School of Economics

Working Paper Series

Evaluating the impact of 20 hours free early childhood education on women's labour market participation

Isabelle Bouchard, Lydia Cheung and Gail Pacheco

2018/05

Evaluating the impact of 20 hours free early childhood education on women's labour market participation

Isabelle Bouchard * Lydia Cheung † Gail Pacheco ‡

July 2018

Abstract

New Zealand introduced a substantial childcare subsidy just over a decade ago, providing 20 hours free early childhood education (ECE) to all three and four year olds. We evaluate the impact of this policy shift on mothers' labour market participation. Using a triple difference strategy and population wide administrative data, we follow mothers' monthly wages from pre-pregnancy to six years post-childbirth. The estimated impact of the ECE reform is a drop in earnings for eligible women, by four to ten percent post-childbirth. Furthermore, most of the reduction occurs prior to the children reaching the age of eligibility. This suggests that the policy may have partially displaced private spending on ECE and that eligible mothers substituted this saving intertemporally.

JEL classification: C21; H40; J13; J22

Keywords: early childhood education; triple differences; mothers' wages and salaries; intertemporal substitution; administrative data

*HEC Montréal. isabelle.4.bouchard@gmail.com

†Auckland University of Technology. lydia.p.cheung@aut.ac.nz

‡Corresponding author. Auckland University of Technology. gail.pacheco@aut.ac.nz. Postal address: Gail Pacheco, New Zealand Work Research Institute, Private Bag 92006, Auckland 1142, New Zealand.

Disclaimer

The results in this paper are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics New Zealand.

The opinions, findings, recommendations, and conclusions expressed in this paper are those of the authors, not Statistics NZ.

Access to the anonymised data used in this study was provided by Statistics NZ under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this paper have been confidentialised to protect these groups from identification and to keep their data safe.

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from `www.stats.govt.nz`.

The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes.

Any person who has had access to the unit record data has certified that they have been shown, have read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

1 Introduction

Public spending on childcare policies is of great interest to policy makers on two fronts. First, childcare subsidies increase access to early childhood education (ECE) opportunities for a wider range of children, especially for those from poor socio-economic households. It is expected that ECE attendance will give the child a greater probability of success in school, which will result in positive future long-term outcomes. Second, it promotes greater female labour market participation. Childcare costs are one of the potential obstacles for women with young children to re-enter the labour market post-childbirth. Its impact on mothers' engagement with the labour market is the focus of this study. More specifically, we contribute to the empirical evidence on the causal impact of childcare policies (namely, subsidies) on mothers' labour market participation (via data on their monthly earnings), for the case study of New Zealand (NZ).

NZ is similar to a number of other developed countries, whereby mothers still experience a labour market wage disadvantage. Interestingly, women across the working age population (15 to 64 year olds) have a relatively high labour force participation rate, eighth highest in the OECD; but when the focus is narrowed to the dominant child-bearing ages (25 to 34 year olds), NZ's ranking falls to the twenty-sixth[1]. Furthermore, close to a third of working women are in part-time employment (Pacheco, Li, and Cochrane (2018)). One of the most cited reasons for women's limited labour market participation is the cost of childcare (Baker, Gruber, and Milligan (2008); Baum (2002); Hegewisch and Gornick (2011); Lefebvre and Merrigan (2008)).

In 2007, the Labour-led government in NZ introduced a policy of 20 hours (per week) free ECE for all three and four year olds in services led by a qualified and registered early childhood teacher. When announced in 2004, the policy was unanticipated by the early childhood community. It brought about a major increase in public funding for ECE (Bushouse (2008)). Within six months of implementation, 76 percent of ECE providers had joined the 20 hours free ECE program (May (2008)), and the number of children enrolled in ECE increased from 165,254 in 2006 to 176,993 in 2008 (Ministry of Education, 2013)[2]. Furthermore, public spending on ECE rose exponentially from $574 million in 2006 to $837 million by 2008, and up to $1.157 billion by 2010.

[1]This is based on 2006 data from the OECD: `https://data.oecd.org/emp/labour-force-participation-rate.htm`

[2]Of course we cannot attribute this increase solely to the ECE subsidy, as there are numerous other factors that could jointly affect ECE participation, such as population growth dynamics, for instance

To assess the impact of the ECE reform, we employ newly linked administrative data sources from the Integrated Data Infrastructure (IDI). This permits a population wide perspective, as well as the ability to link information across sources based on a unique identifier (`snz_uid`) assigned to each individual in the population. The IDI allows identification of all births in NZ via Department of Internal Affairs records, and we are able to link to tax data to follow patterns in mothers' monthly wages over the time period preceding birth, up to six years post-childbirth.

Our empirical approach exploits the temporal variation in childcare coverage induced by the difference in birth years. As the program is implemented simultaneously across the country, we do not have a concurrent comparison group, like untreated states or regions. Instead, we use the birth date of the child to define the treatment and control groups of mothers. A cohort of mothers who give birth earlier, whose children are not eligible, forms the control group; a cohort of mothers who give birth later, whose children are eligible, forms the treatment group. For contemporaneous comparison, we add a second control group—a matched sample of non-mothers. They provide a control for time-specific effects, such as the macroeconomic environment that may affect employment. Our method therefore incorporates triple differences, whereby we compare the earnings of eligible mothers with their own earnings before childbirth, with that of ineligible mothers, and with matched non-mothers.

Our regression estimates reveal a drop in earnings of eligible women, by four to ten percent of pre-pregnancy income, after they give birth. The largest decreases occur when the child is one to two years old. Our interpretation of this finding is that eligible mothers reduce participation in the labour market in an intertemporal fashion, in anticipation of savings in childcare expenditure when their child reaches ages three and four. Thus the ECE reform is likely to have displaced private spending by eligible mothers. This is most significant for middle-earning mothers. In contrast, mothers in the lowest and highest earning quintiles show different patterns of earning changes. Some mothers in the lowest earning quintile is observed to earn more in the years when their children are eligible. Some mothers in the highest earning quintile do not show significant changes in earnings in any year.

Our study makes a number of contributions to the extant literature. We add to the small body of causal evidence that focusses on childcare policies and women's labour market outcomes. Our study is also the first to look at the impact of the 20 hours ECE reform in NZ. Considering the magnitude of this reform, it presents a unique opportunity to understand the related outcomes. Finally, with the use of administrative data, we can evaluate the impact of the ECE reform on both the population as a whole, as well as for a range of disaggregate groups of interest.

The outline of the paper is as follows. Section 2 documents the empirical evidence from the key literature on childcare subsidies and the impact on mothers' labour market outcomes. Section 3 outlines the ECE reform implemented in 2007. Section 4 describes the administrative data sources and methodology we use. Section 5 presents our results. Section 6 compares the magnitude of our estimates with those from other case studies, and mentions potential caveats. Section 7 concludes.

2 Literature Review

In this study we focus on the impact of childcare policies on mothers' labour market participation. The trade-off between costs associated with childcare and the wage earned at work is an important element of the mothers' decision to re-enter the workforce post-childbirth, as well as the extent of participation. Much of the early literature on childcare policies have focussed exclusively on the individual response to cost schedules in a non-experimental setting (Anderson, Binder, and Krause (2002); Anderson, Binder, and Krause (2003); Blau and Currie (2006); Lundin, Mörk, and Öckert (2008); Wrohlich (2004)). The price elasticity estimates uncovered in this space are wide ranging, extending from -0.02 (Wrohlich (2004)) to -0.92 (Kimmel (1998)).

More recently there is a growing body of empirical studies that apply quasi-experimental identification strategies by exploiting exogenous policy changes. For example, Hardoy and Schøne (2015) apply a triple difference approach to evaluate the impact of cheaper childcare in Norway after the implementation of the "Child Care Centre Agreement" in 2003. They compare the labour supply of eligible mothers (before and after childbirth) with the labour supply of ineligible mothers, along with the labour supply of mothers with older children who were also ineligible. They find that the childcare reform result in the employment rate of women with children rising four to five percent.

Another example of reform to childcare subsidisation occurs in Québec, Canada. The Early Childhood Education and Care (ECEC) programme is phased in from 1997 to 2000 with two aims in mind: increasing mothers' labour force participation, and enhancing child development. This involves publicly regulated childcare facilities to start offering spaces at just \$5 per day per child. To evaluate the impact of this reform, Haeck, Lefebvre, and Merrigan (2015) applies a difference-in-differences approach. The treatment group consists of eligible children before and after the reform; the control group consists of comparable children in the rest of Canada. They find that the reform generally have a positive, large and significant effect on the labour force participation of mothers with children one to four years old.

There is contrasting evidence from the Netherlands, suggesting only modest increases in the maternal employment rate following the introduction of the "Law on Childcare" in 2015. Bettendorf, Jongen, and Muller (2015) analyse the impact of this reform, which resulted in substantial increases in childcare subsidies, with a difference-in-differences approach. Their treatment group includes parents affected by the change in childcare costs (i.e. with youngest child up to twelve years old), while the control group are parents not eligible for the subsidies (i.e. with youngest child twelve to eighteen years old). The effect on the mothers' employment rate is just over two percent, and this result is likely confounded by the implementation of targeted earned income tax credits for parents over the same time period.

There is also evidence pointing to no impact on female labour supply following a childcare price shock. For instance, Lundin, Mörk, and Öckert (2008) evaluate the introduction of a cap to childcare prices in Sweden that occurred in 2002 and 2003. This price cap is optional for municipalities; but if implemented, it results in extra state funds to cover the reduced revenue. The authors find that, despite a considerable drop in childcare prices in a number of instances, their difference-in-differences estimates regarding impact on female labour supply are close to zero.

Another example of a study finding minimal impact of subsidised childcare on maternal employment is Havnes and Mogstad (2011). They focus on the expansion of subsidised childcare across Norway in 1975. The authors argue that their results show an instance of subsidised childcare crowding out informal childcare arrangements.

Given the mixed level of findings in prior literature, as well as the paucity of studies that analyse the causal relationship between childcare subsidies and maternal labour market outcomes, our study provides a timely addition to the knowledge in this space. Furthermore, the substantial extent of the childcare reform in NZ provides enhanced motivation for our empirical analysis.

3 Policy Background

The ECE sector in NZ has undergone a significant transformation over the last century, from both policy shifts and changes in the social, economic and cultural landscape. In the first half of the twentieth century, women were encouraged to assume their responsibilities as mothers "at home" (May (2002)). This perspective was reinforced by social policies and public perceptions. During the 1960s and 1970s the perceived role of women began to slowly change as more women entered the workforce. In 1986 there was a major overhaul of the childcare sector, when the government integrated all ECE services under the Ministry of Education (Meade and Podmore (2002)). This step provided the necessary regulatory framework for the ECE sector. In 1988 the entire education system was reviewed and the "Before Five" report was launched. Based on the recommendations within this report, the government introduced bulk funding for ECE services of $2.25 per hour subsidy for up to 30 hours per week for children over two years old.

ECE services are available for children from birth to school entry age, which is on or near their fifth birthday[3] (Meade and Podmore (2002)). It is important to note that there is no public provision of ECE services in NZ, rather it is organised by either private or community-based providers. Services are grouped into two types of models depending on extent of parental involvement: teacher-led services and parent-led services. The former has more regulations, such as the requirement of at least 50 percent of the staff to be registered ECE teachers (Ministry of Education, 2016). With parent-led services, the role of educators or teachers is provided directly by the parents. They operate on various schedules depending on level of interest and commitment of parental volunteers.

The ECE reform is announced in 2004 by the Minister of Finance in his budget speech to parliament. The new program is a shift from a per-hour subsidy approach to 20 hours free, to be implemented in 2007 for all three and four year olds in teacher-led services. This marks a major increase in public funding for ECE and is therefore deemed a substantial policy shift, particularly towards increasing accessibility for all families. When implemented in July 2007, the policy covers up to 6 hours free a day, and up to 20 hours free a week (May (2002)). The funding is available if the child attends an ECE participating in the programme. Thus, it is not a universal entitlement in the sense that providers have the choice of whether to opt into the

[3]Children are eligible to start at school on the day they turn five, and must start before they turn six. Schooling is compulsory from ages 6 to 16.

programme. Within six months of implementation, 76 percent of ECE providers have chosen to join the program (May (2002)), which reimbursed compulsory fees directly to the provider.

Unfortunately, ECE data is not in the IDI prior to 2008. We therefore rely on the Annual ECE Census Summary Report from the Ministry of Education in 2014[4] to illustrate ECE consumption patterns around the relevant time frame. The report provides trends from 2005 to 2014 on a binary indicator of ECE enrolment, and a measurement of intensity with the average number of hours of ECE attendance per week. Among three year olds, in 2006, almost 90% are enrolled in ECE, for an average of about 15 hours per week. Among four year olds, the respective figures are 97% and just under 18 hours per week. Thus, usage of ECE is already quite high in the age group targeted by the ECE reform. The Census Summary also illustrates a general upward trend in the average hours of ECE consumption across all ages under five, between 2005 and 2014, with sharper increases evident around 2007 and 2008 for three and four year olds.

In terms of the expected impact of the ECE policy reform on mothers' labour market outcomes, classic microeconomic theory tells us that it could be either positive or negative. For all eligible mothers who would have purchased ECE services regardless of government subsidy, the policy reduces their out-of-pocket expenses. This represents an increase in real income, if we hold the prices of other consumable items fixed. If we model mothers' decisions on the number of working hours using indifference curves between consumption and leisure (i.e. non-work hours, which can be dedicated to child rearing at home), the policy shifts the optimal choice to one at a higher budget line and indifference curve. The number of hours worked at this new optimum could be more than, equal to, or less than the old optimum, depending on the shape of the indifference curves. If mothers work less in response to an increase in real income, leisure is a normal good; if mothers work more in response instead, leisure is an inferior good. The latter case is more likely if mothers are consuming more hours of ECE services as a result of the policy than they would have otherwise, and thus are able to use these additional child-free hours for work. Still, this classic static theory does not tell us about the timing of the mothers' responses. It is the aim of our empirical study to discover whether non-work hours for new mothers are a normal or inferior good, for different subsets of mothers, and whether their responses to the policy coincide in time with their benefits from the policy.

[4] `http://www.educationcounts.govt.nz/publications/series/annual-early-childhood-education-census`, figures 1.9 and 1.23.

4 Data and Method

Statistics New Zealand's Integrated Data Infrastructure (IDI) is a large research database containing micro-data on individuals and households, encompassing the NZ population. The IDI includes administrative data from a range of government and non-government agencies, and official surveys. All data is confidentialised and individual-level information can be linked across datasets and time (by the researcher) via an individual's unique identifier (`snz_uid`). Within the IDI we are able to identify groups of mothers who are eligible for the policy and groups that are not, based on the timing of their childbirths, and trace any changes in their labour market outcomes.

As the ECE policy is available to all mothers nation-wide upon implementation, there is no contemporaneous control group of mothers within the country to be compared against mothers that benefitted from the policy shift. We therefore use a matched sample of non-mothers as an additional control group, for contemporaneous comparisons. This triple difference approach takes into account difference along three dimensions: (1) between eligible and non-eligible mothers; (2) between incomes before and after childbirth, for the same mothers; and (3) between comparable mothers and non-mothers within the same time period.

We estimate the following model using individual-level data:

$$\begin{aligned} Y_{it} = {} & \alpha_0 + \alpha_1 Eligible_i + \alpha_2 Post_t + \alpha_3 Mother_i \\ & + \alpha_4 Eligible_i * Post_t + \alpha_5 Eligible_i * Mother_i + \alpha_6 Post_t * Mother_i \\ & + \alpha_7 Eligible_i * Post_t * Mother_i + X_{it}'\delta + \tau_t + \varepsilon_{it} \end{aligned}$$

where Y_{it} is the outcome variable of interest, which is women's monthly earnings. $Eligible_i$, $Post_t$, and $Mother_i$ are indicator variables. $Eligible_i = 1$ for all observations where the individual i belongs to the treatment group, $Post_t = 1$ for all observations where the time period t occurs after childbirth[5], and $Mother_i = 1$ if the individual is a mother. X_{it} is a vector of control variables encompassing age, ethnicity, and education, and τ_t are a set of time fixed effects. The parameter of interest is the coefficient of $Eligible_i * Post_t * Mother_i$, which gives the DDD estimator.

[5] For non-mothers, $Post_t = 1$ for all observations sharing the same calendar time as the post-childbirth period of the matched non-mother. We will elaborate on how each mother is matched to a non-mother below.

Our starting point in the IDI is the birth registration dataset from the Department of Internal Affairs (DIA), which contains all births and adoptions. For all births within the periods of our interest, we retrieve the list of mothers' identifiers. We then retrieve other relevant demographic information about these mothers, such as their age, educational attainment, and ethnicity[6], from the Census 2013 data. We then merge in their monthly earnings from the Inland Revenue (IR) dataset. In particular, we use the wage and salary income from the Employer Monthly Schedule (EMS) filed by employers, excluding paid parental leave and various sources of governmental non-employment income, such as unemployment benefits, pension, welfare, or student allowances. Employment earnings from different employers are aggregated to a single observation for each individual-month observation. We focus on wages and salaries because they are the most common source of income, and likely most impacted by childbirth (unlike capital income, for example). One shortcoming of administrative data is the lack of information on the number of hours worked: we cannot distinguish between changes in hourly wages from changes in the number of hours worked.

We construct our study population as follows. We start with the entire population of mothers who give birth between July 2004 and June 2006 for the treatment group, and those between July 2000 and June 2002 for the control group. For these two groups of mothers, we identify *all* the children they have ever given birth to, not limited to those in our treatment and control windows, using the DIA data. From this subset of 198 thousand mothers, we eliminate all mothers who give birth to a stillborn, as well as those with missing demographic variables. We also restrict our analysis to mothers with one or two children, which represent the majority of mothers in NZ. In addition, we exclude mothers who give birth outside the age range of 20-55 years old, to avoid extreme observations. Furthermore, since we are going to compare earnings before and after childbirth, we only include mothers with income observations both before and after childbirth. In particular, in order to match each mother to a non-mother with the most similar income time series before pregnancy, we restrict our attention to mothers who work consecutively from the 21$^{\text{st}}$ month before birth to the 12$^{\text{th}}$ month before birth, inclusive.[7] All aggregated monthly

[6]The ethnicity variable we use is the first ethnicity to which an individual identifies herself. The NZ Census accommodates multiple ethnicities.

[7]When a person is not observed to have income in the IR data in a particular month, the researcher cannot tell whether she is unemployed or out of the labour force. We make the assumption that individuals with periods of zero income of at most one month in length are unemployed, rather than being out of the labour force. This

earnings are normalised to constant June 2006 dollars, using the quarterly seasonally adjusted CPI series provided by Statistics New Zealand. Finally, we eliminate all mothers with extreme monthly income observations exceeding $15,000. This leaves us 45 thousand mothers.

4.1 Summary statistics

For the ease of all future references, we name the relevant groups of mothers as follows, summarised in Table 1. Groups 1N and 1E are mothers of one child. "N" denotes "non-eligible", while "E" denotes "eligible". Mothers of group 1N give birth between July 2000 and June 2002. This cohort of children are all older than four years old when the policy starts, thus they are not eligible for the policy, and serve as the control group. Mothers of group 1E give birth between July 2004 and June 2006. This cohort of children are all three or four years old when the policy starts, thus they are eligible and serve as the treatment group. Groups 2NN, 2NE, 2EE are mothers of two children. The first letter denotes the eligibility of the first child; the second letter denotes the eligibility of the second child. Group 2NN denotes mothers whose second (younger) child is born between July 2000 and June 2002, so that none of the two children are eligible for the policy. Group 2NE denotes mothers whose first (older) child is born between July 2000 and June 2002, and second (younger) child is born between July 2004 and June 2006, so that the first child is not eligible but the second child is. Lastly, group 2EE denotes mothers whose the first (older) child is born between July 2004 and June 2006, so that both children are eligible. In summary, group 1N is an earlier cohort of mothers than group 1E. Similarly, group 2NN is an earlier cohort of mothers than group 2NE, who is an earlier cohort than group 2EE.

Table 1 also presents demographic summary statistics on each group of mothers in our study. Groups 1N and 1E are very close in mean age at childbirth. Among these two groups, group 1E has a higher percentage of mothers with university or post-graduate degrees, and a higher percentage of Asian and Pacific mothers. Mothers in groups 2NN, 2NE, 2EE are fairly similar in their mean age at first and second childbirths; both sets of means fall within a year of each other. Groups 2NN and 2NE are similar in educational attainment, while group 2EE has 6 percent more university graduates and 0.7 percent more with post-graduate degrees. In terms

is a conservative decision rule to make sure that (almost) none of the mothers in our sample is breaking away from the labour force, at the cost of missing some mothers who are unemployed longer but still actively searching. Thus we keep mothers with income gaps of at most one month long between the 21^{st} to the 12^{th} month before birth.

Table 1: Groups of mothers: definitions and summary statistics

Group	1N	1E	2NN	2NE	2EE
First child eligible	×	✓	×	×	✓
Second child eligible	–	–	×	✓	✓
Age at first childbirth	32.63	32.72	30.59	29.78	30.16
	(5.33)	(5.81)	(4.25)	(4.43)	(4.69)
Age at second childbirth	–	–	32.38	33.40	33.16
	–	–	(4.28)	(4.29)	(4.52)
Edu: below HS	.109	.089	.068	.074	.063
Edu: HS	.640	.606	.654	.634	.578
Edu: University	.211	.259	.247	.257	.317
Edu: Post-Grad	.041	.046	.030	.034	.041
Ethnicity: European	.823	.758	.905	.891	.862
Ethnicity: Asian	.075	.133	.038	.050	.077
Ethnicity: Maori	.057	.057	.034	.035	.037
Ethnicity: Pacific Peoples	.044	.052	.023	.024	.024
Unique individuals	7,218	6,120	3,405	12,300	15,873

Notes: N = non-eligible; E = eligible. Means are provided for demographic variables and standard deviations are in parenthesis. Source of data = IDI. Authors' compilation.

of ethnicities, all three groups have similar percentages of Maori and Pacific mothers. However, they have an increasing percentage of Asian mothers and a corresponding, decreasing percentage of European mothers over time.

As evident in Table 1, our treatment and control groups have different demographic profiles. Likely factors at play include the changing ethnicity composition in each population cohort, which is partly influenced by immigration trends; the increase in tertiary education enrollment; and any changes in family planning decisions and timing in having children by these different groups. Their labour market decisions prior to pregnancy, which can be influenced by the broader macroeconomic conditions, would also play a role in contributing to the different group compositions, since we select only those with (more or less) continuous earning histories. In terms of policy effect identification, the difference in demographic composition would be confounding the policy effect if the *only* dimension of comparison was that between the treatment and control groups of 1N and 1E, for example, and if each individual were observed only once (in cross-section data, as opposed to panel data). This is not the case in our study. The three dimensions of comparisons in our triple difference estimate allows us to difference out any constant group fixed effects. In particular, educational attainment level and ethnicity are both recorded in our study as constants within each individual across time. Both the treatment and control groups are compared with its own historic earnings, and earnings of matched non-mothers. Thus, two out of three dimensions of our comparisons are done between groups with fixed education and ethnicity compositions. In addition, we include education and ethnicity indicator variables in all our regressions as control variables.

4.2 Matching non-mothers

Besides using ineligible mothers as control groups, we also identify women with no children (henceforth "non-mothers") as additional control groups. To identify these non-mothers in the IDI, we first use the Census 2013 data, which identifies all women in NZ, then merge in the DIA data, which allows us to exclude all women who had ever given birth or adopted. For this group of non-mothers, we merge in demographic information and IR income data in the same manner as described above for mothers, following similar steps in the construction of the study population.

Next, each mother in our study population is matched to the most comparable non-mother in the following way. For each mother, we first identify the list of non-mothers with identical birth year, birth month, ethnicity, and educational attainment level. The advantage of using administrative data of the entire country's population, as opposed to a much smaller survey dataset, truly shines in this step, as virtually every mother in our study groups has a non-empty list of matches. Each non-mother is then assigned a $t = 0$ (for groups 1N and 1E) or $t_1 = 0$ (for groups 2NN, 2NE, 2EE) to be the same calendar month as the matched mother's first (or only) childbirth. We then extract ten months of pre-childbirth earnings, from t (or t_1) $= -21$ to t (or t_1) $= -12$ inclusive for each mother and non-mother. We compute the total squared difference in earnings for each potential pair. Finally, each mother is matched to the non-mother with the smallest squared difference in earnings among the list, and the matching process is done with replacement.

Matched non-mothers for each group of mothers are named in accordance to the acronyms used for mothers. For example, non-mothers matched to mothers in group 1N are simply called "1N non-mothers." Mothers in group 1N all have $Eligible_i = 0$, so do all "1N non-mothers".

Table 2 shows mean earnings by groups of mothers and matched non-mothers, and by the number of months relative to childbirth(s)[8]. For groups 1N and 1E, let $t = 0$ denote the birth

[8] The mean earnings summary includes instances of zero monthly earnings. We cannot distinguish whether these instances are due to unpaid leave, unemployment, or women exiting the labour force.

month of the only child. For groups 2NN, 2NE, 2EE, let $t_1 = 0$ be the birth month of the first child, and $t_2 = 0$ be the birth month of the second child. Thus, by construction, $t_1 > t_2$ in any given calendar month. The table reports mean monthly earnings for the two years before (first) childbirth; between the two childbirths (for mothers of two children); the twelve months after (the last) childbirth; and the five years thereafter.

Table 2: Mean monthly earnings ($)

group		$-24 \le t < 0$	$0 \le t < 12$	$12 \le t < 72$	unique individuals
1N	mothers	2479.26	940.88	1970.78	7,218
1N	non-mothers	2478.83	2560.58	2803.57	6,810
1E	mothers	2850.03	1043.14	2246.67	6,120
1E	non-mothers	2827.84	2995.62	3193.26	5,793

group		$-24 \le t_1 < 0$	$0 \le t_1$ and $t_2 < 0$	$0 \le t_2 < 12$	$12 \le t_2 < 72$	unique individuals
2NN	mothers	3024.69	1053.67	630.64	1463.50	3,405
2NN	non-mothers	3009.42	3046.60	3067.89	3339.63	3,258
2NE	mothers	2916.45	1529.02	784.01	1658.3	12,300
2NE	non-mothers	2845.72	2888.29	3234.28	3508.07	10,509
2EE	mothers	3014.80	1569.26	789.75	1796.09	15,873
2EE	non-mothers	2924.65	3089.44	3359.38	3545.04	13,050

Notes: All earnings are adjusted by CPI to constant June 2006 dollars. N = non-eligible; E = eligible. Source of data = IDI. Authors' compilation.

Table 2 shows that mothers' and matched non-mothers' mean earnings are extremely similar pre-motherhood. In fact, the difference in means between all paired groups are under NZ$100, which is under five percent for all means in the first column. This is evidence that our matching criteria has achieved its goal in providing a comparable baseline for assessing the effect of motherhood on earnings. The more obvious difference across groups, say between groups 1N and 1E in the first column, is likely due to the difference in calendar time and associated changes in the broader macroeconomic and labour market conditions. We will investigate their difference further in the graphs to follow. The similarity between mothers and their matched non-mothers disappear at childbirth. Mothers' earnings drop markedly after childbirth, which is distinctly absent among non-mothers. For mothers with two children, their mean earnings drop even further after their second childbirth. The marginal decrease in mean earnings from their first child is greater than that from their second child. Their earnings do recover in the period after the second child turns one year old, but not to pre-motherhood levels. In contrast, all groups of non-mothers show a steady increase in their mean earnings over time, likely due to natural career progression that comes with age.

Next, we graph the time series of mean earnings of mother groups 1N and 1E, and their matched non-mothers, by month t in figure 1. (We do not produce the equivalent graph for women in groups 2NN, 2NE, 2EE because we cannot align both t_1 and t_2, i.e. the timing of the two childbirths, because the time gap between the two childbirths differ widely across mothers.) Vertical gray dotted lines mark key months such as childbirth, and the beginning and end of our matching period (from $t = -21$ to $t = -12$).

Figure 1:

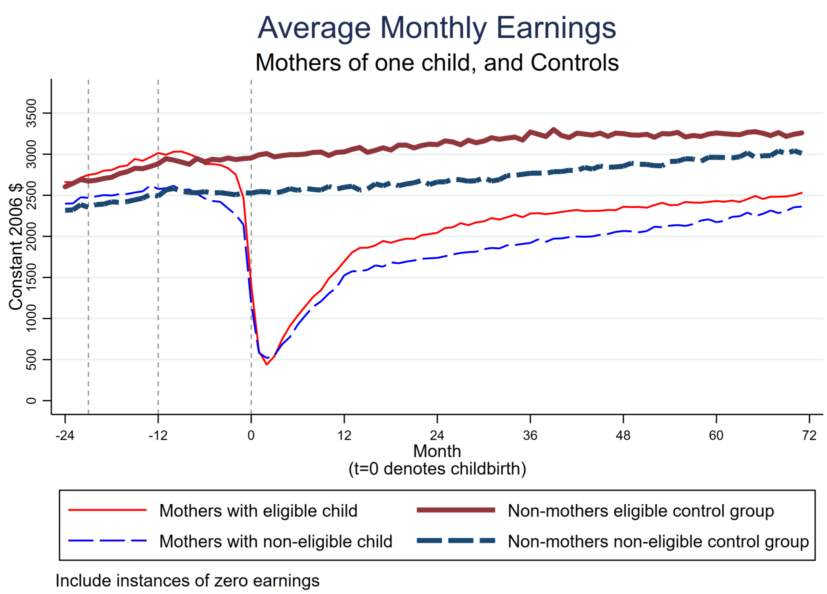

Figure 1 shows that the earnings time series in the matching period are remarkably similar between each group of mothers (thin lines) and their matched non-mothers (thick lines), in terms of both level and trend. This is further evidence that our matching criteria has achieved its goal. As expected, the divergence between mothers and non-mothers is stark after the matching period. Non-mothers see their mean earnings increase at a low but steady rate in the years to follow. In contrast, mothers' earnings begin to drop many months before childbirth[9], reach the minimum two months post-childbirth, and then begin to recover. This trend during the periods of pregnancy and motherhood is largely shared between groups 1N and 1E. The vertical gap between these two lines, and that between the two lines for the equivalent non-mothers,

[9]From our early data exploration, we believe this is a general observation, and not solely an artefact of our choice on the matching window, although it does help guide our choice.

in the pre-pregnancy period is likely due to earnings increasing more rapidly than the CPI in the decade starting 2000. Part of this is likely due to the two groups' different demographic compositions. We also note that this vertical gap is fairly constant between the two groups of non-mothers throughout the years, but less so between the two groups of mothers.

Mothers with eligible child (bright red thin solid line) have higher mean earnings prior to pregnancy, which then drops to almost the same minimum level as the group of mothers with non-eligible child (bright blue thin dashed line) two months post-birth. Thus the former group of mothers experience a drop in mean earnings of greater magnitude from childbirth. They also seem to have a slightly faster rate of earnings recovery within the first year post-birth. The lines for the two groups of mothers seem to have very similar trends after their children turns three years old, when the policy comes into effect for those eligible. We present more detailed regression analysis on the treatment effect of the policy in the next section.

5 Results

After a descriptive examination of earnings time series for various groups of mothers and matched non-mothers, we perform triple difference regression analysis on the ECE policy's effect on earnings. For mothers with one child, the triple difference denote comparisons along these three dimensions: between mothers with eligible and non-eligible child, between before- and after-birth periods for the same mother, and between mother and non-mother for the same time period. These comparisons isolate the effect of the policy by controlling the individual, motherhood, and time effects. For mothers with two children, this logic remains the same, with the slight difference that the three groups of mothers (2NN, 2NE, 2EE) cannot be compared at the same time. This is because these three groups represent different treatments. Comparison between groups 2NE and 2EE holds the effect of the second child constant (eligible in both groups), isolating the effect of ECE policy on the first child. Comparison between groups 2NN and 2NE holds the effect of the first child constant (non-eligible in both groups), isolating the effect of the policy on the second child. In all regression analysis to follow, we present separate sets of results for groups 1N and 1E, 2NN and 2NE, and 2NE and 2EE. We do not further combine these comparison groups to maintain the *ceteris paribus* condition the best we can in each treatment effect analysis.

Tables 3 and 4 show our regression results, for each set of comparison mother group. All re-

Table 3: Triple-Difference on earnings: Mothers in groups 1N, 1E, and matched non-mothers

Post period:	t=12-23	t=24-35	t=36-47	t=48-59	t=60-71	t=12-71
eligible	253.37***	255.70***	275.42***	290.18***	245.92***	265.93***
	(35.50)	(34.82)	(34.40)	(33.89)	(34.79)	(30.90)
post	8.60	-0.84	41.29	109.99***	74.38*	175.46***
	(28.54)	(32.24)	(34.50)	(37.96)	(42.59)	(27.83)
mother	95.46***	95.46***	95.46***	95.46***	95.46***	95.46***
	(29.93)	(29.94)	(29.97)	(29.99)	(30.01)	(30.02)
eligible*post	127.47***	154.55***	161.32***	76.47	117.21**	114.57***
	(27.00)	(30.77)	(37.47)	(48.81)	(54.63)	(28.63)
eligible*mother	-0.16	-0.16	-0.16	-0.16	-0.16	-0.16
	(43.38)	(43.40)	(43.44)	(43.48)	(43.50)	(43.61)
post*mother	-1071.14***	-964.13***	-915.01***	-878.20***	-812.75***	-928.25***
	(26.81)	(29.02)	(31.12)	(33.06)	(34.23)	(27.65)
eligible*post*mother	-189.33***	-123.61***	-122.81**	-66.27	-66.17	-113.64***
	(41.12)	(44.78)	(48.23)	(49.66)	(50.87)	(42.20)
age	335.88***	324.76***	305.26***	290.35***	280.19***	292.99***
	(14.10)	(13.21)	(12.57)	(11.52)	(10.53)	(14.14)
age^2	-4.06***	-3.86***	-3.51***	-3.27***	-3.12***	-3.33***
	(0.23)	(0.21)	(0.20)	(0.18)	(0.16)	(0.21)
ethnic:asian	-604.51***	-615.64***	-610.70***	-615.57***	-624.52***	-471.95***
	(35.23)	(35.30)	(35.81)	(36.04)	(36.16)	(38.14)
ethnic:maori	-258.37***	-277.72***	-287.04***	-293.83***	-295.01***	-223.81***
	(37.78)	(38.11)	(38.69)	(38.93)	(38.99)	(40.84)
ethnic:pacific	-55.53	-27.88	-33.53	-34.55	-61.04*	70.35*
	(36.49)	(36.49)	(37.29)	(37.64)	(36.98)	(39.11)
edu:HS	601.86***	611.48***	641.27***	653.96***	684.84***	656.93***
	(25.84)	(26.32)	(26.56)	(26.78)	(26.39)	(27.38)
edu:Uni	1265.32***	1333.15***	1400.24***	1465.23***	1530.79***	1498.67***
	(35.90)	(36.52)	(36.98)	(37.31)	(37.05)	(38.65)
edu:PG	1861.20***	1915.85***	1987.72***	2040.32***	2119.11***	2178.81***
	(78.38)	(79.63)	(81.67)	(84.28)	(84.14)	(87.41)
unemployment	-58.08***	-53.98***	-37.64***	-25.16**	-53.12***	-34.35***
	(11.99)	(10.88)	(10.50)	(10.49)	(11.09)	(4.95)
constant	-4330.60***	-4226.78***	-4108.61***	-3996.85***	-3680.08***	-3988.39***
	(228.11)	(216.35)	(209.06)	(194.25)	(184.29)	(233.69)
t-dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	572637	572637	572637	572637	572637	1853277
R^2 adjusted	0.1251	0.1162	0.1052	0.1029	0.1002	0.1053

Notes: Reference groups are European ethnicity and no high school qualification. Robust standard errors, clustered by each individual, are provided in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Triple-Difference on earnings: Mothers with two children, and matched non-mothers

Post period:	t=12-23	t=24-35	t=36-47	t=48-59	t=60-71	t=12-71
Groups 2NN, 2NE, and matched non-mothers:						
eligible*post*mother	76.81	44.76	-30.76	5.70	-116.33	-10.09
	(58.50)	(62.31)	(65.91)	(70.72)	(78.89)	(59.68)
Groups 2NE, 2EE, and matched non-mothers:						
eligible*post*mother	-264.95***	-299.76***	-108.95***	7.64	90.59**	-115.09***
	(32.68)	(35.63)	(37.35)	(39.04)	(40.28)	(33.21)

Notes: All variables in table 3 are included in regressions but omitted for brevity. Robust standard errors, clustered by each individual, are provided in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

gressions use the same set of regressors, consisting of the three indicator variables and their interactions, as detailed in Section 4, and control variables, including age, age squared, level of education, ethnicity, unemployment rate, and a full set of t fixed effects. Each table has six columns, for six different post-childbirth periods: when the treatment child is one/two/three/four/five year(s) old; and lastly, the cumulative period when the child is one to five years old. We do not use the period before the treatment child is one year old because mothers usually have zero or very variable earnings immediately after giving birth. We use different post-childbirth periods to explore the possibility that the treatment effect may not be constant across the early years of the child.

We first point out the findings that are common across all regressions, shown in table 3. Firstly, the indicator variables for the four levels of education (with "not finishing high school" as baseline) are always statistically significant. The relative magnitudes of their coefficients are as expected and increasing in educational attainment. Secondly, ethnicity indicators for Asian, Maori, and Pacific (with "European" as baseline) are always negative, with the first two indicators always significant. Thirdly, the age variable is always positive and significant, while the age-squared variable is always negative and significant. This is expected from career paths that progress with age, but at a gradually decreasing rate. Together, they paint this broad picture for working women in our study groups: All else equal, higher levels of educational attainment pays off in earnings; women of non-European ethnicity earn significantly less than the majority European population; age pays off but at a decreasing rate.

We now focus on the interacted indicator variable, $Post * Mother$, which is always negative and significant. This variable gives the difference-in-differences effect of the "motherhood wage penalty", from the comparison of a mother's earnings against her own before giving birth, and against a similar woman with no children. Comparing this coefficient across columns and across tables, we note a strong pattern: when the treatment child is the only child (groups 1N and 1E) or the last of two children (groups 2NN and 2NE), this coefficient becomes *less* negative as the treatment child grows older. In contrast, when the treatment child is the first of two children (groups 2NE and 2EE), this coefficient becomes *more* negative as the treatment child grows older, most likely due to the arrival of the mother's second child.

Turning to our key variable of interest - $Eligible * Post * Mother$, which estimates the impact of the ECE policy reform, this can also be viewed as the change in the motherhood wage penalty brought on by the policy shift. As shown in Tables 3 and 4, when the treatment child is the only child (groups 1N and 1E) or the first child out of two (groups 2NE and 2EE), this variable is negative and significant. Moreover, within each set of comparison, this coefficient becomes less negative as the treatment child grows older. Between groups 1N and 1E, this coefficient increases from -189.33 (dollars per month) when the child is one year old, to -122.81 when the child is three years old. Between groups 2NE and 2EE, this coefficient increases from -264.95 when the child is one year old, to -108.95 when the child is three years old. These numbers represent between a four to ten percent drop in earnings. Interestingly, these negative effects are significant not only when the ECE policy is in effect—when the treatment child is three to four years old—but also before that. When the treatment child is the last child out of two (groups 2NN and 2NE), this triple difference coefficient is not statistically significant.

Why is there a drop in labour market earnings for affected mothers in early years of motherhood? This is potentially because these mothers value time off-work more than earnings, and they value time off-work in the earlier years of their parenthood more than the later years. Thus, non-work hours are a normal good for new mothers, not an inferior good. In summary, mothers elect to smooth out the savings on child care expenses across time by forgoing earnings earlier and taking advantage of the ECE policy later.

Furthermore, the absence of a treatment effect from the comparison between groups 2NN and 2NE largely may be because the treatment effect on the second child is much more varied than that on the first child. This is because the time gap between the two children differs widely across mothers, thus the degree to which mothers have regained their earnings when they have their second child also varies greatly. Since mothers' earnings recovery after their second childbirth depends on their earnings before their second childbirth, the difference in earnings between pre-motherhood and post-second-childbirth is too variable to be statistically significant in our analysis. Alternatively, it is also possible that this is not statistically significant because mothers, when they have their second child, have less need to substitute earnings for time off-work, maybe because they are more experienced as parents, or because the financial burden of raising two children is more pressing than that of raising one child.

5.1 Wage quintiles

Whether mothers can afford to earn less, and the opportunity costs of doing so, depend on mothers' wage levels. These are reasons to suspect that the policy's effect on wages differ across mothers of different wage levels. Thus, we now separate mothers into different wage subgroups, and estimate the triple difference effect on each group separately. Specifically, for each group of mothers, we use their average monthly earnings in the matching period (from t (or t_1) $= -21$ to t (or t_1) $= -12$) to separate them into quintiles. Non-mothers are given the same quintile as the mothers to whom they are matched. For each quintile in each treatment-and-control group, we perform the same triple difference regression as seen in the previous section, using the same set of explanatory variables.

As shown in Tables 5 to 7 we only provide the triple difference estimate (i.e. the coefficient of variable $Eligible * Post * Mother$) from each regression. The pattern on other coefficients remain largely the same as in the previous section[10]. Groups 1N-1E and 2NE-2EE show the greatest number of significant estimates: the vast majority of them are negative, in the years when the treatment children are one and two years old. In contrast, groups 2NN-2NE in table 6 have almost no significant coefficients. These general findings echo the aggregate results from the previous section.

Focussing next on the quintile specific results, we observe that no estimates within the first quintile are significantly negative when comparisong 1N and 1E, or 2NN and 2NE. It is likely that mothers in this quintile are not able to take advantage of the savings in ECE spending because they did not have much ECE consumption prior to the policy reform, and / or they are not able to take a cut in their employment earnings. Interestingly, when comparing the mothers in the comparison group 2NE-2EE in the first quintile, they are the only subgroup with significant positive estimates overall. This is due to an increase in earnings when the child is

[10]These can be provided upon request to the authors.

Table 5: Triple-Difference estimates on earnings, by quintiles: Mothers in groups 1N, 1E, and matched non-mothers

Post period:	t=12-23	t=24-35	t=36-47	t=48-59	t=60-71	t=12-71
Quintile 1	-38.11	38.77	61.48	68.30	70.57	40.20
	(64.83)	(72.70)	(78.38)	(82.17)	(86.82)	(69.04)
Quintile 2	-212.88***	-152.30**	-87.37	-57.18	-55.46	-113.04
	(65.45)	(72.16)	(78.51)	(82.56)	(86.18)	(68.79)
Quintile 3	-251.09***	-197.34***	-177.51**	-80.99	-83.17	-158.02**
	(67.36)	(73.02)	(78.35)	(82.86)	(84.58)	(68.50)
Quintile 4	-322.66***	-221.97**	-260.30***	-216.86**	-203.73**	-245.10***
	(80.45)	(88.52)	(93.35)	(97.89)	(99.72)	(81.56)
Quintile 5	-125.12	-88.41	-153.43	-47.81	-62.28	-95.41
	(132.57)	(144.47)	(157.99)	(159.76)	(162.09)	(134.19)

Notes: All variables in Table 3 are included in these regressions but omitted for brevity. We only provide the estimated coefficient for $Eligible * Post * Mother$ for each quintile regression. Robust standard errors, clustered by each individual, are provided in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Triple-Difference estimates on earnings, by quintiles: Mothers in groups 2NN, 2NE, and matched non-mothers

Post period:	t=12-23	t=24-35	t=36-47	t=48-59	t=60-71	t=12-71
Quintile 1	121.62	87.98	66.56	120.72	43.83	84.27
	(82.67)	(92.06)	(99.83)	(109.13)	(126.52)	(88.42)
Quintile 2	197.22**	142.49*	62.52	105.03	-24.66	97.24
	(77.94)	(84.45)	(89.66)	(96.37)	(109.70)	(79.91)
Quintile 3	57.24	-6.70	-14.75	-76.32	-147.83	-49.73
	(85.80)	(92.36)	(98.12)	(105.39)	(118.54)	(86.25)
Quintile 4	273.38**	193.97*	14.74	10.28	-107.22	72.57
	(108.33)	(117.00)	(121.77)	(138.49)	(148.72)	(109.19)
Quintile 5	-113.00	-15.19	-57.17	125.85	-54.69	-42.56
	(189.19)	(202.41)	(215.84)	(227.32)	(253.09)	(192.99)

Notes: All variables in Table 3 are included in these regressions but omitted for brevity. We only provide the estimated coefficient for $Eligible * Post * Mother$ for each quintile regression. Robust standard errors, clustered by each individual, are provided in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Triple-Difference estimates on earnings, by quintiles: Mothers in groups 2NE, 2EE, and matched non-mothers

Post period:	t=12-23	t=24-35	t=36-47	t=48-59	t=60-71	t=12-71
Quintile 1	39.58	11.00	94.06	180.72***	206.19***	106.31**
	(49.23)	(53.69)	(57.89)	(61.89)	(66.06)	(51.16)
Quintile 2	-186.08***	-214.19***	-110.68**	-14.39	27.64	-99.54**
	(44.48)	(49.30)	(52.20)	(55.70)	(57.08)	(45.36)
Quintile 3	-138.42***	-203.11***	-31.52	33.92	54.72	-56.88
	(48.01)	(51.35)	(54.54)	(57.76)	(59.40)	(46.84)
Quintile 4	-465.92***	-480.60***	-279.95***	-173.07**	25.48	-274.81***
	(63.64)	(68.11)	(72.30)	(76.83)	(79.31)	(63.19)
Quintile 5	-580.99***	-618.90***	-223.67*	3.98	131.83	-257.55**
	(106.82)	(113.66)	(118.68)	(127.02)	(132.21)	(105.10)

Notes: All variables in Table 3 are included in these regressions but omitted for brevity. We only provide the estimated coefficient for $Eligible * Post * Mother$ for each quintile regression. Robust standard errors, clustered by each individual, are provided in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

four and five years old.

In contrast to the lowest quintile results, those in the moddle three quintiles generally experience similar outcomes to our aggregate picture. These mothers experience a drop in earnings as a result of the ECE policy, and this is most prominent when the child is aged one and two years old. Again, indicating potential displacement of private spending on ECE and the savings from this policy used to reduce engagement with the labour market when the child is below preschool age. Moreover, comparison across the second to fourth quintiles show that higher earning mothers cut back their earnings by a bigger magnitude, which suggests that even upper-middle-earning mothers (fourth quintile) value work-free hours more than their bigger opportunity cost of lost income.

This pattern reverses for mothers with one child in the fifth quintile: they do not appear to cut back on earnings. Continuing our logic on balancing the value of non-work time against the opportunity cost of missed earnings, we hypothesise that the latter is now outweighing the former for these high-earning mothers. In addition, it is likely that they also face a larger opportunity cost and general cost with regards to disrupting their career progression. The balance tips back the other way for high-earning mothers with two children, in table 7. They cut back on their earnings when their first child is one to two years old, just like middle-income mothers in the same group. From the childbirth time frames we use to construct group 2NE, mothers in this group have two children who are born two to six years apart. Thus, during this period when the fifth quintile mothers are cutting back on earnings, some of them are pregnant with their second child, or have given birth to them already. These are certainly strong reasons to cut back on work hours, despite their high wages. Other mothers who are not pregnant with their second child yet may still cut back on work in preparation or anticipation of their second pregnancy. Of course, the same can be said for mothers in groups 1N-1E: we cannot rule out the possibility that some may also be hoping for a second pregnancy (that never materialised in the end). In other words, whether a mother who has already had her first child (between July 2000 and June 2002) will eventually be categorised in group 1N or 2NE is an unknown at the time for some mothers. Thus, the difference we now observe between tables 5 and 7 with ex-post data does not come from two cleanly separated mechanisms. Nonetheless, the contrast between the fifth quintile mothers in these two tables is stark enough, that part of it must be driven by the difference in behavior between mothers who are sure to have a second child by the time their first child is 1-2 years old (in table 7), versus mothers who arenot actively trying to have a second child (in table 5). An argument can then be made, that while these two groups of mothers have comparable hourly wages, thus a comparable opportunity cost of working less, the former group of mothers expect roughly double the monetary saving on ECE in the coming few years (because both of her children will benefit from the policy), than the latter group of mothers. This helps tip the decision towards cutting back on earnings.

5.2 Discussion

Overall the impact of the ECE policy equates to a drop of approximately four to ten percent of affected mothers pre-pregnancy average earnings. How does this compare to related studies? The closest comparison is the Dutch reform detailed in Bettendorf, Jongen, and Muller (2015). In that study, they find very modest increases in employment despite the scale of the reform. The Netherlands case also had a lower baseline with regards to female employment and proportion of children in formal childcare, relative to the context in NZ[11]. Thus, it seems reasonable that there was more room for childcare uptake and expansion in maternal labour in the Netherlands than in NZ.

To further place our findings into context, lets consider the monthly cost of ECE in NZ.

[11] See Figure 3 in Bettendorf *et al.*, relative to Figure 1.23 in http://www.educationcounts.govt.nz/publications/series/annual-early-childhood-education-census

According to a 2011 survey[12], the average revenue from an hour of ECE, weighted across different service types, is \$9.55. Using the general CPI, this corresponds to about \$8 in the mid-2000's (i.e. around the time of policy implementation). Thus, if a child were to consume 20 hours of ECE per week, for four weeks in a month, the parents' monthly expenditure would be about \$640 without the government subsidy. All significantly negative estimates in our results are in the same order of magnitude, but smaller in value. This means that mothers are not reducing their income by the maximum possible monetary savings from the policy; instead it roughly ranges between one third to two thirds of it. This could be due to a number of reasons. For instance, it is possible that mothers never intended to reduce their earnings by maximum size of the subsidy, and / or that they never intended to use all 20 free hours of ECE (noting that the average number of ECE hours per week attended by three and four year olds were under 20, at 15 and 18 respectively, when the policy is implemented).

In terms of the caveats we need to consider when interpreting the findings in this study there are three potential issues to keep in mind. First, there is a lack of information on supply of ECE spaces in NZ at the time of the policys implementation. While we know that the majority of ECE providers participated in the programme, there is limited information as to whether demand for ECE services by parents were fully realised. Secondly, there were occasional ad hoc reports of ECE providers using various loop holes with the policy to create additional payment avenues. For instance, imposing a minimum number of enrolment hours that is one hour more than the policys coverage and then charging a high marginal price for the last hour; or creating additional optional charges.

The third and final source of bias that may affect our results is the endogeneity of mothers' fertility choices, if the policy reform acted as an incentive signal and adjusted fertility decisions.Th impact of this endogeneity would be to bias our results towards zero. However, if we consider the timing of the policy announcement, the length of a full-term pregnancy, and the windows we use to define our mother groups, mothers in most of our study groups are unlikely to have had enough time to conceive and give birth to a child in response to the policy's announcement.

The only exception is group 2EE, on their decision to have their second child. This group is

[12] https://www.educationcounts.govt.nz/publications/ECE/income-expenditure-and-fees-of-ece-providers-in-new-zealand

most susceptible to be incentivised by the policy. In our formation of the 2EE group of mothers, we constrain the timing of their first child only (between July 2004 and June 2006), not their second child. Thus, they would have time to conceive and carry their second pregnancy to full term if it were purely incentivised by the policy. In other words, in the counterfactual scenario with no ECE subsidy, some of these mothers would likely be in group 1E instead of group 2EE.

6 Conclusion

In this paper we analyze the effect of the 20 hour free ECE policy in NZ on mothers' earnings. This 2007 policy fully subsidises ECE for three and four year olds for up to 20 hours per week. We use administrative data with earning observations from one year pre-pregnancy to six years post-childbirth. We use a triple difference methodology, by comparing eligible mothers' earnings before and after childbirth, against non-eligible mothers, and against comparable contemporaneous women who are not mothers. We find that most eligible mothers demonstrate a significant reduction in earnings, of four to ten percent of their pre-pregnancy earnings. Thus, leisure (or non-work hours) are a normal good for these mothers, not an inferior good. Moreover, we find that most reduction in earnings happen when the treatment child is one to two years old, before they are old enough to be eligible for the policy.

We hypothesise that mothers are substituting the savings on ECE intertemporally. They value time off-work more in their early years of motherhood more than their later years. Thus, they work less when their treatment child is one to two years old, with the expectation that lost income will be compensated by savings on ECE expenditure when their treatment child is three to four years old.

We further break down the analysis by earning quintiles. Middle-income mothers (second, third, fourth quintiles) show the same pattern as described above. The lowest earning mothers (first quintile) either show no significant reduction in earnings (for those with one child), or a significantly positive increase in earnings during the years when their treatment child is in ECE (for those with two children). The highest earning mothers (fifth quintile) either show no significant reduction in earnings (for those with one child), or the same pattern of reduction when the treatment child is one to two years old (for those with two children). We hypothesise that these differences are likely due to whether the mother can actually afford to cut earnings; whether the policy actually brings about an increased consumption of ECE than the mother would otherwise; and the opportunity cost of lost income and interrupted career trajectory.

From the perspective of public economics, this policy mostly displaces private spending on ECE. It has positive labour market effects for only the lowest earning mothers. We believe that it does not lead more mothers to increase their earnings, through increasing their work hours, because the average consumption of ECE is already close to 20 hours at the policy's implementation. The policy certainly does not bring about a further drastic increase in ECE consumption beyond 20 hours. (This could be due to the slow increase in provision in the early years of the policy.) Thus, the average mother is unlikely to experience a drastic increase in child-free hours, which she might be able to dedicate to work.

From the perspective of child development, this policy is likely to have a neutral (at worst) or positive effect. Unfortunately, our administrative data gives us no information on mothers' hours worked, or the number of hours mothers spend with their children. If we assume the latter does not decrease due to the policy, then it is likely to have some positive effect on parent-child bonding. In addition, the policy eases mothers' financial burden to get back to work, particularly in the early years of motherhood.

References

ANDERSON, D. J., M. BINDER, AND K. KRAUSE (2002): "The motherhood wage penalty: Which mothers pay it and why?," *American Economic Review*, 92(2), 354–358.

——— (2003): "The motherhood wage penalty revisited: Experience, heterogeneity, work effort, and work-schedule flexibility," *Industrial and Labor Relations Review*, 56(2), 273–294.

BAKER, M., J. GRUBER, AND K. MILLIGAN (2008): "Universal child care, maternal labor supply, and family well-being," *Journal of political Economy*, 116(4), 709–745.

BAUM, C. L. (2002): "A dynamic analysis of the effect of child care costs on the work decisions of low-income mothers with infants," *Demography*, 39(1), 139–164.

BETTENDORF, L. J., E. L. JONGEN, AND P. MULLER (2015): "Childcare subsidies and labour supply—Evidence from a large Dutch reform," *Labour Economics*, 36, 112–123.

BLAU, D., AND J. CURRIE (2006): "Pre-school, day care, and after-school care: who's minding the kids?," *Handbook of the Economics of Education*, 2, 1163–1278.

BUSHOUSE, B. K. (2008): "Early childhood education policy in Aotearoa/New Zealand: The creation of the 20 hours free programme," *Fulbright New Zealand*.

HAECK, C., P. LEFEBVRE, AND P. MERRIGAN (2015): "Canadian evidence on ten years of universal preschool policies: The good and the bad," *Labour Economics*, 36, 137–157.

HARDOY, I., AND P. SCHØNE (2015): "Enticing even higher female labor supply: The impact of cheaper day care," *Review of Economics of the Household*, 13(4), 815–836.

HAVNES, T., AND M. MOGSTAD (2011): "Money for nothing? Universal child care and maternal employment," *Journal of Public Economics*, 95(11-12), 1455–1465.

HEGEWISCH, A., AND J. C. GORNICK (2011): "The impact of work-family policies on women's employment: a review of research from OECD countries," *Community, Work & Family*, 14(2), 119–138.

KIMMEL, J. (1998): "Child care costs as a barrier to employment for single and married mothers," *Review of Economics and Statistics*, 80(2), 287–299.

LEFEBVRE, P., AND P. MERRIGAN (2008): "Child-care policy and the labor supply of mothers with young children: A natural experiment from Canada," *Journal of Labor Economics*, 26(3), 519–548.

LUNDIN, D., E. MÖRK, AND B. ÖCKERT (2008): "How far can reduced childcare prices push female labour supply?," *Labour Economics*, 15(4), 647–659.

MAY, H. (2002): "Early Childhood Care and Education in Aotearoa-New Zealand: An overview of history, policy and curriculum," *McGill Journal of Education*, June, 1–15.

MAY, H. (2008): "Towards the right of New Zealand children for free early childhood education," *International Journal of Child Care and Education Policy*, 2(1), 77–91.

Meade, A., and V. N. Podmore (2002): "Early Childhood Education Policy Co-ordination under the Auspices of the Department/Ministry of Education: A Case Study of New Zealand. Early Childhood and Family Policy Series," *UNESCO Early Childhood and Family Policy Series*, 1, 1–34.

Pacheco, G., C. Li, and B. Cochrane (2018): "Empirical evidence of the gender pay gap in New Zealand," *New Zealand Journal of Employment Relations*, forthcoming.

Wrohlich, K. (2004): "Child care costs and mothers' labor supply: an empirical analysis for Germany," Discussion paper, DIW Discussion Papers.