendakur, and Qu 2018) and productivity (Cornelissen, Dustmann, and Schönberg 2017) which can drive earnings potential and hours worked of an individual and therefore labour income. We instrument household income and work hours by taking the traits of the peers of parents' peers, i.e. the parents' "indirect peers".

Using Norwegian administrative data we can observe the workmates and family members as well as the "indirect peers". As is standard practice in the literature, we consider only homogenous peers who had their first child around a similar time and have the same education level as the household. Each workmate and family member has their own set of peers which are used to define the indirect peers of a household. The hours (income) of a household will be instrumented with the characteristics of family of the household's workmates (the workmates of the household's family). To understand the source of variation in the treatment and mediator using this instrumental variables strategy, consider the example of hours. The family's workmates choose how many hours the mother will work soon after the birth of their first child. These indirect peers choices can influence the hours of the family peers, which, in turn, can affect the household's maternal working hours. Such instrumental variables have been used to identify peer effects, but not for identification of the effect of parental behaviour on child outcomes. Similarly to De Giorgi, Pellizzari, and Redaelli (2010), Nicoletti and Rabe (2019) and Nicoletti, Salvanes, and Tominey (2018) our instrumental variables are lagged predictors of the endogenous variables averaged across indirect peers. For example De Giorgi, Pellizzari, and Redaelli (2010)'s main instrumental variable for the peers' enrolment on an economics or business major is the expectation of the indirect peers to enrol on economics or business majors observed two years earlier. In Nicoletti and Rabe (2019) the instrument for the older sibling's test score at age 16 is the test score at 11 averaged across the older sibling's school mates. In our case, we follow Nicoletti, Salvanes, and Tominey (2018) and measure income and labour supply of the indirect peers 1 year after child birth of their first child, which occurred at least one year earlier than the focal parents' first birth. We find that parents of different types of households respond similarly to both instruments, showing that the variation associated to our instruments is not driven by a subgroup of parents.

A potential threat to our identification strategy is caused by unobservable workplace or labour market effects, which can be correlated with the instrumental variables and the focal child's test score outcomes. Our benchmark specification controls for potential unobservables at the work place level (for example work place family friendly practices or access to day care) by estimating the model as deviations from the mean across workmates. Even if the focal household lives in a different neighbourhood and works in a different firm to the indirect peer groups, common unobserved area effects (such as common labour market shocks, location of specific firms, day care facilities) may exist which invalidate our identification assumptions. We show this is not the case by i) controlling for a wider-area labour market fixed effect or the neighbourhood level leave-one-out mean of mothers' labour hours and ii) running a placebo analysis which demonstrates the focal household's hours and income do not respond to a fictitious set of indirect peers belonging to the same labour market and with similar characteristics as the true indirect peers.

Our main findings suggest that a 10-hour increase in the mother's average weekly hours during preschool years leads to a decrease in child's test scores at age 11 and 15 by around 20% of a standard deviation when controlling for the income effect. These direct effects of hours on school test scores are negative but not statistically significantly different to zero in the full sample, meaning that we cannot rule out that a mothers' time is as productive as the counterfactual of formal childcare or time with the father. Is it possible to interpret a negative coefficient on mothers' working hours in the test score equations as evidence of a negative effect on children? The answer is no. The income mediator effect is positive and statistically significant at both age 11 and 15 and compensates fully for any negative direct effect leading to a total effect that is positive and at age 15, statistically significant.

We explore the mechanisms behind the direct effect of mothers' hours, finding a more negative effect in households where mothers have a degree - consistent with evidence that the quality of a mothers' time investments in children differ across mothers' education (Cornelissen, Dustmann, Raute, and Schönberg 2018) - and in households with no debt - suggesting the absence of debt constraints raises the mothers' productivity of time with the child, e.g. by lowering stress. A potential mechanism for the effect of household income during pre-school years on later outcomes is through access to better schools and neighbourhoods.

2 Methodology

2.1 Empirical model

We are interested in estimating the causal effect on the academic achievements of first born children of mothers' working hours in the preschool period. We measure the academic achievements using school test scores observed at ages 11 and 15. For brevity we focus on the longer-run results at age 15 but show results at age 11 in the appendix, discussing any differences in the text. Our aim is to estimate the total average causal effect of mothers hours and to decompose it into the direct effect and the mediator effect through household income.

To estimate such effects we consider the following system of three equations for the child school test score, Y, measured at age 11 or 15, mothers' working hours, H, and household income, I, both measured as averages over the preschool period,

$$\begin{aligned} Y &= \gamma_0^Y + H\gamma_H^Y + I\gamma_I^Y + \mathbf{X}\boldsymbol{\beta}^Y + u^Y, \\ I &= \gamma_0^I + H\gamma_H^I + \mathbf{X}\boldsymbol{\beta}^I + \mathbf{Z}^I\boldsymbol{\rho}^I + u^I, \\ H &= \gamma_0^H + \mathbf{X}\boldsymbol{\beta}^H + \mathbf{Z}^H\boldsymbol{\rho}^H + u^H; \end{aligned} \tag{1}$$

where the coefficients have over-scripts to denote the specific equation they refer to, e.g., γ_0^k is the intercept for equation k, with $k = Y, I$ and H, and under-scripts to denote the corresponding explanatory variable. Therefore, γ_H^k is the coefficient for the mother's work hours H in the equation for child academic achievement when $k = Y$, and for household income when $k = I$; γ_I^Y is the coefficient for household income I in the equation for Y. $\boldsymbol{\beta}^k$ is a vector of coefficients in equation k corresponding to the vector of predetermined family and child characteristics, $\mathbf{X}$, which includes child birth weight and child birth weight squared, child gender, mothers' age at birth, mother and fathers' education, parents' labour participation before the first child birth, fathers' income in the year before birth and child month of birth and year of birth dummies;[9] $\mathbf{Z}^k$ is a vector of instrumental variables that do not directly explain Y and $\boldsymbol{\rho}^k$ is the corresponding vector of coefficients; u^Y, u^I and u^H are error terms which we allow to be correlated with each other.

[9] $\mathbf{X}$ contains an additional variable to ensure the exclusion restrictions for the instrumental variables be valid, namely the test score averaged across family members (see Section 4.1 for more details.)

Table A.1 lists the variables included in each of the equations. All regressions control for mother and fathers' years of schooling, mothers' age at child birth and an indicator for whether she worked before having her child; father income and an indicator for the father working in the year before the child was born as well as child variables including month and year of birth dummy variables, birth weight and birth weight squared and finally the (leave-one-out) mean test score across family peers. To ease the interpretation of the coefficients of model (1), we de-meaned H and I and standardize the child test score Y within the population by the child birth cohort, to have a mean 0 and standard deviation of 1.

The system of equations (1) is our benchmark specification. In Section 2.2 we provide details on the theoretical mechanisms through which mother's work hours and household income can affect child test scores. We then explain how the parameters of the system can be used to decompose the average causal effect of mothers hours into the direct effect and the mediator effect through household income in Section 2.3, while we provide details of the data in Section 3 and the estimation strategy to address the endogeneity issues in model (1) in Section 4.

2.2 Theoretical mechanisms

Well established theoretical models including Cunha and Heckman (2007) and Cunha and Heckman (2008) model child human capital as a function of parental investments of money and time. To interpret our system of equations (1) and discuss the mechanisms through which income and hours may affect child test scores, we draw on the literature to understand how changes in these two preschool inputs map into parental time and monetary investments and therefore into child human capital. Next we explain how the existence of these mechanisms can lead to heterogenous effects of income and hours by formal childcare access, mother's education, gender and by the presence of household debt. Finally, we conclude this section by providing details on how our model (1) incorporates the role of father.

2.2.1 Monetary and time investments in children

A shock to household income can map into a change in parental investments if there are imperfect credit markets and partial insurance against income shocks (Blundell, Pistaferri, and Preston 2008, Blundell, Pistaferri, and Saporta-Eksten 2018, Caucutt and Lochner 2020, Cunha and Heckman 2007). For example, Carneiro and Ginja (2016) estimate that a permanent positive income shock of 10% translates into increased investments in the home learning environment for children by 2% of a standard deviation.

When a mother of a pre-school child increases her work hours, she will inevitably reduce her time investments in the child which, Fiorini and Keane (2014) and Del Bono, Francesconi, Kelly, and Sacker (2016) argue will filter through to child human capital. Alternatively Hsin and Felfe (2014) note that mothers may protect the quality time investments in their children and reduce instead the time spent in activities which do not raise child skills. Therefore an open empirical question is whether there is a negative direct effect of mothers hours on children.

The consequence of the increase in a mothers' hours for the child's skill accumulation will also depend on the relative productivity of the mothers' time compared to the counterfactual childcarer - be it the father, a grandparent or a formal childcare provider. Whilst grandparent care was an important childcare provider prior to the expansion of the formal, subsidised childcare sector in Norway which took place across the 1970s (Havnes and Mogstad 2011), since the 1970s the childcare mode of choice has been predominantly nursery or preschool centres and more recently care of fathers. For this reason we will interpret the direct effect of mothers' hours as the productivity of her hours relative to the alternative of formal childcare or time with the father.

To better understand the availability and quality (i.e. the likely productivity) of formal childcare in Norway, we provide below some details on the Norwegian setting. For children born between 1997 and 2001 - the cohorts considered in our analysis - a strongly subsidized formal day care was available from age 1 up to the start of school at age 6 and it was universal, with very high coverage and high enrolment rates on average across municipalities, but with some variation (Havnes and Mogstad 2015). Day cares in Norway are organized in public or private centers and the day cares are typically divided by age groups such as 1-2, 3-5 (Black, Devereux, Løken, and Salvanes 2014). Private centers are typically owned by nonprofit organizations such as churches and cooperatives, although over the last years for-profit organizations have emerged. Formal childcare is usually available around the workplace of parents or close to their residence.

All types of day care facilities are audited by the municipality and are regulated by national legislation on educational quality and safety. The national regulation imposes a maximum ratio of children to staff of 3 for children aged less than 3 and 6 for older children, and a maximum ratio of children to teachers of 7 for under 3 years old and 11 for over 3 years old children. The central government also regulates universal standards for building and teacher qualifications. Given these national standards, day care centers are very similar in the way they operate across municipalities. Day cares are highly subsidized, with around 40% of day care costs covered directly by the central government, up to one-third by the municipality and the rest by parents. Low-income families receive larger subsidies (Black, Devereux, Løken, and Salvanes 2014). The focus in Norwegian day cares is on developing children's social–emotional development and not on developing academic skills. All in all, formal childcare in Norway is considered to be of high quality with little heterogeneity across sites, although in municipalities with lower access to childcare it is possible that mothers who work are more likely to use informal childcare. We analyse this directly by exploring heterogeneity across childcare access.

In Norway the labour force participation of mothers is high with 88.2% of mothers working when their first child is aged between 1 and 5 for the cohorts of children born between 1997 and 2001. Furthermore, the participation rates are quite similar across level of mother's education with a rate of 85% for mothers with no degree and of 93% for mothers with a degree.

2.2.2 Heterogeneity in productivity of parental inputs

Even if in the majority of areas in Norway childcare coverage is not an issue, there are still some municipalities where formal childcare access is relatively low and where mothers who work would have to use informal childcare. In the empirical section, we provide some evidence on the potential effect of replacing mothers' childcare with informal rather than formal childcare by investigating whether the effect of mothers' hours is heterogeneous between municipalities with low and high childcare access.

Further heterogeneity analysis will be partially motivated by a finding in the literature on child human capital, that the productivity of parental inputs on child human capital may vary across household socio-economic status (SES) and child gender. The productivity of income in driving child outcomes is likely to differ across parental SES if (i) mothers with low and high education have different preferences for monetary investments in children (Boneva and Rauh 2018); (ii) the productivity of monetary investments differs across level of mother's education (Heckman and Mosso 2014); (iii) low educated mothers face credit constraints such that an increase in income may free up resources to invest in their children.

Dahl and Lochner (2012) find the effect of an increase in the EITC on child test scores was highest for particularly low income households, although according to Gregg, Waldfogel, and Washbrook (2006), credit constrained parents spend a positive income shock on essential consumption such as paying bills and buying clothes rather than on investments goods specifically for children. In our empirical analysis we check whether the effect of an increase in income differs between families with and without financial constraints, proxied by the presence of household debt.

Moving to the heterogeneity in the productivity of mothers' time investments in children (relative to formal childcare or fathers' time), the productivity may vary across SES if the quality of the child care provision is heterogeneous across SES. For example if high educated households send children to a childcare of higher quality then the negative direct effect will be smaller for higher educated mothers. However as noted above, in Norway all mothers access a relatively homogenous childcare across all preschool settings because of a common curriculum and fees that are income related to ensure equal access for children from low-income families.

On the other hand the quality of mothers' time investments is potentially increasing across the education level of mothers (see e.g. Cornelissen, Dustmann, Raute, and Schönberg 2018, Hill and Stafford 1974; McLanahan 2004; Guryan, Hurst, and Kearney 2008; Dotti Sani and Treas 2016). Kalil, Ryan, and Corey (2012), Kalil (2015) and Dotti Sani and Treas (2016) document a positive parental educational gradient in time investments and quality of child environment at home. If the productivity of time in child care is lower than the productivity of time spent with a highly educated mother, and vice versa for a low educated mother, then we would expect one extra work hour to have a more negative direct effect on child test scores for women with high education.

There is growing evidence that the production function relating parental investments to child outcomes may differ between boys and girls; but results are quite mixed and there is need for more empirical insights on gender differences in the productivity of inputs. For example Anderson (2008) finds that pre-school interventions implemented within the Abecedarian, Perry and the Early Training Project have beneficial long term effects on girls but not on boys and similarly Hoynes et al. (2016) find that food stamps lead to better long term outcomes for girls than boys. On the contrary, Autor, Figlio, Karbownik, Roth, and Wasserman (2016) find that boys respond to a greater extent to high quality schooling than girls, Bertrand and Pan (2013) document how boys' socio-emotional skills are more sensitive to parental divorce and Fan, Fang, and Markussen (2015) find that the effect of early life maternal employment is to reduce the probability of attaining a degree later in life and this effect is stronger for boys than girls.

The empirical analysis will estimate our model (1) for the full sample of children but we will also investigate potential mechanisms by stratifying the sample by the households' SES (measured by an indicator for the mother having a degree), child gender, formal childcare access, and by the presence of household debt.

2.2.3 The role of fathers

Because fathers' inputs are important for child human capital, fathers enter our empirical model through three main channels. First, the income of fathers can drive mothers' labour hours though inclusion of fathers' income before the child's birth in the set of control variables $\mathbf{X}$. Of course mothers' hours post-birth may react to changes in fathers' income post-birth and in a sensitivity analysis we add household income between ages 1-5 as an additional control in the hours equation. The results show that conditional on our benchmark covariates, household income post-birth does not statistically significantly drive mothers hours and its inclusion does not change our results.

Second, fathers' hours may respond to mothers' hours, as suggested by Goux, Maurin, and Petrongolo (2014). The interpretation of the effect of mothers' hours on household income (γ_H^I) incorporates any reaction of the fathers' hours to an increase in the mothers' hours. Household income, I, is the sum of father's and mother's yearly earnings, $I = (E^F + E^M)$, so the effect of an increase in mother's work hours is given by

$$\gamma_H^I = \frac{\partial I}{\partial H} = \frac{\partial E^M}{\partial H} + \frac{\partial E^F}{\partial H}, \tag{2}$$

where E^M and E^F are the mother's and father's average yearly earnings between age 1 and 5, $\frac{\partial E^F}{\partial H}$ is the father's earnings marginal response to an increase in mother's hours and $\frac{\partial E^M}{\partial H}$ is the marginal effect of an increase in mother's hours on mother's earnings, i.e. the mother's wage rate. The father may respond by working more or less, depending on whether leisure time of the two parents is complementary or substitutable.

Finally as noted above, the productivity of the fathers' time spent with their child in replacement to mother's time is captured by the direct effect of mothers' hours on child outcomes which indeed we interpret as the productivity of the mother's time relative to the counterfactual childcare, which can include the father time or formal childcare.

2.3 Decomposition analysis

In this section we explain how we use the parameters of model (1) to decompose the total effect of mother's work hours in to the direct effect and the mediator effect through household income. Using terminology from the treatment evaluation and mediation analysis, mothers' work hours H is the treatment, the household income I is the mediator and Y is the outcome.

We are interested in evaluating the average effect of increasing the mothers' hours from a pre-treatment level (denoted by h_0) to a post-treatment level (denoted by h_1), on children's school test scores, Y. This effect is called the average total effect in the decomposition literature and the average treatment effect (ATE) in the casual treatment literature. Such effect can be denoted as

$$E(Y_{h_1} - Y_{h_0}) = E(Y_{h_1,I_{h_1}} - Y_{h_0,I_{h_0}}), \tag{3}$$

where Y_{h_1}, Y_{h_0}, I_{h_1} and I_{h_0} denote the outcome and mediator when treatment H is set at the pre- and post- treatment levels h_0 and h_1. The ATE in equation (3) represents the mean change in test score outcomes when hours change from h_0 to h_1, which in turn raises income from I_{h_0} to I_{h_1} and can be decomposed in the following two additive effects:

1. the *direct effect* of mothers' work hours,

$$E(Y_{h_1,I_{h_0}} - Y_{h_0,I_{h_0}}), \tag{4}$$

which is the effect of increasing hours from h_0 to h_1 while removing the mediator effect, i.e. keeping the income fixed at I_{h_0};

2. the *mediator effect* of mothers' work hours through household income,

$$E(Y_{h_0,I_{h_1}} - Y_{h_0,I_{h_0}}), \tag{5}$$

i.e. the effect which is mediated by the change in income from I_{h_0} to I_{h_1} whilst keeping hours at h_0.[10]

We can rewrite the expressions (4)-(5) by using the equations for Y and I in the system of equations (1). By replacing the two outcomes in the direct effect $E(Y_{h_1,I_{h_0}} - Y_{h_0,I_{h_0}})$ with the right hand side of the first equation in (1) and fixing the variables H and I at the corresponding values, we obtain

$$E(Y_{h_1,I_{h_0}} - Y_{h_0,I_{h_0}}) = (h_1 - h_0)\gamma_H^Y, \tag{6}$$

The mediator effect (5) can be rewritten by replacing Y and I with the right hand side of the first and second equations in (1) and fixing the variables H and I at the values denoted in the subscripts,

$$E(Y_{h_0,I_{h_1}} - Y_{h_0,I_{h_0}}) = (h_1 - h_0)\gamma_H^I\gamma_I^Y. \tag{7}$$

Notice that the mediator effect is given by the change in income caused by the increase in mother's hours on income, i.e. the product of $(h_1 - h_0)$ and the causal effect of mothers' hours on household income γ_H^I, multiplied by the effect of household income on child test scores γ_I^Y.

The computation of the direct and mediator effects requires a consistent estimation of the parameters of model (1). We provide details on such estimation in Section (4).

3 Data

We use Norwegian registry data, collected by different administrative units, and linked by Statistics Norway. The data combine information from different registers across time and provide details on children's school test scores, their parental income, education and employment and information to identify where people work and who their family members are.

[10]For decomposition into the direct and indirect (mediator) effect see also Imai, Keele, Tingley, and Yamamoto (2011), Pearl (2012) and Heckman, Pinto, and Savelyev (2013). The direct and mediator effects are also known as the pure or natural direct and indirect effects (see VanderWeele 2013, VanderWeele 2016).

Our analysis focuses on first born children, for the reason that parental investments for the second child can react to changes in the endowments of the first child, a mechanism that we want to rule out in our analysis. For example if the mother returned to work early after having a first child and observed a fall in the child's test scores, she may delay the return for the second child. Each child is linked to their parents using the birth registry, where we also identify first born children.

For parents the annual labour market participation status, hours worked and earnings is recorded in all pre-school years and is based on the tax records. The yearly household income in each of the pre-school years is defined as the sum of the mother's and father's yearly post-tax income (labour earnings), deflated to the year 2020. Post-tax income is calculated as the gross income received, net of any taxes and adding transfers. These transfers include a progressive income tax, child benefits for children up to age 18, unemployment and sickness benefits and any other cash transfers from the social insurance system, general deductions for work related expenditures and finally a regional compensation for living in the Northern most region of Norway. Pre-school income is the mean income across the years 1 and 5 after the first child's birth.

Hours worked are effective hours worked and include all hours that an individual worked in a particular week in November. In a particular year, weekly hours worked is defined from a discrete variable taking the value of 0, 1-19, 20-29 and 30+ hours. We then take the mid-point of each category within each year for the child aged 1-5, whereby hours is defined as 0, 10, 24.5 and 40. We test for the sensitivity of the choice of mid-point values in Section A.2.

Our measure of interest is the pre-school hours worked, which is defined as the mean hours worked between 1 and 5 years after the first child was born and consequently, whilst hours at a specific age take a small number of values, the average across 5 years varies continuously.[11] Parents in Norway are entitled to about one year of fully paid maternity leave after which they return to a contract which specifies work hours. This means that the hours worked measured from one year after the first birth is representative of the usual

[11] Absence of the exact measure of hours may create measurement error in the estimation. The measurement error is unlikely correlated with observed or unobserved variables in our model and therefore any attenuation bias in the OLS estimation tends to cancel in the IV estimation. See Appendix Section A.1 for full details.

hours worked in a particular year. Upon returning to work there is flexibility in Norway such that the post-birth contract may be the same as the pre-birth contract, or may change to allow for part-time work. Figure A.1 plots a histogram of mothers' work hours 1-5 years after birth which shows there is large variation across the distribution. The mean value of household income and mothers' labour hours in pre-school years are approximately 500,000 Kroner (around 55,000 US dollars) and 20 hours, reported in Table 1.

Age 11 test scores are recorded for the population of children born between 1997-2005, whereas the age 15 test scores exist in the education statistics for birth cohorts 1996-2001. We select a common sample across the two outcomes, choosing cohorts born between 1997-2001. Both test score outcomes are constructed by summing grades on Maths and Reading.[12] The test scores at age 11 and 15 are not high stake for the child and therefore more likely to represent child ability as opposed to school quality if schools "teach to the test". The descriptives in Table 1 report the mean and standard deviation of all variables used in our main analysis which focuses on test scores at age 15. Descriptive statistics for test scores at age 11 and for a set of extra controls we consider in our robustness, sensitivity and mechanisms analyses are reported in Table A.2. Test scores at ages 11 and 15 have a mean of 46 and 64 respectively. In our analysis we standardise the test score within the population to have a mean of zero and standard deviation of 1, within each cohort.

From the education and labor market statistics we also derive details on parents' characteristics observed before the birth of their first child, which we use as control variables. In Table 1 we report descriptives for these pre-birth parental variables, which include parents' education measured at child's birth (mothers and fathers have on average 13 years of schooling and around 40% of mothers have a degree), mothers' age at birth (26.5 on average) and labour market variables for the parents, including a dummy variable indicating mothers' and fathers' working status in the year before birth (76% of mothers and 97% of fathers work in the year before birth) and fathers' post-tax income in the year before birth. We also control for the child year and month of birth and a quadratic specification of child birth weight, measured from the birth registry.

[12]We repeated the analysis separately for maths and reading test score outcomes and found no difference to our benchmark. The results are available on request.

Finally, we use Norwegian employer-employee and population registers to identify the plant where individuals work and their family connections in order to define two corresponding sets of workmate and family peers.

A workmate is an individual working in the same company at the same location as the father, who had their first child between 1-5 years before the father and with the same degree status. A father has on average 16 workmate peers. The family peer group is constructed from all sisters and sister-in-laws who gave birth before the household of interest had their first child. Identification of a family peer requires matching for each parent in our data an identifier for their own parents (i.e. the grandparents of the children for whom we observe school test score outcomes) using data on the census dating back to 1967. An individual is classed as a family peer if they gave birth before the mother in question and has the same degree status. Families have on average 2.4 sisters or sister-in-laws.

4 Estimation strategy

The estimation of the direct and mediator effects defined in (6) and (7) requires a consistent estimation of the parameters γ_H^Y and γ_I^Y in the outcome equation and of the parameter γ_H^I in the income equation in the system (1). This implies that we have to deal with the issue of potential endogeneity of the treatment H and mediator I in the outcome equation and of treatment H in the income equation. The endogeneity of H (I) in the outcome equation arrives when there is correlation between the error terms u^H (u^I) and u^Y. This correlation can be caused by unobservable variables which explain both H (I) and Y. Similarly, the endogeneity of H in the income equation arrives if there is correlation between u^H and u^I.[13]

Without taking account of the double endogeneity of H and I, the estimation of the total effect of the treatment and its direct and mediator effects will be biased and inconsistent.[14] Some papers have defined specific assumptions under which it is still possible to estimate consistently the direct effect $[(h_1 - h_0)\gamma_H^Y]$ (see Robins 2003, Acharya, Blackwell, and Sen 2016) or both the direct and mediator effects (see Imai, Keele, Tingley, and Yamamoto 2011

[13] We provide a conceptual diagram for the direct and mediation effects and for the confounding effects of the unobservables in Figure A.2.

[14] See e.g. Rosenbaum (1984); Angrist and Pischke (2008); Bullock, Green, and Ha (2010).

and Heckman and Pinto 2015). The credibility of these assumptions is contentious and depends on the empirical application.

In this paper we avoid to impose such assumptions by relying on instrumental variable estimation to solve both the endogeneity of the mediator, I, and of the treatment, H.[15]

Considering an instrumental variable for hours H and another one for income I that do not explain directly the child outcome Y, we could in theory estimate the regression of Y on H and I by using 2-stage least squares (2SLS) and obtain consistent estimation of the direct effect of H on Y (see Agostinelli and Sorrenti 2021). Nevertheless, such 2SLS estimation does not provide an estimation of the causal effect of hours on income, γ_H^I. Therefore it cannot be used to decompose the total effect of hours H on child outcomes Y into the direct and mediator effects. For this reason we adopt a three-stage least squares (3SLS) estimation of model (1). The first stage consists of the estimation of the reduced form of model (1), i.e. the regression of all endogenous variables on all exogenous explanatory variables and the instruments for mother's work hours and household income. The second stage estimates the structural model (1) by replacing the endogenous variables in the right hand side with their predictions from the first stage. Finally, the third stages uses the residuals from the second stage to estimate the matrix of variances and covariances for the error terms and to apply a feasible generalized least squares estimation of the structural model (1). The 3SLS procedure requires at least one instrument for H and another one for I and we describe our instruments in Section 4.1.[16]

[15] See also Powdthavee, Lekfuangfu, and Wooden 2013, Frölich and Huber 2017, and Huber 2020.

[16] A potential way to simplify the estimation of model (1) is by imposing a zero correlation between the error in the hours equation and the error in the test score equation (see assumption A2 in Dippel, Gold, Heblich, and Pinto 2019) and then using a single instrument for Y rather than an instrument for H and another one for I. This assumption would be satisfied in our application if mother's hours was endogenous in the equation for Y exclusively or at least mainly because of omitted variables that influence household income. While such assumption can be credible in some empirical examples it is not in our application because we can have unobservables, such as maternal psychological health, that can affect negatively both mother's hours and the child's test scores on the top of the effect through income. Indeed in our estimation the correlation between the error terms for the H and I equations is 0.333 with a p-value of 0.000.

4.1 Instrumental variables

Informed by previous papers on peer effects on womens' labour supply decisions,[17] we instrument the mother's labour hours using the average maternal work hours of her family's workmates.

Regarding household income, our model estimates the effect of household income controlling for mothers' labour hours, hence the variation will come mostly from labour market decisions fathers make in the pre-school years. Whilst households cannot directly choose their wage in the labour market, household income can respond to peers through changes in their labour supply, e.g. fathers can decide to increase their working hours if they observe that their peers have costly consumption habits and they decide to conform to these habits (De Giorgi, Frederiksen, and Pistaferri 2020). Furthermore, because there can be peer effects in productivity at work (Cornelissen, Dustmann, and Schönberg 2017), we can expect the labour income of the father's workmates to affect his labour income also through his wage and not just through work hours. In our estimation we instrument household income using the average paternal labour income of workmates' family.

Let $\overline{H}^F$ and $\overline{I}^W$ be the mother's hours and household income averaged across the direct peers of type F, parents' family members, and of type W, father's workmates.[18] The average across peers are computed excluding the parents in question, i.e. the "leave one out" average (see Angrist, 2014). Let $\overline{H}^{FW}$ and $\overline{I}^{WF}$ be the average mother's hours and paternal household income averaged across the indirect peers, which are the mother's family's workmates (denoted FW) and the father's workmates' family (denoted WF) respectively.

The traits of an indirect peer group may be an invalid instrument if the timing of births creates reverse causation in the regressions on the endogenous variables on the instruments. For example, the hours and income of the household may influence the decisions of the direct peers, which in turns influences the decisions of the indirect peers. To ensure the instrumental variables are valid, potential peers of a household are defined to be a peer only if they gave birth to their first child before the focal household. In addition, rather than

[17] See Maurin and Moschion (2004); Mota and Rosenthal (2016); Olivetti, Patacchini, and Zenou (2020); and Nicoletti, Salvanes, and Tominey (2018), which provide clear evidence that female family peers affect women's labour decisions.

[18] For a more precise definition of peers see Section 3.

using directly $\overline{H}^{FW}$ and $\overline{I}^{WF}$ as instruments for H and I, we consider as IVs the income and hours of the indirect peer group measured one year after birth (rather than the 1-5 years after birth) to ensure the IVs are predetermined with respect to H and I. These IVs are denoted $\overline{Z}_H^{FW}$ and $\overline{Z}_I^{WF}$. Precisely, we instrument mother's hours in the pre-school years with the maternal work hours 1 year after the first child birth, averaged across the mother's family's workmates who gave birth to their first child at least 1 year before the mother. We instrument household income in the pre-school years with the paternal earnings 1 year after the first child birth averaged across the father's workmates' family who had their first child at least 1 year before the father. Finally, the instruments are set to zero for households that do not have an "indirect peer", so that we do not need to discard observations where there are no family's workmates or workmates' family.[19]

To be valid instruments, $\overline{Z}_H^{FW}$ and $\overline{Z}_I^{WF}$ must satisfy the relevance and exclusion restrictions. They must be relevant to explain H and I conditional on the control variables, $\overline{Z}_H^{FW}$ must have zero correlation with the error terms u^I and u^Y in the equations for the household income I and the child's outcome Y and $\overline{Z}_I^{WF}$ must have zero correlation with the error term u^Y. We consider each of these assumptions in turn.

In Figure A.3 panel A we describe how variation in the IV for mothers hours drives the focal mother's labour hours. The causal pathway, represented by two horizontal arrows, shows how the focal mother's labour hours, H, is influenced by the average hours worked by her sisters and sister-in-law peers, $\overline{H}^F$. In turn, each of these mothers in the family peer group when making their own labour supply decisions are influenced by the mothers' hours of their workmate peers. The average of the family's workmates' hours 1 year after child birth ($\overline{Z}_H^{FW}$) is the instrumental variable for mothers' hours H. Similarly, in Figure A.3 panel B, the household income I is affected by the average fathers' income of workmate peers of the focal household, $\overline{I}^W$. Each of the workmates were in turn influenced by their family peers (brother and brother in-law) labour income. The instrument for household income I is the

[19] The majority of the households have a non-missing indirect peer group of the family's workmates (56%) and the workmates' family (89%). Only 6% of the sample have neither an indirect peer group of the family's workmates or the workmates' family and therefore for the vast majority of households, variation in their value of one of the IVs contributes towards the estimation. Tables A.3 and A.4 report the mean and standard deviation of the outcome and main controls (child school test scores, hours, income, child gender dummy, mothers' degree and years of schooling) by presence of indirect peers for the IV for hours and income respectively. There are no systematic differences except for hours and income in the pre-school years, which is to be expected given that the IVs are correlated with these variables.

average fathers' income one year after birth of the workmates' family, $\overline{Z}_I^{NW}$. Our regression analysis includes the relevant statistics to confirm that our instrumental variables are strong predictors of hours and income respectively.

Next, we consider the exclusion restrictions from the child test score (Y) equation. The identification assumption is that correlation between the instrument for H and u^Y and between the instrument for I and u^Y must be zero. The zero correlation between $\overline{Z}_H^{FW}$ and u^Y is satisfied if there is no effect going from the family's workmate peers to the child test score, Y, except through the mother's hours H. Similarly, the zero correlation between $\overline{Z}_I^{WF}$ and u^Y is satisfied if there is no effect going from the workmates' family peers to the child test score, Y, except through the household income I.

There are only two potential threats to these exclusion restrictions which are described graphically in Figure A.3 by the pathways from the instrument ($\overline{Z}_H^{FW}$ in panel A and $\overline{Z}_I^{WF}$ in panel B) to the child's test score Y that do not pass through H and I.

Endogenous peer membership would invalidate our identification assumption and exists if individuals sort into their peer groups based in part on unobservable traits which may also explain our outcome of interest, the child's test score. In particular we are concerned that parents' family peers can endogenously sort into work places with unobserved characteristics that are similar to the parents' direct workmates - for example if fathers choose workplaces based partly upon their family friendly work policies, including paternity leave. If this is the case, we could have a correlation between the instrument for hours, $\overline{Z}_H^{FW}$, and the average of Z across the parents' direct workmates, $\overline{Z}_H^{W}$, which drives the test scores of the workmate peers' children, $\overline{Y}^{W}$ and ultimately the child's test score Y (see the pathway from $\overline{Z}_H^{FW}$ to Y through $\overline{Z}_H^{W}$ in panel A).

A solution suggested by Bramoullé, Djebbari, and Fortin (2009) for the endogenous peer membership is to control for a network fixed effect. However as pointed out by Caeyers and Fafchamps (2016) this will induce an "exclusion bias" in the estimation as the fixed effect includes the observation for the household in question. Therefore we control for the workmates' fixed effect by rewriting our model (1) in deviation from the workmates mean but excluding the focal household. Because workmate peers are defined as individuals working in the same plant who gave birth 1-5 years before the focal household, the workmates' mean of each variable will vary within plants across different birth cohorts. Consequently, we exploit variation in the workmates composition across time i.e. variation in the influential workmates across parents working in the same plant but giving birth in different years.

A similar threat to the validity of the instrument for I is caused by a potential correlation between $\overline{Z}_I^{WF}$ and $\overline{Z}_I^{F}$ which leads to the pathway from $\overline{Z}_I^{WF}$ to Y through $\overline{Z}_I^{F}$ and $\overline{Y}^{F}$ in panel B in Figure A.3. This bias may not be large because people cannot choose their families. However as there could be still a correlation between $\overline{Z}_I^{WF}$ and $\overline{Z}_I^{F}$, we include in our benchmark specification $\overline{Y}^{F}$, the test score averaged across the focal child's cousins.[20][21] We test for potential endogeneity of $\overline{Y}^{F}$ using as instrument $\overline{Y}^{NW}$, i.e. the average test scores of the children of the neighbour's workmates and we do not reject its exogeneity. See Section A.2.2 for details including definition of neighbour peers.

The second potential threat to identification is caused by the fact that the instrument for mothers' hours may directly drive test scores Y through the test scores of the family peers' children (i.e. the focal child's cousins), $\overline{Y}^{F}$. This is because the instrumental variable $\overline{Z}_H^{FW}$ affects the average hours worked by mothers in the family peer group ($\overline{H}^{F}$) and therefore the outcomes of the focal child's cousins, $\overline{Y}^{F}$. This is represented in panel A of Figure A.3 by the arrows from $\overline{H}^{F}$ to $\overline{Y}^{F}$ and on to Y. Because we already include $\overline{Y}^{F}$ among the controls, this second threat to the validity of the instrument for hours dissipates. The corresponding threat to the validity of the instrument for I, $\overline{Z}_I^{WF}$, is described in panel B of Figure A.3 by the pathway going from $\overline{Z}_I^{WF}$ to the child test score Y through $\overline{I}^{W}$ and $\overline{Y}^{W}$. However, because we control for the workmates' fixed effect by rewriting our model (1) in deviation from the workmates' leave-one-out mean, this threat disappears.

A parallel discussion relates to the corresponding exclusion restrictions from the income equation, i.e. the assumption that the instrument for hours, $\overline{Z}_H^{FW}$, must have zero correlation with the error term of the income equation, u^I. Figure A.4 shows a possible threat, which is caused again by the potential sorting of parents family peers into group of workmates with unobserved characteristics that are similar to the father's workmates (see the pathway from the $\overline{Z}_H^{FW}$ to I through $\overline{Z}_H^{W}$ in Figure A.4). Nevertheless, because we control for the workmates' fixed effect this threat also vanishes. We assume that fathers' earnings are not

[20] This is again the "leave one out" average (see Angrist, 2014) of the test scores across family members excluding the focal child.

[21] A similar type of strategy has been also suggested by Nicoletti, Salvanes, and Tominey (2018) and von Hinke, Leckie, and Nicoletti (2019).

affected by changes in mothers' work hours in their family (in their sisters and sisters-in-law), $\overline{H}^F$, except through their own partner (mother) changes in hours. For this reason $\overline{H}^F$ can affect I only through H.

It is possible that our instrumental variables do not satisfy the exclusion restriction, if there are unobservables at the workmate or neighbourhood level which are not controlled for in our specification. Our benchmark specification controls for selection into the workplace based on unobserved traits which may correlate with child outcomes (for example family friendly practices) by estimating the model as deviation from the workplace mean, as discussed above. In addition, we rule out unobserved neighbourhood effects by controlling in Section 6.1 for geographical fixed effects at a level of the local labour market and the leave-one-out mean hours of all mothers living in the same neighbourhood, and then by showing that "placebo" indirect peers, defined as unrelated families living within the same labour market with similar observable traits as the true indirect peers, do not explain the focal household hours and income.

A final and more implicit assumption imposed by our strategy is that changes in I caused by variation in the mother's or father's earnings lead to the same effect on child's outcome Y. This may be violated if there is no income pooling and mothers have preferences to invest more in children, which would lead to a slight higher effect on Y of an increase in mother's earnings than in father's earnings. Because the variation of income we use is coming mainly from changes in father's earnings, this would mean that the income effect we estimate might be slightly underestimated.

4.2 Compliers

As our instrumental variables are novel, it is important to understand who are the compliers, i.e. for which group of households do income and hours change in response to the instrumental variables. It may be that the compliers for the instrumental variables for the two endogenous variables - household income and mothers' hours - come from different subgroups which would make comparison of the two sets of estimates difficult. Whilst we cannot test whether compliance is based upon unobservable traits, we analyse whether the effect of the instrumental variables on income and on hours vary across a wide set of observable characteristics. Notice that our instruments and treatments are continuous so that we cannot adopt a standard approach for counting and characterising compliers (see e.g. Angrist and Pischke, 2008). Instead we show how the effect of the instruments on income and hours change across different subgroups of the population, to infer which subgroups are identifying the potential compliers driving our results.

Tables A.5 and A.6 report the first stage coefficients and standard errors on the instrumental variables for income and hours, respectively, across different subgroups of the population. In all these regressions we include the same set of controls as in the benchmark specification.

The variation in household income induced from the instrumental variable comes mostly from the father, as we control for the hours of the mother. Table A.5 shows that all subgroups of fathers respond to the instrumental variable for income - the income level of the workmates' family. The instrumental variable is statistically significant at 1% level and with an effect on income up to 0.128 across the different subgroups of the population. Similarly, in Table A.6, the hours worked for all subgroups of mothers responds statistically significantly at 5% level to their peers of peers - their family's workmates, except for teenager mothers who represent only 6.4% of the full population of mothers.

5 Results

5.1 Regression Results

We report in column 2 of Table 2 the 3SLS estimates of our benchmark model (1) instrumenting H and I with the mean family's workmates' hours and the mean of workmate family's income (see Section 4.1 for more details on the IVs). The set of controls included in the regressions are detailed in Table A.1.[22] To see the benefits of our peers-of-peers approach to identify the effect of H and I on child test scores, in column 1 of Table 2 we also report the OLS estimates.

[22]The full 3SLS estimation results for model (1), including the estimated effects of the control variables, are reported in Table A.7.

Panel a) of Table 2 reports the effect of hours and income in the test scores equation in model (1), γ_H^Y and γ_I^Y. The direct effect of hours on child test scores γ_H^Y estimated using 3SLS in column (2) is -0.023. An increase in mothers' hours by 1 per week in each of the five pre-school years translates into a reduction of test scores at age 15 by 2.3% of a standard deviation although this effect is not statistically different to zero. This direct effect indicates the productivity of the mothers' time relative to time spent with the alternative, for example time with a father or in formal childcare. In Norway formal childcare is publicly provided and subsidized, it has high coverage for the cohorts considered in our sample and is the main alternative to mother's childcare. The result suggests that the mothers' time is as productive as the counterfactual childcare. Note that the OLS estimate is far lower, at 0.002, suggesting that there is a large endogeneity bias in the estimate of the effect of H on Y.

The effect of income on child test scores γ_I^Y in our benchmark model of column (2) is 0.192 and statistically significant at the 5% and even 1% level. The interpretation is that an increase in household income in the pre-school years of the first child by NOK100,000 (approximately 11,000 US dollars and two thirds of a standard deviation) raises test scores at age 15 by 19.2% of a standard deviation. This effect suggests a strong productivity of early parental income on child outcomes at 15. This finding is similar to Carneiro, García, Salvanes, and Tominey (2021) who find that the productivity of early life family income raises child outcomes up to 30 years later. Again, the OLS estimate which does not control for the endogeneity is different, at 0.041.

Looking at the bottom panel of Table 2, our instruments for hours and income are relevant as shown by the F-tests and we do reject at 5% level the exogeneity of hours and income.

5.2 Decomposition analysis

Is it possible to interpret a negative coefficient of mothers' working hours in the test score equation as evidence of a negative effect on children? In order to answer this question we decompose the total effect of a change in mothers' labour hours into the direct effect and the mediator effect defined in equations (6)-(7). Panel b) of Table 2 reports the effect of hours on household income (γ_H^I) and panel c) computes the total effect as the sum of the direct and mediator effects defined in (6) and (7), that are respectively γ_H^Y and ($\gamma_H^I \gamma_I^Y$).

Of course, a change in mothers' labour hours will raise household income hence income can potentially mediate or compensate for the negative direct effect of mothers' working hours by, for example, raising monetary investments in children. As shown in Table 2 the estimated direct effect of hours is -0.023, the estimated mediator effect is (0.258 0.192)= 0.049, leading to the estimated total effect of (0.049-0.023)=0.026. This implies that one hour increase in mother's weekly labour supply leads to a rise of about 2.6% of a standard deviation in the child test score at age 15 which can be decomposed into a direct effect of -2.3% of a standard deviation and a mediator effect of 4.9% of a standard deviation.

Can income compensate for a reduction of mothers' hours in pre-school years? The answer is yes. The positive mediator effect is larger than the negative direct effect, hence the total effect of an increase in mothers' pre-school hours on child outcomes is positive and statistically different to zero.

In Table A.9 column (1) we report the 3SLS estimates of the effects of hours and income on the test score at age 11 and the corresponding decomposition of the total effect. Again we find a negative direct effect of hours but a positive total effect, although not statistically significant, which suggest that any potential negative direct effect is compensated through an increase in income.

To illustrate the contribution of the direct and mediator effects to the total effect of the change in mothers' hours, Figure A.5 plots out the distribution of test scores at ages 11 (panel a) and 15 (panel b). The figure shows the original distribution pre-treatment (labelled "Original") which is the raw distribution in the data. The raw distribution is then shifted by the magnitude of the direct, mediator and total effects caused by an increase in the mother's work hours by 10. In both plots at age 11 and 15 (see left and right panels in Figure A.5) the total effect shifts to the right compared to the original distribution of test scores, as expected from the decomposition analysis. The direct effect of a reduction in the mothers' time investments shifts the distribution of test score outcomes to the left, but this is fully compensated for by the mediator effect, which shifts the distribution right.

To summarize, despite a negative direct effect of mother's hours on child test scores, the total effect of an increase in mothers' pre-school hours is always positive and this positive total effect is even statistically significant at age 15.

6 Sensitivity analysis

6.1 Tests of validity of identification strategy

Crucial for our identification strategy is the assumption that the variation we exploit in the instrumental variables for both mothers' hours and household income is exogenous. Notice that any unobservable that affects our instruments and the corresponding instrumented variables helps our identification rather than threatening it, as long as this unobservable does not affect the child's test score except through the instrumented variables. E.g. if the mother and her family's workmates select themselves into workplaces with similar long working hours requirements, there is no issue in identifying the effect of hours on child's test scores if the unobserved long hours requirements affects Y only through H. On the contrary, if we were interested in estimating the endogenous peers effect of the work hours of mothers' family using our instrument, the family's workmates hours, then we would have an identification issue owing to endogenous selection into the workplace. More in general the issue of unobserved contextual and correlated effects (see Manski 1993 for a definition) is a major issue in identifying endogenous peers effects; but it is not an issue in our context if these unobserved effects drive the child's test score only through the instrumented variables.

Even if the variation in our instrumental variables was driven by unobservable traits of a workplace that explain directly (not through H and I) the child outcomes; our strategy would still be valid because we control for unobserved workplace effects. As it is, our benchmark estimation strategy calculates all variables as deviations from the workplace "leave one out" mean. This also rules out variation coming from households working in the same plant as their family's workmates.

However, at the neighbourhood level, there may be common local environments or shocks which affect the focal household and their family and workmates. For example if jobs in particular industries were concentrated in certain areas, then a shock to one industry may affect simultaneously all peer groups living in the areas with high concentration of such industry. Or, the focal household may live in the same neighbourhood as their family's workmates, sending their children to the same daycare facilities. If the shocks or unobserved characteristics of daycare facilities indirectly affect child test scores through mothers hours or household income, they will help with identification. On the other hand, unobserved common neighbourhood traits that directly explain child test score outcomes would lead our estimates to be inconsistent.

We test whether our results are driven by unobservable common shocks and characteristics at the level of the neighbourhood by firstly controlling for local labour market effects and then running a placebo analysis which randomly assign as peers of peers those from the same neighbourhoods but with common observable traits of the true indirect peers.

The top panel of Table 3 reports the effects of mother's hours and household income on child test scores, i.e. γ_H^Y and γ_I^Y in model (1), while the bottom panel reports the direct effect, mediator effect through income and total effect of a 10-hour increase in mother's hours on child test scores, which are $(10\gamma_H^Y)$ and $[10(\gamma_H^I\gamma_I^Y)]$. Column (1) reports the benchmark results already shown in column (2) in Table 2 with the only difference that we now report the direct, mediator and total effects for an increase in mother's hours by 10 rather than by 1.

Table 3 column (2) reports the corresponding results when additionally controlling for local labour market fixed effects. A labour market is defined as a travel-to-work area and there are 45 in Norway. The results are not statistically different and therefore our benchmark is robust to control for local economic shocks or local characteristics.

We could be worried that high maternal employment in a neighbourhood might lead to less social capital, e.g. because mothers are less engaged in school and educational activities, and this could directly impact child test scores. In Table 3 column (3) we report the results when controlling in our benchmark model for hours worked by mothers in the focal household's neighbourhood – a smaller geographic area than the municipality. Controlling for the neighbourhood specific maternal employment rate leads to results which are again not dissimilar to our benchmark results.

The instrumental variable hours worked of the family's workmates (income of the workmate's family) may have a direct link to the focal child's test scores, for example if the families within the same labour market are exposed to common shocks to daycare facilities. This means that we could pick at random any family within the same wider geographical area and their values of hours and household income (predetermined with respect to the focal household's hours and income) would be directly correlated with the focal household's test scores via a common unobservable effect. In our analysis, we test whether this is true by assigning to each family a fictitious indirect peer group chosen at random who lives in the same neighbourhood and has the same observable traits (measured by mothers' education, age at birth and working status in the year before birth) as the true indirect peer group, but has no real link to the focal family. The traits of these fictitious indirect peers should have no effect on the focal child's test scores if the exclusion restriction holds. The random assignment of fictitious indirect peers is repeated 1000 times with replacement and we analyse the F-statistics for the null hypotheses of zero effects of these new instrumental variables in the hours and income equations. Out of the 1,000 replications, only 2.2% (5.7%) of replications resulted in an F-statistic larger than the rule of thumb of 10 in the income (hours) equations which provides evidence that the effect of our IVs is not caused by common unobserved effects.

6.2 Validity of the instrument for income

While to estimate the direct and mediator effects of hours on school test scores we have to estimate the system of equations (1) using 3SLS, an alternative way to estimate the total effect of mothers' work hours on test scores is by considering a simplified model for test scores which requires only an instrumental variable for mothers' hours and not for household income. Estimating the total effect using 2SLS will lead to a consistent estimate under the assumption that the instrument for hours is valid, whereas the 3SLS estimation will be consistent under the assumption that both the instruments for hours and for income are valid. Assuming that the instrument for hours is valid, results on the total effect of hours that are similar between the two types of estimations would suggest that the instrument for income is valid.

Consider the following simplified model for test scores:

$$Y = \gamma_0^T + H\gamma_H^T + \mathbf{X}\boldsymbol{\beta}^T + u^T, \tag{8}$$

where u^T is the error term, γ_0^T is the intercept, γ_H^T is the total effect of hours H on test scores Y, and $\mathbf{X}$ is a vector of control variables listed in the column labelled "Tests Scores Y" in Table A.1. Of course H may be correlated with the error term u^T and this endogeneity problem can be solved by adopting a two-stage least squares (2SLS) estimation of equation (8), using as IV for hours the mean family's workmates' hours.

The estimate for γ_H^T is 0.003 (standard error 0.000) when using OLS and 0.029 (standard error 0.013) when using 2SLS; and, not surprisingly, the preferred estimation is the 2SLS because we strongly reject the exogeneity of the hours. The total effect γ_H^T computed using 2SLS estimation of model (8) can be directly compared with the total effect $(\gamma_H^Y + \gamma_H^I \gamma_I^Y)$ computed using the 3SLS estimation of model (1), which is 0.026 (see Table 2 panel c). Since the consistency of the 2SLS estimation of model (8) requires the validity of only the instrument for hours, while the consistency of the 3SLS estimation of our benchmark model (1) relies on the validity of both the instruments for hours and income, finding similar results for the total effect of hours suggests that the validity of the instrument for income holds.

6.3 Placebo outcome tests

To further investigate the validity of our instrumental variables we show that our instruments do not have any statistically significant effect on placebo outcomes measured at birth and therefore predetermined with respect to H and I measured between age 1-5. The birth outcomes for the child include birth weight and height, a dummy for transfer of the newborn to the children ward, a dummy for congenital severe malformation and a dummy for severe deformity. Table A.8 shows that the effect of our instruments on each of these placebo outcomes is not statistically different from zero.

6.4 Additional sensitivity analysis

We now summarise additional sensitivity analysis which are explained in detail in appendix Section A.2, which support the validity of our estimation strategy. If H and I during the pre-school years are highly correlated with the same inputs measured when the child is aged 6 or older, our estimates may pick up the effect of inputs after pre-school years, which is not the aim of our paper. In Section A.2.1 we first control for post-preschool inputs of mothers hours and income, which are likely correlated with the pre-school inputs. The results show that our findings are not driven by the later inputs (columns 4 and 5 of Table 3). Similarly our benchmark results are robust to controlling for the incidence of divorce and the number of children measured between ages 6-11 (column 6).

Non-labour income may affect both mother's hours and child test scores but is not considered in the definition of household income. In column 7 our results do not change once we control for non-labour income. Non labour income is calculated as the sum across mothers and fathers of capital income, from stocks, mutual funds, money market funds, bank deposits, bonds etc. plus income from bank accounts and other financial sources. Our results also do not change when we allow household income to enter the equation for mothers hours (column 8) and test the sensitivity of our results to changing the definition of hours (column 9). A further potential bias checked in Section A.2.2 is whether including the family mean test scores as a covariate creates a bad control if this variable is not predetermined with respect to the instrumental variable. Again, our results are robust to this analysis. Finally, there may be nonlinearities in the relationship between hours and/or income and child outcomes. Section A.2.3 allows for a more flexible specification of the test score equation by first including quadratic terms in income and hours, and then estimating linear regression splines which allow, for example, for a differential effect of mothers hours moving from 0 to 1 compared to moving from 20 to 21, and finally by including an interactive effect between hours and income. Looking at these results we conclude that the linear specification is not rejected.

7 Mechanisms

In this section we investigate mechanisms explaining the effect of mothers' hours on child school test scores. In Section 7.1 we consider whether the effect of hours differs by subgroups of the population for which the mechanisms can differ, while in Section 7.2 we assess the role of potential channels through which mother's hours can affect school test scores by estimating whether mothers' hours and household income raise the child's school quality and the value of the house the household owns.

7.1 Heterogeneity analysis

In line with evidence discussed in Section 2.2, of different productivity of parental inputs on outcomes of boys and girls, our analysis next stratifies the sample by the child's gender to explore whether it is true for both sexes that income compensates for the negative effect of mothers' labour hours. The regression estimates and decomposition results reported in columns (2)-(3) of Tables 4 suggest that girls are less sensitive to a change in parental inputs than boys.[23] In particular, in Table 4 top panel we find that the direct effect of mothers' hours on the test score outcomes of girls is not statistically significant whilst mothers' weekly labour hours statistically significantly lowers the boys' test scores at 15 by 4.5% of a standard deviation. The decomposition analysis results in the bottom panel show that the very small negative direct effect of hours for girls is more than compensated by a relatively large mediator effect, which leads to a positive total effect that is statistically significant at 10% level. For boys, the relatively higher negative direct effect of hours is also fully compensated by the mediator effect leading to a total effect that is not statistically different to zero.

Almond and Currie (2011) and Heckman and Mosso (2014) suggest that economic modelling of the effect of parental inputs on child outcomes should allow for a different productivity by the human capital of parents, as discussed in Section 2.2. For this reason, next we consider heterogeneity in the analysis by the education level of the mother - measured by her degree status.

In Table 4, for mothers without a degree (see column 4), an increase in mothers' preschool hours has very little effect (and no statistically significant effect) on the child test score outcomes at age 15. That is, the effect of mothers' time with the child is very similar to the productivity of the counterfactual - either a formal childcare provider or the father, leaving a direct effect which is close to zero. However, as discussed in Section 2.2, the productivity of a highly educated mother's time may be relatively higher and as such there is a more negative direct effect with an hour increase in mothers' work hours lowering child test scores by 2.8% of a standard deviation, although not statistically different to zero (see column 5). Looking at test scores at age 11 the direct effect for mothers with a degree is even more negative, -4.6% of a standard deviation, and statistically significant at 10% level (see columns 4 and 5 in Table A.9).

[23]Table A.9 is the corresponding table for tests scores at age 11 rather than 15.

The productivity of household income for child human capital may be higher for households with low education if the increase in income loosens liquidity constraints in investments in their children, or if there is a diminishing marginal utility of income across the level of income. On the other hand, the productivity may be higher for households with high education, if an increase in income translates into larger increases in investments compared to low educated households (Gregg, Waldfogel, and Washbrook 2006). The results suggest that the productivity of household income is similar across education, whereby an increase of NOK100,000 raises outcomes at age 15 by 20.7% of standard deviation for mothers with a degree and by 17.3% for mothers without a degree.

Looking at the decomposition results by education in the bottom panel of Table 4 columns 4 and 5, the mediator effect of income is larger for mothers with a degree which is expected given that mothers with a degree are more likely to have a higher hourly wage rate. Income fully compensates for any negative direct effect for both mothers with and without a degree, so that the total effect is positive but statistically significant at the 10% level only for mothers with a degree.

To better understand the role of credit constraints on the productivity of income, we look next at the differential impacts of hours and income for parents with and without debt during preschool.[24] Results are reported in Table 4 columns 6 and 7. For households without debt we find a negative effect of hours and a positive effect of income on test scores and both effects are statistically significant at 1% level. Interestingly, for families in debt there is no statistically significant effect of either hours or income. A potential mechanism for the zero effect of hours could be that in households with debt there are higher stress levels and hence mothers time productivity may not be better than the counterfactual; whilst an increase in income is used to pay off the debt and consequently there is no longer a statistically significant effect of raising household income on child outcomes.

Finally, columns 8 and 9 of Table 4 stratify the sample across levels of childcare access. We measure childcare access as the childcare places for children aged 1-6 over the number of children age 1-6 in the municipality of residence in the first year after the child was born. Column 8 considers the subsample with childcare access in the bottom quartile, i.e. families

[24] Debt - or net debt - is calculated as the value of all assets (houses, balances in accounts, commercial buildings) minus the value of all debts (mortgage, bank account, credit cards, etc). 58% of households hold debt between child ages 1-5.

in municipalities which have between 24 and 46 places per 100 children, while column 9 considers the remaining sample. The mean of childcare access is 52 spaces for every 100 children, with a minimum of 0.24 and a maximum 0.88 respectively. We find a negative direct effect of hours only in the sample of households with low childcare access (although it is not statistically significant for either sample). This seems to suggest that the mother's childcare is replaced with a lower quality of informal childcare in municipalities with low childcare access.

7.2 Channels

A channel through which an increase in mother's hours may impact school test scores may be by moving to a better home or a better school, i.e. through improving the school and home environment inputs for the child. To address this, we evaluate the effect of an increase in mother's hours on the house value and on the school quality by replacing the child test score Y in our model (1) in turn with the family's house price and with the child's school quality.[25] The results in column 2 of Table 5 reveals that pre-school family income has a positive effect on the housing value suggesting that income is invested in part to improve the home and neighbourhood quality. The results for school quality in column 3 show firstly a positive coefficient on hours, indicating that working mothers choose to live in areas with better schools and secondly a positive coefficient on income, suggesting that pre-school income can "buy" higher school quality. It seems therefore that two mechanisms through which an increase in mother hours can improve children outcomes are the quality of the school and - through a higher house price - of the neighbourhood.

It could be that pre-school hours or income are correlated with later life household stability, such as the incidence of divorce or the completed family size, and it is these variables which drive our benchmark results. In Table 5 columns 4 and 5 we estimate our benchmark model but replace the dependent variable in the outcome equation with an indicator for divorce after the pre-school years (column 4) and the number of children after the pre-school years (column 5). In the regression results reported in panel a), neither mothers hours or

[25] We measure the house value at 2020 prices in 100,000 Kroner when the child is age 6 and measure school quality by taking the leave-one-out neighbourhood average of school test scores at age 11. Descriptive statistics for these two variables are reported in Table A.2.

household income statistically significantly drive later divorce or number of children. However the total effect of mothers hours on divorce (number of children) is negative (positive) and significant at the 10% level. Can these results explain the positive total effect of mothers hours in our benchmark estimation (see Table 5 columns 1)? An increase of mother's hours by 10 seems to have a positive total effect of 0.405 on number of children. Therefore, fertility decisions do not seem the driver of our main results given that a larger number of children is usually associated with a reduced rather than an increase in monetary investments per child. On the contrary, divorce may be one of the possible channels driving our results. An increase in preschool mother's hours seems to reduce the probability of divorce at 6-11, which ultimately can have a positive effect on children's test scores.

8 Conclusion

The main novelty of our paper is to identify the total, direct and mediated effect of mother's working hours during preschool on child test scores by using a novel instrumental variable estimation which relies on the partially overlapping peer approach. Variation induced using partially overlapping peer methods has previously been exploited to identify causal peer effects. In this paper we take the method further, applying the methodology to identify the effect of an individual or household's behaviour on children's outcomes. Our strategy can be applied to answer other research questions on the causal effect of individual behaviours which may be affected by peers, e.g. the consumption of health care such as the decision of immunise a child, or workers behaviours such as decisions on the timing of retirement.

Does income compensate for mothers working during preschool years? By using the overlapping peers approach and Norwegian administrative data covering the full population of first born children between 1997 and 2001, we decomposed the total effect of an increase in mothers' labour hours on children into the causal direct effect through a reduction in her time investments, and the causal mediator income effect. We find that any negative effect of an increase in mothers' hours worked on child outcomes at age 11 and 15 is compensated by the increase in household income. The estimated total effect suggests that a 10-hour increase in the mother's average weekly hours in pre-school years leads to a rise in test scores at age 15 by 26% of a standard deviation.

An increase in mothers' hours worked leads to changes in the time allocation of children by replacing the time a mother spends with her child with alternative childcare time. This may create a potential decrease of the total time the child spends in educational, play and other activities that are important for child development. An increase by 10 in the mother's worked hours per week during preschool years seems to lead to changes in time investments, conditional on household income, causing a decrease in child's test scores at age 11 and 15 by around 20% of a standard deviation. We find that the direct effect is negative but not statistically significant different to zero, therefore suggesting that there are no differences in the productivity of the mothers' time relative to time spent in an alternative childcare. This conclusion seems plausible in the Norwegian context where we can believe that the alternative to mother's time, which is usually formal childcare or time with the father, is of high quality. However, if we had a strong belief that mother's time is better than any alternative childcare and we formally tested for such hypothesis using a one-tailed test, then we would not reject that the direct effect of mother's hours is negative, i.e. that mother's time is at least as productive as the alternative time confirming results in some previous papers.

A paper with a similar research aim to ours is Agostinelli and Sorrenti (2021) who estimate in the US the role of mothers' labour hours and household income in driving child cognitive and behavioural skills. A common finding across the two papers is a positive effect of income and a negative effect of mothers' work hours for child development. Agostinelli and Sorrenti (2021) find that the direct effect of an increase by 100 of the mother's yearly hours (about 2 hours per week) leads to a reduction in maths and reading test scores by about 3% of a standard deviation. The equivalent direct effect in our benchmark estimation is about 4% of a standard deviation and therefore very comparable. The difference is that whilst we find that income fully compensates for the increased labour market hours of the mother, in Agostinelli and Sorrenti (2021) the negative effect of hours is not fully cancelled out by the positive effect of income. This difference is likely caused by the fact that our instrument for income identifies variation in income coming from parents at different points of the income distribution, whereas Agostinelli and Sorrenti (2021) identify changes in income for the lower end of the income distribution, i.e. those eligible for EITC.

Our work is also related to previous papers that have estimated the role of parental time investments in early life using structural models, including Bernal (2008) and Del Boca, Flinn, and Wiswall (2014). Unlike these papers we do not estimate the effect of the productivity of mothers' time investments, but estimate the effect relative to a counterfactual form of childcare. Assuming that working mothers replace their time investments with formal childcare time, our results are comparable to Bernal (2008) who finds that the direct effect of one year of full time work and child care use during preschool leads to a reduction in child test scores by about 13% of a standard deviation. Assuming instead that the mothers' time gets replaced with the fathers' time, our results are comparable to Del Boca, Flinn, and Wiswall (2014), who find that the productivity of father's time is lower than the mother's time for both passive and active time spent with the child in the age period 1-5.

Finally, looking at the potential mechanisms through which hours may lead to better test scores, we find evidence that increasing mothers hours raises both the housing value and school quality, suggesting that school and neighbourhood inputs are likely to be relevant mechanisms of the effect of hours.

References

Acharya, A., M. Blackwell, and M. Sen (2016): "Explaining causal findings without bias: Detecting and assessing direct effects," *American Political Science Review*, 110(3), 512–529.

Agostinelli, F., and G. Sorrenti (2021): "Money vs . Time: Family Income, Maternal Labor Supply, and Child Development," Discussion Paper 273, University of Zurich, Department of Economics.

Aklin, M., and P. Bayer (2017): "How Can We Estimate the Effectiveness of Institutions? Solving the Post-Treatment versus Omitted Variable Bias Dilemma," *Working paper.*

Almond, D., and J. Currie (2011): "Human Capital Development Before Age Five," *In D. Card and O. Ashenfelter (eds.) Handbook of Labor Economics*, 4(15), 1315–1486.

Angrist, J. D. (2014): "The perils of peer effects," *Labour Economics*, 30, 98–108.

Angrist, J. D., and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020): "Estimating the production function for human capital: results from a randomized controlled trial in Colombia," *American Economic Review*, 110(1), 48–85.

Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2016): "School quality and the gender gap in educational achievement," *American Economic Review*, 106(5), 289–95.

Baker, M., and K. Milligan (2010): "Evidence from maternity leave expansions of the impact of maternal care on early child development," *Journal of Human Resources*, 45(1), 1–32.

——— (2015): "Maternity leave and children's cognitive and behavioral development," *Journal of Population Economics*, 28(2), 373–391.

Bernal, R. (2008): "The Effect of Maternal Employment and Child Care on Children's Cognitive Development," *International Economic Review*, 49(4), 1173–1209.

Bernal, R., and M. P. Keane (2011): "Child care choices and children's cognitive achievement: The case of single mothers," *Journal of Labor Economics*, 29(3), 459–512.

Bertrand, M., and J. Pan (2013): "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior," *American Economic Journal: Applied Economics*, 5(1), 32–64.

Black, S. E., P. J. Devereux, K. V. Løken, and K. G. Salvanes (2014): "Care or cash? The effect of child care subsidies on student performance," *Review of Economics and Statistics*, 96(5), 824–837.

Blau, B. F. D., and L. M. Kahn (2013): "Female Labor Supply : Why Is the United States Falling Behind ?," 103(3), 251–256.

Blau, F. D., and A. J. Grossbergm (1992): "Maternal labour supply and children's cognitive development," *Review of Economics and Statistics*, pp. 474–481.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption inequality and partial insurance," *American Economic Review*, 98(5), 1887–1921.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2018): "Children, time allocation, and consumption insurance," *Journal of Political Economy*, 126(S1), S73–S115.

Boneva, T., and C. Rauh (2018): "Parental Beliefs about Returns to Educational Investments - The Later the Better?," *Journal European Economic Association*, 16(6), 1669–1711.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150(1), 41–55.

Bullock, J. G., D. P. Green, and S. E. Ha (2010): "Yes, but what's the mechanism?(don't expect an easy answer).," *Journal of Personality and Social Psychology*, 98(4), 550.

Caeyers, B., and M. Fafchamps (2016): "Exclusion Bias in the Estimation of Peer Effects," *NBER working paper*, (22565).

Carneiro, P., I. L. García, K. G. Salvanes, and E. Tominey (2021): "Intergenerational mobility and the timing of parental income," *Journal of Political Economy*, 129(3), 757–788.

Carneiro, P., and R. Ginja (2016): "Partial insurance and investments in children," *The Economic Journal*, 126(596), F66–F95.

Carneiro, P., K. V. Løken, and K. G. Salvanes (2015): "A Flying Start? Maternity Leave Benefits and Long-Run Outcomes of Children," *Journal of Political Economy*, 123(2), 365–412.

Caucutt, E. M., and L. Lochner (2020): "Early and late human capital investments, borrowing constraints, and the family," *Journal of Political Economy*, 128(3), 1065–1147.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2018): "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance," *Journal of Political Economy*, 126(6), 2356–2409.

Cornelissen, T., C. Dustmann, and U. Schönberg (2017): "Peer Effects in the Workplace," *American Economic Review*, 107(2), 425–56.

Cunha, F., and J. Heckman (2007): "The Technology of Skill Formation Author(s): Source: T," *American Economic Review*, 97(c), 31–47.

Cunha, F., and J. J. Heckman (2008): "Formulating , Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *The Journal of Human Resources*, 43(4), 738–782.

Dahl, B. 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(5), 1927–1956.

Dahl, G. B., K. V. Løken, M. Mogstad, and K. V. Salvanes (2016): "What is the case for paid maternity leave?," *Review of Economics and Statistics*, 98(4), 655–670.

Danzer, N., and V. Lavy (2013): "Parental leave and children's schooling outcomes: Quasi-experimental evidence from a large parental leave reform," Discussion paper, National Bureau of Economic Research.

De Giorgi, G., A. Frederiksen, and L. Pistaferri (2020): "Consumption network effects," *The Review of Economic Studies*, 87(1), 130–163.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010): "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics*, 2(2), 241–75.

Del Boca, D., C. Flinn, and M. Wiswall (2014): "Household choices and child development," *Review of Economic Studies*, 81(1), 137–185.

Del Bono, E., M. Francesconi, Y. Kelly, and A. Sacker (2016): "Early maternal time investment and early child outcomes," *The Economic Journal*, 126(596), F96–F135.

Deuchert, E., M. Huber, and M. Schelker (2019): "Direct and indirect effects based on difference-in-differences with an application to political preferences following the Vietnam draft lottery," *Journal of Business & Economic Statistics*, 37(4), 710–720.

DIPPEL, C., R. GOLD, S. HEBLICH, AND R. PINTO (2019): "Mediation Analysis in IV Settings With a Single Instrument," Discussion paper, Mimeo.

DOTTI SANI, G. M., AND J. TREAS (2016): "Educational gradients in parents' child-care time across countries, 1965–2012," *Journal of Marriage and Family*, 78(4), 1083–1096.

DUSTMANN, C., AND U. SCHÖNBERG (2012): "Expansions in maternity leave coverage and children's long-term outcomes," *American Economic Journal: Applied Economics*, 4(3), 190–224.

ERMISCH, J., AND M. FRANCESCONI (2013): "The effect of parental employment on child schooling," *Journal of Applied Econometrics*, 28, 796–822.

FAGERENG, A., M. MOGSTAD, AND M. RØNNING (2021): "Why do wealthy parents have wealthy children?," *Journal of Political Economy*, 129(3), 703–756.

FAN, X., H. FANG, AND S. MARKUSSEN (2015): "Mothers' Employment and Children's Educational Gender Gap," Discussion paper, National Bureau of Economic Research.

FIORINI, M., AND M. P. KEANE (2014): "How the allocation of children's time affects cognitive and noncognitive development," *Journal of Labor Economics*, 32(4), 787–836.

FRÖLICH, M., AND M. HUBER (2017): "Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables Series B Statistical methodology," Discussion Paper 8280, IZA.

GOUX, D., E. MAURIN, AND B. PETRONGOLO (2014): "Worktime regulations and spousal labor supply," *American Economic Review*, 104(1), 252–76.

GREGG, P., J. WALDFOGEL, AND E. WASHBROOK (2006): "Family expenditures post-welfare reform in the UK: are low-income families starting to catch up?," *Labour Economics*, 13(6), 721–746.

GURYAN, J., E. HURST, AND M. KEARNEY (2008): "Parental Education and Parental Time with Children," *Journal of Economic Perspectives*, 22(3), 23–46.

Havnes, T., and M. Mogstad (2011): "No child left behind: Subsidized child care and children's long-run outcomes," *American Economic Journal: Economic Policy*, 3(2), 97–129.

——— (2015): "Is universal child care leveling the playing field?," *Journal of Public Economics*, 127, 100–114.

Heckman, B. J., R. Pinto, and P. Savelyev (2013): "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103(6), 2052–2086.

Heckman, J. J., and S. Mosso (2014): "The economics of human development and social mobility," *Annual Review of Economics*, 6(1), 689–733.

Heckman, J. J., and R. Pinto (2015): "Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs," *Econometric Reviews*, 34(1-2), 6–31.

Hill, C. R., and F. P. Stafford (1974): "Allocation of Time to Preschool Children and Educational Opportunity.," *Journal of Human Resources*, 9(3).

Hsin, A., and C. Felfe (2014): "When does time matter? Maternal employment, children's time with parents, and child development," *Demography*, 51(5), 1867–1894.

Huber, M. (2020): "Mediation Analysis," in *Handbook of Labor, Human Resources and Population Economics*, ed. by G. L. Organization. Springer International Publishing.

Huebener, M., D. Kuehnle, and C. K. Spiess (2019): "Parental leave policies and socio-economic gaps in child development: Evidence from a substantial benefit reform using administrative data," *Labour Economics*, 61, 101754.

Imai, K., L. Keele, D. Tingley, and T. Yamamoto (2011): "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies," *American Political Science Review*, 105(4), 765–789.

Kalil, A. (2015): "Inequality begins at home: The role of parenting in the diverging destinies of rich and poor children," in *Families in an era of increasing inequality*, pp. 63–82. Springer.

Kalil, A., R. Ryan, and M. Corey (2012): "Diverging destinies: Maternal education and the developmental gradient in time with children," *Demography*, 49(4), 1361–1383.

Lee, L.-f., X. Liu, and X. Lin (2010): "Specification and estimation of social interaction models with network structures," *The Econometrics Journal*, 13(2), 145–176.

Lewbel, A., S. Norris, K. Pendakur, and X. Qu (2018): "Consumption Peer Effects and Utility Needs in India," *Boston College Working Paper*.

Liu, Q., and O. N. Skans (2010): "The duration of paid parental leave and children's scholastic performance," *The BE Journal of Economic Analysis & Policy*, 10(1).

Macmillan, L., and E. Tominey (2020): "Parental Inputs and Socio-Economic Gaps in Early Child Development," Discussion Paper 20-04, CEPEO.

Manski, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60(3), 531–542.

Maurin, E., and J. Moschion (2004): "The Social Multiplier and Labour Market Participation of Mothers," *American Economic Journal: Applied Economics*, 1(1), 251–272.

McLanahan, S. (2004): "Diverging destinies: How children are faring under the second demographic transition," *Demography*, 41(4), 607–627.

Mota, N., M. F. P. E., and S. Rosenthal (2016): "Neighbourhood Effects, Peer Classification, and the Decision of Women to Work," *IZA Discussion Paper*, 9985.

Nicoletti, C., and B. Rabe (2019): "Sibling spillover effects in school achievement," *Journal of Applied Econometrics*, 34(4), 482–501.

Nicoletti, C., K. G. Salvanes, and E. Tominey (2018): "The Family Peer Effect on Mothers' Labor Supply," *American Economic Journal: Applied Economics*, 10(3), 206–234.

Olivetti, C., E. Patacchini, and Y. Zenou (2020): "Mothers, peers, and gender-role identity," *Journal of the European Economic Association*, 18(1), 266–301.

Pearl, J. (2012): "The causal mediation formula—a guide to the assessment of pathways and mechanisms," *Prevention Science*, 13(4), 426–436.

Powdthavee, N., W. N. Lekfuangfu, and M. Wooden (2013): "The marginal income effect of education on happiness: Estimating the direct and indirect effects of compulsory schooling on well-being in Australia," Discussion Paper 7365, IZA.

Rasmussen, A. W. (2010): "Increasing the length of parents' birth-related leave: The effect on children's long-term educational outcomes," *Labour Economics*, 17(1), 91–100.

Robins, J. M. (2003): "Semantics of causal DAG models and the identification of direct and indirect effects," *Oxford Statistical Science Series*, pp. 70–82.

Rosenbaum, P. R. (1984): "The consequences of adjustment for a concomitant variable that has been affected by the treatment," *Journal of the Royal Statistical Society: Series A (General)*, 147(5), 656–666.

VanderWeele, T. (2016): "A unification of mediation and interaction: A 4-way decomposition (Vol 25, pg 749, 2014)," *Epidemiology*, 27(5), E36–E36.

VanderWeele, T. J. (2013): "A three-way decomposition of a total effect into direct, indirect, and interactive effects," *Epidemiology (Cambridge, Mass.)*, 24(2), 224.

von Hinke, S., G. Leckie, and C. Nicoletti (2019): "The use of instrumental variables in peer effects models," *Oxford Bulletin of Economics and Statistics*, 81(5), 1179–1191.

Tables

Table 1: Descriptive statistics: Full sample of first births

Variable	Mean	SD
Test score age 15	64.413	20.981
Endogenous variables		
Household income	5.200	2.323
Mothers' hours	19.836	13.320
Instrumental variables		
Father earnings: workmates' family	1.977	1.316
Hours: family's workmates	19.950	11.860
Covariates		
Mothers' years schooling	13.336	2.392
Mothers' degree	0.391	0.488
Mothers' age birth	26.519	4.625
Working before birth	0.762	0.426
Mother participation 1-5	0.882	0.322
Father earnings before birth	2.891	3.019
Father participation year before birth	0.965	0.185
Fathers' education	12.850	2.461
Child month of birth	6.418	3.387
Child year of birth	1998.983	1.409
Child birth weight	3512.875	565.689
Mean family test score 15	68.615	25.720
Peer group sizes		
Coworkers	16.356	34.915
Family	2.397	1.220
Observations	64,762	64,762

Notes: Data source, Norwegian administrative data, first-born children born in 1997-2001. Income is the yearly household net income in NOK100,000 at 2020 prices and hours is the mother's weekly work hours, both are averaged across years 1-5 after the first child birth.

Table 2: Estimation results

	(1) OLS	(2) 3SLS
a) Test score equation		
Mothers' hours γ_H^Y	0.002***	-0.023
	(0.000)	(0.018)
Household income γ_I^Y	0.041***	0.192***
	(0.002)	(0.047)
b) Income equation		
Effect of hours on income γ_H^I	0.042***	0.258***
	(0.001)	(0.045)
c) Decomposition		
Direct effect (effect of hours on test scores) γ_H^Y		-0.023
		(0.018)
Mediator effect ($\gamma_H^I \gamma_I^Y$)		0.049***
		(0.015)
Total effect of hours on test scores ($\gamma_H^Y + \gamma_H^I \gamma_I^Y$)		0.026**
		(0.013)
Observations	64,762	64,762
Endogenous equation statistics		
F-statistic IV hours		33.66
F-statistic IV income		122.94
Endogeneity test score equation p-value		0.000
Endogeneity income equation p-value		0.000

Notes: The results in columns (1) and (2) are computed using the OLS and 3SLS estimation respectively of our benchmark model 1. Panel a) reports the effect of hours and income on school test scores; in panel b) the effect of hours on income; in panel c) direct and mediator effects defined as γ_H^Y and ($\gamma_H^I \gamma_I^Y$) along with the total effect. Hours in column (2) are instrumented with the mean family's workmates' hours, income in column (2) is instrumented with the mean workmates family's income. Standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is school test scores at age 15 measured in standard deviations. Income is the yearly household net income in NOK 100,000 at 2020 prices and hours is the mother's weekly work hours, both are averaged across years 1-5 after the first child birth. All remaining controls are measured at or before the child birth (see Table 1). Endogeneity of H (I) p-value is the p-value of the correlation between the error term for the H (I) equation and the Y equation. Data sources: Norwegian administrative data, first-born children born in 1997-2001.

Table 3: Sensitivity and validity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	a) Estimation results								
				Controls					
	Benchmark	Labour market	Local maternal hours	Income 6-11	Hours 6-11	Children & divorce 6-11	Non labour income	Income in hours eqn	Change hours definition
Mothers' hours	-0.023	-0.020	-0.021	-0.026	-0.030	-0.027	-0.024	-0.023	-0.024
	(0.018)	(0.020)	(0.018)	(0.019)	(0.027)	(0.018)	(0.018)	(0.018)	(0.020)
Household income	0.192***	0.195***	0.192***	0.264***	0.192***	0.195***	0.193***	0.191***	0.192***
	(0.047)	(0.054)	(0.047)	(0.066)	(0.047)	(0.048)	(0.047)	(0.049)	(0.048)
	b) Decomposition results for 10 hours increase								
Direct	-0.231	-0.205	-0.207	-0.263	-0.304	-0.268	-0.240	-0.230	-0.245
	(0.182)	(0.201)	(0.178)	(0.185)	(0.267)	(0.184)	(0.181)	(0.184)	(0.197)
Mediator	0.495***	0.508***	0.494***	0.536***	0.704***	0.501***	0.500***	0.494***	0.522***
	(0.149)	(0.167)	(0.147)	(0.166)	(0.232)	(0.150)	(0.149)	(0.152)	(0.162)
Total	0.263**	0.303**	0.287**	0.273*	0.400*	0.233*	0.259**	0.263**	0.277**
	(0.132)	(0.129)	(0.131)	(0.152)	(0.205)	(0.129)	(0.130)	(0.132)	(0.139)
Observations	64,762	64,181	64,762	64,760	64,762	64,511	64677	64,762	64,762

Notes: Column (1) benchmark 3SLS estimates; column (2) controls for labour market fixed effects; column (3) benchmark + neighbourhood level mean hours worked of mothers 1 year after birth ; column (4) benchmark + control for household income averaged across ages 6-11; column (5) benchmark + control for mothers' hours averaged across ages 6-11; column (6) benchmark + control for number of children and incidence of divorce child aged 6-11; column (7) benchmark + control for non-labour income (capital income from stocks, mutual funds, money market funds, bank deposits, bonds, interest) in pre-school ages 1-5; column (8) benchmark + control for pre-school income in the hours equation; column (9) measures hours as 14 (rather than 10.5) if mothers report working 1-19. Panel a) reports regression results from model 1 and panel b) the direct and mediator effects defined as γ_H^Y and $(\gamma_H^I \gamma_I^Y)$, and then the total effect of mothers hours on school test scores $(\gamma_H^Y + \gamma_H^I \gamma_I^Y)$. Standard errors in parentheses in panel a) computed using the 3SLS estimation and in panel b) by delta method. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is school test scores at age 15 measured in standard deviations. Income is the yearly household net income in NOK100,000 at 2020 prices and hours is the mother's weekly work hours, both are averaged across years 1-5 after the first child birth. Panel a) coefficient estimates from the Y-equation in 1 and panel b) decomposition analysis from equations 3, 6. All remaining controls are measured at or before the child birth (see Table 1). Data sources: Norwegian administrative data, first-born children born in 1997-2001.

Table 4: Heterogeneity

	(1) Benchmark	(2) Girls	(3) Boys	(4) No degree	(5) Degree	(6) No debt 1-5	(7) Debt 1-5	(8) Low childcare access	(9) High childcare access
a) Estimation results									
Mothers' hours	-0.023	-0.003	-0.045*	-0.013	-0.028	-0.076***	0.010	-0.055	0.004
	(0.018)	(0.024)	(0.026)	(0.025)	(0.026)	(0.025)	(0.025)	(0.122)	(0.018)
Household income	0.192***	0.175**	0.215***	0.173**	0.207***	0.329***	0.101	0.283	0.088**
	(0.047)	(0.068)	(0.065)	(0.080)	(0.056)	(0.076)	(0.070)	(0.429)	(0.043)
b) Decomposition results for 10 hours increase									
Direct	-0.231	-0.031	-0.447*	-0.128	-0.283	-0.763***	0.096	-0.554	0.038
	(0.182)	(0.236)	(0.258)	(0.247)	(0.257)	(0.250)	(0.253)	(1.223)	(0.179)
Mediator	0.495***	0.325**	0.685***	0.361*	0.636***	0.836***	0.220	0.569	0.241*
	(0.149)	(0.155)	(0.250)	(0.187)	(0.234)	(0.263)	(0.159)	(0.962)	(0.126)
Total	0.263**	0.295*	0.238	0.233	0.353*	0.074	0.316*	0.015	0.279*
	(0.132)	(0.179)	(0.198)	(0.182)	(0.193)	(0.209)	(0.177)	(0.386)	(0.159)
Observations	64,762	33,107	31,655	39,448	25,314	26,595	37,192	15,220	45,569

Notes: Column (1) benchmark 3SLS estimates; column (2) sample of girls; column (3) sample of boys; column (4) sample of mothers with no degree; column (5) sample of mothers with a degree column (6) sample with no debt (total assets - total debts) ages 1-5; column (7) sample with some debt ages 1-5; column(8) households in low childcare access municipalities (24-46 places available per 100 children aged 1-6, or bottom quartile) and column (9) households not in low childcare municipalities (places available per 100 children aged 1-6). Standard errors in parentheses in panel a) computed using the 3SLS estimation and in panel b) by delta method. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is school test scores at age 15 measured in standard deviations. Income is the yearly household net income in NOK100,000 at 2020 prices and hours is the mother's weekly work hours, both are averaged across years 1-5 after the first child birth. Panel a) coefficient estimates from equation 1 and panel b) decomposition analysis from equations 3, 13 and 15. All remaining controls are measured at or before the child birth (see Table 1). Data sources: Norwegian administrative data, first-born children born in 1997-2001.

Table 5: Mechanisms

	(1) Benchmark	(2) House value age 6	(3) School quality	(4) Later divorce as outcome	(5) Number children by 11
			a) Estimation results		
Mothers' hours	-0.023	0.112	0.108***	-0.003	0.029
	(0.018)	(0.539)	(0.034)	(0.004)	(0.023)
Household income	0.192***	3.580**	0.284***	-0.016	0.043
	(0.047)	(1.474)	(0.072)	(0.012)	(0.058)
		b) Decomposition results for 10 hours increase			
Direct	-0.231	1.119	1.080***	-0.031	0.291
	(0.182)	(5.389)	(0.335)	(0.044)	(0.229)
Mediator	0.495***	8.927**	0.725***	-0.042	0.113
	(0.149)	(3.971)	(0.226)	(0.031)	(0.153)
Total	0.263**	10.046**	1.805***	-0.074**	0.405**
	(0.132)	(4.269)	(0.360)	(0.034)	(0.177)
Observations	64,762	64,151	64,762	64,656	64512

Notes: Column (1) benchmark 3SLS estimates; column (2) dependent variable in outcome eqn replaced with house value at 2020 prices measured at age 6 (in 100,000 Kroner); column (3) dependent variable is school quality at age 11, measured as the neighbourhood average of a national test score at age 11; column (4) dependent variable is incidence of divorce or separation between ages 6-11; column (5) dependent variable is number of children by age 11. Standard errors in parentheses in panel a) computed using the 3SLS estimation and in panel b) by delta method. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is school test scores at age 15 measured in standard deviations. Income is the yearly household net income in NOK100,000 at 2020 prices and hours is the mother's weekly work hours, both are averaged across years 1-5 after the first child birth. Panel a) coefficient estimates from equation 1 and panel b) decomposition analysis from equations 3, 13 and 15. All remaining controls are measured at or before the child birth (see Table 1). Data sources: Norwegian administrative data, first-born children born in 1997-2001.

A Online Appendix

A.1 Estimation in presence of measurement errors

In our application we measure mother's work hours in each of the 5 years after her first childbirth. These variables are subject to measurement error. This is because for all mothers we observe their working hours in November of the considered year after their childbirth. This implies that the number of hours worked Δ years after childbirth by women who gave birth in January of the year t is observed in November of the year $(t+\Delta)$, i.e. $[12\ \Delta\ +10]$ months after childbirth, while for women giving birth in December of the year t we observe their labour supply only $[12\ \Delta\ -1]$ months after childbirth. Henceforth we define our outcome variable as the mother's working hours Δ years and 6 months after childbirth, where $\Delta = 1, ..., 5$. This implies that the working hours for women who give birth in May of the year t is correct, but the working hours for women who do not give birth in May will be subject to measurement error and will be probably overestimated for women giving birth between December and May and underestimated for women giving birth between June and November. This is especially true for the first years after childbirth where female labour supply is subject to more change than in later years.

Furthermore, we do not observe the exact number of hours worked, but we know whether the mother works 0, between 1 and 19, 20 and 29 or 30 or more hours per week. By rounding the working hours to 0 for non-working mothers and to 10, 24.5 and 40 for working mothers, we can use this "rounded" variable and quantify and compare differences between mothers in term of hours.

The measurement errors caused by the rounding and by the month of observation affect not only H, but also the corresponding average of hours for the family (sisters and sister-in-laws), $\overline{H}^F$, and for the workmates $\overline{H}^W$. We do not have any reason to believe that such measurement errors be correlated with any of observed and unobserved variables in our model. For this reason, in the following we assume that H follows the model

$$H = TH + Dmonth\,\eta + e \tag{9}$$

where TH is the true working hours, $Dmonth$ is a row vector of 12 dummy variables indicating the month of birth of the child, η is the column vector of corresponding coefficients and e is a classical measurement error which is independently and identically distributed across individuals, independent of the true value TH and independent of the explanatory variables and error terms in our model of interest.

Under this modified classical measurement error model, the error on H would cause an attenuation bias of the effect of H on Y when using ordinary least squares estimation. However, because we use an instrumental variable approach this bias cancels out. The instrument $\overline{Z}_H^{FW}$ is the leave-one-out average of work hours over family's workmates one year after childbirth and observed one year or more before the focal mother gives birth. Similarly to H, the hours for each of the family's workmates follow model (9) so that the average hours across family's workmate will be equal to:

$$\overline{Z}^{FW} = \overline{TZ}_H^{FW} + \overline{Dmonth}\,\eta + \overline{e}, \tag{10}$$

where $\overline{TZ}_H^{FW}$ is the true average of work hours across the family's workmates, and $\overline{Dmonth}$ and $\overline{e}$ are the leave-one-out averages of $Dmonth$ and e over the family's workmates. Because the error e is independently and identically distributed across individuals and the leave-one-out average $\overline{e}$ excludes the focal mother, $\overline{e}$ and e are independent. In conclusion, measurement errors for the hours worked do not cause any inconsistency for our three-stage least squares estimation.

A.2 Additional Sensitivity Analysis

A.2.1 Different sets of control variables

There is likely to be a high correlation between parental inputs in the 1-5 years after birth of a first child and parental inputs from age 6 onwards. Whilst we may expect household income and mothers' hours to vary year-on-year, the level of the inputs in the pre-school years is likely highly correlated with inputs from age 6 onwards.

To address a concern that our estimates of the effect of pre-school parental inputs on children may pick up the effect of inputs that occur *after* the child starts school, we estimate the mean household income and mean mothers' hours between years 6-11 and control for each in turn in regression of test score outcomes at age 15.[26] These are clearly endogenous controls, but our analysis reported in Table 3 shows that our results are not statistically different to our benchmark estimates once we condition on income (column 4) or hours (column 5) after the pre-school years. Moreover, the decomposition analysis are again similar to our benchmark results suggesting that our results are robust to removing the influence of income or hours observed after pre-school years.[27] Similarly, we do not find differences in the results when we add in our benchmark model the incidence of divorce and the number of children between ages 6-11[28] (see column 6).

In column 7, we estimate our benchmark model controlling for non-labour income[29] average between age 1 and 5 in an attempt to check whether including this specific component of the household income may lead to different results and again we do not find any relevant change.

[26] Income and hours data is not available past the child's age of 13 in some cohorts, so for simplicity we control for income or hours up to age 11.descriptives of these variables are reported in Table A.2.

[27] Another way to think of this problem is to ask whether income and mothers hours for the child aged 6-11 change in response to the same inputs measured during pre-school years. We repeated our benchmark analysis but replacing the test score outcome with H or I measured at ages 6-11. There was a statistically significant effect of income in pre-school years on income in the subsequent years suggesting this is a potential mechanism. However importantly, even controlling for income age 6-11, our benchmark results hold.

[28] Likewise to income and hours, we consider divorce and number of children up to age 11 (rather than age 15) due to data limitations. See Table A.2 for descriptive statistics.

[29] Non labour income is calculated as the sum across mothers and fathers of capital income, from stocks, mutual funds, money market funds, bank deposits, bonds etc plus income from bank accounts and other financial sources. In Table A.2 we report the mean and standard deviation of this variable.

As discussed in Section 2.2, whilst our benchmark specification controls for fathers' income measured pre-birth, our results are robust to allowing for mothers' pre-school hours to react to household income post-birth, as shown in Table 3 column 8. The reason the results are similar is that household income does not statistically significantly drive mothers' hours.[30]

Because the mother's weekly hours observed in a specific year is a discrete variable taking the value of 0, 1-19, 20-29 and 30+ hours and we replaced it with a variable taking the midpoint of each bin, we are concerned whether this midpoint choice is appropriate. For this reason we checked whether our chosen midpoints are close to the mean for the corresponding bin using the Norwegian Labour Force Survey (LFS) in 2007. We found that our midpoints of 10.5 (1-19), 24.5 (20-29) ad 40 (30+) and the corresponding mean values in the LFS, which are 14, 24 and 38 respectively, are approximately equal except for the bin 1-19. Therefore, we run a sensitivity analysis where the midpoint for the bin 1-19 was replaced with 14 rather than 10.5. The results are reported in Table 3 column 9 and suggest that there are no changes with respect to the benchmark specification results. We have also re-estimated the full set of heterogeneity analysis from Table 4 with the different midpoint for bin 1-19 and results (available on request) are almost identical.

A.2.2 Testing exogeneity of the family mean test scores

In our empirical specification we control for the mean test score outcomes of the focal household's family members, in order to remove any potential bias from unobservable traits correlated with the family network and with child outcomes. By the nature of the peer group definition, this control is predetermined with respect to the test score outcomes of the focal child, however it is potentially not predetermined with respect to the instrumental variable in which case it represents a bad control.

It is possible to create an instrumental variable for the family mean test scores at ages 11 and 15 by exploiting an indirect peer group different to those used to construct the IVs in our benchmark specification - the family mean test score defined for the workmates of the focal household's neighbours. A neighbourhood is defined by the postcode of residence, and a relevant neighbourhood peer consists of all neighbours who gave birth between 1-5

[30]The coefficient on household income in the hours equation is -0.019 with standard error 0.627.

years prior to the focal household with the same level of degree status of the mother. On average, there are around 2500 households within a postcode, and 63 neighbourhood peers. The intuition for the IV for the family mean test score is that the family mean is partially determined by child ability within the neighbourhood, which itself is partially determined by child ability of the workplace peers. We append our estimation by including an equation for the mean family test scores and test for the exogeneity of the variable by examining the correlation between the error term in the test score equation and the mean family test score equation. The F-statistics testing the strength of the IV for mean family test score is very high at 791 at age 15, and the p-value for the correlation between the error terms from the test score equation and the mean family test score equation is 0.551 which indicates that any correlations between errors is not statically different to zero. This suggests that the IVs for the mean family test scores are strong predictors and that we cannot reject the null hypothesis that the mean family test score is exogenous. For this reason, the mean family test scores does not seem to be a bad control.

A.2.3 Non-linearity in the effect of income and hours on test scores

Our benchmark model specification assumed that income and hours affect child test scores linearly. We firstly relax that assumption by including as regressors in the child test score equations at age 15 the square of household income and mothers' hours worked. Reported in panel a) of Table A.10, the coefficients for hours squared and income squared in the test score equation are not statistically significant at 5% and not even at 10% level. Furthermore, Panel b) of Table A.10 shows that the decomposition estimates are almost identical to our benchmark.

Next we test for a non-linear effect of hours on test score outcomes in a more flexible specification. It may be that the marginal effect of an increase mothers' working hours on children is very different when comparing mothers moving from a low number of hours, to an increase starting from a higher level of hours. To test this, we use linear regression splines in mothers' working hours, with knots defined at the quartiles of mothers' working hours.[31]

[31] We use a control function approach for the estimation of all our non-linear models. In a first step a regression of hours (income) is estimated on the full set of controls and the IV. Next we estimate jointly the test score and income equations, where we include both residuals (hours and income) in the test score equation and the residual for hours in the income equation.

The 4 quartiles for mother's hours are 1.285, 5.918, 13.149 and 19.836. Table A.11 reports in panel a) the regression results, where the coefficients represent the differential effect of hours at a particular quartile relative to the first, while panel b) reports the decomposition results for an increase of mothers' hours by 10. There is neither evidence of a non-linear effect of hours on child test score nor is there any change to our conclusion that the income completely compensates for the negative direct effect of an increase in mother's hours.

In a similar way, we test for a non-linear effect of income on test scores by using a linear regression spline across deciles of the income distribution. The 10 deciles for household income in NOK100,000 are 0.165, 0.386, 0.601, 0.816, 1.046, 1.308, 1.627, 2.055, 2.727 and 5.200. Table A.12 results suggest that there is no non-linearity in the effect of income on child test scores. In the decomposition results reported in Table A.13 we draw the same conclusion that the total effect of an increase in mothers' labour hours is not statistically different across the distribution of income.

Next we test for the presence of an interactive effect of hours and family income on child test scores. The return to household income may change across the distribution of hours or vice versa. Therefore, we include the interaction term between mother's hours and income in the test score equation, i.e.

$$Y = \gamma_0'^Y + H\gamma_H'^Y + I\gamma_I'^Y + I \cdot H\gamma_{I,H}'^Y + \mathbf{X}\boldsymbol{\beta}'^Y + u'^Y \tag{11}$$

where $\gamma_{IH}'^Y$ denotes the coefficient on the interaction term between I and H.[32]

Adding the interaction term $(I \cdot H)$ to the test score equation results in changes in the direct and mediator effects and in an additional term to the total effect, which we term the *interaction effect* of mothers' work hours and household income, given by the equation

$$E(Y_{h_1,I_{h_1}} - Y_{h_1,I_{h_0}} - Y_{h_0,I_{h_1}} + Y_{h_0,I_{h_0}}), \tag{12}$$

which differs from zero only if there are both a mediation and an interaction effect, i.e. if both $\gamma_H'^I$ and $\gamma_{I,H}'^Y$ are not zero.

[32]Notice that in the above and following equations we use the same notation as in (1) but we add a superscript prime for all parameters and error terms.

Adding the interaction term to the test score equation will add the term $(h_1 - h_0)E(I_{h_0})\gamma'^{Y}_{I,H}$ to the direct effect which becomes

$$E(Y_{h_1,I_{h_0}} - Y_{h_0,I_{h_0}}) = (h_1 - h_0)\gamma'^{Y}_{H} + (h_1 - h_0)E(I_{h_0})\gamma'^{Y}_{I,H}, \tag{13}$$

where the average counterfactual income $E(I_{h_0})$, using the right hand side of the equation for I in (1), can be expressed as

$$E(I_{h_0}) = \gamma'^{I}_{0} + h_0\gamma'^{I}_{H} + E(\mathbf{X})\boldsymbol{\beta}'^{I} + E(\mathbf{Z}^I)\boldsymbol{\rho}'^{I} + E(\mathbf{W}^I)\boldsymbol{\eta}'^{I}. \tag{14}$$

As in our estimation we use de-meaned explanatory variables and h_0 equal to the average hours worked, $E(I_{h_0}) = 0$ and the second right hand side addend in equation (13) cancels out and the direct effect is simply given by the product of the coefficient on hours in the test score equation γ'^{Y}_{H} and the change in mothers' hours worked $(h_1 - h_0)$.

Adding the interaction term to the test score equation will lead to a change in the mediator effect into

$$E(Y_{h_0,I_{h_1}} - Y_{h_0,I_{h_0}}) = (h_1 - h_0)\gamma'^{I}_{H}[\gamma'^{Y}_{I} + h_0\gamma'^{Y}_{I,H}]. \tag{15}$$

which has the additional term $\gamma'^{Y}_{I,H}$ corresponding to the interaction between I and H that now contributes to the productivity of household income.

Finally, replacing Y and I in the interaction effect (12) with the right hand side of the corresponding equations in (1), we can show that

$$E(Y_{h_1,I_{h_1}} - Y_{h_1,I_{h_0}} - Y_{h_0,I_{h_1}} + Y_{h_0,I_{h_0}}) = (h_1 - h_0)^2\gamma'^{I}_{H}\gamma'^{Y}_{I,H}. \tag{16}$$

We use formulas (13)-(16), to compute the decomposition analysis of the total effect of an increase of 10 hours in mother's hours into the direct, mediator and interaction effects and results are reported in Table A.14.

Reported in Table A.14 are the regression estimates (panel a) and decomposition results (panel b) when including the additional interaction term. The contribution of the interaction effect to the total effect is zero. Consequently, including the additional interaction term does not lead to a statistically significant change in the direct, indirect or total effects of mothers' labour hours in pre-school years on outcomes.

A.3 Appendix Figures

Figure A.1: Density of mothers' hours worked 1-5 years after birth

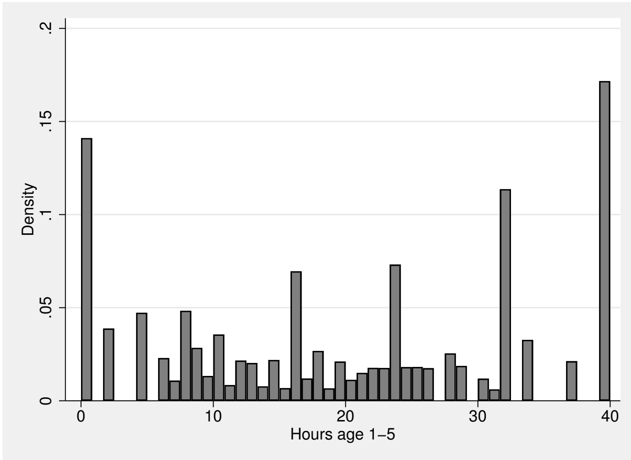
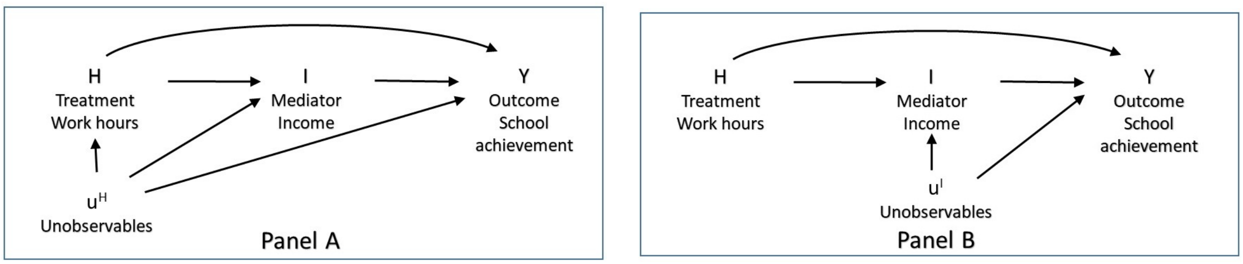
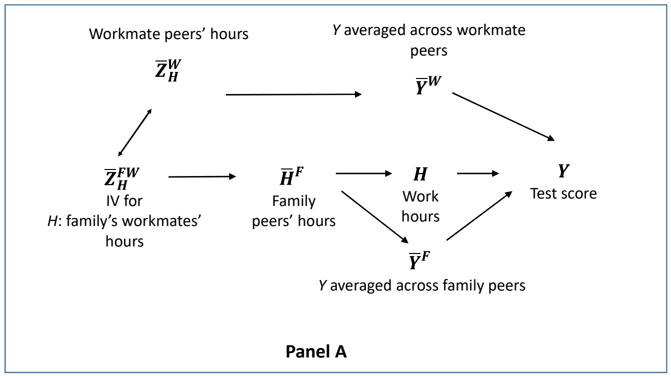


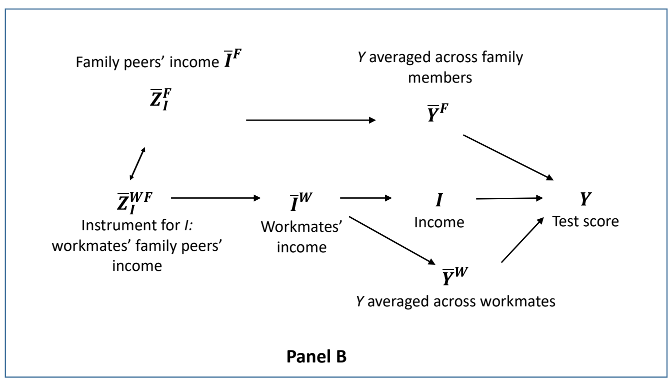
Figure A.2: Conceptual diagrams of the mediation pathways



Notes: In each of the two panels the curved arrow represents the direct pathway from H to Y, and the two horizontal arrows represent the mediated pathway from H to Y through I. Panel A shows the confounding effect of the unobservables u^H that causes endogeneity of the treatment, H, in the equation for Y and for I. Panel B shows the confounding effect of u^I that causes endogeneity of the mediator I in the equation for Y.

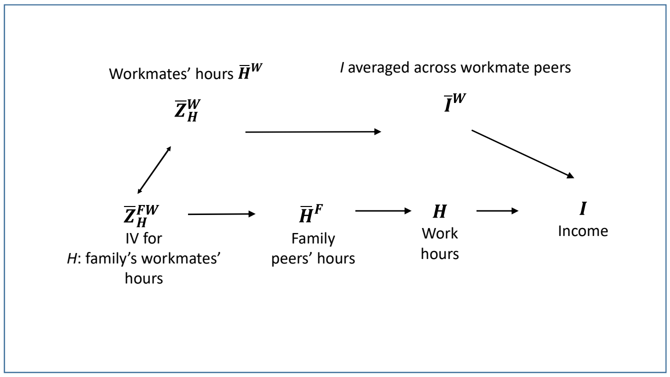
Figure A.3: Graphical representation of threats to validity of IV strategy





Notes: Our model estimates the effect of mothers hours (H) and household income (I) on child outcomes (Y). The IV for hours (income) $\overline{Z}_H^{FW}$ ($\overline{Z}_I^{WF}$) is the hours (father's income) 1 year after birth of the indirect peers which influence hours (income) through the direct family (workmate) peer mean hours $\overline{H}^F$ (income $\overline{I}^W$). Our model is estimated as deviations from the workplace mean, therefore it controls for $\overline{Z}_H^W$ and $\overline{Y}^W$ in Panel A and $\overline{Y}^W$ in Panel B. We control additionally for $\overline{Y}^F$ to rule out the link in Panel A between family peer hours and Y through the family mean test scores. A family or workmate is defined to be a peer if they gave birth to their first child before the focal household. Consequently, $\overline{Z}_H^{FW}$ is predetermined to $\overline{H}^F$, $\overline{H}^F$ is predetermined to H and $\overline{Z}_H^W$, $\overline{Y}^W$, $\overline{Y}^F$ and H are predetermined to Y in Panel A. Similarly, $\overline{Z}_I^{WF}$ is predetermined to $\overline{I}^W$, $\overline{I}^W$ is predetermined to I and $\overline{Z}_I^F$, $\overline{Y}^F$, $\overline{Y}^W$ and I are predetermined to Y in Panel B.

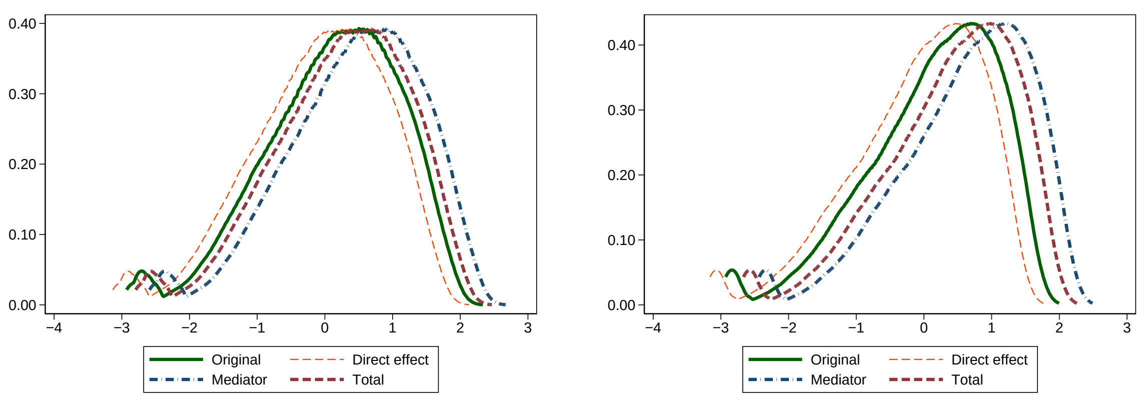
Figure A.4: Graphic representation of confounding effects in the income equation



Notes: One equation in our model estimates the effect of mothers hours (H) on household income (I). The IV for hours (income) $\overline{Z}_H^{FW}$ is the hours 1 year after birth of the indirect peers which influence hours through the direct family peer mean hours $\overline{H}^F$. This instrument may directly affect I through selection into workplace based on unobservables. We control for this in our model by estimating the model as deviations from the workplace mean, controlling for $\overline{Z}_H^W$ and $\overline{I}^W$.

Figure A.5: Decomposition analysis: Full sample (benchmark) results

a) Test score density at age 11 b) Test score density at age 15



Notes: The figure illustrates how the distribution of child test scores is shifted by the direct, mediator and total effects. While "Original" denotes the density of the test score with no changes; the "Direct effect", "Mediator" and "Total" depict the densities shifted by the direct, mediator and total effect, respectively, of an increase in the mother's hours by 10.

A.4 Appendix Tables

Table A.1: Control and instrumental variables in model (1)

Control variables and IVs	Test scores Y	Income I	Work hours H
Endogenous variables			
Mean household income age 1-5	Yes		
Mean hours worked age 1-5	Yes	Yes	
Mother variables			
Years of schooling	Yes	Yes	Yes
Age at 1st child	Yes	Yes	Yes
Work before 1st child	Yes	Yes	Yes
Father variables			
Income before 1st child	Yes	Yes	Yes
Participation before 1st child	Yes	Yes	Yes
Years of schooling	Yes	Yes	Yes
Child variables			
Month of birth dummies	Yes	Yes	Yes
Year of birth dummies	Yes	Yes	Yes
Birth weight level and squared	Yes	Yes	Yes
Mean family test score	Yes	Yes	Yes
Instrumental variables			
Father earnings of workmates' family		Yes	
Hours of family's workmates			Yes

Notes: Our model also considers two extra variables which are the 'leave one out' average of test scores across co-workers, $\overline{Y}^W$, and the 'leave one out' average of test scores across family members, $\overline{Y}^F$. The first of these two variables is controlled for by expressing each equation in model (1) in deviation from the workers' 'leave one out' mean, whereas the second variable is directly included.

Table A.2: Supplementary descriptive statistics

Variable	Mean	SD	N
Test score age 11	45.913	16.016	64,762
Income age 6-11	712180.5	431,911	64760
Hours age 6-11	23.649	14.272	64762
Number of children up to age 11	2.117	1	64512
Divorce up to age 11	0.040	0.162	64,656
Non labour income age 1-5	21217.06	71,179	64677
Debt age 1-5	0.583	0.493	63787
House value age 6	2233497	2159694	64151
Childcare spaces per child age 1-6	0.522	0	60789

Table A.3: Statistics reported by presence of indirect peers of family's workmates.

	(1)		(2)		(3)
	No indirect peer		Indirect peer		
	Mean	sd	Mean	sd	Difference
Age 15 test score	64.414	20.979	64.412	20.983	-0.001
Household income 1-5	5.103	2.336	5.325	2.300	0.222***
Hours 1-5	19.463	13.325	20.317	13.299	0.854***
Gender male	0.507	0.500	0.516	0.500	0.009*
Mothers' degree	0.389	0.488	0.393	0.488	0.003
Mothers' years of schooling	13.334	2.414	13.339	2.363	0.005
Observations	28265		36497		64762

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Statistics reported by presence of indirect peers of workmates' family.

	(1)		(2)		(3)
	No indirect peer		Indirect peer		
	Mean	sd	Mean	sd	Difference
Age 15 test score	64.436	20.679	64.410	21.018	-0.026
Household income 1-5	5.168	2.022	5.204	2.358	0.035
Hours 1-5	20.914	13.101	19.702	13.341	-1.211***
Gender male	0.501	0.500	0.513	0.500	0.012
Mothers' degree education	0.389	0.488	0.391	0.488	0.002
Mothers' years of schooling	13.308	2.314	13.340	2.401	0.031
Observations	7132		57630		64762

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Income equation: Analysing compliers.

	Total sample	Teen dad	Dad in 20s	Dad in 30s	Dad in 40s
	(1)	(2)	(3)	(4)	(5)
IV income	0.061***	0.091**	0.044***	0.061***	0.128***
	(0.005)	(0.038)	(0.007)	(0.009)	(0.031)
Observations	64,762	832	32,407	28,006	3,517

	Dad compulsory education	Dad degree education	Girls	Boys
	(6)	(7)	(8)	(9)
	0.054***	0.088***	0.068***	0.052***
IV income	(0.011)	(0.013)	(0.008)	(0.007)
Observations	62,476	15,195	17,049	31,655

	Dad income <=p(25)	p(25)-p(50)	p(50)-p(75)	p(75)+
	(10)	(11)	(12)	(13)
IV income	0.052***	0.037***	0.023***	0.063***
	(0.010)	(0.008)	(0.009)	(0.013)
Observations	16,226	16,206	16,145	16,185

Notes: The coefficient and standard error on the instrumental variable in the equation for income I is reported, across subgroups of parent characteristics. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. p(25), p(50), p(75) denote the 25th, 50th and 75th percentiles.

Table A.6: Hours equation: Analysing compliers.

	Total sample	Mum compulsory education	Mum high school education	Mud degree + education	
	(1)	(2)	(3)	(4)	
IV hours	0.024***	0.026***	0.021***	0.025***	
	(0.004)	(0.009)	(0.007)	(0.006)	
Observations	64,762	12,314	24,527	25,314	
	Teen mum	Mum 20s	Mum 30 +	Girls	Boys
	(5)	(6)	(7)	(8)	(9)
IV hours	-0.011	0.017***	0.029***	0.024***	0.023***
	(0.015)	(0.005)	(0.008)	(0.006)	(0.006)
Observations	4,146	44,541	15,892	31,655	33,107
	No work pregnancy	Work pregnancy	Dad compulsory education	Dad high school education	Dad Degree education
	(10)	(11)	(12)	(13)	(14)
IV hours	0.027***	0.022***	0.031***	0.019***	0.022***
	(0.008)	(0.005)	(0.009)	(0.006)	(0.007)
Observations	15,439	49,323	15,195	29,506	17,049

Notes: The coefficient and standard error on the instrumental variable in the equation for hours H is reported, across subgroups of parent characteristics. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Full benchmark regression estimates

Equation	(1) Test score 15	(3) Income	(4) Hours
Mothers' hours	-0.023	0.258***	
	(0.018)	(0.045)	
Household income	0.192***		
	(0.047)		
Instrumental variables			
Father earnings of neighbour's workmates		0.060***	
		(0.005)	
Hours 1 year after birth of family's neighbour peers			0.023***
			(0.004)
Covariates			
Male	-0.099***	0.044*	-0.080
	(0.008)	(0.025)	(0.095)
Child birth weight	0.042***	0.002	0.093**
	(0.004)	(0.013)	(0.047)
Child birth weight squared	-0.004*	0.014**	-0.058**
	(0.002)	(0.007)	(0.025)
Mother years schooling	0.078***	-0.087**	0.750***
	(0.011)	(0.034)	(0.023)
Mother work before 1st child	0.173	-1.510***	8.625***
	(0.137)	(0.388)	(0.116)
Mother age 1st child	0.008	-0.008	0.459***
	(0.006)	(0.021)	(0.011)
Father income before 1st child	-0.035***	0.246***	0.016
	(0.012)	(0.005)	(0.019)
Father participation before 1st child	-0.030	0.198	2.912***
	(0.045)	(0.147)	(0.251)
Father years schooling	0.040***	0.141***	0.129***
	(0.007)	(0.008)	(0.023)
Individual Ivs			
Mean child test scores of family peers	0.071***	-0.026	0.385***
	(0.010)	(0.034)	(0.107)
Observations	64,762	64,762	64,762

Notes: The regressions control additionally for child year and month of birth dummies. Standard errors computed using the 3SLS estimation in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Placebo: reduced form effect of IVs on birth outcomes

	Child birth weight	Height	Transfer to children ward	Congenital malformation	Severe deformity
IV for hours	0.0004	-0.001	-0.0001	0.00004	0.00003
	(0.0003)	(0.001)	(0.0001)	(0.00006)	(0.00005)
IV for income	0.005	0.012	0.0005	0.001	0.0002
	(0.003)	(0.008)	(0.002)	(0.001)	(0.0005)
Observations	64,762	61,815	33,675	64,762	64,762

Table A.9: Estimation results: Test scores at age 11

	(1) Benchmark	(2) Girls	(3) Boys	(4) No degree	(5) Degree	(6) No debt 1-5	(7) Debt 1-5	(8) School quality	(9) Low childcare access	(10) High childcare access
a) Estimation results										
Mothers' hours	-0.020	0.003	-0.044*	0.003	-0.046*	-0.060**	0.013	0.109***	0.142	0.005
	(0.018)	(0.024)	(0.026)	(0.025)	(0.027)	(0.024)	(0.025)	(0.033)	(0.208)	(0.015)
Household income	0.132***	0.102	0.162**	0.148*	0.116**	0.168**	0.084	0.284***	-0.529	0.051
	(0.048)	(0.069)	(0.065)	(0.081)	(0.055)	(0.072)	(0.072)	(0.072)	(0.636)	(0.041)
b) Decomposition results for 10 hours increase										
Direct	-0.204	0.027	-0.445*	0.030	-0.462*	-0.604**	0.125	1.095***	1.419	0.050
	(0.182)	(0.236)	(0.259)	(0.252)	(0.268)	(0.237)	(0.254)	(0.333)	(2.085)	(0.146)
Mediator	0.338**	0.189	0.513**	0.308*	0.356*	0.443**	0.180	0.724***	-1.403	0.117
	(0.136)	(0.137)	(0.229)	(0.185)	(0.191)	(0.204)	(0.158)	(0.224)	(1.466)	(0.097)
Total	0.134	0.216	0.068	0.338*	-0.106	-0.162	0.305*	1.819***	0.016	0.167
	(0.129)	(0.177)	(0.190)	(0.194)	(0.181)	(0.217)	(0.176)	(0.357)	(0.934)	(0.152)
Observations	64,762	33,107	31,655	39,448	25,314	26,595	37,192	64,762	30,403	30,386

Notes: Column (1) benchmark 3SLS estimates; column (2) sample of girls; column (3) sample of boys; column (4) sample of mothers with no degree; column (5) sample of mothers with a degree column (6) sample with no debt (total assets - total debts) ages 1-5; column (7) sample with some debt ages 1-5; column(8) households in low childcare access municipalities (24-46 places available per 100 children aged 1-6, or bottom quartile) and column (9) households not in low childcare municipalities (places available per 100 children aged 1-6). Standard errors in parentheses in panel a) computed using the 3SLS estimation and in panel b) by delta method. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is school test scores at age 11 measured in standard deviations. Income is the yearly household net income in NOK100,000 at 2020 prices and hours is the mother's weekly work hours, both are averaged across years 1-5 after the first child birth. Panel a) coefficient estimates from equation 1 and panel b) decomposition analysis from equations 3, 13 and 15. All remaining controls are measured at or before the child birth (see Table 1). Data sources: Norwegian administrative data, first-born children born in 1997-2001.

Table A.10: Sensitivity: Nonlinearity in income and hours in the test score equations, quadratic specification

a) Regression estimates	
Mothers' hours	-0.021
	(0.017)
Household income	0.193***
	(0.045)
Hours squared	0.000003
	(0.00002)
Income squared	-0.0009
	(0.001)
b) Decomposition results for 10 hours increase	
Direct	-0.208
	(0.171)
Mediator	0.495***
	(0.126)
Total	0.287**
	(0.132)
Observations	64,762

Notes: Equations for test scores at age 15 estimation of the joint model (1) estimated using a control function approach. Standard errors computed using 3SLS in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The school test scores at age 11 and 15 are measured in standard deviations, income is the yearly income in NOK100,000 at 2020 constant prices and hours is the mother's weekly work hours. Both income and hours are averaged between 1 and 5 years after the first child birth and are de-meaned. All remaining controls are measured at or before the child birth (see Table 1). Data sources: Norwegian administrative data, first-born children born in 1997-2001.

Table A.11: Sensitivity: Nonlinearity in hours in the test score equations, splines in hours

a) Regression estimates	
Mothers' hours: Effect at the 1st quartile	-0.022
	(0.017)
Mothers' hours: Differential effect at 2nd quartile	0.006*
	(0.003)
Mothers' hours: Differential effect at 3rd quartile	-0.001
	(0.003)
Mothers' hours: Differential effect at 4th quartile	-0.004*
	(0.002)
Household income	0.189***
	(0.045)
b) Decomposition results for 10 hours increase	
Direct effect of hours at the 1st quartile	-0.221
	(0.171)
Direct effect of hours at the 2nd quartile	-0.164
	(0.173)
Direct effect of hours at the 3rd quartile	-0.175
	(0.172)
Direct effect of hours at the 4th quartile	-0.212
	(0.171)
Mediator effect of hours	0.487***
	(0.126)
Total effect of hours at the 1st quartile of hours	0.266**
	(0.132)
Total effect of hours at the 2nd quartile of hours	0.323**
	(0.134)
Total effect of hours at the 3rd quartile of hours	0.311**
	(0.133)
Total effect of hours at the 4th quartile of hours	0.274**
	(0.132)
Observations	64,762

Notes: Equations for test scores at age 15 estimation of the joint model (1) estimated using a control function approach. Standard errors computed using 3SLS in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The school test scores at age 15 are measured in standard deviations, income is the yearly income in NOK100,000 at 2020 constant prices and hours is the mother's weekly work hours. Both income and hours are averaged between 1 and 5 years after the first child birth and are de-meaned. All remaining controls are measured at or before the child birth (see Table 1). Data sources: Norwegian administrative data, first-born children born in 1997-2001.

Table A.12: Sensitivity: Nonlinearity in income in the test score equations, splines in income

Regression estimates	
Mothers' hours	-0.021
	(0.017)
Income effect of the 1st decile	0.166***
	(0.047)
Differential effect of income at the 2nd decile	0.007
	(0.026)
at the 3rd decile	0.089
	(0.056)
at the 4th decile	-0.102
	(0.068)
at the 5th decile	0.125
	(0.474)
at the 6th decile	0.062
	(0.484)
at the 7th decile	-0.181
	(0.117)
at the 8th decile	0.035
	(0.055)
at the 9th decile	-0.000
	(0.036)
at the 10th decile	-0.023
	(0.019)
Observations	64,762

Notes: Equations for test scores at age 15 estimation of the joint model (1) estimated using a control function approach. Standard errors computed using 3SLS in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The school test scores at 15 are measured in standard deviations, income is the yearly income in NOK100,000 at 2020 constant prices and hours is the mother's weekly work hours. Both income and hours are averaged between 1 and 5 years after the first child birth and are de-meaned. All remaining controls are measured at or before the child birth (see Table 1). Data sources: Norwegian administrative data, first-born children born in 1997-2001.

Table A.13: Sensitivity: Decomposition, splines in income

Decomposition results for 10 hours increase	
Direct effect of hours	-0.207
	(0.171)
Mediator effect through income at the 1st decile	0.427***
	(0.126)
at the 2nd decile	0.444***
	(0.134)
at the 3rd decile	0.671***
	(0.171)
at the 4th decile	0.409***
	(0.158)
at the 5th decile	0.731
	(1.148)
at the 6th decile	0.890***
	(0.267)
at the 7th decile	0.423***
	(0.155)
at the 8th decile	0.512***
	(0.142)
at the 9th decile	0.511***
	(0.132)
at the 10th decile	0.453***
	(0.126)
Total effect of hours at the 1st income decile	0.220*
	(0.132)
at the 2nd income decile	0.237*
	(0.140)
at the 3rd income decile	0.464***
	(0.175)
at the 4th income decile	0.202
	(0.164)
at the 5th income decile	0.523
	(1.149)
at the 6th income decile	0.683**
	(0.269)
at the 7th income decile	0.216
	(0.160)
at the 8th income decile	0.305**
	(0.147)
at the 9th income decile	0.304**
	(0.138)
at the 10th income decile	0.246*
	(0.132)
Observations	64,762

Notes: Equations for test scores at age 15 estimation of the joint model (1) estimated using a control function approach. SEs computed using 3SLS in parentheses. Test scores measured in standard deviations, income yearly income in NOK100,000 at 2020 prices, hours is mother's weekly work hours. Income and hours are averaged between 1 and 5 years after the first child birth. Controls from Table 1.

Table A.14: Sensitivity: Allowing for interaction between hours and income in the test score equation

	(1)
a) Regression estimates	
Mothers' hours	-0.021
	(0.017)
Household income	0.192***
	(0.045)
Hours*Income	0.000
	(0.000)
b) Decomposition results for 10 hours increase	
Direct	-0.207
	(0.171)
Mediator	0.495***
	(0.126)
Interaction	0.000
	(0.000)
Total	0.287**
	(0.132)
Observations	63,022

Notes: Equations for test scores at age 15 estimation of the joint model (1) using a estimated control function approach. Standard errors computed using 3SLS in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The school test scores at age 11 and 15 are measured in standard deviations, income is the yearly income in NOK100,000 at 2020 constant prices and hours is the mother's weekly work hours. Both income and hours are averaged between 1 and 5 years after the first child birth and are de-meaned. All remaining controls are measured at or before the child birth (see Table 1). Data sources: Norwegian administrative data, first-born children born in 1997-2001.