Pre-analysis plan for an information experiment about the Swedish EITC

September 9, 2020

Abstract

In March 2018 we conducted an experiment where new parents were sent information about the Swedish Earned Income Tax Credit (EITC). We will soon order the data necessary to evaluate how this treatment affected their economic behaviour. In this plan we present our hypotheses and specify how we will test them. We also provide background information about the Swedish EITC and the experiment we conducted.

Economic models of labour supply usually assume perfect, or at least constant, knowledge about tax policy, even though most economists would agree that we should expect stronger responses when policy is well-known. This neglect is understandable, given that no one knows *how* important such knowledge is. With exception for the experiment conducted by Chetty and Saez (2013), in which they offered households a two-minute explanation of the American EITC, we are not familiar with any other information experiments on tax policy and labour supply. From the viewpoint of empirical research, this should not come as a surprise; major policies are usually known by the public, which makes information experiments difficult to conduct, and minor reforms yield too small responses.

To fill this gap, we have conducted an experiment in which we sent half of all new parents information about the Swedish EITC, in Swedish called *Jobbskatteavdraget*. We have chosen this empirical setting because we believe it is a case where an information campaign can have substantial effects. First, labour supply responses can be slow, given that people who want to increase their income first may need to find a job, advance within the firm or even get a longer education. Most people on parental leave, on the other hand, are already employed, meaning that they usually have the opportunity to get back to work as soon as they prefer to do so. Second, surveys have shown that most people are not aware of the Swedish EITC, despite the large size of the tax reduction (Riksrevisionen 2009). Moreover, some of the incentives we inform about are unintended side effects of the reform, which we assume that even fewer are aware about. Taken together, this setting is therefore a most-likely case for when additional information may alter the optimizing behaviour of individuals.

The letters were delivered in March 2018. This pre-analysis plan outlines the steps that will be taken next: data collection, hypotheses to be tested, as well as the specific analyses to be done. As a background, we will first present the details of the Swedish EITC, followed by a description of how the experiment was conducted.

The Swedish EITC

The EITC was implemented in 2007 and has thereafter been modified in 2008, 2009, 2010, 2014, 2016 and 2019. All numbers presented in this plan refer to the rules that were in place 2018. The Swedish EITC has a complicated construction, and is calculated as the municipal tax rate multiplied with the difference between a special amount (särskilt belopp), which is a function of a person's earned income, and the basic deduction (grundavdrag), which is a function of a person's total non-capital income. Both functions are defined in constant prices, with adjustments being made on an annual basis. The functions can be written as equations or illustrated graphically as in Figure 1. Needless to say, it is close to impossible for people to have accurate knowledge about the details of this tax reduction.

There was much debate when the EITC was first introduced. The introduction of the EITC was the flag-ship reform of the center-right government that took office in the fall of 2006. This, together with the large size of the tax reduction, means that we should expect the EITC to be relatively well-known. Yet, according to a survey conducted in 2009, only 40 percent of the Swedish population knew about the tax reduction (Riksrevisionen 2009), indicating that an information experiment potentially could have a substantial effect.

What we assume to be even less known is how the tax reduction affects some of the economic incentives facing new parents. The construction of the EITC means that it will be more beneficial for two people with a shared economy to distribute earned income equally between them. The effect is such that for a pair of new parents where both earn around 35 000 SEK per month, their joint EITC would be about 16 000 SEK larger if they split their time at work equally (over a calendar year), compared to if one of them was on leave for the whole period while the other one was at work. Seeing how this is an unintended side effect of the reform, it has not been a part of the public debate and we would not expect the subjects in our experiment to be aware of these effects when they receive our letter.

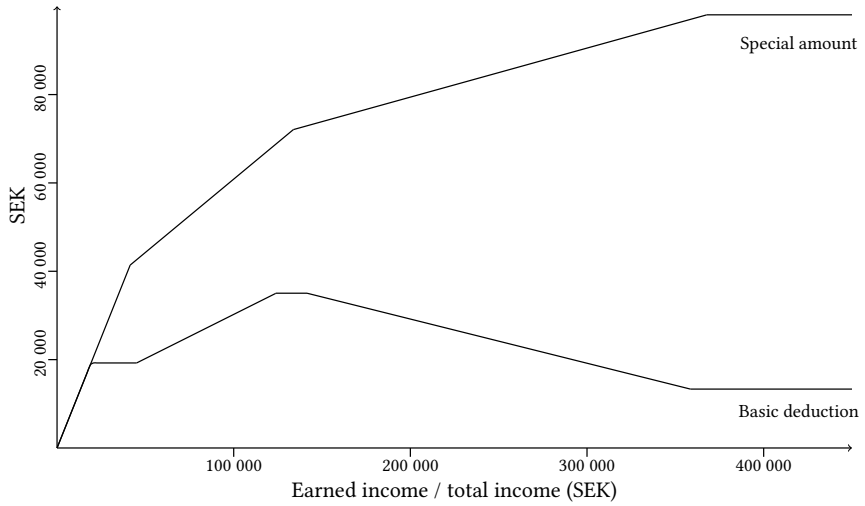
It should be noted that when thinking about how the EITC affects the incentives for someone choosing a mix of work and paid parental leave, the kink points in the upper curve (the special amount) are much more important than the ones in the lower curve (the basic deduction). The reason is that the Swedish parental insurance covers almost 80 percent of the income loss from not working. Hence, if a person decides to get back to work a month earlier, it may increase the person's earned income with 25 000, but the total income with only 5 000, effectively making any changes in the basic deduction much smaller than the ones in the special amount. In our information letter, we therefore chose to show the EITC as a function of the earned income, while holding other forms of income constant.

The experiment

We started planning this experiment during 2016 and in November of that year the necessary funding to conduct it was kindly provided to us by the Institute for Evaluation of Labour Market and Education Policy (IFAU).

Prior to conducting the experiment, we obtained an approval from the Regional Ethical Review Board in Uppsala (dnr 2017/026, with supplement dnr 2017/026/1).[1] The application to the Ethical Review Board consisted of a project description and

[1] This is the Swedish equivalence of IRB.



The special amount is a function of earned income and the basic deduction is a function of total non-capital income. The EITC is defined as the municipal tax times the difference between the special amount and the basic deduction. All numbers refer to the income year 2018.

Figure 1: The construction of the EITC

an overview of the data material that we would use in our upcoming analysis. The experiment was furthermore approved by an internal review board (Uppdragsråd) at Statistics Sweden. This means that the overall research design was set prior to conducting the actual experiment.

In March 2018 Statistics Sweden delivered our treatment, which consisted of a letter which was sent to half of all mothers who had given birth during the preceding eight-month period. The letter included information about the Swedish EITC and how the size of this tax credit increases with the size of an individual's earned income. Specifically, the letter highlighted the effect of sharing parental leave on the household's total tax credit. Important to note is that the letter provided accurate information about the EITC, and the experiment does not contain any element of deceit. We informed the recipients that the letter was part of a project at Uppsala University where the aim is to provide information regarding how household income is affected when a person is absent from work during a calendar year. However, the letter did not reveal that we would later use registry data to study the recipients' behaviour, since such information could result in a Hawthorne effect. The fact that we did not obtain informed consent to analyse registry data was made explicit in our application to the Regional Ethical Review Board and is something we have in common with other research projects that use Swedish registry data. The letter furthermore contained contact information to the principal investigator and information that the letter was sent out in cooperation with Statistics Sweden (SCB).

The distribution of the letter was carefully timed. At first glance it may seem as if the treatment was delivered rather late, given that the first quarter had already passed when the letter was received. However, the economic incentives we inform about make it particularly beneficial for an average earner to work about 4–5 five months during 2018, leaving 4–5 months between the letter and the expected behavioural response. This should leave just enough time for families to adapt their plans for the second half of the year (especially after returning from vacation over the summer months), while

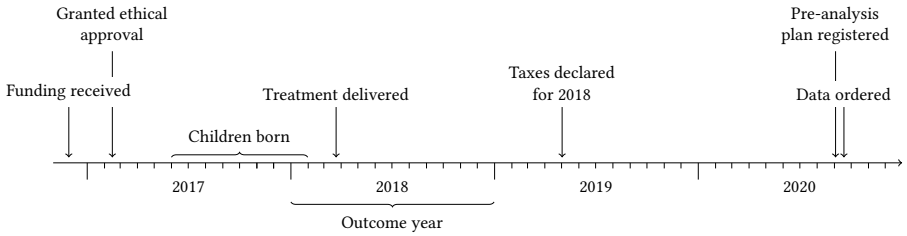


Figure 2: Timeline of the study

simultaneously not leaving so much time that the information is completely forgotten.

The sampling frame consisted of all mothers to children born between 1 June 2017 and 31 January 2018. These 74,667 individuals were randomly allocated to a treatment group ($n = 37,334$) and a control group ($n = 37,333$). After the randomization was conducted, 151 individuals were removed from the list of letter recipients, either because the child had a protected identity (92), the mother had a protected identity (7), the mother had emigrated (2), the mother was deceased (1), Statistics Sweden did not have a registered postal address (1) or because the mother was under 18 years old (48). Hence, the letter was sent to 37,183 of the 37,334 individuals in the treatment group. Because this check was made post treatment, we will not exclude individuals from the study because they were not sent a letter. However, it should be noted that some individuals in the sampling frame may not show up in the final data that we will order, for example because they emigrated or deceased before the population registry for 2018 was established. Importantly, such attrition should not be affected by the treatment assignment and therefore will not bias the results.

The letters' address field had the name of both parents, unless information about the biological father was missing, and the letter was sent to the mother's address. The construction of the sampling frame, the randomization into the treatment group and the delivery of the letter was carried out by Statistics Sweden. At this point, the researchers did not have any data access.

The letter sent out was based on calculations where we assumed an average municipal tax rate and non-work income of 150 000 SEK, which allowed us to express the EITC as only a function of the earned income. The letter included a graph which showed the size of the EITC for different income levels (see attachment at the last page of this plan) and a text which said that if you during 2018 have an earned income of 135 000 SEK you get 15 000 more in EITC than if you do not work at all during the year. We then included a sentence written in bold with an example of a person with a wage corresponding to the mother's wage (based on her income during previous years), saying that it would very beneficial for this person to work at least x months during 2018 (the specified number of months was based on the wage). The letter also informed that the EITC continues to increase with income, but that the payoff is smaller after that threshold, and that this means that a couple typically receive more in EITC if they share the parental leave equally during a calendar year compared to if one parent would stay at home for the whole period.

If any of the parents during 2015 or 2016 had registered an earned income larger than 350 000 SEK or 600 000 SEK, respectively, the letter also included information about at what income level the EITC no longer increases (incomes larger than 365 000 SEK) and when it starts decreasing (incomes larger than 615 000 SEK).

Theoretical expectations

Recent work suggests that information is crucial for people's ability to adjust their labour supply in response to the tax system, and that it is especially important for responses at the intensive margin (Chetty 2012; Chetty et al. 2013). Based on this, we would expect the information in our letter to increase the labour supply in the treatment group. We expect this response to be larger at the intensive margin, but we will investigate both types of responses. At the intensive margin, we would primarily expect an effect among mothers, because they tend to be the ones who stay at home for the longest period, also during the year post birth. On the extensive margin, we are just as likely to see a behavioural response among fathers, because the EITC affects the incentives to have a job in the same way for men and women.

Regarding the extensive margin, it should be noted that the Swedish EITC increases the incentives to work in particular for people who have the possibility to work maybe 3–4 months or more. For smaller amounts of work, the EITC usually does not make that much of a difference. The reason is that the EITC does not begin until the earned income exceeds the basic deduction, which usually is somewhere between 20 000 and 35 000 SEK for people out of work.

H1 Receiving the letter increases mothers' labour supply at the intensive margin.

H2 Receiving the letter increases labour supply at the extensive margin.

Two common policy goals is to achieve a high labour supply and to reduce economic inequalities related to parenting. However, depending on how the response among mothers and non-employed affect their partners, the two hypotheses above could be true without the treatment affecting aggregate supply. And similarly, an increase in labour supply among mothers does not necessarily imply that parental leave becomes more equally distributed within the couple. From a policy perspective, we may therefore be more interested in how an information campaign like this would affect total labour supply, on the one hand, and the distribution of parental leave on the other.

H3 Receiving the letter increases the couple's joint labour supply.

H4 Receiving the letter makes parental leave more equally distributed within the couple.

Empirical specification

Data

The data we plan to use consist of Swedish registry data on primarily labor market-related variables merged with a treatment indicator saying if an individual belonged to the control or treatment group. All data, including this indicator, will be ordered from Statistics Sweden, in an order made specifically for this project.

It should be noted that when we submit this pre-analysis plan, no one have had the possibility to work with this indicator of treatment status; not anyone of us, nor any other researcher. Up until this point, the data is in the hands of Statistics Sweden.[2] We

[2] Statistics Sweden needs to formally decide ("sekretessprövning") whether they can provide the data for research purposes. Their decision is based on the content of the Ethical Review Application discussed above.

Table 1: Variables at our disposal

Name	Description	Years
LopNr	ID	–
AterPNr	Marker for reused ID	–
LopNrBarn	ID of biological children	–
Brev	Treatment indicator	–
FodArManBarn	Year and month of birth (children)	–
FodAr	Year of birth (everyone)	–
ForPeng_NDag	Days with parental allowance	2000–2013
ForPeng_Ndag_MiDAS	Days with parental allowance	2014–2018
ForvInk	Earned income	2000–2018
Kon	Biological sex (binary)	–
Kommun	Municipality of residence	2017—2018
Ssyk4_2012_J16	Occupation	2017—2018
SsykAr_J16	Year for info about occupation	2017—2018
SsykStatus_J16	Occ. congruence with Nov. work	2017—2018
YrkStalln	Employment status	2017–2018
SyssStat	Labour market status	2017–2018
AstSNI2007	Industry of workplace	2017—2018
AntalSys	Employees at workplace	2017–2018
Sun2000niva	Highest level of education	2017
Kallkod	Source for education	2018
ExamAr	Year when edu. was finished	2018
LoneInk	Wage income	2018
Tjomf	Percent of full-time	2018
ManL	Monthly wage	2018
Glon/Grlon	Monthly wage excl. “rörliga tillägg”	2018
Uppr	Enumeration factor	2018
Nsvar	Enumeration factor	2018
LillaN	Enumeration factor	2018

have recently received a cost proposal from Statistics Sweden and we will not accept their offer until after this plan has been submitted. After that, it will take up to five weeks before we get access to the data.

The variables we will have access to are listed in Table 1. The data cover everyone residing in Sweden during 2018, as defined by the RTB register. We will here define the variables we will use in our main analyses.

We define annual earned income (I) as the variable ForvInk multiplied by 100, which transforms it into SEK per year.

We define an individual's *potential income* (I^*) as the maximum value of his or her relative income (compared to the national mean) during the three years preceding 2018, multiplied by the average income during 2018; or 180 000 SEK (corresponding to a monthly wage of 15 000, or approximately 75 % employment at a minimum wage), whatever is larger. This variable is intended as an estimate for what a person's income would be if he or she was working full time for a year.

$$I_i^* = max\left(180000, \frac{I_{i,15} \times \hat{I}_{18}}{\hat{I}_{15}}, \frac{I_{i,16} \times \hat{I}_{18}}{\hat{I}_{16}}, \frac{I_{i,17} \times \hat{I}_{18}}{\hat{I}_{17}}\right) \tag{1}$$

We calculate the estimated number of days at work (D_i) as a person's earned income (I) divided by his or her potential income (I^*), multiplied by 365. We also add a restriction that the time at work must not exceed the length of a year minus the time with paid parental leave (P_i), where the latter is measured with the variable ForPeng_NDag_MiDaS.

$$D_i = min\left(\frac{I_i}{I_i^*} \times 365, 365 - P_i\right) \tag{2}$$

For the extensive margin, we define employment using the register variable SyssStat. This variable is measured in November of each year and can take three values: Employed (1), Not employed but with registered taxable income (5), and Not employed and no registered taxable income (6). Category 5 implies that the person was not employed during the month of measurement (i.e. November) but has had work income at some other point during the year. Our employment variable is thus defined as

$$E_i = \begin{cases} 1 & \text{if SyssStat} = 1, 5 \\ 0 & \text{if SyssStat} = 6 \end{cases} \tag{3}$$

We define previous parental leave as the number of paid days of parental leave (ForPeng_NDag or ForPeng_NDag_MiDaS) taken out during the year of birth and the consecutive year of the person's most recent child born between 2000 and 2015. If he or she has got no previous children, the variable is coded as zero and we will include a corresponding binary indicator for missingness in all analyses where this variable is present.

Years of education is derived from the register variable Sun2000Niva, taken from LISA 2017. The register variable consists of a three-digit code which we translate into years of education according to the following scheme:

Sun2000niva	Years of education
<200	7
200–299	9
310–319	10
320–329	11
330–339	12
410–419	13
520–529	14
530–539	15
540–549	16
550–559	17
600–629	18
640–649	20

Imputation

With the two exceptions described below, we will – after conducting the data operations described above – replace missing values on binary and continuous variables with their global mean (based on data in the sample frame) and impute categorical variables with the global mode.[3] If data is missing for more than 10 percent of the observations, a binary indicator for missingness will be added to the vector of controls. The first exception is our measure of previous parental leave, which we have already described how we will handle. The second exception is that observations will be excluded if data is missing on the treatment or the outcome variable.

Vector of control variables

To reduce the amount of residual variation in the dependent variables, and thus increase the precision of our effect estimates, we will include a vector of control variables in all models. Some variables are related to the child, like birth month and the sex of the child. These variables will enter the regression in the same way for all hypotheses. Other variables are related to the parents. These ones will be calculated differently depending on the hypothesis we are testing. When testing H1, H3 and H4, we will include one variable for the mother and another variable for the father. When testing H2, we will include one variable for the index person and another variable for his or her partner. If there is none, a binary indicator for missingness will be included, and all variables set to zero.

Variables in the vector of controls
Variables related to the child
Birth month of the child (FE)
Sex of the child
Variables related to the parents
Years of education
Potential income
Year of birth
Employed 2017
Employment status 2017 (FE)
Previous parental leave
Binary indicators of missingness

Balance test

We will present a table where we compare averages for all binary and continuous variables included in the vector of covariates. The balance tests will be two-tailed with a .05 p-value threshold. No measures will be taken if imbalances are found, beyond what is already described in this pre-analysis plan (i.e., including the control vector in all analyses).

[3] Whether a variable is categorical or continuous is determined by how they are specified in the vector of covariates that we present in this next section.

Visualization

We believe that the pre-registration of hypothesis tests is fundamental to reduce the share of type I-errors in published work. However, we also acknowledge the risk that only analysing pre-specified specifications may lead us to ignore unforeseen responses that are of economic importance. This risk is especially large in projects like this one, where there are few prior examples to learn from, and it is far from obvious what kind of response we should expect. To strike a balance in this trade-off, we will combine pre-specified hypothesis tests with a visual analysis where we put less emphasis on statistical significance.

To visualize the response to our letter, we will plot the two density functions for earned income divided by treatment status. We will also plot the density-difference, either as the simple difference between the two kernel densities or using the least-squares density-difference (LSDD) estimator developed by Sugiyama et al. (2013). We will provide a confidence interval for this difference. However, the probability of the difference being outside this interval at at least one part of the distribution is much larger than .05 under the null hypothesis, and we want to emphasize the importance of making a combined assessment of the size of the difference, the error bound, and whether the difference is a logical response to the information that was provided in the letter.

We will provide correct p-values for the observed areas of excess mass above or below the control group density. Such p-values can be obtained through randomization inference and states the probability of obtaining at least one area of at least that size if the null hypothesis is true. We will also provide the result from the Kolmogorov–Smirnov-test, which uses the largest observed difference in the *cumulative* density functions as its test statistic. However, none of these p-values directly answer to questions of economic importance, and we therefore do not want to reduce this analysis to a matter of a single statistical test.

Testing H1

The first hypothesis states that receiving the letter increases mothers' labour supply at the intensive margin. The mother's labour supply will be operationalized as her earned income during 2018, which we measure using the variable ForvInk. The hypothesis can thus be tested by regressing earned income (I_i) on an indicator for belonging to the treatment group (T_j) and a vector of covariates ($\boldsymbol{\chi_i}$), while restricting the sample to mothers. We use i as the index letter for individuals and j as the index letter for couples. H_1 implies that $\beta_1 > 0$:

$$I_i = \alpha + \beta_1 \times T_j + \boldsymbol{\beta\chi_i} + \epsilon_i \tag{4}$$

Testing H2

The second hypothesis states that the letter increases labour supply at the extensive margin. It can be tested in the same way as H_1, but switching income for employment. This time we include both mothers and fathers. H_2 implies that $\beta_1 > 0$:

$$E_i = \alpha + \beta_1 \times T_j + \boldsymbol{\beta\chi_i} + \epsilon_i \tag{5}$$

Testing H3

The third hypothesis states that receiving the letter increases the couple's joint labour supply. The couple's joint labor supply is operationalized as the total of their individual earned incomes, and the equation to be estimated can therefore be written as below, with H_3 implying that $\beta_1 > 0$:

$$I_j = \alpha + \beta_1 \times T_j + \boldsymbol{\beta \chi_j} + \epsilon_j \tag{6}$$

Testing H4

The fourth hypothesis states that receiving the letter makes parental leave more equally distributed within couples. We define parental leave as everything that is not work, and use the absolute difference in days at work as our outcome. We can then regress this difference on the treatment indicator, with H_4 implying that $\beta_1 < 0$:

$$|D_j^M - D_j^F| = \alpha + \beta_1 \times T_j + \boldsymbol{\beta \chi_j} + \epsilon_j \tag{7}$$

Hypothesis testing and multiple comparisons

We will use robust standard errors. When testing H2 they will be clustered at the household level. All our four hypotheses will be tested using one-tailed tests. When adjusting p-values for the number of hypotheses, we will use the free step-down resampling method of Westfall and Young (1993). In our case, the procedure consists of the following steps:

1. Sort the p-values corresponding to the four hypotheses in ascending order, from p_1 to p_4, where $p_r \leq p_{r+1}$.
2. Conduct 100 000 simulations under the null hypothesis of no treatment effects and save the p-values from each simulation.
3. In each simulation, sort the p-values (p_r^*) in the same order as the original tests, and enforce monotonicity by defining p_r^{**} as the smallest p-value among this test as well as the tests that originally resulted in a larger p-value ($p_r^{**} = min(p_r^*, p_{r+1}^*, ..., p_4^*)$).
4. For each test, calculate the adjusted p-value (p_r^a) as the share of simulations where $p_r > p_r^{**}$. Or alternatively, compute the four threshold p-values corresponding to $\alpha = .05$ by finding the 5th percentile of p^{**} for each test.
5. Enforce monotonicity again, so that larger unadjusted p-values always correspond to larger adjusted p-values ($p_r^a = max(p_1^a, p_2^a, ..., p_r^a)$).

If the tests are independent of each other, this procedure is identitical to Holm's step-down algorithm. But if the tests are positively correlated, this procedure will be more powerful.

Estimation sample

Our full sample consists of the 74 667 mothers that were included in the sampling frame when the treatment was delivered, together with the biological fathers of the children, minus anyone who is not included in the data we will order. In addition, we also have access to data covering the whole Swedish population. These data will primarily be used for benchmark analyses not specified above.

Software

All data will be analysed with the software package Stata 15.

Additional analyses

In addition to the carefully specified hypothesis tests above, we will conduct a large number of other analyses. Some of them are briefly presented below, but we will adjust and expand this set of analyses in response to what the hypothesis tests show, to comments we get from colleagues and reviewers, as well as to new ideas we get along the way. As we choose to preserve our degrees of freedom for these analyses, we also acknowledge the risk that we – consciously or not – let the outcome of statistical tests affect what results we present, leading us to the recommendation that statistical significance must be interpreted with some caution.

To put the results from H1–H4 into perspective, we will use two kinds of benchmarks. For one thing, we will compare our results to elasticities estimated in the relevant literature, but we will also try to estimate behavioural responses to the EITC in absence of information. Since we only have observational data for these estimates, the responses cannot be causally identified, but there are three sources of variation in the incentives from the EITC that we will analyse in a difference-in-difference design. First, there is a lot of variation of time, given that the EITC was implemented step-wise under 2007–2010 and 2014, as well as phased out at higher incomes after 2016. Second, the incentives differ substantially between households depending on their level of income and how it is distributed between the spouses. And third, the incentives also differ depending on what time of the year a child is born.

We will also conduct a large number of analyses where we test the robustness of our results or how the effects differ between subgroups. Below we list some of them.

- People who are self-employed tend to have more control over their income. Partly because they can decide for themselves how much they want to work, and partly because they can sometimes choose how much to pay themselves as wages, regardless of how much they worked or what profit their business made. It is therefore interesting to run the main analysis separately for self-employed and others.

- We have chosen to include the full sample in our main analyses, even though the incentives for a behavioural response are much stronger in some groups than others. We will therefore conduct an analysis where we restrict the sample to individuals where we would expect the largest effect. The selection will be based on criteria like potential income and when the child was born.

- We will create a theoretical model for the spouses' labour supply. Based on that model, we will determine expected responses and see how well it corresponds to the actual behaviour we observe.

Bunching

Another interesting aspect is to analyse whether the treatment has increased bunching at the kink points mentioned in the letter. That kind of optimizing behaviour is dependent on detailed knowledge about the tax system, which most subjects did not have before they received our letter. Information is thought to be so important in this regard that the amount of bunching has previously been used as a proxy for knowledge about the tax system (Chetty et al. 2013). We will look at three different kinks. The one where we would expect the highest probability of finding an effect is at 135 000, which is the kink we highlighted in every letter. When it seemed relevant (previous income above 350 000 and 600 000 SEK, respectively), the letters also included the information that earned incomes above 365 000 SEK does not qualify for additional EITC, and that the EITC will start to decrease after earned incomes above 615 000 SEK.

One of the key issues when quantifying bunching per se is to estimate what the income distribution would look like in the absence of a kink. We draw on Chetty et al. (2011) and create this counterfactual distribution by fitting a polynomial to the income distribution while omitting observations close to the kink. Specifically, we first recode the income variable so that it measures the deviation from the kink point, group the data into bins each covering 1000 SEK, and exclude all observations more than 100 000 SEK away from the kink (e.g., incomes smaller than 35 000 or larger than 235 000). Separately for the control and treatment group, we then fit a seventh-degree polynomial to the income distribution while omitting observations close to the kink point (e.g., incomes between 125 000 and 145 000). This fitted function then serves as our counterfactual distributions.

Borrowing on the notation in Chetty et al. (2011), we regress the number of individuals in each bin (C_j) on a seventh-degree polynomial of the income relative to the kink, and a set of bin-specific indicators for belonging to the small window around the kink ($\mathbf{1}[Z_i = l]$). The size of this window will be determined based on visual inspection of the distribution.

$$C_i = \sum_{k=0}^{7} \beta_k \times Z_i^k + \sum_{l=-10}^{10} \gamma_l \times \mathbf{1}[Z_i = l] + \epsilon_i \tag{8}$$

The predicted values from the polynomial then serves as our initial estimate of the counterfactual distribution and the sum of the coefficients for the binary indicators serve as our initial estimate of the excess mass around the kink. However, this calculation overestimates the amount of bunching because the counterfactual distribution does not include the excess number of individuals around the kink. Or put differently, the area under the counterfactual is smaller than the area under the empirical distribution (if there is bunching). To correct for this, we use the same iterative procedure as described by Chetty et al. (2011).

References

Chetty, Raj (2012). "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply". *Econometrica* 80.3, 969–1018.

Chetty, Raj and Emmanuel Saez (2013). "Teaching the tax code: Earnings responses to an experiment with EITC recipients". *American Economic Journal: Applied Economics* 5.1, 1–31.

Chetty, Raj, John N Friedman, Tore Olsen, and Luigi Pistaferri (2011). "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records". *The Quarterly Journal of Economics* 126.2, 749–804.

Chetty, Raj, John N Friedman, and Emmanuel Saez (2013). "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings". *American Economic Review* 103.7, 2683–2721.

Riksrevisionen (2009). *Jobbskatteavdraget*. Report RiR 2009:20, Riksrevisionen.

Sugiyama, Masashi, Takafumi Kanamori, Taiji Suzuki, Marthinus Christoffel du Plessis, Song Liu, and Ichiro Takeuchi (2013). "Density-difference estimation". *Neural Computation* 25.10, 2734–2775.

This section has been removed to preserve the anonymity of the researchers.

23 mars 2018

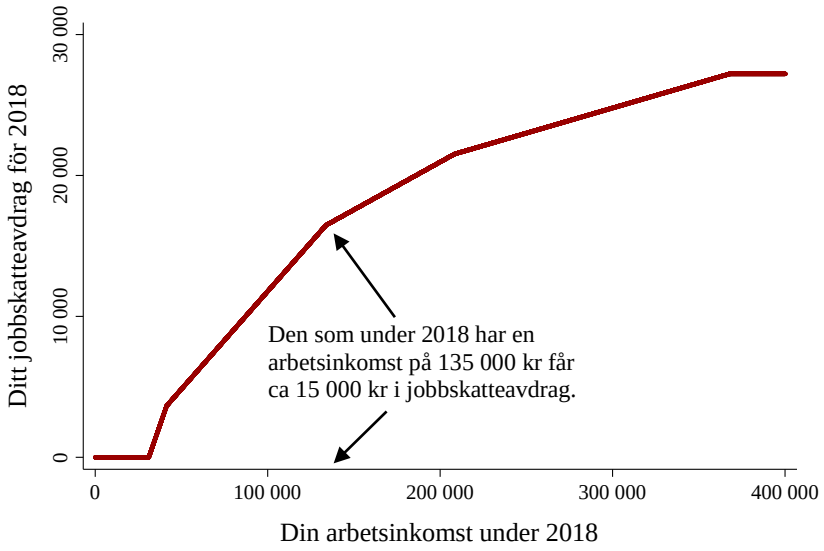
Till nyblivna föräldrar

Har ni funderat på hur fördelningen av föräldraledigheten påverkar hur stort jobbskatteavdrag ni får? Det handlar faktiskt om ganska stora summor. I ett projekt vid informerar vi nyblivna föräldrar om hur hushållskassan påverkas av att en person är ledig från arbetet under ett helt kalenderår.

I Sverige får alla som arbetar ett jobbskatteavdrag. Storleken på avdraget ökar snabbt med arbetsinkomster som under ett kalenderår uppgår till 135 000 kr. Med en sådan årslön får man ungefär 15 000 kr mer i jobbskatteavdrag jämfört med den som är ledig hela året. **För en person som tjänar 35 000 per månad är det därför mycket lönsamt att arbeta minst 4 månader under 2018.**

Storleken på jobbskatteavdraget fortsätter att öka med inkomsten även efter 135 000 kr, men då betydligt långsammare. I regel innebär det att ett par som delar lika på föräldraledigheten under ett kalenderår får ett större jobbskatteavdrag än ett par som låter den ena föräldern vara hemma hela året.

Figuren nedan visar hur storleken på jobbskatteavdraget varierar med din arbetsinkomst under 2018.[1] Det är kanske någonting ni vill ta hänsyn till när ni planerar hur mycket ni ska arbeta under resten av året?



[1] Exakt hur stort jobbskatteavdrag du får beror på dina individuella förutsättningar, men uppskattningarna vi presenterar stämmer relativt väl för alla under 65 år.