Estimating the effects of universal transfers: new ML approach and application to labor supply reaction to child benefits

Filip Premik*

University of Minnesota (PhD Candidate), and FAME|GRAPE

Abstract

This paper studies the evaluation of large-scale government programs in which violations of identification assumptions may invalidate use of standard econometric methods. Instead of relying on a potential outcomes framework, I obtain counterfactual outcomes from a general discrete choice model. The is estimated non-parametrically using Generalized Random Forest estimator. I apply the method to study the effects of introducing a universal child benefit program in Poland and show that it led to a 2–4 percentage points decrease in labor supply among eligible females resulting mainly from changes in perceived trade-offs and beliefs that discouraged labor market participation.

Keywords: program evaluation, child benefit programs, female labor supply, generalized random forest
JEL classification: I38, J22, C14, C25, C54

*filip.premik@gmail.com. The author thanks to participants of Applied Microeconomics Workshop at the University of Minnesota, Warsaw International Economic Meeting, Metody Ilościowe w Badaniach Ekonomicznych Conference for valuable comments. I gratefully acknowledge the support of the Polish National Science Centre (grant #2016/23/N/HS4/03637). The remaining errors are mine.

1 Introduction

Large scale government programs aiming at improving life situation of individuals have become an important part of public sector around the world. These programs take various forms, including direct nonequivalent transfers, or tax credits. Regardless of the form, they constitute a significant cost to the taxpayers. Therefore, policymakers are increasingly interested in understanding how such policies affect behavior of individuals in the economy to measure the effectiveness of the programs and seek directions for potential improvements.

Empirical quantification of the effects of large scale government programs often poses a challenge for economists because such programs are typically universal, which means that all individuals satisfying some general conditions participate in a program. For example, child benefit programs are often addressed to all families raising their children. As a result, two major identification problems arise. First, it is difficult find individuals who do not satisfy participation requirements but are otherwise comparable to individuals who benefit from the policy. In other words, an economist may fail to find a suitable control group whose behavior would proxy counterfactual behavior of eligible individuals had they not receive the support. Second, large scale programs are likely to produce general equilibrium effects which may confound the measurement of policy effects.

This paper introduces a simple framework to evaluate effects of large scale government programs on individual's decisions that overcomes these obstacles. Instead of relying on the presence of a suitable control group, I approximate individuals' counterfactual outcomes using a flexible choice model. Using minimal assumptions concerning the decision environment, the model allows individuals' decisions to depend on a broad range of determinants, including observed and unobserved state variables known by individuals at the decision time, and their beliefs regarding future outcomes, in an unrestricted manner. The decision model is locally identified through a set of conditional moment restrictions and is estimated nonparametrically using Generalized Random Forest estimator (Athey et al., 2019). I employ a data driven approach exploiting large amounts of information regarding individuals' socio-economic background to approximate individuals' decisions as close as possible. Datasets including such information are increasing available for the researchers evaluating the impacts of large scale government policies.

My framework shares flexibility of reduced form approaches (by conditioning on potentially a large set of state variables and not requiring functional form assumptions) and benefits from appealing interpretation of the underlying choice model, which is a typical aspect of structural models, without necessity to impose restrictive assumptions on the expectations or future evolution of state variables.

The choice model delivers a set of decision rules describing expected choice of each individual given the observed state variables, including period, eligibility status and demographics. I use the estimated decision rules to decompose changes in the outcome of interest between any two periods. There are three components of this decomposition. First component describes changes in the outcomes of interest that are a result of changes in the way individuals make their decision, holding fixed their observed characteristics in the first period. It accommodates relative changes in payoffs resulting from various choices as well as evolution of future beliefs. The intuition behind this effect is somewhat close to the standard average treatment effects analyzed in standard reduced form approaches to the program evaluation, though there is no direct mapping between these two objects. Second component measures changes resulting from adjustments in individual characteristics, holding fixed the decision rule. This effects accommodates self-selection mechanisms, which are ruled out in standard reduced form approaches to the program evaluation. Lastly, the third component is a residual term summarizing the part of variation that cannot be explained by the model.

Following dynamics of the two former components before and after the government intervention sheds light on how the program affected outcome of interest. The model provides a convenient framework that allows an economist to identify and subsequently shut down the variation attributed to the program to simulate the counterfactual paths of evolution of the outcome of interest, had the program not been introduced. In turn, verifying whether the estimated variation in residual term is statistically negligible provides a convenient model specification test.

I apply my framework to evaluate the effects of a large scale child benefit program in Poland on female labor supply. At the cost of approximately 2% GDP yearly, the program *Family 500 Plus* (henceforth: *P500*, or the *intervention*) aims to improve the situation of families upbringing kids and increase long term fertility. Starting from 2016, families upbringing two or more minors receive a monthly nonequivalent transfer of approximately 20% of a median wage per second and any further child. A feature of Polish labor market is that labor supply adjustments in hours worked play rather a minor role. Thus, I focus on changes in labor force participation among Polish women in response to the program *P500.* The decision model effectively boils down to a general binary choice setting.

Female labor force participation is shaped by dynamics of two forces, inflows to and outflows from labor force. The former includes women moving from inactivity into either an active search for employment or working. The latter is related to women who leave the labor force by either quitting their jobs or terminating search if unemployed. Building on empirical regularity suggesting that the nature of women's labor supply decisions is fundamentally different depending on whether she does or does not participate in the labor force, I estimate a model and obtain the decomposition separately for inflows and outflows. The time-series variation allows me to identify three possible effects of the program *P500.* I simulate counterfactual participation paths using time series of flows in which I shut down the variation attributable to the intervention, relying on the fact that the time series of labor force participation is a function of the flows.

The results of my analysis indicate that the program *P500* led to a decrease in female labor force participation, which was driven by discouraging activation of women outside of the labor force which occurred as a result of changes in females' trade-offs and future beliefs. There is also evidence on self-selection out of the labor force that increased the outflow rates among a subset of demographics. These direct effects have been propagating and accumulating over time, leading to 2 percentage point drop in the labor force participation rate among the eligible woman in two years and 4 percentage points after four years after introduction of the program.

The decrease in labor force participation may have affected employers who facing increased difficulty in maintaining the staff improved the working conditions. If this is true, one would expect an implied decrease in outflow rates driven by changes in economic environment. Such an outflow is confirmed by the decomposition. Hence, the program *P500* impacted the labor supply also in a less direct manner. Removing this effect in a counterfactual labor force participation path mitigates the further propagation of initial shocks resulting in 2 percentage point estimate of the total effect after four years since introduction of the program.

Most relevant to my study are papers investigating the effects of universal child benefit programs. Economic theory suggests such transfers may have detrimental effects on labor supply, particularly among women (Moffitt, 2002). The results of my paper explain mechanisms driving this regularity.

Schirle (2015); Koebel and Schirle (2016) show that Canadian Universal Child Care Benefit decreases labor supply of married women. Baker et al. (2021) provide an overview over a few reforms of Canadian child benefit system showing reduction in child poverty and no evidence on labor supply response on both extensive and intensive margin. González (2013) investigates universal child benefit program in Spain and finds a decrease in maternal labor force after childbirth. A common denominator of these studies (and many more, for a review of literature evaluating the labor supply effects of child benefits and other family-related welfare programs see Moffitt (2002) and Immervoll et al. (2007)) is the reduced form approach taken as a tool to describe changes in labor supply as a result of a benefit program. In my paper, I explicitly model women's decision rule which allows me to avoid the restrictive assumptions regarding data generating process from the Roy's potential outcomes model. In addition, my results have an appealing interpretation of a micro founded model.

Another strand of the literature on evaluating impacts of large child support programs uses structural modeling as a tool to answer the research questions. Blundell et al. (2000) study Working Families' Tax Credit program in UKUsing a structural model of labor supply with childcare costs they showed increased labor force participation as a response to the program. Stephens Jr and Unayama (2015) investigate the effects of Japanese child benefit system on household wealth accumulation. A fully specified structural model requires a number of assumptions regarding agent expectations and law of motion of state variables. There are also computational constraints limiting the number of state variables. In turn, my simple choice model framework does not impose strong assumptions on the structure of decision problem. Moreover, it uses machine learning techniques to which allow to tractably condition woman's decisions on a large number of observed state variables.

My study adds to the discussion concerning effects of the program *p500* on various sectors of Polish economy. Magda et al. (2018) use difference-in-difference approach to provide an early evaluation of the effects on the female labor supply. They find treatment effects implying 2-3 percentage points drop in the female labor force supply as a result of introducing the program. My estimates are consistent with these findings. This is because within a short time after the program has been introduced, its main impacts came through the channel of changes in economic environment, which can be captured by the traditional methods. However, my paper extends this study by applying a method that abstracts from parallel outcomes assumptions and allows me to simulate counterfactual paths of labor force participation as if the program has not been introduced to study longer term effects of the program. Myck (2016) and Myck and Trzciński (2019) utilize a microsimulation model to evaluate ex-ante potential effects of *P500*. Their model relies on a discrete choice model of labor supply in which a household with two adults chooses labor supply for both of them. These authors use household budget survey data on pre-treatment period to simulate the effects of introducing the program. Their results indicate a drop in the labor force supply of roughly 150 thousand women, or approx. 2% of economically active women, which again is similar to my findings focusing on early stage program evaluation. The simulations are obtained in the short-run and in partial equilibrium, that is they ignore potential changes in the wage structure and working conditions. My approach allows for implicit consideration of these effects. Finally, Paradowski et al. (2020) applies difference-in-difference framework to show a substantial reduction in poverty and inequality indices among Polish household after introduction of the program.

This paper applies machine learning methods to study labor force participation. In a related setting, Cengiz et al. (2021) use similar tools to predict which individuals are likely to be affected by the minimum wage reforms. Sigurdsson (2019) uses forest based estimators in studying labor supply responses to temporary variation in wages, exploiting exogenous variation in a tax cut. Angrist and Frandsen (2019) study performance of machine learning algorithms in causal studies, illustrating it with an example concerning effects of college characteristics on wages. My study differs from these papers by using the machine learning algorithm to estimate a flexible structural choice model and then simulate counterfactual decisions.

The remainder of the paper is structured as follows. Section 2 describes the policy design and data. Section 3 presents the model and explains its use in program evaluation exercise. Section 4 introduces details of the estimation routine. Empirical results are discussed in section 5. Section 6 measures the effects of the program on aggregate labor force participation. Section 7 concludes.

2 Program Design and Data

2.1 The Program Family 500Plus

The program Family 500Plus (henceforth *P500*) provides a universal child benefit for each second and further child aged 0-17 in a household. In addition, there was an income threshold for eligibility of the first child until 2019, when the program has been extended to all children in the household. The benefit comes as a monthly non-equivalent payment of roughly 20-25% of the net average wage (PLN500, or approximately US$130) per eligible child. The program's main goals are to improve the financial well-being of families upbringing children and stimulate fertility in the long run. The program constitutes a significant financial effort to the government budget, at the cost of approximately 1.5-2% of GDP yearly.

In order to obtain the benefit, an eligible household is supposed to apply for it at the local administration. Table 1 indicates high participation rates among the eligible. Approximately 95% of households with two or more children below 18 are obtaining the benefit. Extending eligibility for the first child in 2019 shifted the participation rate among single-child households to a comparable level.

The program has been announced in the first quarter of 2016, and the first payments arrived in the next quarter. However, some regions experienced delays in the distribution of the initial payments, which have been eventually distributed in the second half of 2016. For this reason, I divide the timeline into three general periods. The pre-intervention period includes all years up to 2015. At that time, any influence of the program can be ruled out. I refer to 2016 as a transition year because that was a period in which the program has been announced and gradually introduced. 2017 and subsequent years belong to the post-intervention period.

2.2 Data Sources

Data comes from the Labor Force Survey conducted by Central Statistical Office in Poland. In each quarter, approximately 30,000 households are interviewed using a detailed questionnaire concerning their labor market outcomes. The sample is representative for the population and constitutes a rotating panel. Each household is interviewed 4 times. The first two waves are collected in two consecutive quarters. The third wave is collected after a year after the first, and the fourth follows in the quarter right after the third. For example, if a household enters the sample in 2016Q1, then it is re-interviewed in 2016Q2, 2017Q1, and 2017Q2. In each wave of the survey, the responses of all adult members of the household are recorded.

I restrict my attention to the subsample of females of age between 20 and 60. Typically, individuals in Poland leave the schooling system in the 19th year after birth. The lower threshold allows me to abstract from schooling and birth date effects. In turn, Polish women are eligible for retirement at the age of 60, which motivates the choice of the upper threshold. Figure 1 illustrates this reasoning. Most of the economic activity and child upbringing are performed by women not younger than 20 and not older than 60.

The data does not allow me to verify the eligibility of the first child in a household. Therefore, I focus on the labor force participation decisions of females that do have children below 18 (shortly: childless) who are not eligible, and females with two or more children below 18 (shortly: $\geq$2 children) who are eligible to receive the benefit at least for one child. In addition, since nearly all eligible women actually participate, I focus on the distinction between eligible and ineligible females. This is motivated by the fact that it is impossible to determine which of the eligible individuals observed before the introduction of the program *P500* would actually participate given the short panel dimension in my data.

2.3 Labor Force Flows

A woman is a member of the labor force in a given period if she works or is actively searching for a job. Labor force participation is determined by labor market flows. Given the structure of my data, I measure inflows in a given period as the share of females who are in the labor force and were not there in the previous period. Analogously, I define outflows in a given period as the share of women who are not currently in the labor force but were there in the previous period. Figure 2 depicts the quarterly time series of the flows. Differences in their dynamics suggest that the inflows are driven by other economic processes than the outflows.

Before the *P500* was introduced, the inflows for both groups followed roughly the same trend, which mostly stabilized after 2016, as shown in figure 2. The introduction of *P500* has coincided with a significant drop in the inflows among eligible women. The average inflows among the eligible females decreased by 2.7 percentage points after the introduction of the program, as shown in table 2. The change in inflows among the ineligible is also negative but an order of magnitude smaller. Changes in the outflows among eligible and ineligible females are of opposite signs but low magnitude. However, the measure of participating females is much larger than inactive (as shown on figure 3). Therefore, even relatively smaller changes in the outflows may translate into significant shocks to the aggregate labor supply.

Figure 3 presents trends in levels of labor force participation. The eligible women participation rates remain roughly constant in the pre-intervention period, drop by approximately 3 percentage points in 2016-2017, and stabilize at the end of the observation window. Participation rates among females without children below 18 are steadily increasing across the sample window.

2.4 Predictors of Female's Labor Market Decisions

The data comes with a detailed description of the household's socio-economic background and labor market activities. This information is crucial for predicting female's labor market decisions and subsequently evaluating the effects of the program *P500*. I classify available variables into several groups.

First, I consider a set of household-level variables describing the household composition, number of earners, basic demographics, and month in which the interview has been taken - a particularly important covariate that controls for seasonal variation in labor force participation. Second, I consider a range of demographic characteristics of the woman, including age, marital status, dummies for their spouse's, parents', and children's presence in the household. This group contains also the number of children below 18, which is a fundamental variable driving the program eligibility. The third group describes female's employment status. It provides a comprehensive description of the job (for the employed), reasons for not having a job (for the jobless), and past working experience. The fourth group summarizes a woman's job search effort (including intensity, duration, and type of searched job) and the fifth describes her educational background.

In this paper, I focus on women's decisions which are likely to depend also on the outcomes of other members of their household. For example, they typically share responsibility for financial well-being with their spouses. To account for that, I consider another group of variables that describe spouse's outcomes, conditional on their presence in the household. Specifically, I take into account spouse's employment situation, job search, and educational background. Female's decisions may also depend on parental support. Guided by this fact, I consider a subset of variables describing mother's and father's sources of income and subjective evaluation of their labor market status.

Table **??** provides a brief summary of available predictors of female's labor market decisions, and table 4 provides a detailed description of these variables. In total, I take a set of 379 observed state variables to the estimation.

3 The Model

In this section, I present a general discrete choice model of woman's decision of whether to be a part of the labor force.

3.1 Decision Rule

Time is discrete and indexed by t. In period t, a woman chooses $y \in \{0, 1\}$ conditionally a finite set of state variables that are known by her at the time the decision is taken. In my application, I model inflows into and outflows from the labor force separately. In analyzing inflows, $y = 1$ describes woman's decision to enter the labor force. In analyzing outflows, $y = 1$ denotes her decision to leave the labor force.

There are two distinct types of state variables. First, the decision is affected by a set of state variables s that is observed by both woman and econometrician. These variables are henceforth called the *observed* state variables. Second, the woman exploits information that is not available for the econometrician, denoted by ε, which I refer to as unobserved state variables. The unobserved state variables are drawn from a joint distribution $F_t(\varepsilon|s)$, which may depend on the observed state variables s and time t.

Finally, the decision is also affected by the set of beliefs about the evolution of state variables in the future, denoted by $G_t(\varepsilon', s'|\varepsilon, s)$. Both F_t and G_t are assumed to have finite first moments.

Woman's payoff function in period t depends on her choice, values of the state variables and beliefs:

$$v_t(y, s, \varepsilon; G_t) \equiv v_t(y, s, \varepsilon)$$

where the equality holds because G_t is defined as a function of s and ε. I assume that the payoff function v_t is measurable. This specification allows for an arbitrary scheme of discounting future outcomes and beliefs. In particular, it is not necessary to assume that the decision-maker has rational expectations.

The value of the decision problem at time t can be written as:

$$V_t(s,\varepsilon) = \max_{y\in\{0,1\}} \Big\{ v_t(y,s,\varepsilon) \Big\} \tag{1}$$

The optimal policy function is:

$$y_t(s,\varepsilon) = \mathbf{1}\Big[v_t(1,s,\varepsilon) \geq v_t(0,s,\varepsilon) \Big] \tag{2}$$

Using assumptions of the model, I derive the probability that the woman chooses $y = 1$ conditionally on the set of observed state variables by integrating out the unobserved state variables:

$$\begin{aligned} \mathrm{P}[y=1|s,t] = \int\limits_{\varepsilon: v_t(1,s,\varepsilon)-v_t(0,s,\varepsilon)\geq 0} \mathrm{d}F_t(\varepsilon|s) \equiv \\ \equiv \varrho_t(s) \end{aligned} \tag{3}$$

$\varrho_t(s)$ is a conditional choice probability and describes woman's decision rule given s and plays a fundamental role in my analysis. The estimated decision rules serve for generating counterfactual outcomes describing woman's choices at various t and s. I use these counterfactual outcomes to evaluate the impacts of the program *P500*.

3.2 Decomposing Differences in Choice Probability

Evaluating the effects of the child benefit program is essentially asking how a woman changed her labor force participation in a response to the benefits. To simplify the exposition, suppose there are only two periods: $t \in \{0, 1\}$ denoting pre- and post-intervention periods respectively. In general, the derived decomposition holds for any pair of consecutive periods.

The expected change in woman's labor force participation decisions between periods 1 and 0 can be decomposed into two elements:

$$\begin{aligned} \mathbb{E}_{\varepsilon}\big[y_1(s_1, \varepsilon_1) - y_0(s_0, \varepsilon_0)\big] &= \varrho_1(s_1) - \varrho_0(s_0) = \\ &= \underbrace{\varrho_1(s_0) - \varrho_0(s_0)}_{\beta(s_0)} + \underbrace{\varrho_1(s_1) - \varrho_1(s_0)}_{\gamma_1(s_1, s_0)} \end{aligned} \tag{4}$$

First, $\beta(s_0)$ describes changes in woman's conditional choice probabilities between pre- and post-intervention periods holding fixed the pre-intervention vector of observed state variables. This parameter summarizes the inter-temporal changes in the functional form of the within-period payoff functions, including the effects of changes in individual beliefs regarding the future. For example, consider a woman who is unemployed but actively searches for a job in the pre-intervention period, and has two or more children below 18 (that means is eligible for receiving the benefit). In the post-intervention period, the additional income from the program *p500* may magnify the significance of disutility from a potentially costly job search process in the per-period payoff, because the salary income becomes less necessary to sustain the household. If the woman expects the transfers to arrive regularly in the future, her probability of continuing job search would drop significantly without a change in any of the state variables s.

The introduction of the *P500* has followed closely its announcement, so it is unlikely that individuals would be able to adjust their pre-intervention characteristics in anticipation of receiving the transfer. That makes s_0 *exogenous* in the standard reduced form language. The standard program evaluation literature studies changes in the decision function conditioned in exogenous pre-intervention variation. Typically, researchers estimate the average treatment effects derived from Roy's potential outcomes framework (Roy, 1951). Although my approach lies within a dynamic discrete choice framework and there is no direct mapping between $\beta(s_0)$ and any of the Roy-style average treatment effects, the intuition behind both effects is similar. To emphasize this similarity, I refer to $\beta(s_0)$ as the *treatment* parameter.

Second, $\gamma_1(s_1, s_0)$ describes changes in woman's conditional choice probabilities between pre- and post-intervention periods holding fixed the decision rule. This parameter describes changes resulting from the inter-temporal shift in individual characteristics, including outcomes of self-selection mechanisms. For example, consider an ineligible woman with one child. Suppose she derives high utility from staying out of labor force and raising her child. If she believes that the benefit program will be sustained in a long term, she may decide to give birth to another child, self-selecting to the program. The benefit would provide additional financial means that would lower her probability of being in the labor force without any change to the functional form of her payoff function.

The realizations of observed state variables in the post-intervention period may be affected by the intervention itself. By adjusting elements of s_1, a woman may increase her probability of receiving a benefit, which in turn affects her labor force participation. In order to reflect this self-selection mechanism, I label $\gamma_1(s_1, s_0)$ the *selection* parameter. Self-selection effects are not identified within the standard approaches of policy evaluation. In my framework, it is straightforward to measure their impact on woman's choice.

3.3 Sample Decomposition and Specification Test

The decomposition given by equation (4) is complete if the researcher can observe the true optimal policy functions $\varrho_t(\cdot)$. In a real-world setting, this object has to be estimated from the data. In finite samples, there may not be enough variation to fully average out the impact of unobserved state variables. Define the resulting error as:

$$\hat{\xi}(s_1, s_0) \equiv \big(\overline{y}_1(s_1) - \overline{y}_0(s_0)\big) - \big(\hat{\varrho}_1(s_1) - \hat{\varrho}_0(s_0)\big) \tag{5}$$

where $\overline{y}_t(s_t)$ is the sample average of the outcome variable at time t among individuals with realization of state variables s_t. Since this error refers to the variation in the unobservables, I refer to it as an *residual* parameter.

Having defined the residual parameter, I propose the exact decomposition of sample averages:

$$\begin{aligned} \overline{y}_1(s_1) - \overline{y}_0(s_0) &= \hat{\varrho}_1(s_1) - \hat{\varrho}_0(s_0) + \hat{\xi}(s_1, s_0) = \\ &= \underbrace{\hat{\beta}(s_0)}_{\text{treatment}} + \underbrace{\hat{\gamma}_1(s_1, s_0)}_{\text{selection}} + \underbrace{\hat{\xi}(s_1, s_0)}_{\text{residual}} \end{aligned} \tag{6}$$

If the model specification reflects the true data-generating process, the variation in observed choices resulting from $\hat{\xi}(s_1, s_0)$ should be negligible. That suggests a simple specification test with the null hypothesis $H_0 : \hat{\xi}(s_1, s_0) = 0$.

3.4 Remarks

So far I have derived the decomposition for each pair of the observed state variable pairs (s_1, s_0). In practice, it may be convenient to analyze the effects at higher level of aggregation. Both equations (4) and (6) are easily averaged over the dimensions of s. In most of the empirical part results, I average the estimated effects over nearly all variables in s, distinguishing solely between two categories for the number of woman's children below 18 driving the program eligibility. The specification test follows the same intuition on the aggregated level.

The decomposition can be obtained for any two periods in the data. In particular, having a few pre- and post-intervention periods one may construct a time series of the *treatment*, *selection* and *residual* effects. This approach allows for uncovering potential long-term trends in the data. In the empirical part, I estimate quarterly time series of the decomposition to identify changes in the women's labor supply that are likely attributed to the intervention.

My approach permits also evaluating the validity of identification in the earlier studies on *P500* which relies on the standard DID framework to identify the causal parameters of the Roy's potential outcomes framework. The potential outcomes can be viewed as decision rules with and without the presence of the treatment. To uncover the causal effect, one needs to assume that both the treated and the control use the same decision rule in absence of the treatment. Moreover, they need to remain unchanged over time up to a common trend. In my framework, I directly estimate the decision rule allowing for arbitrary dependence on observed state variables, including program eligibility (or participation) and time. This opens up a possibility to verify whether classical DID assumptions hold in a given context, including SUTVA (Lechner et al., 2011). As a natural extension, my approach permits evaluating the validity of identification strategies pursued in the earlier studies on the program *P500*, for example Magda et al. (2018).

I treat woman's problem as dynamic, allowing for arbitrary forms of beliefs and transitions. My framework relies on the assumption that conditionally on s one can integrate out all of the unobserved heterogeneity, though the knowledge about the functional form of its distribution is not required. This is a standard practice in structural modeling, where usually some distributional assumptions are required. On the contrary, in reduced-form approaches, one may abstract from specifying the distribution of unobserved heterogeneity, yet it has to be independent from s. In my approach, this independence assumption can be relaxed.

4 Estimation

The conditional choice probabilities $\varrho_t(s)$ reflecting woman's decision rules at different values of state variables and time are key primitives of the model that need to be estimated from the data. For each t and s, they are point-identified through a conditional moment restriction:

$$\mathbb{E}\Big[y - \varrho | t, s\Big] = 0 \tag{7}$$

Estimation based on conditional moment restriction is often subject to the curse of dimensionality, which effectively limits the analysis to very few state variables. In this paper, I estimate conditional choice probability function $\varrho_t(\cdot)$ using the Generalized Random Forest estimator developed by Athey et al. (2019, GRF) that allows me to condition women's decisions on a large set of state variables without facing the curse of dimensionality. Although GRF relies on a machine learning algorithm, it is shown to produce consistent and asymptotically normal estimates of the conditional choice probabilities. This appealing feature enables statistical inference, making GRF particularly suitable for applications in applied economics.

Appendix 7 contains an intuitive description of the mechanics behind the GRF estimator.

4.1 Estimating Conditional Choice Probabilities

Motivated by the findings from the descriptive analysis, I estimate separate models of inflows and outflows. In a model of the inflows, I estimate the probability of a woman being in the labor force in questionnaire waves 2 or 4, conditionally on being out of the labor force in questionnaire waves 1 or 3, respectively. Analogously, in a model of the outflows, I estimate the probability of a woman being out of the labor force in questionnaire waves 2 or 4, conditionally on being in the labor force in questionnaire waves 1 or 3, respectively.

In all model specifications, a woman conditions her choice on a set of observed state variables

s, which cannot be a result of the decision. To account for this, I exploit the rotating panel structure of the survey. I focus on inter-quarter changes in the labor force participation decisions. Specifically, I condition women's choices regarding labor force participation observed in questionnaire waves 2 and 4 on the responses given in waves 1 and 3, respectively. Quarter-lagged state variables cannot be a result of the decision and are the most relevant source of variation relevant for the women's choices available in my data.

The empirical strategy relies on uncovering the underlying female's decision rule regarding labor force participation. In the real world, a decision regarding labor force participation usually takes into account a series of various factors describing the current life situation of a woman. In order to approximate the optimal policy as closely as possible, I consider a high dimensional set of observed state variables. The main idea is to avoid making arbitrary choices regarding which variables available in the questionnaire to include in the model.

I do not model explicitly the joint decisions in the household. In the empirical part, I condition women's decisions on the characteristics of other household members in a previous period. This approach does not preclude joint decisions in the data-generating process, because all of the conditioning variables refer to the past.

In theory, the GRF framework can incorporate any non-linearity pattern in the way the observed state variables affect the outcome variable, at the cost of quickly increasing forest size and resulting computational complexity. With a forest large enough, it is sufficient to estimate one model of female's labor force participation that pools together data from all periods, previous quarter employment, and program eligibility statuses. In practice, this is not a convenient approach due to the high computational complexity of the algorithm. Therefore, I place a-priori restrictions to help the algorithm perform well without the necessity of growing a large-sized forest. I estimate a separate forest for each combination of period, treatment eligibility (females without children below 18 and females with at least two children below 18), and labor force status (in or out) in the preceding quarter.

The GRF routine produces estimates of conditional choice probabilities[1]. Given program eligibility and initial labor force status, I obtain $\varrho_t(s)$ for any period t and vector of observed state characteristics s. I aggregate the estimated conditional choice probabilities by averaging over all dimensions in s using survey population weights. I obtain counterfactual conditional choice probabilities by using the estimated model in period t to predict the outcomes using observations from period $s \neq t$.

The parameters β, γ and ξ are functions of the counterfactual conditional choice probabilities obtained for the same individuals. That makes it difficult to derive appropriate standard errors. However, as conditional choice probabilities obtained from the GRF procedure are asymptotically normal, bootstrap techniques are expected to perform well. For each period t, I repeatedly draw a sample of N_t individuals with replacement and estimate the decision model. A distribution of parameters obtained by repeating this procedure is expected to converge to the true sampling distribution of the effects of interests as the number of repetitions goes large. All statistical inference performed in the empirical part of this paper is based on 150 bootstrap replications per decision model.

I estimate the decision models in two separate fashions. I start with a more general setting in which I distinguish only two periods: pre-intervention (2014-2015; $t = 0$) and post-intervention (2017-2019; $t = 1$) and refer to it as a *pre-post* model. Next, I turn to investigate quarterly dynamics in parameters of decomposition (6). In order to control for seasonal variation in labor force flows and increase estimation precision through increased sample size, I estimate a series of models in a quarterly rolling observation window. That means a decision model for a quarter q is estimated using data on quarters $q - 3$ to q. I refer to this approach as *quarterly* models.

[1] I use `R` package `grf` developed by Tibshirani et al. (2020).

5 Results

I use pre-post model estimates to summarize the overall changes in parameters of the decomposition (6) between pre- and post-intervention periods and to study heterogeneous impacts among women with different demographics. Pre-post models provide also a description of the most important predictors of the labor force flows and changes in their significance after the intervention. In turn, the quarterly models give, more precise insights into the timing of adjustments in labor force flows and help to identify the variation resulting from the introduction of the program.

5.1 Pre-Post Effects

I summarize the main changes in the female labor market flows using the pre-post framework. Components of decomposition (6) are estimated separately for each cell defined by eligibility (women with two or more children below 18 and women without children below 18) and flow (inflows and outflows) indicators. I decompose changes in the observed flow rates into three elements, describing variation resulting from changes in women's decision rule (*treatment* parameter), their observed characteristics (*selection* parameter), and residual factors (*residual* parameter). I also report differences between these estimates between eligible and ineligible females.

Table 5 presents estimated parameters in the pre-post framework. A key driving force affecting changes in the inflows among eligible women is the *treatment* channel. The pre-intervention population of females with two or more children below 18 decrease their inflow rate in the post-intervention period by 2.14 percentage points solely as a result of changes in their decision rule, that is, the functional form of per-period payoff function and beliefs regarding the future evolution of state variables. This channel does not affect inflows to the labor force among women without children below 18, as indicated by the low and statistically insignificant parameter estimate.

Changes in women's observed state variables described by the *selection* channel affected mostly the outflows. The selection parameter estimates indicate over 1 percentage point drop in the post-intervention period outflows among both eligible and ineligible females resulting from changes in their observed characteristics.

Pre-post models of the flows pass the specification test. The estimated *residual* parameters are statistically zero, which indicates that the variation in women's choices driven by unobserved state variables has been successfully integrated out. The ξs can be also interpreted in terms of model goodness of fit. By definition, they are residuals between the observed changes in inflow rates and changes predicted by the model. Low and statistically insignificant estimates of the *residual* parameters (table 5) imply that the model explains the data satisfactorily well.

5.2 Quarterly Effects

An important question is which of the described changes in labor force flows can be attributed to the intervention. To answer this question, I analyze the dynamics in parameters based on the *quarterly* approach. Figure 4 presents the time series of estimates.

The *pre-post* estimation reveals an approximately 2 percentage point drop in the average inflow rate among the eligible women driven by changes in women's decision rule after the introduction of the program (table 5). The *quarterly* estimates show that all of these changes occur within a few quarters following the announcement and introduction of the program. I interpret these changes as a direct impact of the program on women's decision rules. They show that the economic impact of the program has been quickly internalized. The arrival of benefits discouraged labor activization among women with two or more children, shifting their inflow rates down already in 2016. In turn, the introduction of the program is not likely to affect the decision rules of ineligible females. This presumption is confirmed by the fact that the *treatment* parameter estimates for the group of ineligible females indicate no changes in inflows driven by changes in the decision rule from the second half of 2015 till the end of my sample.

Estimates of *selection* parameters in the decomposition of the inflows show that neither eligible nor ineligible women changed their labor market activization rates due to changes in their observed characteristics. This result allows me to rule out the hypothesis of self-selection to the program on the inflows margin.

The *treatment* parameters in outflows decomposition diverge for the first time in mid-2018 after an increase in parameters for ineligible females. It is unlikely that this effect is related to the program *P500*. Economic theory does not predict that child benefits would change the economic environment or individual beliefs among ineligible females that would support increased outflows from the labor force. First, there is no direct effect of the program on women who do not obtain the transfers. Second, the indirect effects would rather be associated with a decrease in outflows. As shown above, the benefits discouraged inflows to the labor market among women with 2 or more children in 2016. The implied shrinkage of the labor supply is likely to make it more difficult for employers to keep the current staff and hire new suitable employees. As a result, employers are likely to improve the job conditions for existing employees or new hires, which in turn would limit the outflows. This is exactly what happens afterward. The *treatment* parameters in the outflows decomposition are decreasing in both groups between mid-2018 and 2019 in a nearly parallel fashion. The mechanism described above is consistent with the *treatment* channel, as the improved employment conditions affect the economic environment and future beliefs accommodated in women's decision rules. There should not be differences in the dynamics of this effect between the eligible and ineligible women, which is confirmed by my results. Therefore, I conclude that the drop in *treatment* parameters in the outflow rate changes decomposition at the end of the sample is likely to be a result of the *P500*. Given the parallel nature of changes, they cannot be captured by standard program evaluation methods.

The estimated *selection* parameters in the decomposition of outflows are roughly constant throughout the sample for both eligible and ineligible women. However, the time series of estimates for women with two or more children has two spikes above the trend after the introduction of the program *P500*. These spikes are likely to reflect the impacts of the intervention. Initially, a measure of women self-selected themselves out of the labor force after having received their first payments in the second quarter of 2016. Intuitively, these women are likely to be experiencing strong disutility from work and were kept on the labor market by financial constraints before receiving the benefits. Next, in the fourth quarter of 2017, there came another increase in outflows among the eligible triggered by the *selection* channel. The lagged reaction to the program introduction is intuitive. It takes time to adjust some characteristics. For instance, periods of notice make the process of quitting a job longer. The quitting process itself may encompass a gradual decrease in working hours. Moreover, some females may have postponed their quitting decision in order to sustain increased income of wage and the benefit for some periods to accumulate funds or repay debts. These factors are likely to be captured as changes in the observed state variables and therefore contribute to the *selection* channel. Given the fact that these effects are driven by changes in post-intervention observed state variables, they cannot be captured by standard program evaluation methods.

The *residual* parameters in both decompositions and among both eligible and ineligible women are statistically insignificant throughout the whole sample, supporting my choice of specification. The model performs well also at the quarterly level.

5.3 Heterogeneous Effects

The elasticity of labor supply with respect to the benefit is likely to vary with observed state variables. Intuitively, women with high earnings or wealth would not react strongly to an additional income of 20% of the median wage. In turn, the benefit may be a significant job search discouragement for women with low levels of achieved education. The first to provide evidence on heterogeneous impacts of the program *P500* were Magda et al. (2018). Following their intuition, I decompose changes in inflows and outflows on subsamples generated by the value of chosen observed state variables. I consider woman's education level, size of the city she lives in, her age and marital status, number of eligible children, and age of the youngest child. The differences in estimated parameters between eligible and ineligible women are presented in table 6. The restricted subsamples of eligible women in the last two categories are compared to the pooled sample of all ineligible females.

My results extend Magda et al. (2018) findings in terms of the direction of the effects among demographics. Consistently with the aggregate effects, I observe significant variation in *treatment* parameters in inflows decomposition and *selection* channel for the outflows. The *treatment* channel consistently leads to higher decreases in inflows among the eligible females. However, the magnitude of these effects varies. The program *P500* discouraged most strongly women with two or more children in large cities and those that are divorced or have never been married. Both results are intuitive. First, female labor market activity is significantly higher in the largest cities in Poland. Since the cities tend to offer more and better job opportunities, it is easier to find a job. That results in larger pre-intervention inflows rates, generating a higher base for drops as a result of the program. Second, women without financial support contributed by their spouses face higher pressure for their own income. This pressure is weakened by the benefits, leading to stronger effects in subgroups defined by this demographic. In turn, changes in the decision rule induced by the program have the weakest impact on the inflows among women raising more than four children below 18, women raising infants and toddlers, and higher age categories (40-49). The parental duties among the first two groups tend to require more effort. Eligible women that belong to these subpopulations are more likely to opt-out of labor force participation regardless of the benefit program.

The *selection* mechanisms in shaping the dynamics in outflows are more heterogeneous. The strongest effect was observed among females whose youngest child is below three. Their outflows increase by nearly 1.5 percentage points due to changes in their observed characteristics compared to the analogous change in the outflow rate among the ineligible. This result is in line with economic intuition because, for mothers of toddlers, the trade-off between work and home duties is the most pronounced. Changes in women's observed characteristics increased relatively the eligible women's outflows also among subpopulations with lower levels of achieved education, living in smaller cities and in higher age categories. In turn, changes driven by the *selection* channel decrease the outflow rate of women with two or more children below 18 relative to women without children among the youngest and divorced females. In this case, the benefits probably supported daycare payments, enabling eligible women to sustain their jobs.

5.4 Important Predictors

Previous sections indicate that changes in the observed female characteristics induce changes in their labor market flows and consequently labor force participation. In this section, I investigate which of the observed state variables are the strongest predictors of labor market flows and how this classification changes between the pre-intervention period and the post-intervention period.

The random forest algorithm provides a simple framework to evaluate the predictive power of particular covariates in explaining the outcome by comparing the *split significance* measure across observed state variables. It summarizes the intensity with which the algorithm exploited information in each covariate to predict the outcome variable. The construction of the measure is described in appendix 7.

For each combination of eligibility status and type of flow, I present ten most important variables according to the four measures based on the split significance: most important predictors for the flows in pre- and post-intervention periods, and predictors whose importance increased and decreased the most. The split significance decreases very fast in the first few covariates, which is why analysis of the top ten variables is sufficient to point out the most important features. The results are presented in tables 7 - 10.

The algorithm choice of the strongest predictors for the labor force flows is consistent with economic theory and ad-hoc choices by researchers in the empirical literature. Among others, work experience, level of education, and age appear consistently as the main predictors in most of the specifications. Their importance does not change significantly between the pre- and post-intervention periods. Similarly, the length of unemployment affects the probability of entering the labor force conditional on being out of the labor force a quarter before, and the current job description affects the probability of moving out of the labor force among initially employed.

Changes in split significance measure reveal some effects of the program *P500*. The variable indicating *family duties as subjective evaluation of current labor status* shows up among the top 10 predictors to increase their importance after the intervention, both in predicting inflows and outflows among women with two or more children. Moreover, the variable indicating *benefits as the main source of income* observes the largest increase among predictors of outflows among the eligible women. Notably, these variables do not show up as important predictors among ineligible women neither in the pre-intervention nor in the post-intervention period. This result confirms that the benefits play important role in shaping eligible females' decisions in the post-intervention period[2].

Changes in the classification of the most important predictors of flows reveal also other interesting impacts of the program. In predicting inflows among women with two or more children, three spousal characteristics observe a significant increase in split significance measure: *monthly wage*, *declared willingness to work more in order to earn more*, and *learned profession*. All of these variables refer to the actual and potential level of income that the

[2] *Necessity to provide care as a reason to quit the job* experiences the largest drop in split significance measure in predicting inflows among the eligible women after the introduction of the *P500*, and the highest increase among the ineligible woman respectively. To explain this seemingly counter-intuitive result note that this variable refers mainly to care provision to the sick and elderly.

spouse contributes to the household account. That suggests the benefits do not support women's independence in labor decisions. In contrast, it ties their decisions more closely to their spouse's income potential. Notably, these state variables do not play an important role and do not observe increased importance in predicting inflows among ineligible women.

Finally, the variables with the strongest increase in the split significance measure in predicting outflows both among the eligible and ineligible are features related to job safety, including *permanent horizon of employment* (equivalent to tenure), and *public institutions as employers*. This result is consistent with the hypothesis which postulates that employment safety contributes to the decrease in outflows through the *treatment* channel in 2018-2019.

6 Measuring the Effect on Aggregate Labor Force Participation

In the previous sections, I distinguish three potential channels of how the program *P500* may have impacted the labor market flows. First, the benefits may have changed the relative profitability of costly job search, discouraging labor market activization among eligible women outside of the labor force. Most of these changes occurred in quarters directly following the program introduction. Second, eligible women are likely to self-select themselves out of the labor market. These mechanisms vary among women, affecting only a range of demographics. Third, as it became more difficult to keep the staff and hire new suitable workers, employers may have improved working conditions for the already employed. This in turn is likely to be the driving force of the decreasing trend in *treatment* parameters for outflows for both the eligible and the ineligible women at the end of the sample. All of the effects mentioned above cause shifts in levels of labor market flows that interact and accumulate each period affecting future labor market participation rates.

In this section, I investigate the impact of the *P500* on the aggregate labor supply. The law of motion of labor force participation y_t is given by:

$$\mathrm{P}[y_t = 1] = \underbrace{\mathrm{P}[y_t = 1|y_{t-1} = 0]}_{\text{inflow rate}} \cdot \mathrm{P}[y_{t-1} = 0] + \underbrace{\mathrm{P}[y_t = 1|y_{t-1} = 1]}_{\text{(negative) outflow rate}} \cdot \mathrm{P}[y_{t-1} = 1] \tag{8}$$

The flow rates can be further decomposed into:

$$\mathrm{P}[y_t = 1|y_{t-1} = y] = \underbrace{\mathrm{P}[y_{t-1} = 1|y_{t-2} = y]}_{\text{lagged flow rate}} + \underbrace{\mathrm{P}[y_t = 1|y_{t-1} = y] - \mathrm{P}[y_{t-1} = 1|y_{t-2} = y]}_{\text{change in flow rate}} \tag{9}$$

for $y \in \{0, 1\}$. Given the decomposition (6), the change in flows can be written as:

$$\text{change in flow rates}_{t,t-1} = \beta^{flow}(s_{t-1}) + \gamma_1^{flow}(s_t, s_{t-1}) + \xi^{flow}(s_t, s_{t-1}) \tag{10}$$

The parameters in the decomposition of changes in flow rates are already estimated, and previous sections indicate which part of their variation may be attributed to the program. To evaluate the effects of *P500* on aggregate labor force participation, I simulate paths of labor force participation by shutting down the variation in flow changes that is related to the intervention in these channels. I focus on the quarterly approach to modeling and use population weights to obtain estimates at the aggregate level.

First, I eliminate the *treatment* channel in predicting inflows among the eligible females by assuming that $\beta(s_{t-1}^{eligible})$ takes values of $\beta(s_{t-1}^{ineligible})$ in periods 2016Q2-2017Q1 (channel (a)). Second, I turn off the *selection* channel in outflow rate change decomposition among the eligible by setting the respective parameters to their counterparts estimated in the model for ineligible females in periods 2016Q2 and 2017Q4 (channel (b)). Third, I model a situation in which the drop in *treatment* parameters in the outflows decomposition does not occur. Specifically, I set both βs in periods 2018Q2-2019Q2 to their averages over the preceding quarters after the introduction of the program, that is 2016Q2-2017Q4 (channel (c)). Only the impact of the channel (a) can be captured using the standard program evaluation methods, as it focuses on changes conditional on the pre-intervention set of state variables.

Channels (a) and (b) describe a direct impact of the program on labor force outcomes, and so they do not predict changes in the behavior of ineligible women. Channel (c) is related to indirect influence, as it concerns responses to direct changes. The likelihood that channel (a) is caused by the program is the largest, as it has an intuitive direction, occurred soon after the introduction, and was large enough not to be a reflection of sample error. In turn, channel (c) requires most assumptions regarding the market environment and agent reactions that are not directly a part of my model. I proceed with three simulation scenarios. In Scenario (1) I analyze what would happen to the labor force participation rates if only the effect (a) can be attributed to the program. In Scenario (2), I add the variation resulting from effect (b). Finally, in Scenario (3) I assume that all the effects are attributable to the intervention.

Table 11 shows the effects of the *P500* on labor force participation. I report the differences between simulated and realized paths evaluated in 2017Q4 and 2019Q4. The former describes the immediate impacts of the intervention and can be compared with the results of other studies, in particular with Magda et al. (2018). The latter summarizes the overall effects of the program. The complete time series of projected labor participation rate paths are shown in figure 5. The simulated flows are shown in figures 6 and 7.

The results indicate that in absence of the program, the labor force participation rate among eligible women would be 2-2.5 higher at the end of 2017. These direct effect estimates are similar to those obtained by Magda et al. (2018). The changes in flows interact and accumulate each period, leading to further evolution of intervention impacts on labor force participation. At the end of 2019, the estimates suggest a 3.5-4.3 drop in labor force participation among eligible women, solely due to the propagation of the direct (immediate) shocks, as shown by Scenarios (1) and (2). These effects are estimated with satisfactory precision.

Accounting also for the indirect effects of the program summarized by channel (3) mitigates the counterfactual increases in labor force supply after 2017Q4, suggesting two percentage points difference between the projected and realized paths of labor force participation rate at the end of 2019 as a result of the intervention. However, it is imprecisely estimated as suggested by large t-Statistic in table 11 or wide 95% confidence interval in figure 5. This comes from the fact that based on assumptions on the economic environment that are not explicitly accounted for in the model, the estimated path comprises accumulated counterfactual predictions from multiple periods. The uncertainty at each data point interacts with others and propagates into subsequent periods. Scenario (3) predicts also an effect on the ineligible females, though it is small in magnitude, attributing a 0.5 drop in the labor force participation among the ineligible women to the intervention. This estimate is also statistically insignificant. At the same time, the 95% confidence interval is not very wide, suggesting that the program did not have substantial indirect effects on the population of women that do not have eligible children.

7 Conclusion

In this paper, I propose a simple but comprehensive approach in analyzing effects of the introduction of large scale government program and apply the framework to study how a child benefit program impacts women labor supply. I develop and estimate a general discrete choice model of women's decisions that explains labor force participation flows. Changes in flows are decomposed into components related to changes in woman's decision rule, her observed characteristics and residual factors. By shutting down variation triggered by the program, I simulate counterfactual flow rate paths. Using the law of motion for the aggregate labor supply, I recover impacts of the program on labor force participation.

My estimates reveal a moderate but statistically significant drop in labor force participation among the eligible woman as a result of the program. There are several forces contributing to this effect. The strongest is related to discouraging labor force entry immediately after its introduction by affecting females' perceived trade-offs and beliefs. Some demographics experienced increased exit as a result of self-selection, mainly within first two years. Lastly, changing economic conditions on the labor market are likely to mitigate the negative effects by limiting outflows from the labor market. Changes in flow rates induced by the program are not large on quarterly basis, but accumulate over time leading to significant changes in the labor force participation.

The results suggest that in order to mitigate negative effects of the introduction of a large scale child benefit program on labor supply, the government should focus on supporting entering the labor force among initially inactive woman. Unfortunately, this may be a difficult task, as the discouragement effect results from the fact that the benefits themselves make it less profitable to make search effort. As related literature suggests, other designs of child support mechanisms – with a leading example of tax credits – may fulfill goals set to the program 500+ without creating negative incentives affecting labor supply.

References

Angrist, J. and B. Frandsen (2019): "Machine labor," Tech. rep., National Bureau of Economic Research.

Athey, S., J. Tibshirani, S. Wager, et al. (2019): "Generalized random forests," *The Annals of Statistics*, 47, 1148–1178.

Baker, M., D. Messacar, and M. Stabile (2021): "The Effects of Child Tax Benefits on Poverty and Labor Supply: Evidence from the Canada Child Benefit and Universal Child Care Benefit," Tech. rep., National Bureau of Economic Research.

Blundell, R., A. Duncan, J. McCrae, and C. Meghir (2000): "The labour market impact of the working families' tax credit," *Fiscal studies*, 21, 75–104.

Breiman, L. (2001): "Random forests," *Machine learning*, 45, 5–32.

Cengiz, D., A. Dube, A. S. Lindner, and D. Zentler-Munro (2021): "Seeing Beyond the Trees: Using Machine Learning to Estimate the Impact of Minimum Wages on Labor Market Outcomes," Tech. rep., National Bureau of Economic Research.

González, L. (2013): "The effect of a universal child benefit on conceptions, abortions, and early maternal labor supply," *American Economic Journal: Economic Policy*, 5, 160–88.

Immervoll, H., H. J. Kleven, C. T. Kreiner, and E. Saez (2007): "Welfare reform in European countries: a microsimulation analysis," *The Economic Journal*, 117, 1–44.

Koebel, K. and T. Schirle (2016): "The differential impact of universal child benefits on the labour supply of married and single mothers," *Canadian Public Policy*, 42, 49–64.

Lechner, M. et al. (2011): *The estimation of causal effects by difference-in-difference methods*, Now.

Magda, I., A. Kiełczewska, and N. Brandt (2018): "The Effects of Large Universal Child Benefits on Female Labour Supply," *IZA Discussion Paper No. 11652*.

Moffitt, R. A. (2002): "Welfare programs and labor supply," *Handbook of public economics*, 4, 2393–2430.

Myck, M. (2016): "Estimating Labour Supply Response to the Introduction of the Family 500+ Programme," *Centre For Economic Analysis (CenEA) Working Paper Series WP01/16*.

Myck, M. and K. Trzciński (2019): "From Partial to Full Universality: The Family 500+ Programme in Poland and its Labor Supply Implications," *ifo DICE Report*, 17, 36–44.

Paradowski, P. R., J. Wolszczak-Derlacz, and E. Sierminska (2020): "Inequality, poverty and child benefits: Evidence from a natural experiment," Tech. rep., LIS Working Paper Series.

Roy, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford economic papers*, 3, 135–146.

Schirle, T. (2015): "The effect of universal child benefits on labour supply," *Canadian Journal of Economics/Revue canadienne d'économique*, 48, 437–463.

Sigurdsson, J. (2019): "Labor supply responses and adjustment frictions: A tax-free year in iceland," *Available at SSRN 3278308*.

Stephens Jr, M. and T. Unayama (2015): "Child benefit payments and household wealth accumulation," *The Japanese Economic Review*, 66, 447–465.

Tibshirani, J., S. Athey, and S. Wager (2020): *grf: Generalized Random Forests*, r package version 1.2.0.

Generalized Random Forest

Generalized Random Forest estimator builds on the intuition developed in the local maximum likelihood literature. For each s, $\varrho_t(s)$ is a solution to estimating equation (7) in which the sample averages are obtained using a set of appropriate weights. These weights measure *similarity* in terms of the observed state variables between the target observation s and remaining of observations. Traditionally, they have been based on kernel functions, which are prone to dimensionality curse. In GRF, similarity weights are obtained adaptively, inheriting appealing features of the random forest algorithm.

Random forest (Breiman, 2001) is an algorithm that serves to produce predictions regarding some outcome given a (large) set of covariates. It delivers predictions that themselves are averages of predictions generated by simpler predictive algorithms, called trees. A tree algorithm is based on partitioning the dataset in a recursive way. At each stage, the dataset is split into two subsamples. The splitting rule heuristically chooses a covariate and a threshold, classifying observations into different subsamples depending on whether the value of the covariate is below or above the threshold. The splitting process repeats until a required number of splits is performed or another stopping criterion holds. The goal of these sample splits is to cluster observations that are *similar* with respect to some measure. The standard tree algorithms cluster observations sharing similar values of the outcome. The trees used in GRF perform splits to maximize heterogeneity in the target functional $\varrho_t(\cdot)$.

The GRF approach produces a set of similarity weights for each observation i with characteristics s. For each tree, if observation j with characteristics s' falls to the same final leaf[3] as s, it is assigned a number equal to 1 over the number of all observations that end up in the same leaf. Otherwise, it is assigned 0. The forest weight for j in predicting i's choice probability, denoted by $\alpha_j(s)$, is given by average of the assigned numbers over all trees. Therefore, the forest weights are obtained by averaging neighborhoods produced by different

[3] See **?** for a detailed treatment of tree methods in machine learning.

trees. They add up to 1 and by construction provide a measure of similarity with the target observation s.

Having obtained forest-based similarity weights for a target observation s, the predicted CCP $\hat{\varrho}_t(s)$ satisfies:

$$\hat{\varrho}_t(s) = \arg\min_{\varrho} \Big|\Big| \sum_{j=1}^{N} \alpha_j(s)\big(y_{jt} - \varrho\big) \Big|\Big|_2 \tag{11}$$

GRF exploits a so called *honest* sampling scheme in estimating (growing) trees. As a result, the predicted probabilities are consistent for the population conditional choice probabilities and asymptotically normal, which make GRF particularly useful for econometric applications. Based on the random forest algorithm, GRF is designed to deal with high dimensional datasets and provides additional advantage in handling missing data. This is because the exact values of covariates are redundant in growing a tree. Therefore, to handle the missing data it is sufficient to just label them as a distinct category. All of these appealing features of the GRF framework motivate its use in estimating the optimal policy in woman's labor force participation decisions.

Split Significance Measure

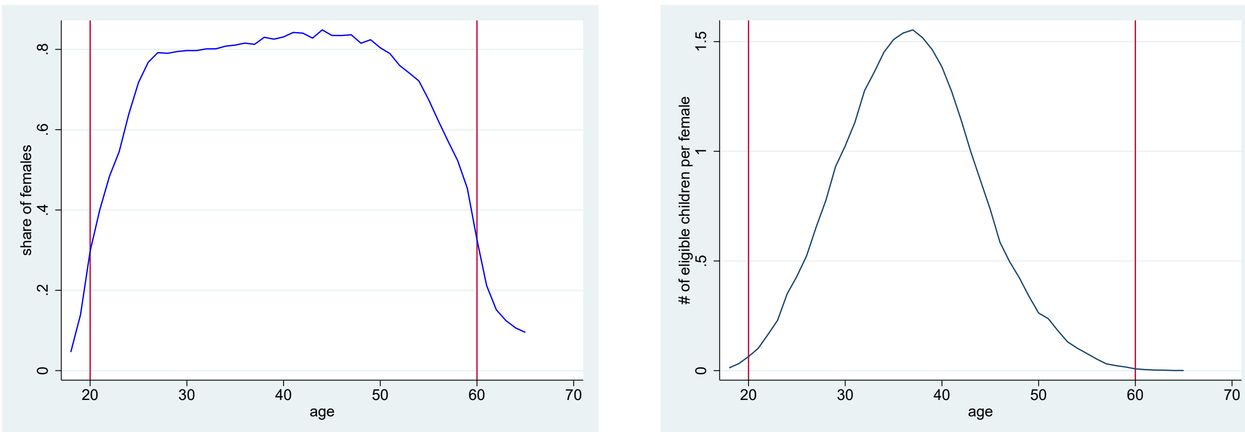
The random forest algorithm provides a simple framework to evaluate predictive power of particular covariates in explaining the outcome. Intuitively, a good measure of variable importance is a count of how many times an observed the variable is used to perform a data split in the forest. Let $k_{\max}$ be the maximal weighted sum of splits observed in the sample and k_j be the weighted number of splits for a covariate s_j. The *split significance* describes the relative importance of variables:

$$\text{split significance}_j = 100 - 100 \cdot \frac{k_j}{k_{\max}}$$

The larger the split significance measure is the smaller impact covariate j has in predicting the outcome, as compared to the most important predictor. Split significance equal to 100 implies that a covariate is not used in the prediction. Note that this measure is agnostic about the direction of the correlation between the predicted outcome and observed state variable.

Graphs and Tables

Figure 1: Labor force participation and children bearing in the life cycle.



(a) female labor force participation in the life cycle

(b) number of children below 18 per female in life cycle

Table 1: Program Participation by the Number of Children

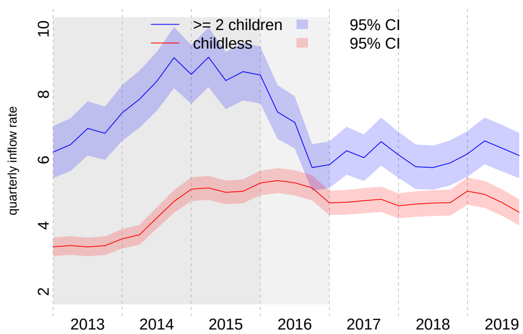
2 children	more than 2 children	1 children	
2017Q1 - 2019Q4		2017Q1-2019Q1	2019Q4
.943	.968	.234	.899

Table 2: Labor force participation - inflows and outflows.

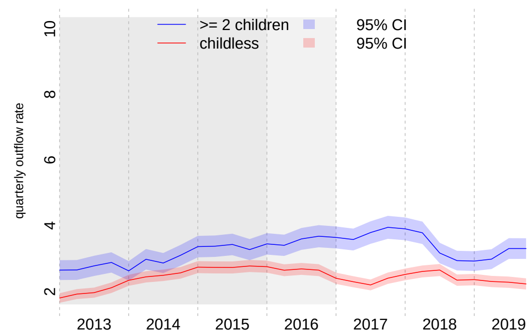
	>=2 children		childless	
	inflows $y_q^0 = 0$	outflows $y_q^0 = 1$	inflows $y_q^0 = 0$	outflows $y_q^0 = 1$
post-int. (2017-2019)	6.202	3.393	4.648	2.325
pre-int. (2014-2015)	8.901	3.188	4.893	2.669
difference	-2.699	.205	-.245	-.343

Inflows and outflows expressed in percentage points.

Figure 2: Labor force participation - inflows and outflows.



(a) y_q^0: quarterly rate of inflows to labor force, percentage points. Dark gray background indicates pre-intervention periods, light gray indicate the transition period.



(b) y_q^1: quarterly rate of outflows from labor force, percentage points. Dark gray background indicates pre-intervention periods, light gray indicate the transition period.

Table 3: Choice of the Observed State Variables – Summary

	female	husband	mother	father
household level covariates	✓			
female demographics	✓			
employment status	✓	✓	✓*	✓*
job search	✓	✓		
education	✓	✓		

* only a selection of variables in the group is chosen.

Figure 3: Labor force participation of females (ages 20-60) by number of eligible children (below 18).

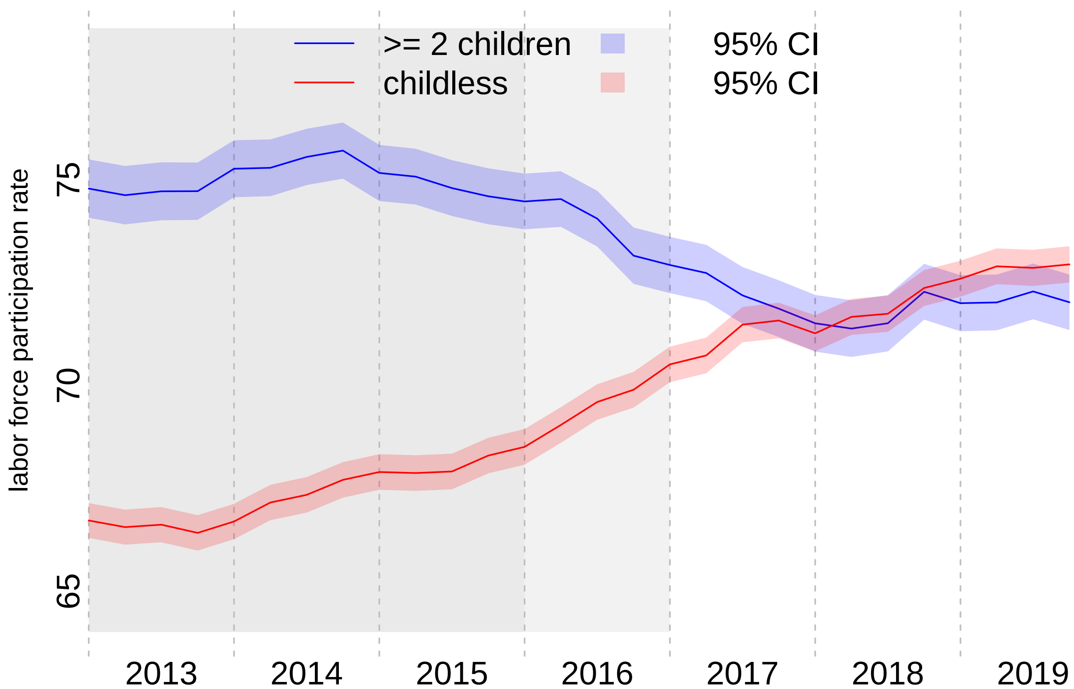


Table 4: List of covariates used to predict female's labor force participation decision.

code	description		type
	main description	detailed description	
	household level covariates (group hh)		
hh1	# of individuals in the household:	all	int
hh2		in the labor force	int
hh3		above 65	int
hh4		having a job	int
hh5_1	household lives in a city of population:	more than 100k	bin
hh5_2		50k-100k	bin
hh5_3		20k-50k	bin
hh5_4		10-20k	bin
hh5_5		less than 10k	bin
hh5_6	household lives in a rural zone		bin
hh6_1-16	voivodship		bin (16)
hh7_1-12	month		bin (12)
	female demographics (group i)		
i1	age		int
i2_1	marital status	unmarried	bin
i2_2		married	bin
i2_3		widowed	bin
i2_4		divorced or separated	bin
i3	lives with:	husband/wife	bin
i4		mother	bin
i5		father	bin
i6	# of female's children	below 18	int
i7		18 or more	int
i8	# of child-years of all female's children until reaching 18		int
i9	# of other children in the household		int
i10	# of child-years of other children until reaching 18		int
i11	age of the youngest child below 18		int
i12	age of the oldest child below 18		int
i13	age of the youngest child above 18		int
	employment status (groups w, hw, fw, mw)		
w1_1	worked as usually		bin
w1_2	worked in limited amount of time		bin
w1_3	# of hours usually worked		int
w2_1	limited time due to the maternal duties		bin
w2_2	limited time due to the working system		bin
w3_1	worked as individual entrepreneur		bin
w3_2	worked as salaried employee		bin
w4_1	received at least 50\% of the salary while not working	yes	bin
w4_2		no	bin
w5_1	working place	public institution	bin
w5_2		private institution	bin
w6_1	working horizon	permanent	bin
w6_2		temporary because cannot find permanent job	bin

Continued on next page

Table 4 – continued from previous page

code	description		type
	main description	detailed description	
w7	duration of temporary job		int
w8_1	works in a shift system	yes	bin
w8_2		no	bin
w9_1	paid overtime	yes	bin
w9_2		no	bin
w10	monthly wage		cont
w11			cat
w12_1	company's headquarters in the community	yes	bin
w12_2		no	bin
w13_1	company's headquarters in Poland	yes	bin
w13_2		no	bin
w14	# of employees in the workplace		ordered
w15_1	works full-time		bin
w15_2	works part-time		bin
w16_1	works part-time because of necessity of care provision		bin
w16_2	works part-time because cannot find a full-time job		bin
w17_1	wants to work more in order to earn more	yes	bin
w17_2		no	bin
w18_1-w18_10	occupation (2 digit)		bin x 10
w19	total working experience (years)		int
w20_1	has an additional job	yes	bin
w20_2		no	bin
w21_1	used to work in the past		bin
w21_2	has never worked		bin
w22	# of months since lost the job		int
w23_1	reasons for losing a job	pension	bin
w23_2		disability benefit	bin
w23_3		bankruptcy of the employer	bin
w23_4		unsatisfactory working conditions	bin
w23_5		necessity to provide care	bin
w23_6		end of temporary job	bin
w23_7		personal/family	bin
w23_8		illness	bin
w23_9		others	bin
w24	# years worked in the previous job		int
w25_1	main source of income	salaried worker	bin
w25_2		farmer	bin
w25_3		entrepreneurship	bin
w25_4		pension	bin
w25_5		disability benefit	bin
w25_6		unemployment benefit, other benefits	bin
w25_7		dependent	bin
w26_1	other source of income	salaried worker	bin
w26_2		farmer	bin
w26_3		entrepreneurship	bin

Continued on next page

Table 4 – continued from previous page

code	description		type
	main description	detailed description	
w26_4		pension	bin
w26_5		disability benefit	bin
w26_6		unemployment benefit, other benefits	bin
w26_7		dependent	bin
w27_1	subjective evaluation of current labor status	employed	bin
w27_2		unemployed	bin
w27_3		student	bin
w27_4		pensioner	bin
w27_5		disability	bin
w27_6		family duties	bin
w27_7		other inactivity	bin
w28_1	subjective evaluation of labor status in the previous year	employed	bin
w28_2		unemployed	bin
w28_3		student	bin
w28_4		pensioner	bin
w28_5		disability	bin
w28_6		family duties	bin
w28_7		other inactivity	bin
w29	learned profession is the actual profession		bin
	job search (groups s, hs)		
s1_1	searched for a job within past month	yes	bin
s1_2		no and has not found a job	bin
s2	reservation wage		cont
s3_1	reasons for not looking for a job	failed to find bfore	bin
s3_2		student	bin
s3_3		care provision	bin
s3_4		other personal reasons	bin
s3_5		pension	bin
s3_6		others	bin
s3_7		has already found one	bin
s4	reasons for not looking for a job	intensity of care provision requirement (6)	ordered
s5_1	can start a job within 2 weeks	yes	bin
s5_2		no	bin
s6	working or studying before start searching		bin
s7	intensity of search (0-14)		ordered
s8	duration of search (months)		int
s9_1	type of job searched	full time	bin
s9_2		part time	bin
s10_1	registered as unemployed	yes	bin
s10_2		no	bin
s11_1	reasons for job search	fear from losing the current job	bin
s11_2		wants to work more	bin
s11_3		wants to work less	bin
s11_4		wants better employment conditions	bin
s11_5		want a permanent job	bin

Continued on next page

Table 4 – continued from previous page

code	description		type
	main description	detailed description	
	education (groups e, he)		
e1	student		bin
e2_1	mode of study	full-time	bin
e2_2		part-time	bin
e3_1 - e3_10	learned profession		bin x 10
e4_1	education achieved	no	bin
e4_2		elementary	bin
e4_3		vocational	bin
e4_4		secondary without diploma	bin
e4_5		secondary with diploma	bin
e4_6		postsecondary	bin
e4_7		undergraduate	bin
e4_8		master degree	bin
e4_9		PhD degree	bin
e5	years since education completed		int

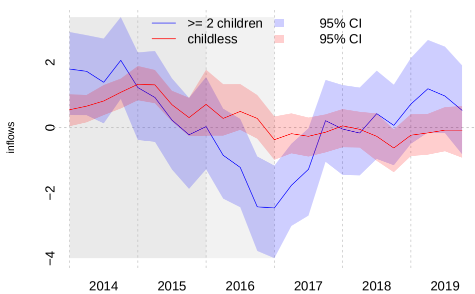
variables: bin: dummy indicator, int: integer valued, ordered: ordered valued, cont: continuous valued. Gray rows indicate choice of variables for mother and father.

Table 5: Flows to the labor force – estimates of pre- and post-intervention differences.

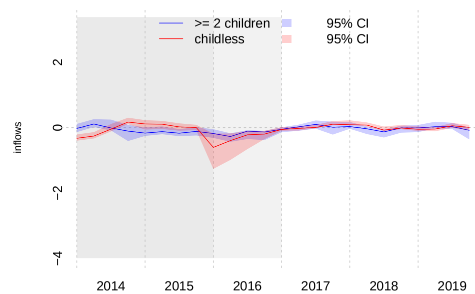
	treatment	*selection*	*residual*
	$\hat{\beta}(s_0) = \hat{\varrho}_1(s_0) - \hat{\varrho}_0(s_0)$	$\hat{\gamma}_1(s_1, s_0) = \hat{\varrho}_1(s_1) - \hat{\varrho}_1(s_0)$	$\hat{\xi} = (\overline{y}_1 - \overline{y}_0) - (\hat{\varrho}_1(s_1) - \hat{\varrho}_0(s_0))$
		Inflows to the Labor Force	
>= 2 children	-2.142	-0.142	-0.415
	$(-4.283)^{***}$	(-1.791)	(-0.851)
childless	0.149	-0.3	-0.095
	(0.299)	$(-3.79)^{***}$	(-0.458)
difference	-2.292	0.158	
	$(-4.123)^{***}$	(1.458)	
		Outflows from the Labor Force	
>= 2 children	1.189	-1.152	0.168
	$(5.644)^{***}$	$(-16.458)^{***}$	(0.952)
childless	0.907	-1.318	0.067
	$(4.308)^{***}$	$(-18.835)^{***}$	(0.761)
difference	0.281	0.166	
	(1.205)	$(2.005)^{*}$	

The table presents estimated parameters of the decomposition (6) in pre-post setting. In the inflows part, ϱ denotes the conditional probability of being in the labor force conditionally on being out a quarter before. In the outflows part, ϱ denotes the conditional probability of being out of the labor force conditionally on being in a quarter before. $t = 0$ and $t = 1$ denote pre-intervention (2014-2015) and post-intervention (2017-2019) periods respectively. t-Statistics based on 150 bootstrap replications presented in the parentheses. Stars denote *** p-val.<0.001, **, p-val.<.01, * p-val.<0.05.

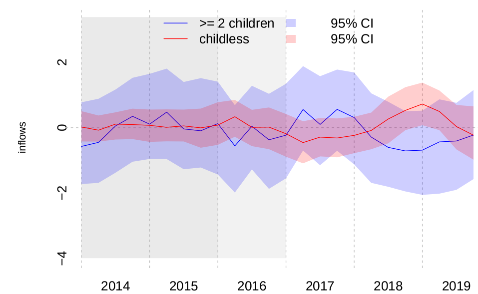
Figure 4: Estimated Parameters of Decomposition (6), Time Series Approach.



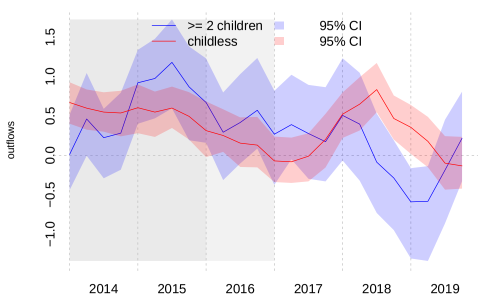
(a) inflows: *treatment* parameters: $\hat{\beta}(s_{t-1}) \equiv \hat{\varrho}_t(s_{t-1}) - \hat{\varrho}_{t-1}(s_{t-1})$



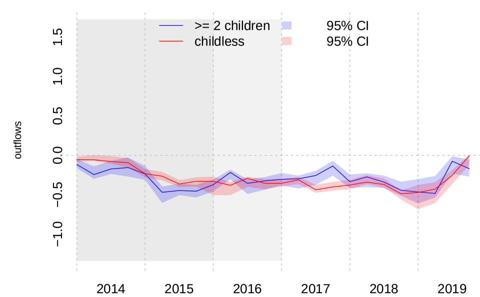
(b) inflows: *selection* parameters: $\hat{\gamma}_t(s_t, s_{t-1}) \equiv \hat{\varrho}_t(s_t) - \hat{\varrho}_t(s_{t-1})$



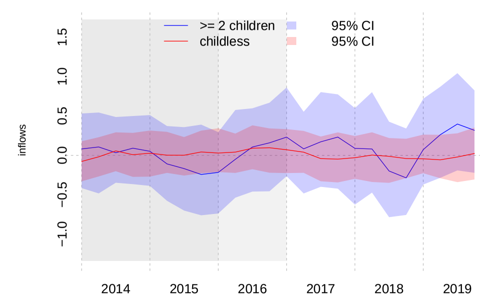
(c) inflows: *idiosyncratic* parameters: $\hat{\xi}_t(s_t, s_{t-1}) \equiv (\overline{y}_1 - \overline{y}_0) - (\hat{\varrho}_1(s_1) - \hat{\varrho}_0(s_0))$



(d) outflows: *treatment* parameters: $\hat{\beta}(s_{t-1}) \equiv \hat{\varrho}_t(s_{t-1}) - \hat{\varrho}_{t-1}(s_{t-1})$



(e) outflows: *selection* parameters: $\hat{\gamma}_t(s_t, s_{t-1}) \equiv \hat{\varrho}_t(s_t) - \hat{\varrho}_t(s_{t-1})$



(f) outflows: *idiosyncratic* parameters: $\hat{xi}_t(s_t, s_{t-1}) \equiv (\overline{y}_1 - \overline{y}_0) - (\hat{\varrho}_1(s_1) - \hat{\varrho}_0(s_0))$

Table 6: Heterogenous Impacts - differences in parameters between the eligible and ineligible.

	inflows		outflows	
	treatment	*selection*	*treatment*	*selection*
	$\hat{\beta}(s_0) = \hat{\varrho}_1(s_0) - \hat{\varrho}_0(s_0)$	$\hat{\gamma}_1(s_1, s_0) = \hat{\varrho}_1(s_1) - \hat{\varrho}_1(s_0)$	$\hat{\beta}(s_0) = \hat{\varrho}_1(s_0) - \hat{\varrho}_0(s_0)$	$\hat{\gamma}_1(s_1, s_0) = \hat{\varrho}_1(s_1) - \hat{\varrho}_1(s_0)$
		education		
tertiary	−2.181 (−1.77)	−0.193 (−0.939)	0.362 (1.886)	−0.008 (−0.137)
secondary	−2.354 (−3.722)***	0.029 (0.269)	0.311 (1.048)	0.454 (4.716)***
vocational or lower	−2.101 (−3.839)***	−0.011 (−0.092)	0.323 (0.816)	0.546 (3.546)***
		size of the city		
>= 100k	−2.933 (−3.757)***	0.369 (2.336)*	0.447 (1.904)	0.055 (0.677)
20-100k	−2.318 (−3.799)***	−0.128 (−0.752)	0.254 (0.965)	0.367 (3.535)***
< 20k	−2.119 (−3.484)***	0.121 (0.761)	0.285 (1.027)	0.595 (5.936)***
rural	−2.113 (−3.684)***	0.146 (1.203)	0.146 (0.579)	−0.029 (−0.284)
		woman's age		
20-29	−2.19 (−2.591)**	−0.159 (−1.135)	0.958 (1.578)	−0.628 (−3.088)**
30-39	−1.931 (−2.494)*	0.999 (5.139)***	0.4 (1.627)	0.07 (0.829)
40-49	−1.472 (−1.848)	0.822 (4.674)***	−0.129 (−0.635)	0.514 (5.822)***
		woman's marital status		
married	−1.976 (−3.57)***	0.17 (1.295)	0.45 (1.799)	0.051 (0.587)
never married	−2.84 (−3.349)***	0.36 (1.813)	0.856 (1.346)	−0.241 (−0.951)
divorced	−3.457 (−3.472)***	−0.303 (−1.57)	0.433 (0.856)	−0.643 (−3.457)***
		# of eligible children		
2	−2.494 (−4.218)***	0.112 (0.964)	0.268 (1.173)	0.193 (2.378)*
3	−1.95 (−3.49)***	0.298 (2.541)*	0.235 (0.856)	0.24 (2.347)*
>= 4	−1.481 (−2.474)*	0.313 (2.261)*	0.819 (1.711)	−0.539 (−2.37)*
		age of the youngest child		
0-3	−1.566 (−2.003)*	0.55 (2.713)**	0.376 (0.8)	1.437 (10.785)***
4-6	−2.314 (−3.297)***	−0.035 (−0.261)	0.491 (1.345)	0.259 (2.423)*
7-12	−2.511 (−4.393)***	0.177 (1.599)	0.388 (1.477)	−0.067 (−0.637)
13-17	−2.172 (−3.33)***	0.15 (1.15)	0.102 (0.468)	0.143 (1.808)

The table presents differences between estimated parameters of the decomposition (6) in pre-post setting for the eligible and ineligible, by subgroups defined by covariates. In the inflows part, ϱ denotes the conditional probability of being in the labor force conditionally on being out a quarter before. In the outflows part, ϱ denotes the conditional probability of being out of the labor force conditionally on being in a quarter before. $t = 0$ and $t = 1$ denote pre-intervention (2014-2015) and post-intervention (2017-2019) periods respectively. t-Statistics based on 150 bootstrap replications presented in the parentheses. Stars denote *** p-val.<0.001, **, p-val.<.01, * p-val.<0.05.

Table 7: 10 most important covariates in predicting labor force participation according to the split significance measure – women with 2 or more children, conditionally on being out of labor force a quarter before.

	who	description	details
		Top post-intervention predictors of the inflows	
1	woman	subjective evaluation of labor status in the previous year	employed
2	woman	can start a job within 2 weeks	yes
3	woman	total working experience (years)	
4	woman	# of months since lost the job	
5	woman	# years worked in the previous job	
6	woman	occupation	missing
7	woman	learned profession	specialists
8	woman	years since education completed	
9	woman	subjective evaluation of current labor status	unemployed
10	husband	monthly wage	
		Top 10 pre-intervention predictors of the inflows	
1	woman	total working experience (years)	
2	woman	# years worked in the previous job	
3	woman	can start a job within 2 weeks	yes
4	woman	occupation	missing
5	woman	reasons for losing a job	necessity to provide care
6	woman	subjective evaluation of current labor status	unemployed
7	woman	subjective evaluation of labor status in the previous year	employed
8	woman	# of months since lost the job	
9	woman	registered as unemployed	no
10	woman	# of child-years of all female's children until reaching 18	
		Top 10 predictors of the inflows to increase their importance	
1	woman	learned profession	specialists
2	woman	subjective evaluation of labor status in the previous year	employed
3	woman	# of months since lost the job	
4	woman	years since education completed	
5	husband	monthly wage	
6	woman	education achieved	master degree
7	woman	occupation - current or most recent	specialists
8	husband	wants to work more in order to earn more	yes
9	husband	learned profession	specialists
10	woman	subjective evaluation of current labor status	family duties
		Top 10 predictors of the inflows to decrease their importance	
1	woman	reasons for losing a job	necessity to provide care
2	woman	registered as unemployed	no
3	woman	subjective evaluation of labor status in the previous year	unemployed
4	household	voivodship	lubelskie
5	woman	total working experience (years)	
6	household	household lives in a city of population:	more than 100k
7	woman	reasons for not looking for a job	failed to find before
8	woman	# of child-years of all female's children until reaching 18	
9	woman	occupation - current or most recent	simple task workers
10	husband	has an additional job	no

Table 8: 10 most important covariates in predicting labor force participation according to the split significance measure – childless women, conditionally on being out of labor force a quarter before.

	who	description	details
		Top post-intervention predictors of the inflows	
1	woman	can start a job within 2 weeks	yes
2	woman	age	
3	woman	subjective evaluation of current labor status	unemployed
4	woman	# of months since lost the job	
5	woman	registered as unemployed	yes
6	woman	subjective evaluation of labor status in the previous year	unemployed
7	woman	reasons for losing a job	necessity to provide care
8	woman	years since education completed	
9	woman	main source of income	dependent
10	husband	years since education completed	
		Top 10 pre-intervention predictors of the inflows	
1	woman	subjective evaluation of current labor status	unemployed
2	woman	registered as unemployed	yes
3	woman	can start a job within 2 weeks	yes
4	woman	subjective evaluation of labor status in the previous year	unemployed
5	woman	age	
6	woman	# of months since lost the job	
7	woman	years since education completed	
8	woman	# years worked in the previous job	
9	woman	main source of income	dependent
10	husband	years since education completed	
		Top 10 predictors of the inflows to increase their importance	
1	woman	reasons for losing a job	necessity to provide care
2	woman	age	
3	woman	can start a job within 2 weeks	yes
4	woman	# of months since lost the job	
5	husband	years since education completed	
6	woman	marital status	unmarried
7	woman	reasons for not looking for a job	pension
8	woman	main source of income	dependent
9	father	main source of income	salaried worker
10	husband	total working experience (years)	
		Top 10 predictors of the inflows to decrease their importance	
1	woman	subjective evaluation of labor status in the previous year	employed
2	woman	registered as unemployed	yes
3	woman	subjective evaluation of current labor status	unemployed
4	woman	lives with:	mother
5	woman	occupation - current or most recent	sevices and retail
6	woman	# years worked in the previous job	
7	mother	subjective evaluation of labor status in the previous year	employed
8	woman	reasons for not looking for a job	failed to find before
9	woman	reasons for losing a job	pension
10	husband	main source of income	pension

Table 9: 10 most important covariates in predicting labor force participation according to the split significance measure – women with 2 or more children, conditionally on being in the labor force a quarter before.

	who	description	details
		Top post-intervention predictors of the outflows	
1	woman	# of hours usually worked	
2	woman	# of employees in the workplace	
3	woman	subjective evaluation of current labor status	employed
4	woman	# of months since lost the job	
5	woman	# years worked in the previous job	
6	woman	has an additional job	no
7	woman	intensity of search (0-14)	
8	woman	subjective evaluation of labor status in the previous year	employed
9	woman	duration of search (months)	
10	woman	wants to work more in order to earn more	no
		Top 10 pre-intervention predictors of the outflows	
1	woman	# of hours usually worked	
2	woman	# of employees in the workplace	
3	woman	# of months since lost the job	
4	woman	subjective evaluation of current labor status	employed
5	woman	reservation wage	
6	woman	searched for a job within past month	no and has not found a job
7	woman	# years worked in the previous job	
8	woman	intensity of search (0-14)	
9	woman	can start a job within 2 weeks	yes
10	woman	works full-time	
		Top 10 predictors of the outflows to increase their importance	
1	woman	main source of income	unemployment benefit, other benefits
2	woman	has an additional job	no
3	woman	working horizon	permanent
4	woman	subjective evaluation of current labor status	family duties
5	woman	worked as salaried employee	
6	woman	working place	public institution
7	household	# of individuals in the household:	having a job
8	woman	searched for a job within past month	yes
9	woman	wants to work more in order to earn more	no
10	woman	duration of search (months)	
		Top 10 predictors of the outflows to decrease their importance	
1	woman	reservation wage	
2	woman	can start a job within 2 weeks	yes
3	woman	searched for a job within past month	no and has not found a job
4	woman	main source of income	dependent
5	woman	used to work in the past	but not anymore
6	woman	# of employees in the workplace	
7	woman	registered as unemployed	no
8	woman	subjective evaluation of current labor status	unemployed
9	woman	# of months since lost the job	
10	woman	works full-time	

Table 10: 10 most important covariates in predicting labor force participation according to the split significance measure – childless women, conditionally on being in the labor force a quarter before.

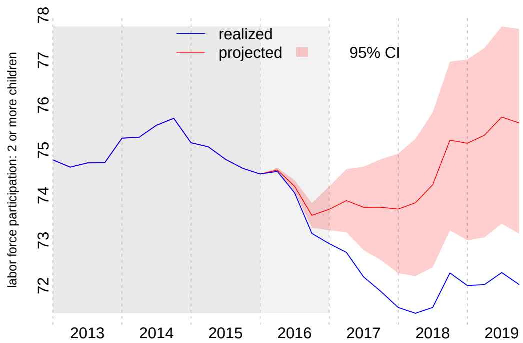
	who	description	details
		Top post-intervention predictors of the outflows	
1	woman	# of hours usually worked	
2	woman	subjective evaluation of current labor status	employed
3	woman	searched for a job within past month	no and has not found a job
4	woman	intensity of search (0-14)	
5	woman	duration of search (months)	
6	woman	worked as usually	
7	woman	works full-time	
8	woman	has an additional job	no
9	woman	subjective evaluation of labor status in the previous year	employed
10	woman	# of employees in the workplace	
		Top 10 pre-intervention predictors of the outflows	
1	woman	# of hours usually worked	
2	woman	# of employees in the workplace	
3	woman	# of months since lost the job	
4	woman	searched for a job within past month	no and has not found a job
5	woman	subjective evaluation of current labor status	employed
6	woman	duration of search (months)	
7	woman	intensity of search (0-14)	
8	woman	worked as usually	
9	woman	reservation wage	
10	woman	used to work in the past	but not anymore
		Top 10 predictors of the outflows to increase their importance	
1	woman	subjective evaluation of current labor status	employed
2	woman	has an additional job	no
3	woman	works full-time	
4	woman	main source of income	salaried worker
5	woman	intensity of search (0-14)	
6	woman	searched for a job within past month	yes
7	woman	works in a shift system	no
8	household	# of individuals in the household:	having a job
9	woman	worked as usually	
10	woman	working horizon	permanent
		Top 10 predictors of the outflows to decrease their importance	
1	woman	# of months since lost the job	
2	woman	# of employees in the workplace	
3	woman	reservation wage	
4	woman	used to work in the past	but not anymore
5	woman	can start a job within 2 weeks	yes
6	woman	main source of income	dependent
7	woman	subjective evaluation of current labor status	unemployed
8	woman	registered as unemployed	yes
9	woman	# years worked in the previous job	
10	woman	subjective evaluation of labor status in the previous year	unemployed

Table 11: Effects on the labor force participation – simulation results.

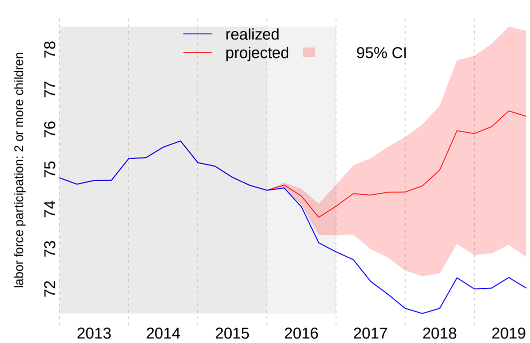
	the eligible		the ineligible	
	2017Q4	2019Q4	2017Q4	2019Q4
	Scenario (1)			
labor force participation	−1.868 (−3.109)**	−3.583 (−3.034)**	—	—
inflow rates	−1.917 (−2.683)**	−1.917 (−2.683)**	—	—
outflow rates	—	—	—	—
	Scenario (2)			
labor force participation	−2.557 (−3.421)***	−4.306 (−2.824)**	—	—
inflow rates	−1.917 (−2.683)**	−1.917 (−2.683)**	—	—
outflow rates	0.198 (1.594)	0.102 (0.553)	—	—
	Scenario (3)			
labor force participation	−2.557 (−3.421)***	−2.051 (−1.182)	—	−0.576 (−1.454)
inflow rates	−1.917 (−2.683)**	−1.917 (−2.683)**	—	—
outflow rates	0.198 (1.594)	−0.9 (−2.282)*	—	0.151 (0.967)

The table presents results of program evaluation exercise, that is the differences between the realized and simulated paths of female labor force participation at a given point in time. t-Statistics based on 150 bootstrap replications presented in the parentheses. Stars denote *** p-val.<0.001, **, p-val.<.01, * p-val.<0.05.

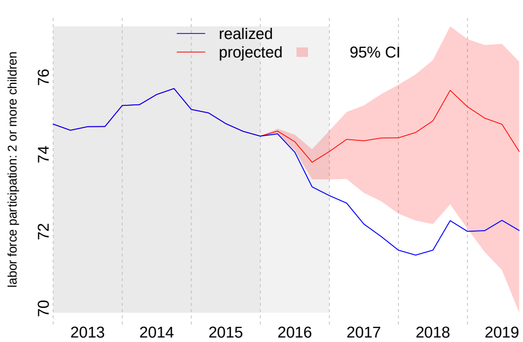
Figure 5: Effects on the labor force participation – simulation results.



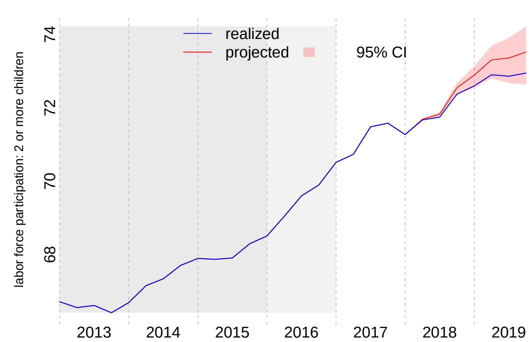
(a) Effects on the eligible - Scenario (1)



(b) Effects on the eligible - Scenario (2)

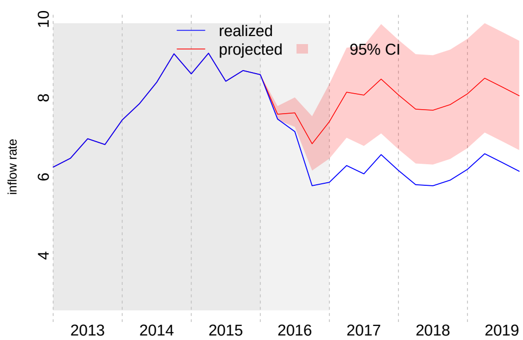


(c) Effects on the eligible - Scenario (3)

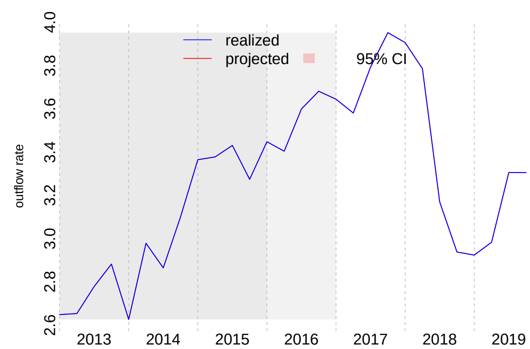


(d) Effects on the ineligible - Scenario (3)

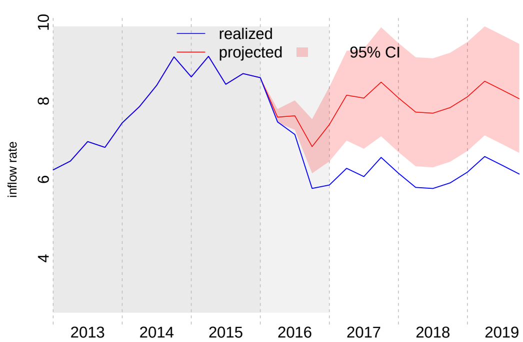
Figure 6: Effects on the flows – simulation results.



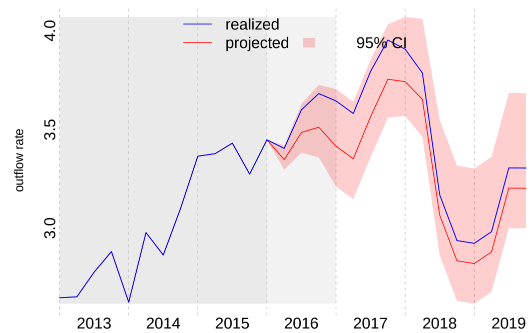
(a) Inflow rate dynamics among the eligible - Scenario (1)



(b) Outflow rate dynamics among the eligible - Scenario (1)

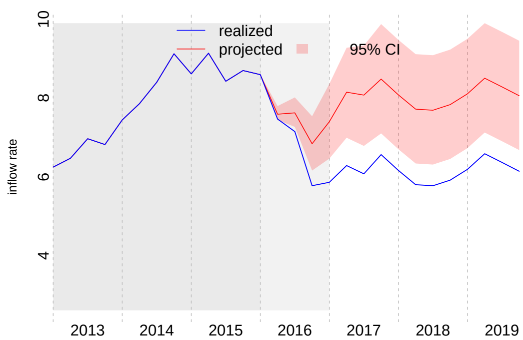


(c) Inflow rate dynamics among the eligible - Scenario (2)

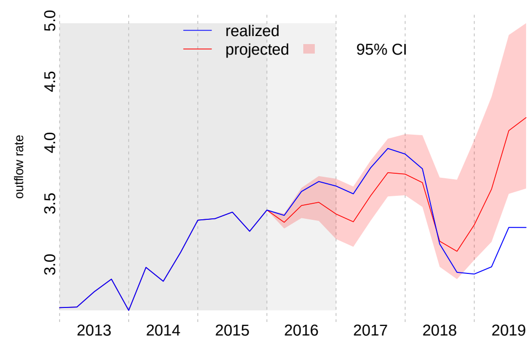


(d) Outflow rate dynamics among the eligible - Scenario (2)

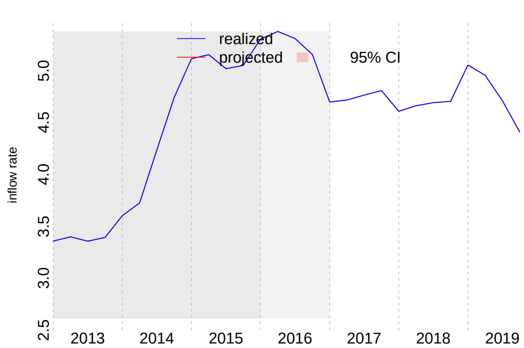
Figure 7: Effects on the flows – simulation results.



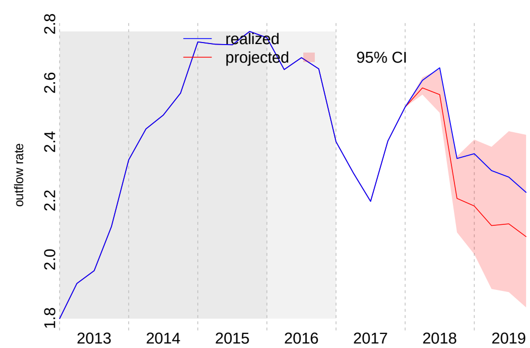
(a) Inflow rate dynamics among the eligible - Scenario (3)



(b) Outflow rate dynamics among the eligible - Scenario (3)



(c) Inflow rate dynamics among the ineligible - Scenario (3)



(d) Outflow rate dynamics among the ineligible - Scenario (3)